



Otimize os custos das cargas de trabalho da Microsoft em AWS

AWS Orientação prescritiva



AWS Orientação prescritiva: Otimize os custos das cargas de trabalho da Microsoft em AWS

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

As marcas comerciais e imagens de marcas da Amazon não podem ser usadas no contexto de nenhum produto ou serviço que não seja da Amazon, nem de qualquer maneira que possa gerar confusão entre os clientes ou que deprecie ou desprestigie a Amazon. Todas as outras marcas comerciais que não pertencem à Amazon pertencem a seus respectivos proprietários, que podem ou não ser afiliados, patrocinados pela Amazon ou ter conexão com ela.

Table of Contents

Introdução	1
Visão geral do	1
Público	1
Como usar este guia	2
Resultados de negócios desejados	3
Jornada de otimização de custos	4
Principais recomendações para a otimização de custos	7
Visão geral	7
Principais recomendações	7
Otimização e Avaliação de Licenciamento da AWS	9
Visão geral	9
Opções de avaliação	10
Avaliação completa	10
Escopo de workloads	11
Coletar dados	11
Analisar dados	12
Planejar as próximas etapas	14
Avaliação de impacto	15
Próximas etapas	16
Recursos adicionais	16
Windows no Amazon EC2	17
Automatizar as programações de parada e inicialização	18
Visão geral do	18
Estudos de caso	18
Cenário de otimização de custos	19
Recomendações de otimização de custos	21
Recursos adicionais do	34
Workloads do Windows do tamanho certo	35
Visão geral do	35
Cenário de otimização de custos	35
Recomendações de otimização de custos	36
Recomendações	46
Recursos adicionais do	46
Selecionar o tipo de instância certo para workloads do Windows	47

Visão geral do	47
Recomendações de otimização de custos	48
Próximas etapas	57
Recursos adicionais do	59
Trazer licenças para workloads do Windows e SQL Server	59
Visão geral do	59
Hosts dedicados do Amazon EC2	60
AWS opções de licenciamento	64
Portabilidade de licenças do Windows Server	65
Cenário de otimização de custos	66
Recomendações de otimização de custos	73
Recursos adicionais do	73
Otimizar gastos com o Windows no Amazon EC2	74
Visão geral do	74
Conceitos básicos do Savings Plans	74
Cenário de otimização de custos	81
Recomendações de otimização de custos	85
Recursos adicionais do	87
Monitore os custos usando AWS ferramentas	87
Visão geral do	87
Recomendações de otimização de custos	87
Recursos adicionais do	91
SQL Server	92
Escolher uma solução de alta disponibilidade e recuperação de desastres	93
Visão geral do	93
Grupos de disponibilidade Always On do SQL Server	94
Instâncias de cluster de failover Always On do SQL Server	96
SIOS DataKeeper	98
Grupos de Disponibilidade Always On	100
Grupos de disponibilidade distribuídos	101
Envio de logs	102
AWS Database Migration Service	104
AWS Elastic Disaster Recovery	105
Comparação de custos	106
Recomendações de otimização de custos	111
Recursos adicionais do	112

Compreender o licenciamento do SQL Server	112
Visão geral do	112
AWS opções de licenciamento	113
Impacto nos custos de trazer licenças	114
Otimização de licença	114
Recomendações de otimização de custos	115
Recursos adicionais do	46
Selecionar a instância certa do EC2 para as workloads do SQL Server	121
Visão geral do	122
Comparação de custos	122
Cenário de otimização de custos	124
Recomendações de otimização de custos	125
Recursos adicionais do	129
Consolidar instâncias	129
Visão geral do	129
Cenário de otimização de custos	130
Recomendações de otimização de custos	131
Recursos adicionais do	132
Comparar edições do SQL Server	133
Visão geral do	133
Impacto do custo	134
Recomendações de otimização de custos	136
Recursos adicionais do	142
Avaliar a edição SQL Server Developer	142
Visão geral do	142
Impacto do custo	143
Recursos adicionais do	46
Avaliar o SQL Server no Linux	146
Visão geral do	146
Impacto do custo	147
Recomendações de otimização de custos	148
Recursos adicionais do	149
Otimizar as estratégias de backup do SQL Server.	150
Visão geral do	150
Backup no nível do servidor usando snapshots habilitados para VSS	150
Backup do SQL Server usando AWS Backup	153

Backup em nível de banco de dados	154
Recomendações de otimização de custos	164
Recursos adicionais do	167
Modernizar bancos de dados do SQL Server.	168
Visão geral do	168
Ofertas de banco de dados	168
Comparação entre o Amazon RDS e o Aurora	169
Recomendações de otimização de custos	171
Recursos adicionais do	176
Otimizar o armazenamento para o SQL Server	177
Visão geral do	177
Tipos, performance e custo de armazenamento SSD para o Amazon EBS	178
Otimização geral de custos de SSDs para o Amazon EBS	179
Recursos adicionais do	181
Otimizar o licenciamento do SQL Server usando o Compute Optimizer	181
Visão geral do	182
Recomendações de otimização de custos	182
Configurar o Compute Optimizer	183
Recursos adicionais do	185
Otimizar o dimensionamento do SQL Server usando o Compute Optimizer	185
Visão geral do	185
Configurar o Compute Optimizer	186
Recursos adicionais do	186
Analise Trusted Advisor as recomendações para cargas de trabalho do SQL Server	187
Visão geral do	187
Recomendações de otimização de custos	187
Configure Trusted Advisor	188
Recursos adicionais do	189
Contêineres	190
Migrar aplicações do Windows para contêineres	191
Visão geral do	191
Benefícios de custo	191
Recomendações de otimização de custos	193
Próximas etapas	197
Recursos adicionais do	197
Otimizar os custos AWS Fargate das tarefas no Amazon ECS	197

Visão geral do	197
Benefícios de custo	198
Recomendações de otimização de custos	198
Próximas etapas	205
Recursos adicionais do	205
Obter visibilidade dos seus custos do Amazon EKS	205
Visão geral do	205
Benefícios de custo	206
Recomendações de otimização de custos	206
Próximas etapas	210
Recursos adicionais do	210
Redefinir a plataforma das aplicações do Windows com o App2Container	211
Visão geral do	211
Benefícios de custo	212
Recomendações de otimização de custos	212
Próximas etapas	212
Recursos adicionais do	213
Armazenamento	214
Amazon EBS	214
Migrar volumes do Amazon EBS de gp2 para gp3	215
Modificar snapshots do Amazon EBS	219
Excluir volumes do Amazon EBS não anexados	222
Amazon FSx	225
Escolher o armazenamento adequado de arquivos SMB	226
Habilite a deduplicação de dados na Amazon FSx	231
Entenda a fragmentação de dados no FSx Windows File Server	234
Entenda o uso do volume de HDD na Amazon FSx	238
Usar uma única zona de disponibilidade	241
AWS Storage Gateway	244
Gateway de Arquivos do Amazon S3	244
Amazon FSx File Gateway	244
Impacto do custo	245
Recomendações de otimização de custos	247
Recursos adicionais do	250
Active Directory	251
Active Directory autogerenciado no Amazon EC2	251

Visão geral do	251
Impacto do custo	251
Recomendações de otimização de custos	252
Recursos adicionais do	257
AWS Managed Microsoft AD	257
Visão geral do	257
Impacto do custo	257
Recomendações de otimização de custos	257
Recursos adicionais do	259
AD Connector	259
Visão geral do	259
Impacto do custo	260
Recomendações de otimização de custos	260
Recursos adicionais do	261
.NET	262
Refatorar para o .NET Modern e migrar para o Linux	263
Visão geral do	263
Impacto do custo	263
Recomendações de otimização de custos	264
Considerações e recursos adicionais	265
Containerizar aplicações .NET	266
Visão geral do	266
Impacto do custo	266
Recomendações de otimização de custos	268
Recursos adicionais do	271
Usar instâncias e contêineres do Graviton	271
Visão geral do	271
Impacto do custo	271
Recomendações de otimização de custos	273
Recursos adicionais do	274
Suporte à escalabilidade dinâmica para aplicações .NET Framework estáticas	275
Visão geral do	275
Impacto do custo	280
Recomendações de otimização de custos	281
Recursos adicionais do	283
Use o armazenamento em cache para reduzir a demanda do banco de dados	283

Visão geral do	283
Impacto do custo	283
Recomendações de otimização de custos	284
Recursos adicionais do	291
Considerar o .NET sem servidor	291
Visão geral do	291
Impacto do custo	292
Recomendações de otimização de custos	292
Recursos adicionais do	297
Considerar bancos de dados com propósito específico	297
Visão geral do	297
Impacto do custo	301
Recomendações de otimização de custos	304
Recursos adicionais do	305
Próximas etapas	306
Histórico do documento	307
Glossário	308
#	308
A	309
B	312
C	314
D	317
E	321
F	323
G	325
H	326
eu	328
L	330
M	331
O	336
P	338
Q	341
R	342
S	345
T	349
U	350

V	351
W	351
Z	352
.....	cccliv

Otimize os custos das cargas de trabalho da Microsoft em AWS

Bill Pfeiffer, Chase Lindeman e Kevin Sookhan, Amazon Web Services (AWS)

Outubro de 2025 ([histórico do documento](#))

Visão geral do

Este guia fornece recomendações, melhores práticas e estratégias para ajudar a otimizar os custos de suas cargas de trabalho da Microsoft em AWS. O guia também inclui conhecimentos básicos da AWS, técnicas de otimização de custos e arquiteturas de referência para ajudar você a criar e automatizar workloads econômicas e de alta performance que atendam aos seus objetivos de negócios. Coletivamente, essa orientação é denominada Microsoft on AWS Cost Optimization (MACO). As orientações da MACO foram desenvolvidas por especialistas do setor e é baseada em cenários do mundo real.

Este guia aborda as seguintes workloads da Microsoft:

- Windows na Amazon Elastic Compute Cloud (Amazon EC2)
- SQL Server
- Contêineres
- Armazenamento
- Active Directory
- .NET

Público

Este guia é destinado a arquitetos, engenheiros, administradores, diretores CTOs, tomadores de decisões técnicas e AWS parceiros. É útil, mas não necessário, ter experiência prévia e um conhecimento básico de AWS faturamento, tecnologias da Microsoft e administração de AWS sistemas.

Como usar este guia

Você pode usar este guia para planejar e implementar sua jornada MACO para a nuvem. Recomendamos que você leia este guia do início ao fim para obter uma compreensão abrangente das opções e abordagens para otimizar os custos de suas cargas de trabalho da Microsoft em AWS. Você pode revisar as seguintes seções de workloads com base nas necessidades da sua organização:

- [Windows na Amazon EC2](#)
- [SQL Server](#)
- [Contêineres](#)
- [Armazenamento](#)
- [Active Directory](#)
- [.NET](#)

Important

As amostras de código fornecidas neste guia são apenas para fins de demonstração. É uma prática recomendada testar todo o código em um ambiente de desenvolvimento antes de usá-lo em um ambiente de produção. Antes de implementar qualquer código, recomendamos que você teste seu código em pequenos lotes e, em seguida, revise as alterações de custo resultantes do código usando o [AWS Cost Explorer](#). Isso pode ajudar você a solucionar casos de exceção e outras questões que podem se tornar problemáticas posteriormente.

Important

Os exemplos de preços neste guia são baseados nos preços no momento da publicação. Os preços estão sujeitos a alterações. Além disso, seus custos podem variar de acordo com suas Região da AWS AWS service (Serviço da AWS) cotas e outros fatores relacionados ao seu ambiente de nuvem.

Resultados de negócios desejados

Este guia pode ajudar você e sua organização a alcançar os seguintes resultados comerciais:

- Saiba como usar a Otimização e Avaliação de Licenciamento da AWS (AWS OLA) para avaliar e otimizar os ambientes on-premises e na nuvem atuais, com base na utilização real de recursos, no licenciamento de terceiros e nas dependências de aplicações.
- Desenvolva um caso de negócios para a otimização de custos usando a Calculadora de Modernização da AWS para workloads da Microsoft.
- Otimize os custos para suas workloads específicas da Microsoft, incluindo workloads para Windows no Amazon Elastic Compute Cloud (Amazon EC2), SQL Server, contêineres, armazenamento, Active Directory e .NET.

Jornada de otimização de custos

O escopo, o tempo e o caminho específico de sua jornada de migração para a nuvem dependem de seus objetivos de negócios, requisitos técnicos e outros fatores. Esta seção fornece um exemplo de uma jornada de migração para a nuvem que se concentra no [gerenciamento financeiro da nuvem com a AWS](#) e segue as recomendações e as práticas recomendadas da MACO. Você pode usar esse exemplo para compreender como projetar uma jornada de migração para a nuvem para workloads da Microsoft.

As seguintes tarefas de alto nível ilustram a abordagem que uma organização pode adotar para implementar as recomendações e as práticas recomendadas da MACO:

- Estabelecer uma estratégia de marcação e habilitar tags de alocação de custos definidas pelo usuário. Para obter mais informações, consulte o whitepaper da AWS [Best Practices for Tagging AWS Resources](#).
- Defina orçamentos com base em aplicações, equipes ou departamentos. Para obter mais informações, consulte [Managing your costs with AWS Budgets](#) no Guia do usuário do Gerenciamento de Faturamento e Custos da AWS.
- Realize uma Otimização e Avaliação de Licenciamento da AWS (AWS OLA) para acelerar a economia. Para obter mais informações, consulte [Avaliação de otimização e licenciamento da AWS](#) na documentação da AWS.
- Traga a sua própria licença (BYOL) das workloads do Windows e do SQL Server usando Amazon Elastic Compute Cloud Dedicated Hosts. Para obter mais informações, consulte a seção [Trazer licenças para workloads do Windows e SQL Server](#) deste guia.
- Otimize seu licenciamento do SQL Server na AWS. Para obter mais informações, consulte a seção [Compreender o licenciamento do SQL Server](#) deste guia.
- Selecione o tipo certo de instância para workloads do Windows. Para obter mais informações, consulte a seção [Selecionar o tipo de instância certo para workloads do Windows](#) deste guia.
- Selecione o tipo certo de instância para workloads do SQL. Para obter mais informações, consulte a seção [Selecionar a instância certa do EC2 para as workloads do SQL Server](#) deste guia.
- Migre o Amazon Elastic Block Store (Amazon EBS) de gp2 para gp3. Para obter mais informações, consulte a seção [Migrar volumes do Amazon EBS de gp2 para gp3](#) deste guia.
- Controle as workloads com o Agendador de Instâncias do EC2 na AWS. Para obter mais informações, consulte a seção [Automatizar as programações de parada e inicialização](#) deste guia.

- Elimine os custos do SQL Server para workloads que não sejam de produção usando o SQL Server Developer Edition. Para obter mais informações, consulte a seção [Avaliar a edição SQL Server Developer](#) deste guia.
- Use uma única zona de disponibilidade para o Amazon FSx para Windows File Server para workloads de desenvolvimento e testes. Para obter mais informações, consulte a seção [Usar uma única zona de disponibilidade](#) deste guia.
- Dimensione corretamente suas workloads do Windows usando o AWS Compute Optimizer. Para obter mais informações, consulte a seção [Workloads do Windows do tamanho certo](#) deste guia.
- Otimize os gastos com o Windows no Amazon EC2 usando Savings Plans. Para obter mais informações, consulte a seção [Otimizar gastos com o Windows no Amazon EC2](#) deste guia.
- Habilite a deduplicação de dados no FSx para Windows File Server. Para obter mais informações, consulte a seção [Habilite a deduplicação de dados na Amazon FSx](#) deste guia.
- Use a fragmentação de dados para sistemas de arquivos no FSx para Windows File Server. Para obter mais informações, consulte a seção [Entenda a fragmentação de dados no FSx Windows File Server](#) deste guia.
- Otimize suas estratégias de backup do SQL Server. Para obter mais informações, consulte a seção [Otimizar as estratégias de backup do SQL Server](#) deste guia.
- Faça com que as aplicações estáticas do .NET Framework sejam compatíveis com a escalabilidade dinâmica. Para obter mais informações, consulte [Suporte à escalabilidade dinâmica para aplicações .NET Framework estáticas](#) neste guia.
- Use microsserviços do .NET sem servidor. Para obter mais informações, consulte a seção [Considerar o .NET sem servidor](#) deste guia.
- Migre suas aplicações do Windows para contêineres. Para obter mais informações, consulte a seção [Containerizar aplicações .NET](#) deste guia.
- Use o [AWS Compute Optimizer](#) para dimensionar corretamente os contêineres Windows em execução no AWS Fargate para Amazon Elastic Container Service (Amazon ECS). Para obter mais informações, consulte a seção [Habilitar o Compute Optimizer](#) deste guia.
- Refatore para o .NET Modern e migre para o Linux. Para obter mais informações, consulte a seção [Refatorar para o .NET Modern e migrar para o Linux](#) deste guia.
- Aproveite as instâncias e contêineres do Graviton. Para obter mais informações, consulte a seção [Usar instâncias e contêineres do Graviton](#) deste guia.
- Modernize bancos de dados do SQL Server. Para obter mais informações, consulte a seção [Modernizar bancos de dados do SQL Server](#) deste guia.

- Projete a infraestrutura do Active Directory. Para obter mais informações, consulte a seção [Active Directory](#) deste guia.

Para obter mais informações sobre a jornada do cliente focada no gerenciamento financeiro da nuvem com a AWS, consulte o whitepaper da AWS [Cloud Financial Management capability](#).

Principais recomendações para a otimização de custos

Visão geral


A otimização de custos é um dos pilares do [AWS Well-Architected Framework](#) e desempenha um papel fundamental em seus planos de migração para a nuvem. Você encontrará recomendações para otimizações de custos em todo este guia, mas esta seção destaca as recomendações de maior impacto. Você pode implementar essas recomendações rapidamente, e elas terão um impacto significativo em sua organização. Essas recomendações podem ajudar a estabelecer as bases para todo o seu esforço de otimização de custos.

Principais recomendações

A tabela a seguir lista as principais recomendações para as otimizações de custo de maior impacto. A coluna “Dificuldade de implementar” classifica cada otimização com base em uma escala do que é mais fácil de implementar (1) até o que é mais difícil de implementar (5). A coluna “Economia estimada” mostra uma estimativa baseada em porcentagem de quanto sua organização pode economizar em cada otimização recomendada.

Otimizações	Dificuldade de implementar	Economia estimada
Workloads do Windows do tamanho certo	3	25%
Trazer licenças para workloads do Windows e SQL Server	3	30%
Avaliar a edição SQL Server Developer	2	20%
Compreender o licenciamento do SQL Server	2	Até 50%
Automatizar as programações de parada e inicialização	3	Até 40%

Otimizações	Dificuldade de implementar	Economia estimada
Selecionar o tipo de instância certo para workloads do Windows	1	10 a 30%
Refatorar para o .NET Modern e migrar para o Linux	5	10 a 20%
Otimizar gastos com o Windows no Amazon EC2	3	Até 20 a 40%
Migrar volumes do Amazon EBS de gp2 para gp3	4	Até 20%

 Important

A economia estimada na tabela anterior se aplica a cada domínio técnico individual, não ao gasto geral da AWS em uma conta. Por exemplo, você pode implementar o Agendador de Instâncias em vários tipos e tamanhos de ambiente que podem alterar a economia potencial. As estimativas se aplicam especificamente aos custos de instâncias do Amazon EC2 e não implicam nenhuma economia geral em outros Serviços da AWS. Essas estimativas são fornecidas como um indicador, não como garantia.

Os especialistas em MACO estão disponíveis para falar mais detalhadamente sobre otimizações de custos. Para marcar uma reunião para se aprofundar em seu caso de uso, entre em contato com a equipe da sua conta ou envie um e-mail para optimize-microsoft@amazon.com.

Otimização e Avaliação de Licenciamento da AWS

Visão geral

Use a [Otimização e Avaliação de Licenciamento da AWS \(AWS OLA\)](#) para avaliar e otimizar os ambientes on-premises e na nuvem atuais, com base na utilização de recursos, no licenciamento de terceiros e nas dependências de aplicações. Você pode usar a AWS OLA para ajudar sua organização a criar uma estratégia de migração e licenciamento que proporcione a redução de custos à medida que você migra para a AWS ou avalia as workloads existentes da Microsoft na AWS. Uma AWS OLA também pode ajudar você a obter o seguinte:

- Entenda a performance das aplicações, os contratos e as implantações existentes.
- Dimensione corretamente seus recursos.
- Desenvolva um roteiro para a Nuvem AWS.
- Reduza ou elimine os custos usando os investimentos atuais e pagando apenas pelo que você usa.

Recomendamos que você faça da AWS OLA a primeira etapa em sua [jornada de otimização de custos](#). Você pode trabalhar com a AWS Partner Network para concluir uma AWS OLA. Os parceiros ajudarão você a coletar dados de avaliação e fornecerão recomendações para otimizar seus custos de licenciamento e instâncias.

O diagrama a seguir fornece uma visão geral do processo de avaliação.



Opções de avaliação

Você pode escolher entre duas opções da AWS OLA para suas workloads da Microsoft na AWS:

- **Versão Lite:** neste caso de uso, todas as suas workloads estão no VMware. Você pode fornecer à AWS uma saída do [RVTools](#). Em seguida, a AWS pode oferecer um tempo de resposta de 1 a 5 dias. Essa abordagem usa informações pontuais extraídas diretamente do VMware vCenter para desenvolver recomendações de dimensionamento e oferecer opções de preços sob demanda.
- **Versão Full:** neste caso de uso, você tem um ambiente misto em execução em diferentes provedores de nuvem, servidores físicos e servidores virtuais. A AWS usa agentes do sistema operacional para coletar dados de uso de 14 a 30 dias. Isso permite que a AWS tome decisões fundamentadas sobre o dimensionamento de instâncias com base nos padrões de uso da sua aplicação. A AWS usa várias ferramentas de terceiros, como o Cloudamize, para concluir a análise. A AWS trabalha com sua AWS Partner Network para ajudar a fornecer a avaliação final do custo total de propriedade (TCO) com várias opções de preços que levam em consideração modelos de preços e diferentes arquiteturas.

Avaliação completa

A avaliação completa da AWS OLA é iniciada com uma chamada telefônica de uma hora. Durante essa chamada, a AWS ajuda você a determinar a infraestrutura ideal da AWS para ser compatível com sua migração, escolher um método de coleta de dados e estabelecer uma linha do tempo para a conclusão. A implementação de ferramentas de descoberta em sua organização depende do método de coleta de dados, do tamanho da sua organização e das ferramentas que sua organização usa para gerenciar sua frota de servidores. Normalmente, são necessárias duas semanas para coletar dados de uso.

O processo completo da AWS OLA leva de 30 a 45 dias e consiste nas seguintes fases:

- Escopo de workloads
- Coletar dados
- Analisar dados
- Planejar as próximas etapas

Escopo de workloads

Primeiro, a AWS trabalha com você e sua equipe para determinar o escopo da avaliação. Isso geralmente é dividido por tipo de ambiente (por exemplo, não produção e produção). O escopo inclui a localização das workloads. Podem ser workloads que você está migrando para a AWS, workloads que já estão em execução na AWS (por exemplo, a AWS OLA para o Amazon EC2) ou workloads em execução em outros provedores de nuvem.

Coletar dados

Em seguida, a AWS implanta ferramentas para ajudar na descoberta de recursos e coletar dados de performance de seus servidores. Essas ferramentas vêm em quatro opções de implantação:

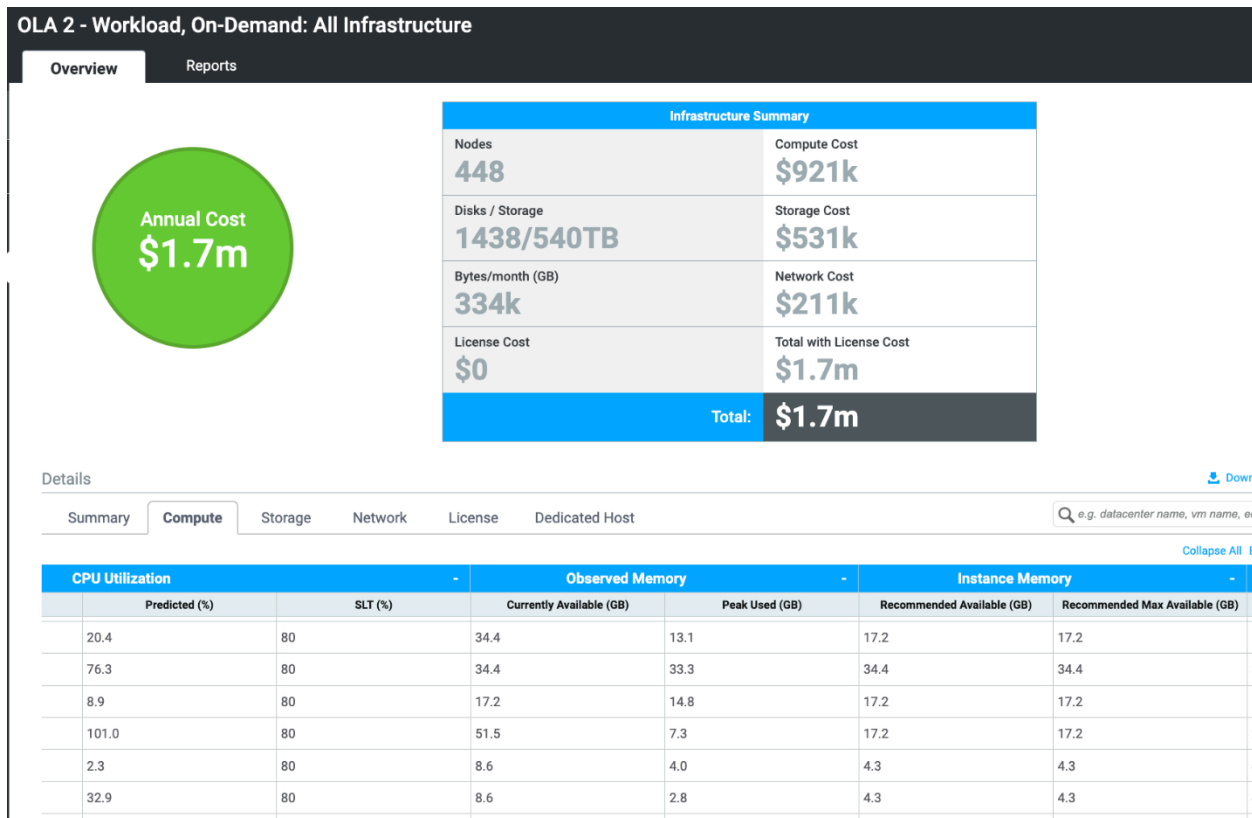
- Ferramentas que podem consultar o hipervisor (requer somente credenciais do VMware vCenter ou Hyper-V)
- Agentes que podem ser implantados em máquinas físicas ou virtuais
- Detecção sem agente usando SSH, Windows Remote Management (WinRM) ou Windows Management Instrumentation (WMI), dependendo do seu ambiente e sistema operacional
- Coleta e análise de dados de arquivo simples

Para a implantação de suas ferramentas, você pode misturar e combinar cada opção e consolidar os resultados. É fundamental garantir que qualquer opção escolhida não sobrecarregue seus recursos de TI. A AWS se esforça para tornar o processo de avaliação o mais completo possível. Além de uma breve chamada telefônica para ajudar na configuração, a equipe da AWS OLA e os arquitetos de soluções especialistas da Microsoft prepararão a análise do custo total de propriedade (TCO) e as recomendações para análise.

A coleta de dados geralmente leva de duas a três semanas quando a utilização da CPU, a utilização da RAM, o throughput do armazenamento, a IOPS e o throughput da rede são analisados. De preferência, essa coleta ocorre durante os horários de pico do mês comercial (por exemplo, durante os relatórios financeiros de final de mês). A AWS deseja capturar o pico de uso porque isso fornece boas amostras estatísticas de qual deveria ser o tamanho certo da instância da AWS, ao mesmo tempo em que garante que a performance possa exceder o que está disponível on-premises. A AWS mescla métricas de utilização com heurísticas de performance de várias gerações de processadores para definir exatamente a quantidade de CPU e RAM que uma determinada workload exige. Essas metas geralmente são menores do que as alocadas on-premises. Isso não apenas reduz

o custo computacional de acordo com o tamanho da instância, mas também otimiza os custos de licenciamento.

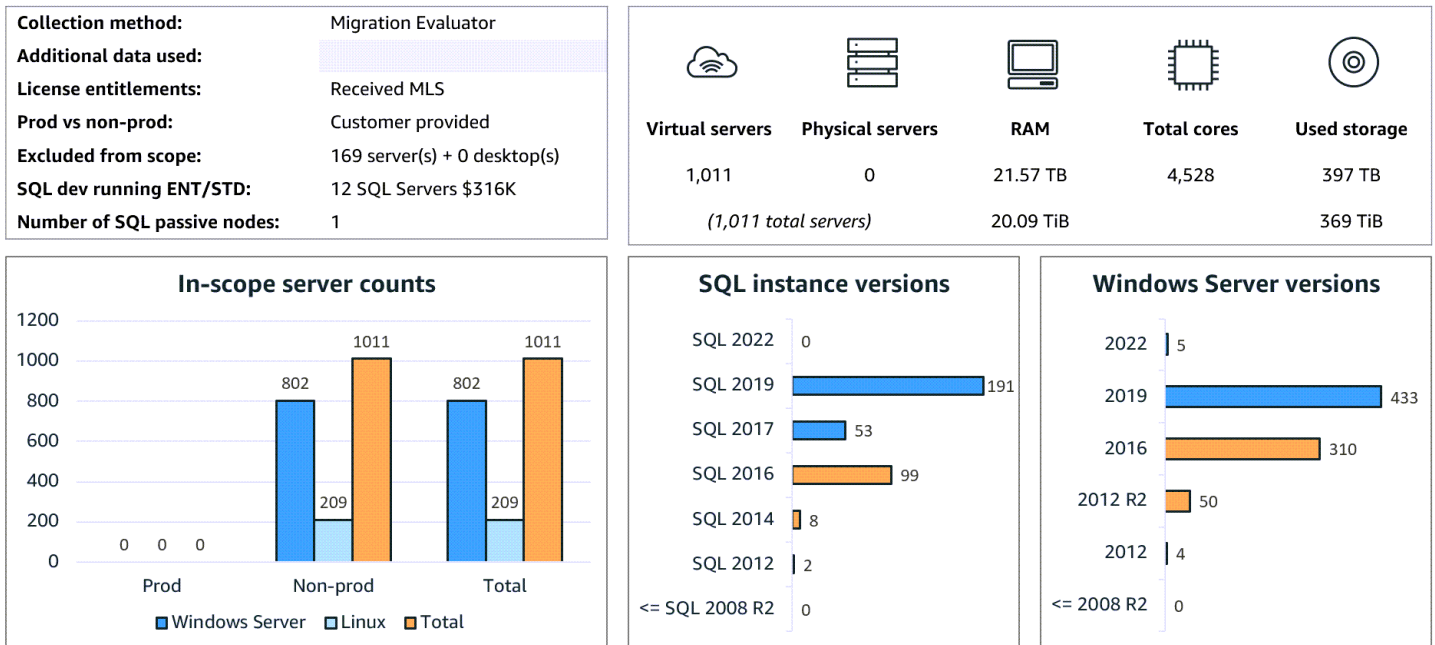
A exibição do painel a seguir mostra um exemplo de custos de infraestrutura que podem ser capturados por uma avaliação.



Analisar dados

A AWS faz uma apresentação resumida após a conclusão da coleta de dados. A AWS analisa os dados, resume as descobertas e, então, faz recomendações para o uso on-premises e a migração para a nuvem. Você pode reduzir os custos de computação e licenciamento examinando oportunidades de consolidação, ganhos de elasticidade (em que as workloads podem ser desativadas ou ajustadas sazonalmente), oportunidades de SKU correto (por exemplo, a edição SQL Server Enterprise está em uso, mas os requisitos de recursos e o uso de recursos sugerem que a edição SQL Server Standard é adequada). Para produtos como o SQL Server, que são licenciados pelo núcleo, geralmente faz sentido do ponto de vista financeiro colocar as workloads em uma instância de computação mais cara. Ou seja, se o perfil da CPU e a proporção de RAM para vCPU tiverem um efeito líquido na redução do número de núcleos licenciados para os casos de uso com licença incluída e traga a sua própria licença (BYOL).

Confira a seguir um exemplo de análise com base nos dados coletados pela avaliação.



Os cenários comuns de otimização incluem a identificação de oportunidades de otimização de recursos da AWS e economias em licenças de terceiros.

Exemplos de oportunidades de otimização de recursos da AWS:

- Evite o excesso de provisionamento para picos de uso.
- Evite especificar em excesso e subutilizar recursos.
- Dimensione suas instâncias corretamente e migre para as mais novas gerações de instâncias do EC2.
- Economize nos custos operacionais migrando para bancos de dados gerenciados.

Exemplos de economia de licenças de terceiros:

- Reduza os núcleos necessários para executar a mesma workload.
- Remova a edição Enterprise do SQL Server e os pacotes de complementos adicionais desnecessários.
- Remova os servidores zumbis e substitua o hardware desatualizado.
- Use as opções BYOL e com licença incluída para reduzir futuros contratos comerciais.
- Modernize para soluções de código aberto e nativas da nuvem.

Planejar as próximas etapas

Por fim, a AWS usa os dados de performance coletados para estimar o tamanho e o custo específicos da workload. A AWS também pode analisar de forma agregada seu ambiente com escopo e fornecer uma análise quantitativa. Isso pode ajudar você a determinar se a melhor opção é uma atualização on-premises ou uma migração para a AWS. Você pode criar um caso de negócios econômico na nuvem usando o resumo da análise de TCO (conforme mostrado no exemplo a seguir) fornecido ao final de uma AWS OLA.

	Option 1: Amazon EC2 shared	Option 1a: Amazon EC2 shared + power management	Option 2: Amazon EC2 mixed	Option 2a: Amazon EC2 mixed + power management
<i>Option details: compute</i>	100% Reserved Instances (RIs)	RIs + on-demand power management	100% RIs	RIs + on-demand power management
<i>Option details: Microsoft licenses</i>	WS LI and SQL BYOL	WS LI and SQL BYOL	WS BYOL or LI+SQL BYOL	WS BYOL or LI+SQL BYOL
Compute costs¹				
Year 1 compute cost	\$414,546	\$482,623	\$504,019	\$513,941
Year 1 vendor license included cost	\$392,858	\$244,415	\$9,804	\$4,783
	\$807,404	\$727,038	\$513,823	\$518,724
<i>Total compute savings in year 1, compared to Option 1</i>	—	10% (\$80,366)	36% (\$293,581)	36% (\$288,680)
Storage and networking costs²				
Annual estimated storage cost	\$336,494	\$336,494	\$336,494	\$336,494
Annual estimated networking cost	\$41,455	\$41,455	\$41,455	\$41,455
	\$377,949	\$377,949	\$377,949	\$377,949
Microsoft license costs**				
WS/CIS annual Software Assurance (SA) + current SPLA/Subs cost	\$0	\$0	\$0	\$0
WS/CIS license + SA + SPLA/Subs true-up cost	\$0	\$0	\$0	\$0
SQL annual SA + current SPLA/Subs cost	\$0	\$0	\$0	\$0
SQL license SA + current SPLA/Subs true-up cost	\$0	\$0	\$0	\$0
	\$0	\$0	\$0	\$0
Total estimated costs	\$1,185,353	\$1,104,987	\$891,772	\$896,673
<i>Annual TCO savings in year 1, compared to Option 1</i>	—	7% (\$80,366)	25% (\$293,581)	24% (\$288,680)

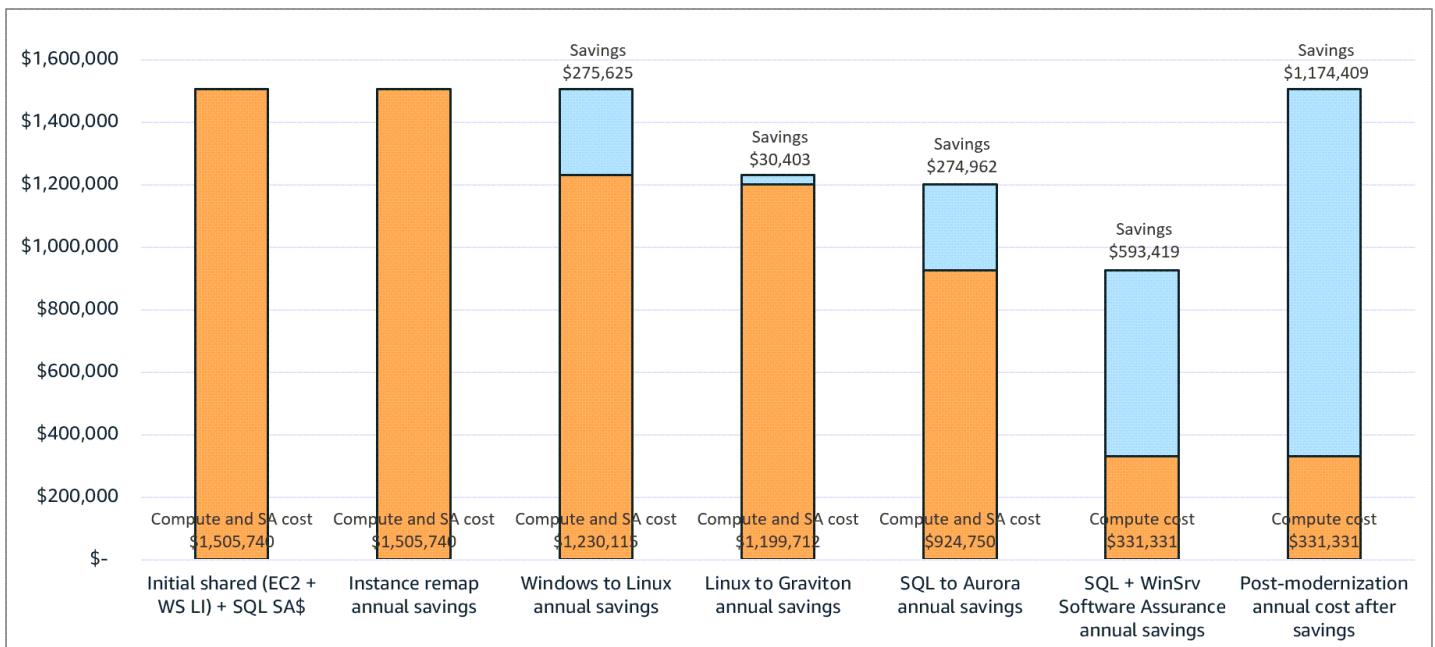
¹ Pricing model used: 3-year, no upfront RI

² Software Assurance and true-up costs provided by Microsoft

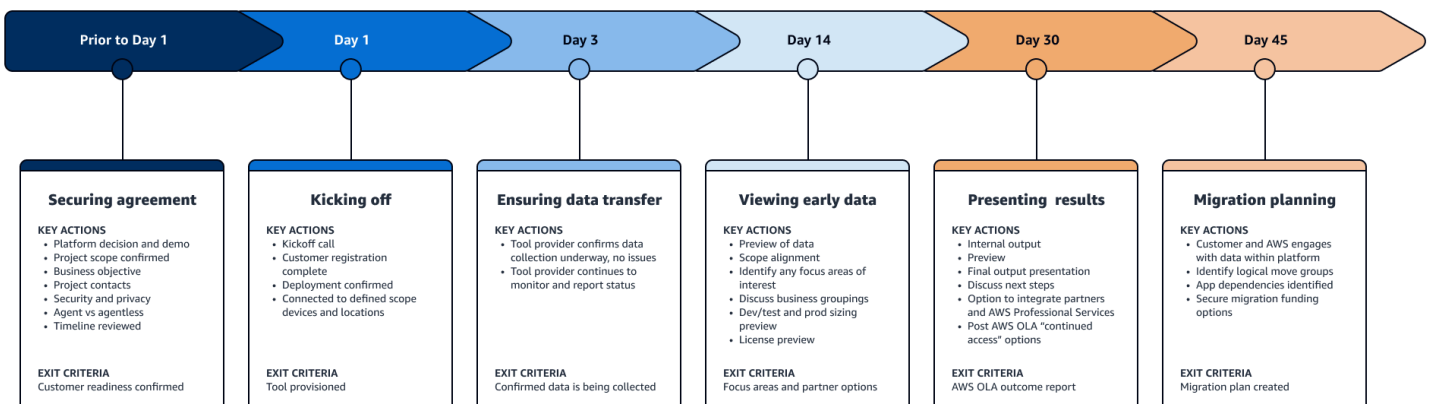
Uma AWS OLA também fornece insights sobre o impacto que a modernização pode ter em suas workloads existentes, fazendo sugestões como as seguintes:

- Mudar para um sistema operacional Linux.
- Adicionar suporte de aplicações para processadores ARM (AWS Graviton).
- Mover as workloads do SQL Server para o Amazon Aurora.
- Remover a garantia de software transferindo as workloads do Windows e do SQL Server para tecnologias de código aberto.

O diagrama a seguir mostra a economia de custos que pode ser obtida por meio de técnicas de modernização, como migrar do Windows para o Linux ou do SQL Server para o Aurora.



O processo completo da AWS OLA leva aproximadamente 45 dias para ser concluído. O diagrama a seguir mostra um exemplo de linha do tempo.



Se você tiver um ambiente VMware puro e puder fornecer resultados do RVTools, poderá reduzir essa linha do tempo para uma semana útil. Além disso, a AWS pode analisar um arquivo simples que inclui dados de ativos e de utilização, como média e pico de CPU e média e pico de RAM.

Avaliação de impacto

O cliente médio normalmente experimenta uma redução de custo de 20 a 30% com o esforço de dimensionamento correto. O dimensionamento correto combina a workload de origem com as instâncias de melhor tamanho da AWS com base nos dados de uso. Esses ajustes de dimensionamento correto não apenas reduzem o custo mensal do ambiente da AWS, mas

frequentemente resultam em economias em outras partes da organização. Por exemplo, um ganho de 20 a 30% no licenciamento do Windows ou do SQL Server pode reduzir o próximo true-up com a Microsoft ou liberar licenciamento para aplicações adicionais de linha de negócio. Geralmente, é na consolidação e no dimensionamento correto das workloads do SQL Server que os ganhos financeiros mais significativos são obtidos.

A AWS pode ajudar você a categorizar os sistemas em buckets de modernização. Alguns sistemas são legados e não são financeiramente viáveis de usar, enquanto outros podem ser modernizados em contêineres ou aplicações sem servidor, em que as economias mais significativas são obtidas. A conversa com sua equipe da AWS passa de tópicos generalizados sobre o que a nuvem permite para discussões mais específicas sobre como e por que workloads específicas devem ser modernizadas. A AWS também ajuda você a analisar possíveis oportunidades de inovação.

Próximas etapas

Se você está iniciando sua jornada de otimização de custos para workloads da Microsoft que estão sendo executadas em ambientes on-premises ou na AWS, interaja com sua equipe da conta da AWS e solicite uma AWS OLA. Os membros da equipe da AWS podem responder às suas perguntas e ajudar a decidir se uma AWS OLA é, em última análise, a escolha certa para você e sua organização. Como alternativa, você pode [solicitar um AWS OLA on-line](#).

Recursos adicionais

- [Avaliação de otimização e licenciamento da AWS](#) (documentação da AWS)
- [AWS re:Invent 2022 - How to save costs and optimize Microsoft workloads on AWS \(ENT205\)](#) (YouTube)

Windows no Amazon EC2

O [Amazon Elastic Compute Cloud \(Amazon EC2\)](#) é uma plataforma de computação em nuvem altamente flexível e escalável, ideal para executar suas workloads do Windows. Você pode usar o Amazon EC2 para implantar, gerenciar e escalar suas workloads do Windows Server na infraestrutura segura, confiável, altamente disponível e adaptável da Nuvem AWS. Considere os seguintes benefícios principais de executar workloads do Windows no Amazon EC2:

- **Escalabilidade:** o Amazon EC2 permite que você escale facilmente suas workloads do Windows para acomodar as mudanças nos requisitos. Você pode criar rapidamente novas instâncias do EC2 para lidar com o aumento da demanda e, com a mesma facilidade, encerrar as instâncias quando elas não forem mais necessárias. Você paga apenas pelos recursos que realmente utilizar.
- **Flexibilidade:** o Windows no Amazon EC2 é compatível com uma ampla variedade de tipos de instância que são projetados para atender a vários requisitos de workload, desde instâncias de uso geral até instâncias otimizadas para computação ou memória. Essa flexibilidade garante que você possa escolher o melhor tipo de instância para suas aplicações específicas baseadas em Windows, maximizando a performance e minimizando os custos.
- **Segurança** — AWS fornece várias camadas de segurança para suas cargas de trabalho do Windows, incluindo firewalls de rede, criptografia de dados e controles de acesso seguro. Isso significa que você pode confiar que suas aplicações e dados estão protegidos, sem deixar de ter controle total sobre suas configurações e definições de segurança.
- **Eficiência de custos** — O modelo de pay-as-you-go preços permite que você pague somente pelos recursos que você usa, eliminando a necessidade de investimentos iniciais em hardware e software. Esse modelo também permite que você otimize seus custos, reduza os gastos de capital e aumente a eficiência operacional. É um modelo de preços ideal para empresas de todos os portes.

Esta seção do guia abrange os seguintes tópicos:

- [Automatizar as programações de parada e inicialização](#)
- [Workloads do Windows do tamanho certo](#)
- [Selecionar o tipo de instância certo para workloads do Windows](#)
- [Trazer licenças para workloads do Windows e SQL Server](#)
- [Otimizar gastos com o Windows no Amazon EC2](#)

- [Monitore os custos usando AWS ferramentas](#)

Automatizar as programações de parada e inicialização

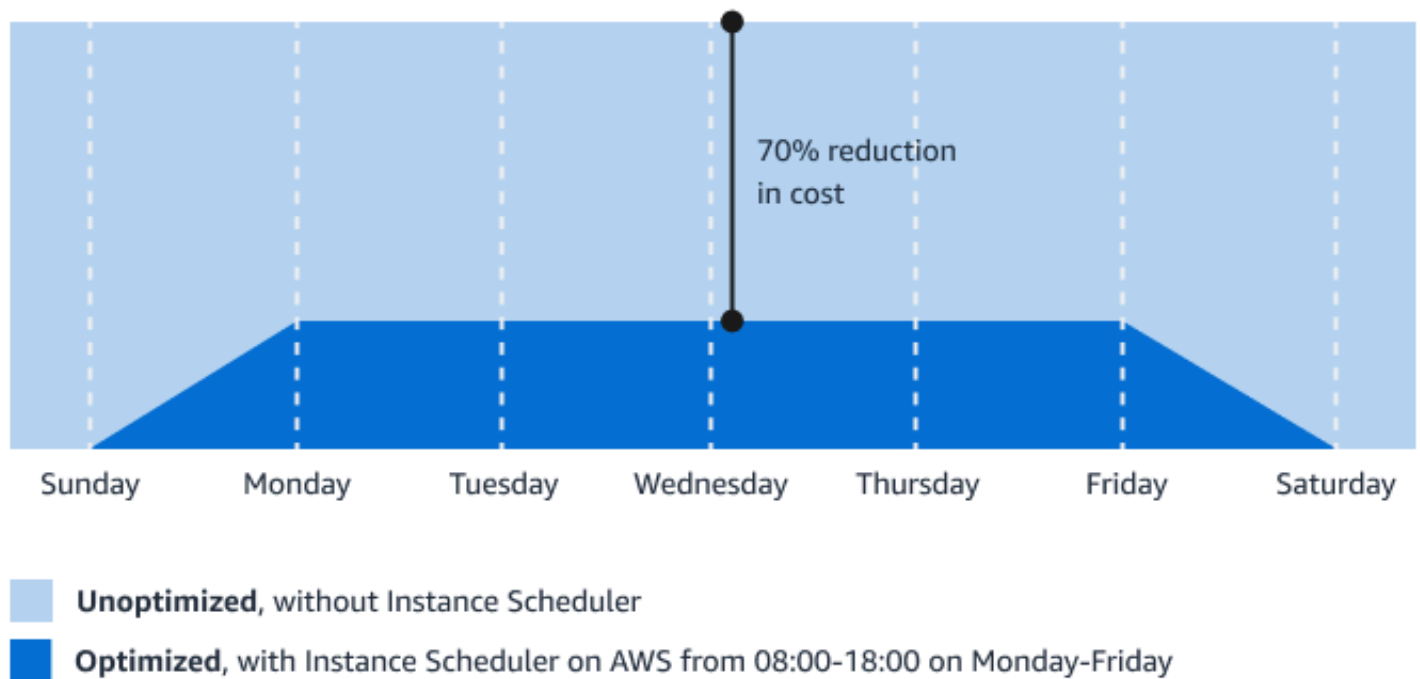
Visão geral do

O [Agendador de Instâncias na AWS](#) pode ajudar você a reduzir os custos operacionais automatizando a inicialização e a interrupção das instâncias do [Amazon EC2](#) e do [Amazon Relational Database Service \(Amazon RDS\)](#). Se você deixar todas as suas instâncias em execução em plena utilização contínua, poderá acabar pagando por recursos que não estão sendo usados. O Agendador de Instâncias ativado AWS permite que você desative as instâncias em momentos em que elas não são necessárias, como fora do horário comercial, fins de semana ou outros períodos em que o uso é baixo. Isso pode gerar uma economia significativa com o tempo.

O Agendador de Instâncias ativado AWS também oferece agendamento de instâncias entre contas, marcação automatizada e a capacidade de configurar horários ou períodos usando uma interface de linha de comando ou a janela de manutenção. [AWS Systems Manager](#) Esses recursos podem ajudar você a gerenciar suas instâncias com mais eficiência e precisão, rastrear e alocar custos em diferentes projetos ou equipes.

Estudos de caso

Considere o exemplo de uma empresa que usa o Instance Scheduler AWS em um ambiente de produção para interromper automaticamente as instâncias fora do horário comercial todos os dias. Se essa empresa deixar todas as suas instâncias em execução em plena utilização, ela poderá obter uma economia de até 70% nas instâncias que só são necessárias durante o horário comercial normal. O gráfico a seguir mostra como a utilização semanal é reduzida de 168 horas para 50 horas.

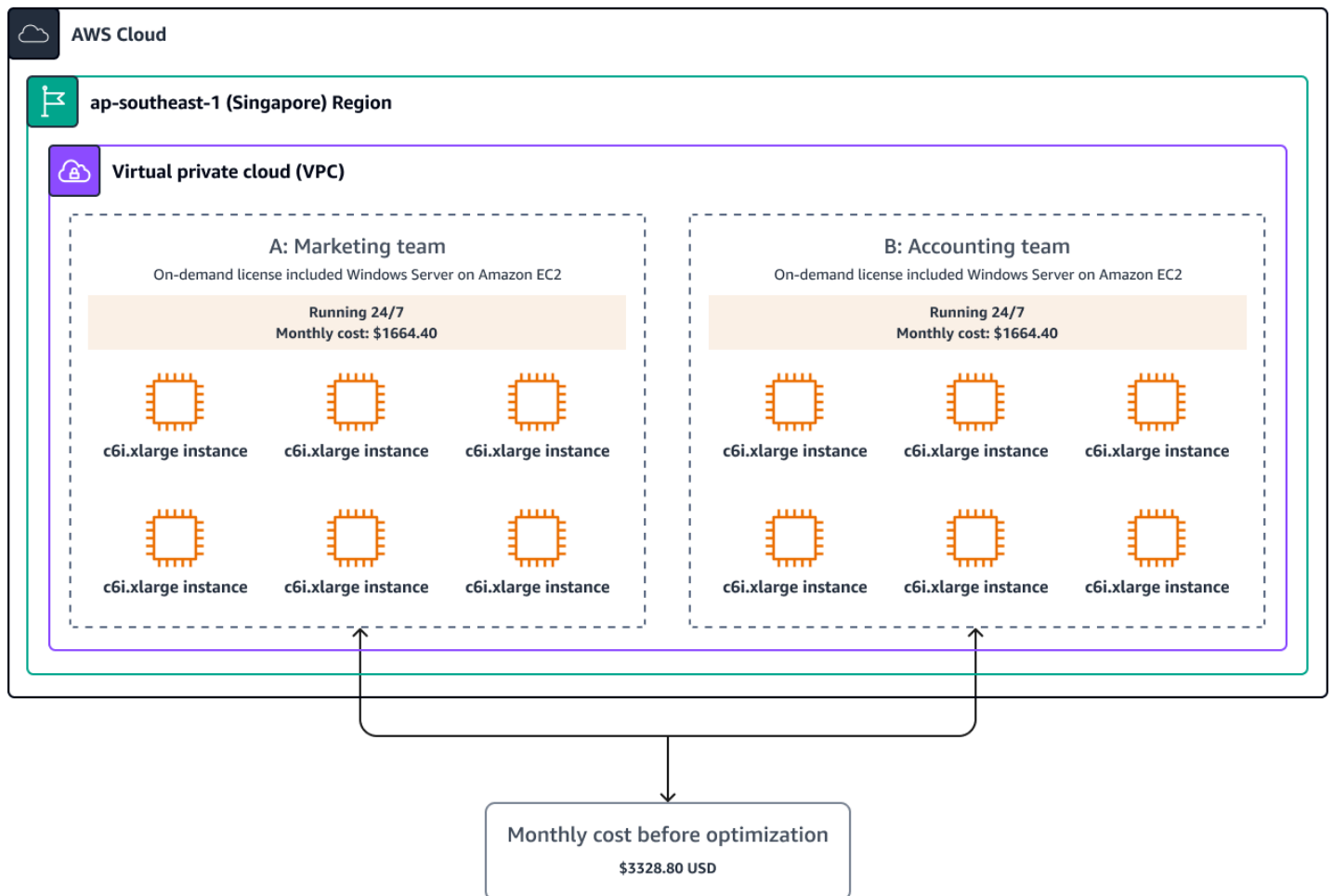


Considere outro exemplo. A empresa de energia elétrica Jamaica Public Service Company Limited (JPS) migrou seu banco de dados para o Amazon RDS. Agora, a JPS usa o Amazon EC2 para hospedar serviços de API e executar outras aplicações. Para o JPS, o Instance Scheduler on AWS se tornou a principal ferramenta para gerenciar ambientes que não são de produção. A JPS usou o Instance Scheduler AWS para reduzir os custos de desenvolvimento e gerenciar instâncias do EC2 com base nas necessidades da equipe e nos horários de trabalho. Isso ajudou a JPS a reduzir os custos em 40%. Para obter mais informações, consulte o estudo de AWS caso O [serviço público da Jamaica migra com eficiência para a nuvem e reduz os custos em 40% usando AWS](#) o Agendador de Instâncias.

Cenário de otimização de custos

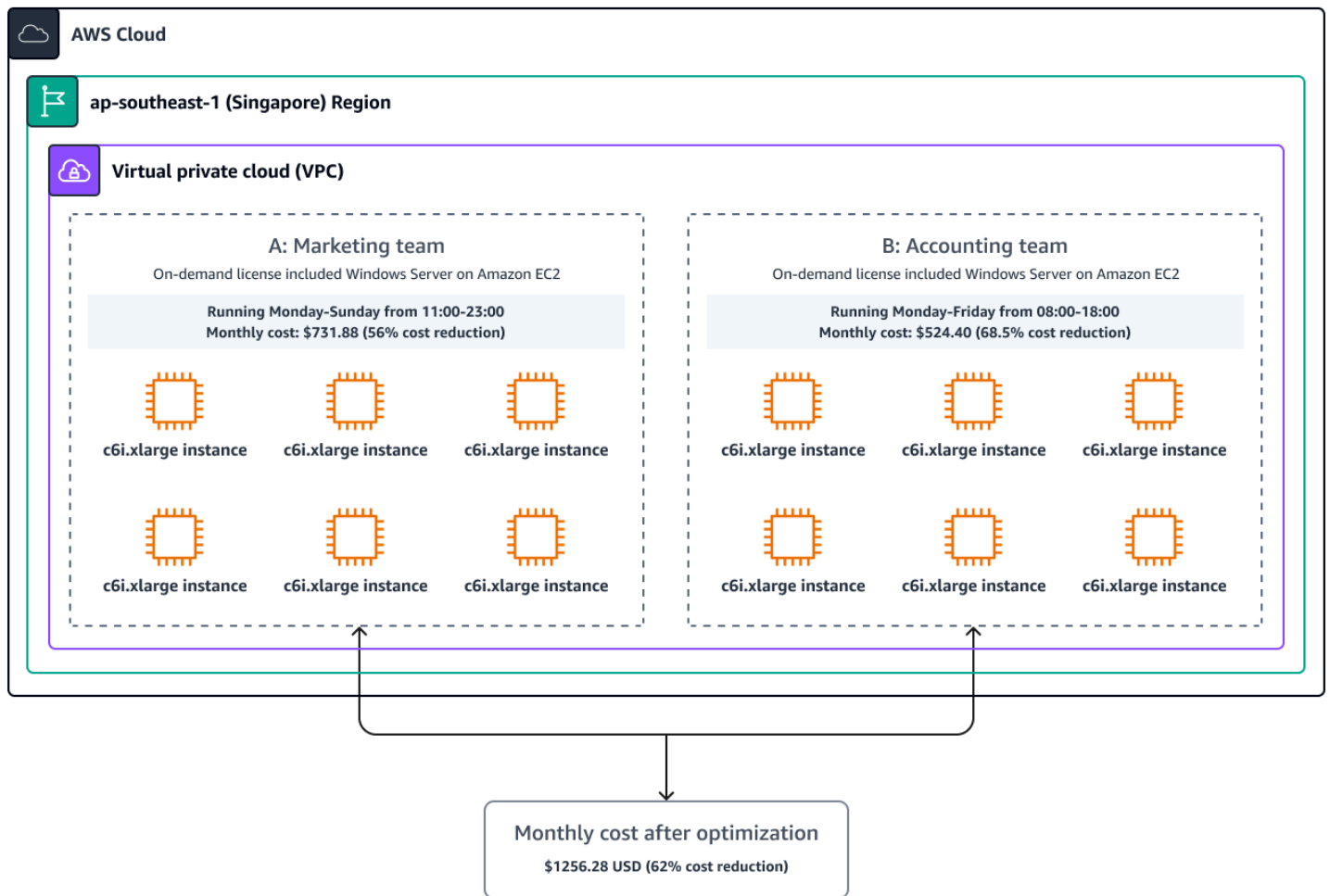
O cenário de exemplo a seguir ajuda a ilustrar as vantagens de custo de usar o Instance Scheduler no. AWS Nesse cenário, uma grande empresa de varejo em Singapura implanta dois ambientes Windows no Amazon EC2. O primeiro ambiente, conhecido como workload A, é utilizado pela equipe de marketing para analisar as transações em tempo real na loja enquanto as lojas estão abertas. O segundo ambiente, conhecido como workload B, é reservado para a equipe de contabilidade, que trabalha somente durante o horário comercial normal. O cronograma operacional atual de ambos os ambientes (24 horas por dia, 7 dias por semana) não é ideal, considerando os padrões de uso atuais, e requer otimização para reduzir os custos operacionais da empresa.

O diagrama a seguir mostra o custo mensal antes da otimização.



Por exemplo, há 31 dias no mês de março, dos quais 23 são dias úteis. Se a equipe de marketing usar o Instance Scheduler AWS e operar suas instâncias somente quando necessário (ou seja, por 321 horas por mês em vez de 730 horas por mês), ela poderá economizar 932,52 USD por mês. Isso equivale a uma redução de 56% nos custos operacionais. A equipe de contabilidade também pode obter vantagens significativas, com o tempo de uso da instância caindo de 730 horas para 230 horas por mês. Isso resulta em uma redução de USD 1.140, ou 68,5%. A empresa poderá economizar um total combinado de USD 2.072,52 por mês (o equivalente a uma redução de 62%) ou USD 24.870,24 por ano.

O diagrama a seguir mostra o custo mensal após a otimização.



Note

O preço desse exemplo foi determinado usando a [AWS Calculadora de Preços](#) em março de 2023.

Recomendações de otimização de custos

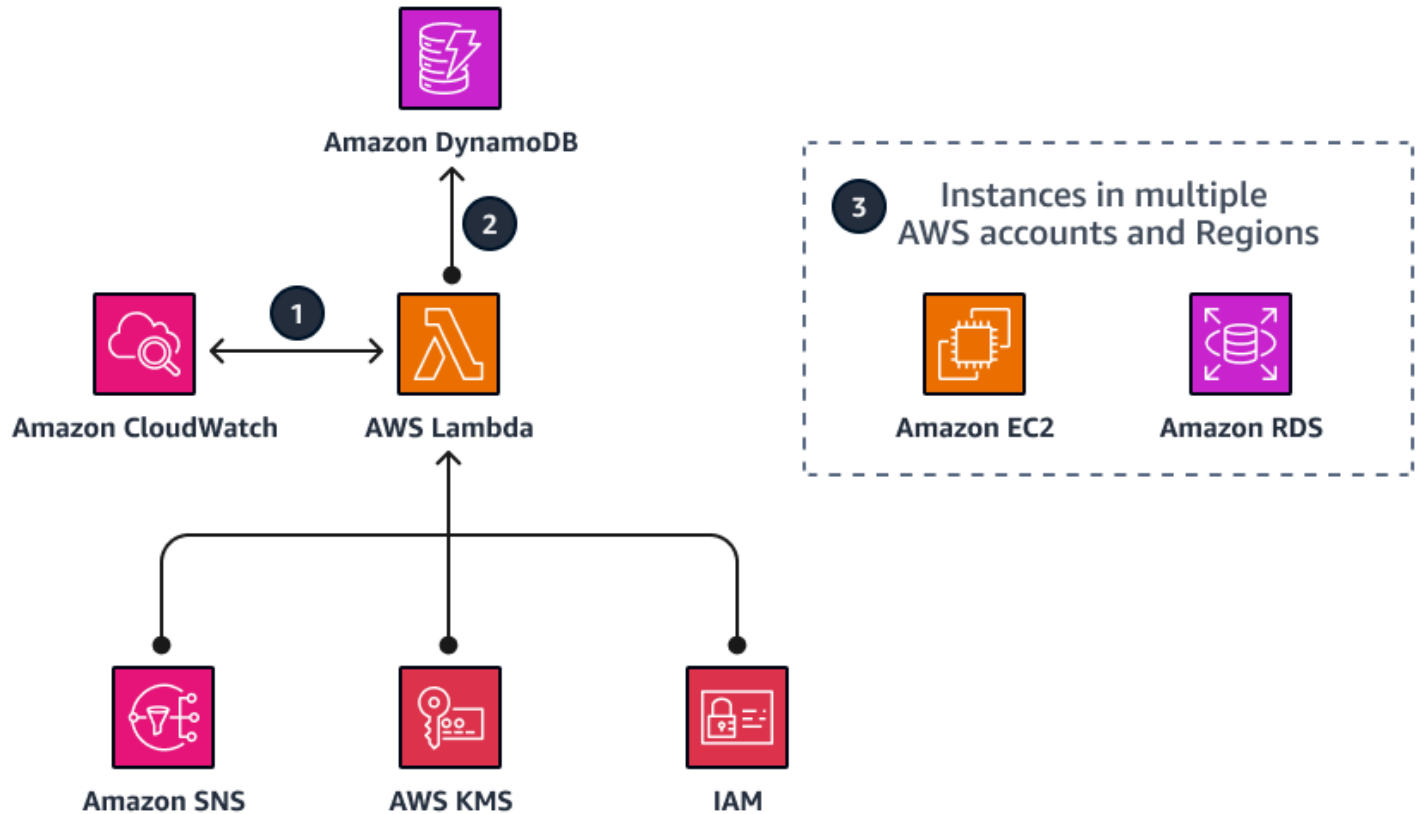
Esta seção explica como implantar e configurar o Agendador de Instâncias na AWS com base no cenário de exemplo abordado na seção anterior Cenário de otimização de custos. Recomendamos que você execute as próximas etapas a seguir para otimizar seus custos usando o Agendador de Instâncias em AWS:

1. Inicializar a pilha do Agendador de Instâncias
2. Configurar períodos

3. Configurar as programações

4. Marcar instâncias de

O diagrama de arquitetura a seguir mostra o que é criado na pilha do Instance Scheduler. Nuvem AWS



O diagrama mostra as seguintes etapas do fluxo de trabalho:

1. Um AWS CloudFormation modelo configura um CloudWatch evento da Amazon em um intervalo definido por você. Esse evento invoca uma AWS Lambda função. Durante a configuração, você define Regiões da AWS as contas e. Você também define uma tag personalizada que o Instance Scheduler on AWS usa para associar agendamentos a instâncias aplicáveis do Amazon EC2, instâncias do Amazon RDS e clusters.
2. Os valores de configuração da programação são armazenados no Amazon DynamoDB, e a função do Lambda os recupera sempre que é executada. Em seguida, você pode aplicar a tag personalizada às instâncias aplicáveis.
3. Durante a configuração inicial do Agendador de Instâncias, você define uma chave de tag para identificar as instâncias aplicáveis do Amazon EC2 e do Amazon RDS. Quando você cria

uma programação, o nome que você especifica é usado como o valor da tag que identifica a programação que você deseja aplicar ao recurso marcado.

Inicializar a pilha do Agendador de Instâncias

Esta seção mostra como iniciar a CloudFormation pilha do Agendador de Instâncias em AWS

Note

Você é responsável pelo custo do Serviços da AWS usado ao executar o Instance Scheduler no AWS. A partir de janeiro de 2023, o custo de execução dessa solução com configurações padrão na região us-east-1 é de aproximadamente USD 9,90 por mês para cobranças do Lambda, ou menos se você tiver um crédito de uso mensal de nível gratuito do Lambda. Para obter mais informações, consulte a seção Custo do [Agendador de Instâncias no Guia de AWS Implementação](#) na Biblioteca de AWS Soluções.

Para executar a pilha do Agendador de Instâncias, conclua as etapas a seguir.

1. Faça login no [Console de gerenciamento da AWS](#) e escolha [Launch solution](#) (modelo disponível para download) para iniciar o `instance-scheduler-on-aws.template` CloudFormation modelo.

Note

Também é possível [fazer download do modelo](#) para usá-lo como ponto de partida para a sua própria implantação.

2. Por padrão, esse modelo é iniciado na região Leste dos EUA (Norte da Virgínia). Para iniciar o Agendador de Instâncias em outra região, use o seletor de região na barra de navegação do console.

Note

O exemplo usa a região Ásia-Pacífico (Singapura).

3. Na página Criar pilha, na seção Pré-requisito - Preparar modelo, verifique se a opção Modelo está pronto está selecionada. Na seção Origem do modelo, verifique se a opção de URL do Amazon S3 está selecionada.
4. Verifique se o URL correto do modelo aparece na caixa de texto URL do Amazon S3 e escolha Avançar.
5. Na página Especificar detalhes da pilha, atribua um nome para a sua pilha de soluções. Para obter informações sobre limitações de nomenclatura de caracteres, consulte [Limites do IAM e do STS](#) na documentação AWS Identity and Access Management (IAM). O nome da pilha para o exemplo deste guia é denominado MyInstanceScheduler.

Note

O nome da pilha não pode ter mais de 28 caracteres.

6. Em Parâmetros, revise os parâmetros do modelo e modifique-os conforme necessário.
7. Escolha Avançar. Na página Configurar opções de pilha, selecione Avançar.
8. Na página Revisar, verifique e confirme as configurações. Marque a caixa de seleção confirmando que o modelo criará recursos do IAM.
9. Selecione Criar para implantar a stack.

Configurar períodos

Depois de implantar o CloudFormation modelo, a solução cria uma tabela do DynamoDB que contém exemplos de regras e cronogramas de períodos que você pode usar como referência para criar suas próprias regras e cronogramas de períodos personalizados. Para ver um exemplo de configuração de período, consulte [Sample schedules](#) na documentação do Agendador de Instâncias na AWS .


Para concluir a etapa desse cenário, você deve gerar períodos que correspondam a cada workload e atendam às suas necessidades específicas. Por exemplo:

```
Period 1 (Workload A):  
  Name: retail-hours  
  Days: Monday to Sunday  
  Hours: 1100 - 2300  
Period 2 (Workload B):  
  Name: office-hours  
  Days: Monday to Friday
```

Hours: 0800 - 1800

Para configurar os períodos, siga as seguintes etapas:

1. Faça login no console do [DynamoDB](#) e verifique se você está na mesma região em que lançou o modelo para CloudFormation o Instance Scheduler. AWS
2. No painel de navegação, escolha Tabelas e selecione a tabela chamada ConfigTable.
3. Escolha Explorar itens da tabela.
4. Para criar um período para o horário comercial, selecione o período para o item horário comercial.
5. Na página Editar item, altere o valor de begintime para 0800 e de endtime para 1800. Mantenha o valor padrão em vigor para os dias da semana.

 Note

Os valores de begintime e de endtime determinam quando as instâncias devem ser inicializadas e interrompidas, enquanto o valor de weekdays determina a quais dias da semana essa programação se aplica (de segunda a sexta-feira, neste exemplo).

6. Escolha Salvar alterações.
7. Para duplicar o período office-hours e usá-lo para criar um novo período para o horário de varejo, selecione period para o item office-hours. Em seguida, no menu Ações, escolha Duplicar item.
8. Modifique os atributos para atender às suas necessidades. Os atributos a seguir são usados para atender aos requisitos do cenário de exemplo:

```
type: period
name: retail-hours
begintime: 11:00
description: Retail hours
endtime: 23:00
weekdays: mon-sun
```

9. Selecione Create Item (Criar item).
10. No ConfigTableDynamoDB, identifique os dois períodos que você acabou de criar listados nas listas de itens.

Configurar as programações

No contexto do Instance Scheduler on AWS, os cronogramas se referem à aplicação de um ou mais períodos e ao fuso horário relevante. Essas programações são então atribuídas às suas instâncias como tags. Esta seção mostra como criar duas programações (mostradas abaixo) para acomodar os padrões de tempo variáveis dos dois exemplos de workloads e, em seguida, associar as agendas aos períodos que você criou na seção anterior.

```
Schedule 1:  
  Name: singapore-office-hours  
  Period: office-hours  
  Timezone: Asia/Singapore  
Schedule 2:  
  Name: singapore-retail-hours  
  Period: retail-hours  
  Timezone: Asia/Singapore
```

Para criar e configurar as programações, conclua as seguintes etapas:

1. Faça login no console do [DynamoDB](#) e verifique se você está na mesma região em que lançou o modelo para CloudFormation o Instance Scheduler. AWS
2. No painel de navegação, escolha Tabelas e selecione a tabela chamada ConfigTable.
3. Escolha Explorar itens da tabela.
4. Para duplicar a programação do horário comercial do Reino Unido e usá-la para criar uma nova agenda para seu horário de expediente (horário comercial de Cingapura, por exemplo), selecione a programação para o uk-office-hours item. Em seguida, no menu Ações, escolha Duplicar item.
5. Modifique os atributos para atender às suas necessidades. Os atributos a seguir são usados para atender aos requisitos do cenário de exemplo:

```
type: schedule  
name: singapore-office-hours  
description: Office hours in Singapore  
periods: office-hours  
timezone: Asia/Singapore
```

6. Selecione Create Item (Criar item).
7. Repita as etapas de 4 a 6 para criar uma programação para o horário de varejo de Singapura usando os seguintes valores de atributos:

```
type: schedule
name: singapore-retail-hours
description: Retail hours in Singapore
periods: retail-hours
timezone: Asia/Singapore
```

8. No ConfigTableDynamoDB, identifique as duas programações e os dois períodos que você criou.

Marcar instâncias de

Depois de estabelecer suas programações, você deve usar tags para alocá-las às instâncias específicas que você deseja usar. Você pode usar o editor de tags no [AWS Resource Groups](#) para gerar e atribuir tags às suas instâncias do Amazon EC2.

1. Faça login no [Console de gerenciamento da AWS](#) e verifique se você está na mesma região em que lançou o CloudFormation modelo anteriormente.
2. Abra o [console do Resource Groups](#). No painel de navegação, expanda Marcação e escolha Editor de tags.
3. Na seção Localizar recursos para marcar, em Regiões, escolha suas regiões. Em Tipos de recursos, escolha Amazon EC2 ou Amazon RDS. Esse cenário se concentra nas instâncias do Amazon EC2 na workload A. A equipe de marketing está usando a workload A na região de Singapura. Os recursos dessa workload já estão marcados com uma chave Department e um valor Marketing. Você pode usar essa tag para pesquisar as instâncias.
4. Escolha Recursos de pesquisa.
5. Selecione as instâncias que você deseja incluir na programação na lista de resultados da pesquisa e escolha Gerenciar tags dos recursos selecionados.
6. Na seção Editar tags de todos os recursos selecionados, escolha Adicionar tag para adicionar as tags de programação do Agendador de Instâncias às suas instâncias do EC2. Você pode usar as chaves e os valores de tags que correspondem às programações (criadas anteriormente no DynamoDB).
7. Em Chave de tag, adicione Programação. Em Valor da etiqueta, insira singapore-retail-hours.
8. Escolha Revisar e aplicar alterações de tag.
9. Para aplicar a tag a todas as instâncias do EC2 que você selecionou, escolha Aplicar alterações a todas as selecionadas.
- 10 Repita as etapas de 3 a 9 para qualquer programação adicional que você queira aplicar.

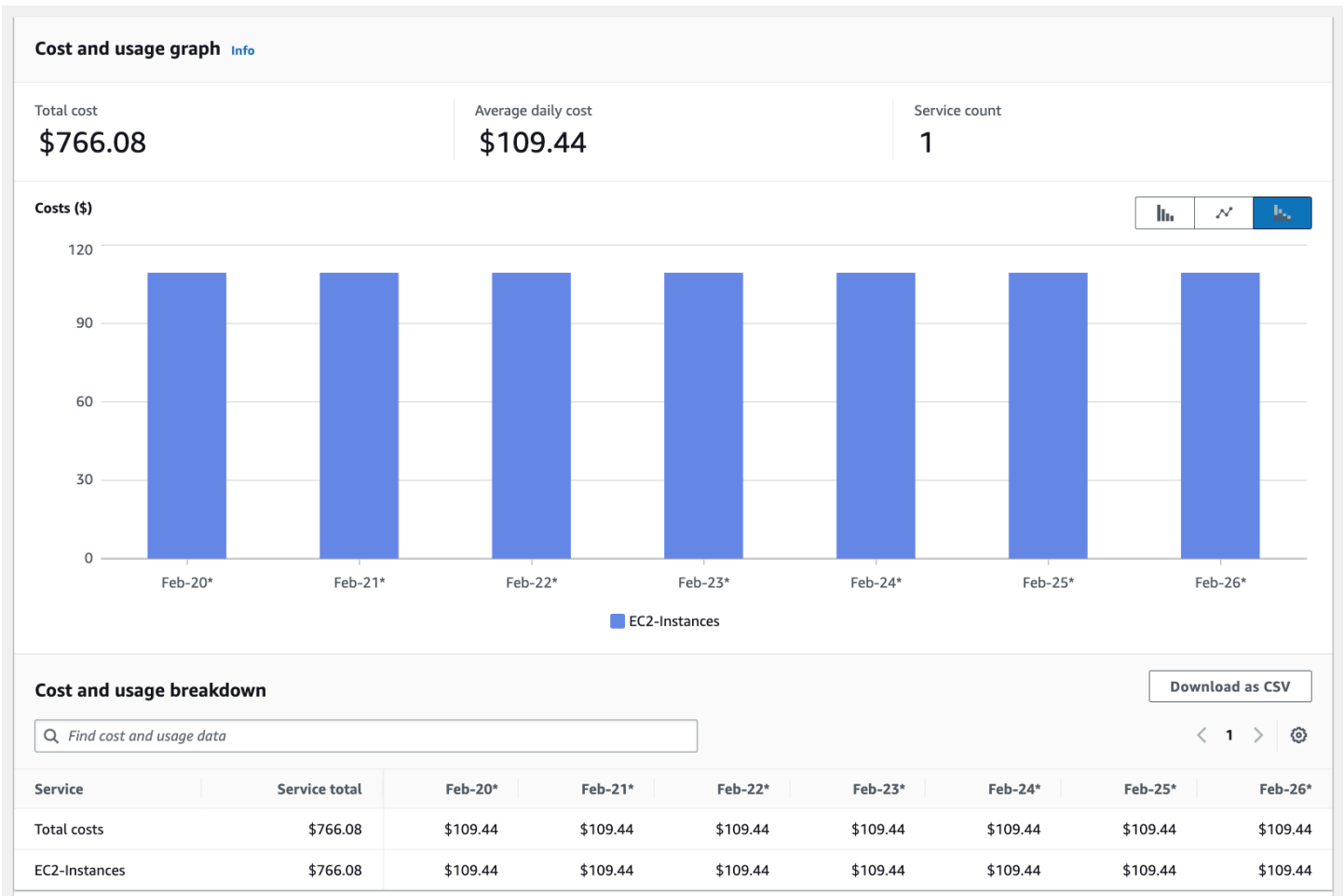
Resultados da validação

Recomendamos que você use [AWS Cost Explorer](#) para medir os benefícios de custo de usar o Instance Scheduler on AWS. É possível usar o Cost Explorer para fazer o seguinte:

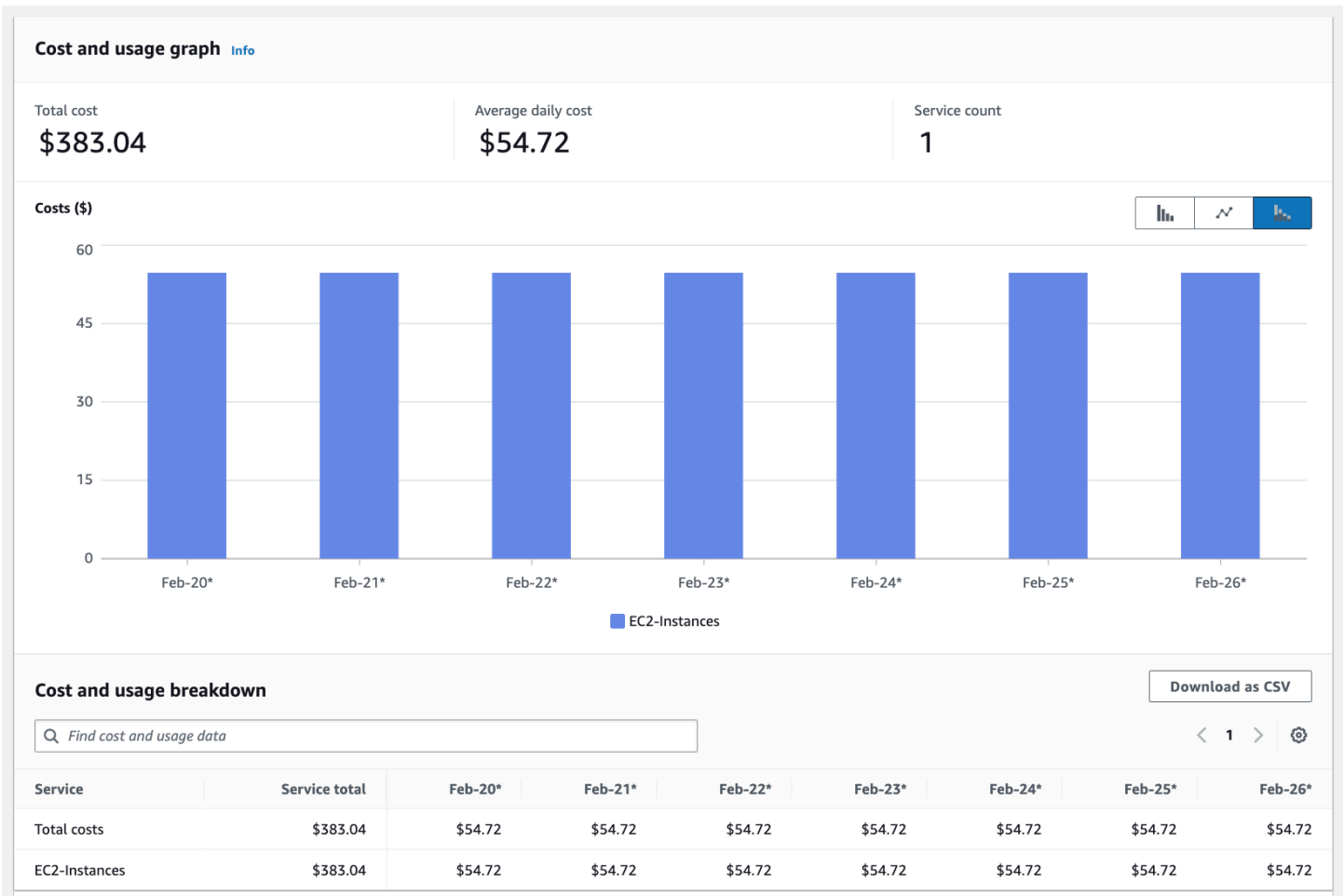
- Visualizar e analisar os custos associados às suas instâncias do EC2, incluindo instâncias gerenciadas pelo Agendador de Instâncias.
- Filtrar sua visualização do Cost Explorer por tags para que você possa se concentrar em workloads específicas e obter uma visão granular da economia de custos obtida com o uso do Agendador de Instâncias.
- Obter insights sobre o impacto financeiro de usar o Agendador de Instâncias.
- Identifique oportunidades para maior otimização de custos e tome decisões baseadas em dados para otimizar seus AWS gastos.

Os gráficos a seguir ilustram o custo operacional da workload A e da workload B durante um período de sete dias (de segunda a domingo) antes da otimização usando o Agendador de Instâncias.

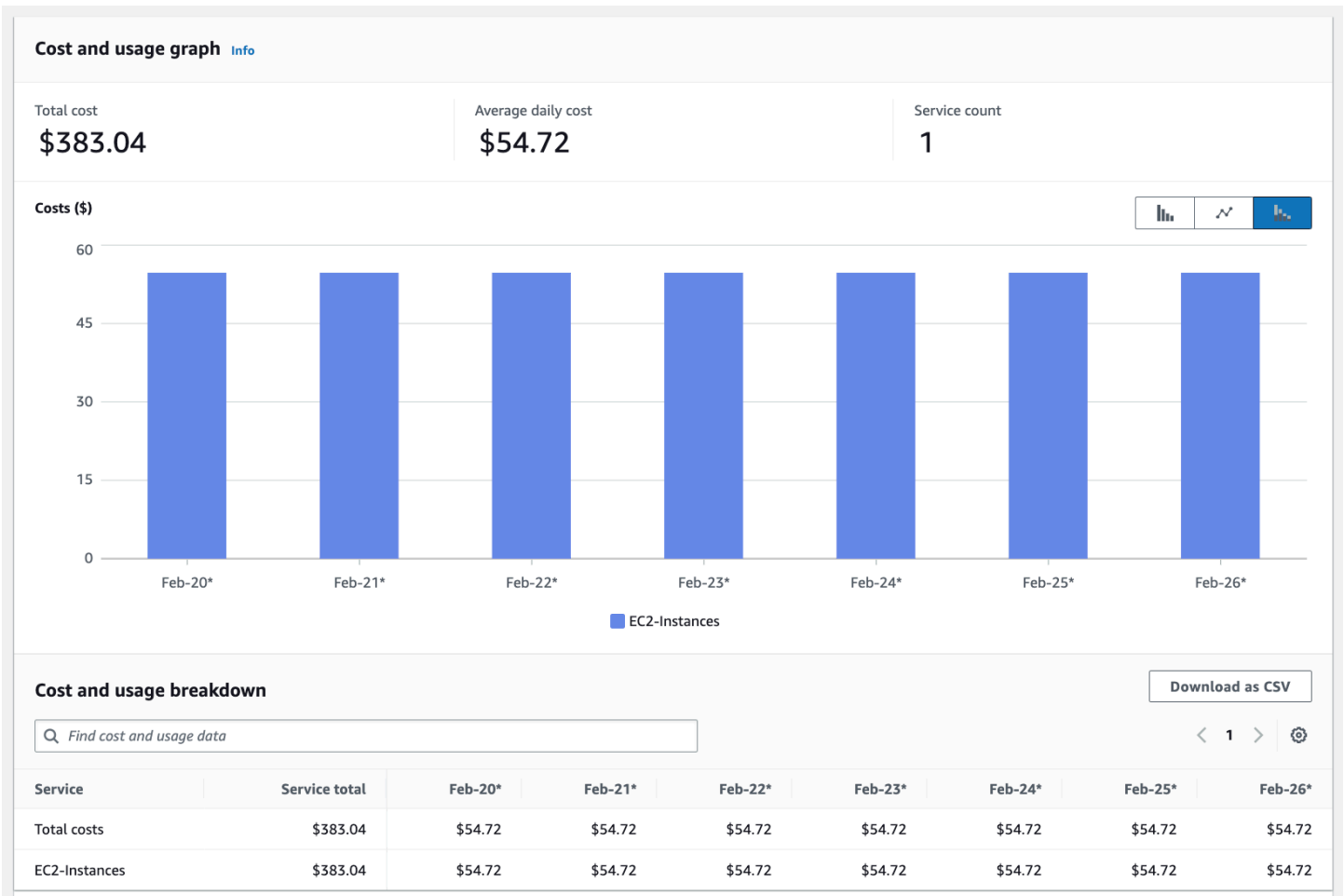
Despesas totais combinadas das workloads A e B



Despesas da workload A

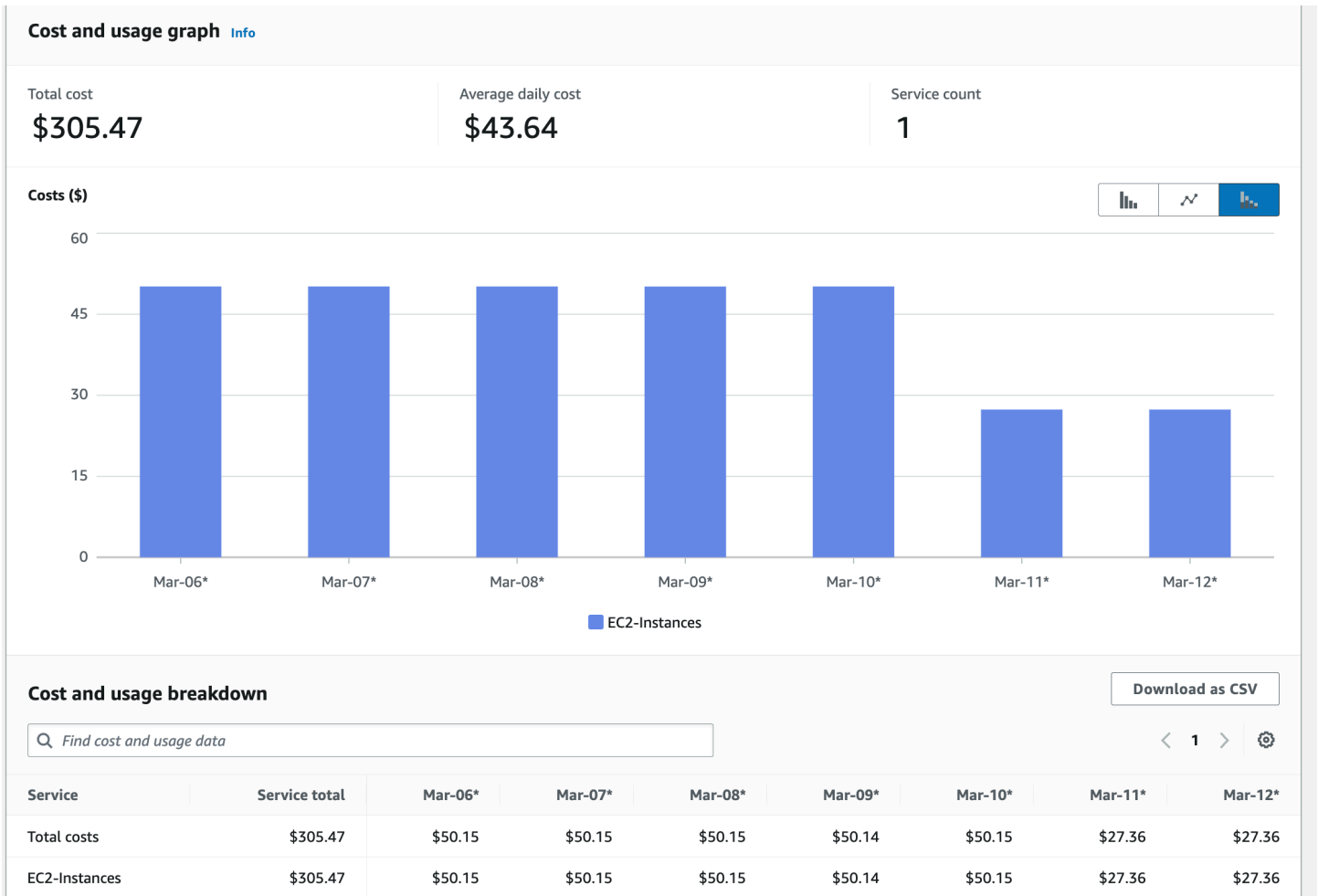


Despesas da workload B

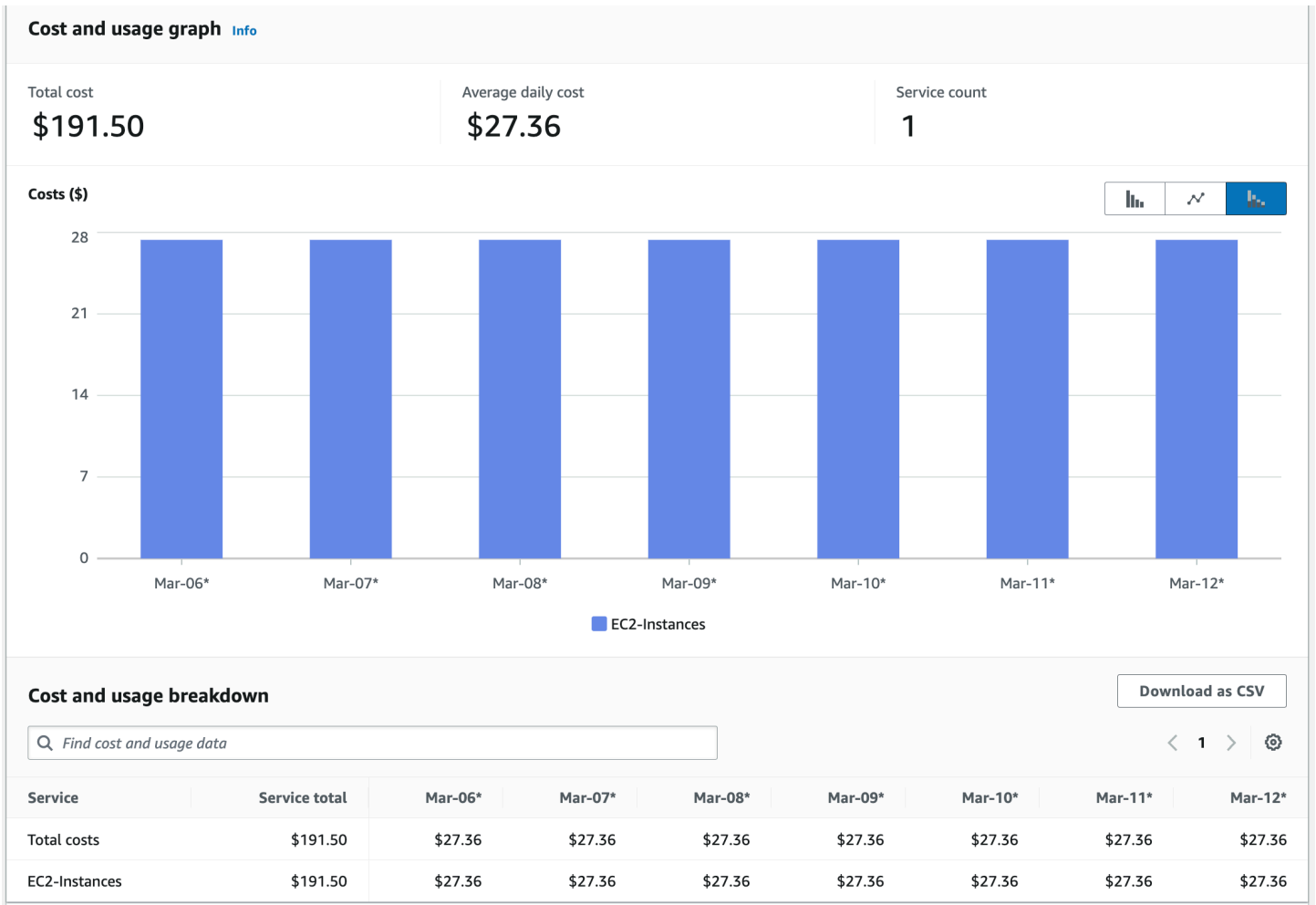


Nesse cenário, o Cost Explorer mostra as reduções de custo resultantes da implementação do Agendador de Instâncias na AWS. Os gráficos a seguir mostram os custos operacionais da workload A e da workload B por um período de sete dias (segunda a domingo) após a otimização.

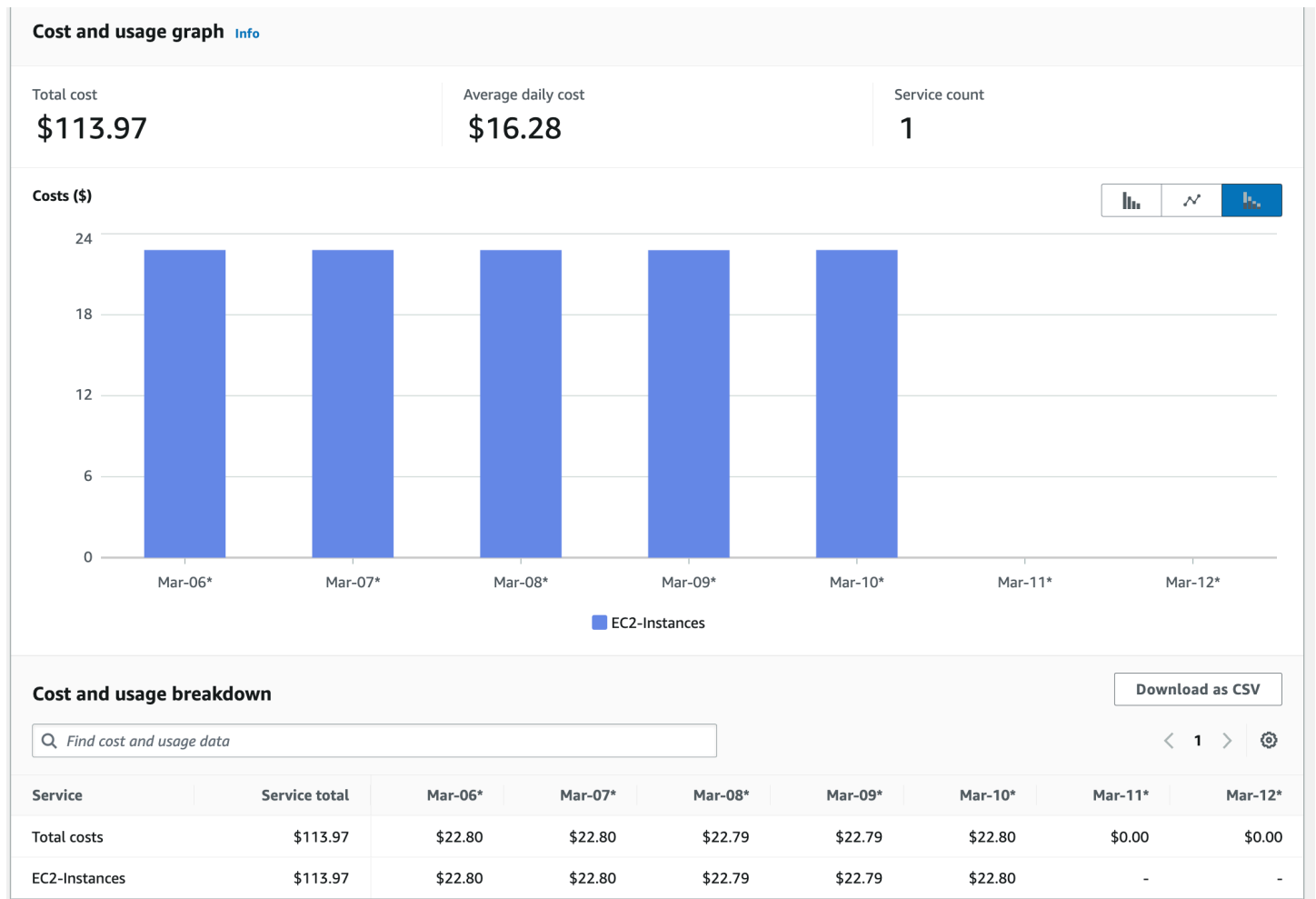
Despesas totais combinadas das workloads A e B



Despesas da workload A



Despesas da workload B



Recursos adicionais do

- [Automatize a inicialização e a interrupção de AWS instâncias](#) (Agendador de instâncias na AWS documentação)
- [De volta ao básico: Usando um programador de instâncias para controlar os custos de recursos do Amazon EC2 e do Amazon RDS](#) () YouTube
- [Marcando seus AWS recursos](#) (Guia do usuário de AWS recursos de marcação)
- [Analisando seus custos com AWS Cost Explorer](#) (Gerenciamento de Faturamento e Custos da AWS documentação)

Workloads do Windows do tamanho certo

Visão geral do

O dimensionamento correto é uma das ferramentas mais potentes para economizar custos. AWS oferece vários métodos para coletar informações de dimensionamento correto, desde a análise de possíveis cargas de trabalho usando uma [Avaliação de AWS Otimização e Licenciamento \(AWS OLA\)](#) até a revisão de cargas de trabalho existentes usando [AWS Cost Explorer](#)

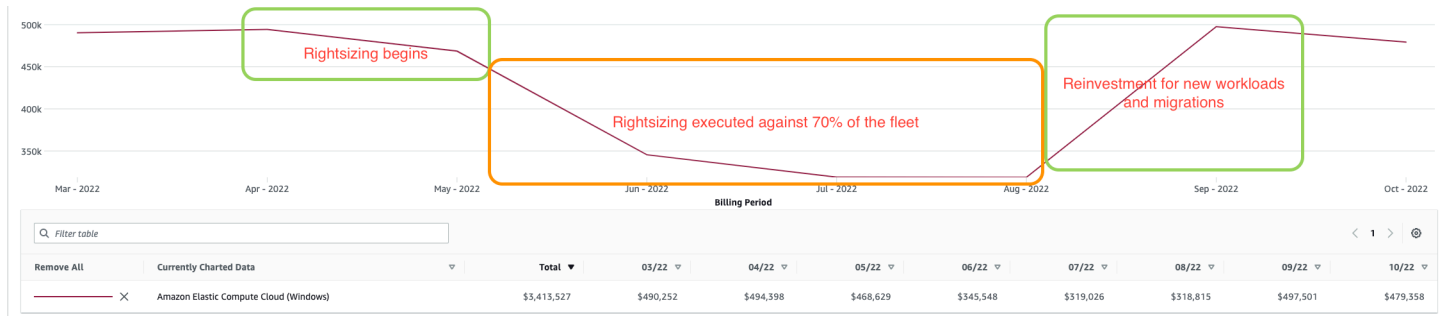
Esta seção mostra como usar o [AWS Compute Optimizer](#) para identificar oportunidades de dimensionamento correto do Amazon EC2. O Compute Optimizer ajuda a evitar o provisionamento excessivo e o subprovisionamento dos seguintes tipos de recursos: AWS

- Tipos de instâncias do [Amazon Elastic Compute Cloud \(Amazon EC2\)](#)
- [Volumes do Amazon Elastic Block Store \(Amazon EBS\)](#)
- Serviços [do Amazon Elastic Container Service \(Amazon ECS\)](#) em AWS Fargate
- [AWS Lambda funções com base nos dados de utilização fornecidos pela Amazon CloudWatch](#)

Cenário de otimização de custos

Avaliar a eficácia do dimensionamento correto pode ser um desafio, pois os esforços de dimensionamento correto podem ser direcionados a uma aplicação específica, a uma equipe ou a toda a organização. Por exemplo, considere uma organização que migra milhares de instâncias para AWS, com 90% de sua frota composta por cargas de trabalho do Windows. A organização pode usar o Compute Optimizer para analisar sua frota e descobrir um superprovisionamento significativo em suas contas e Regiões da AWS. Depois, ela pode usar o [AWS Systems Manager Automation](#) para dimensionar corretamente sua frota por meio de várias janelas de manutenção. Como resultado, a organização consegue ajustar o tipo de instância do tamanho certo para 70% de sua frota e obtém uma economia de custos de 35%.

O painel a seguir ilustra a economia obtida ao longo de vários meses quando esse exemplo de organização implementou estrategicamente as recomendações de dimensionamento correto do Compute Optimizer. O objetivo era operar suas workloads existentes da forma mais eficiente possível, a fim de retomar uma migração paralisada de um data center de colocalização próximo ao final do contrato.



Recomendações de otimização de custos

Recomendamos que você execute as próximas etapas a seguir para otimizar seus custos usando o Compute Optimizer:

- Habilitar o Compute Optimizer
- Habilitar a coleta de métricas de memória para nós do Windows
- Aplicar as recomendações do Compute Optimizer
- Marcar instâncias para obter o dimensionamento correto
- Ative a etiqueta de alocação de custos para funcionar com ferramentas de AWS faturamento
- Implemente recomendações de dimensionamento corretas com automação AWS Systems Manager
- Considerar métodos alternativos de redimensionamento
- Revisar os custos antes e depois no Cost Explorer

Habilitar o Compute Optimizer

Você pode habilitar o [Compute Optimizer](#) no nível da organização ou de uma única conta no AWS Organizations. A configuração de toda a organização fornece relatórios contínuos para instâncias novas e existentes em toda a sua frota para todas as contas de membros. Isso permite que o dimensionamento correto seja uma atividade recorrente em vez de uma point-in-time atividade.

Nível da organização

Para a maioria das organizações, a forma mais eficiente de usar o Compute Optimizer é no nível organizacional. Isso fornece visibilidade de várias contas e várias regiões em sua organização e centraliza os dados em uma fonte para análise. Para habilitar essa funcionalidade no nível da organização, faça o seguinte:

1. Faça login na sua [conta gerencial do Organizations](#) com um perfil que tenha as [permissões necessárias](#), e opte por incluir todas as contas dessa organização. A organização deve ter [todos os recursos habilitados](#).
2. Depois de habilitar a conta gerencial, você pode entrar na conta, ver todas as outras contas de membros e consultar suas recomendações.

Note

É uma prática recomendada configurar uma [conta de administrador delegado](#) para o Compute Optimizer. Isso permite que você exerça o princípio do privilégio mínimo. Dessa forma, você pode minimizar o acesso à conta gerencial da organização e, ao mesmo tempo, fornecer acesso ao serviço de toda a organização.

Nível de conta única

Se você tem como alvo uma conta com altos custos, mas não tem acesso a ao AWS Organizations, você ainda pode habilitar o Compute Optimizer para essa conta e região. Para saber mais sobre o processo de inclusão, consulte [Getting started with AWS Compute Optimizer](#) na documentação do Compute Optimizer.

Habilitar a coleta de métricas de memória para nós do Windows


As métricas de memória fornecem ao Compute Optimizer as métricas essenciais necessárias para fazer recomendações bem fundamentadas de dimensionamento correto em sua organização. Isso se deve à análise de CPU, memória, rede e armazenamento que é conduzida antes de oferecer uma recomendação.

Para passar métricas de memória das instâncias do Windows EC2 para o Compute Optimizer, você deve habilitar CloudWatch o agente e configurar as métricas de memória a serem coletadas a cada 60 segundos. Não há custo adicional para usar métricas de memória com CloudWatch.

Ative o CloudWatch agente e configure as métricas de memória

Faça o download do [ComputeOptimizearquivo.yml](#). Você pode usar esse arquivo para habilitar a coleta de memória para todas as instâncias da sua conta. O arquivo de modelo gera os seguintes componentes:

- [AWS Systems Manager Armazenamento de parâmetros](#) — armazena a configuração do CloudWatch agente necessária para coletar métricas de memória.
- AWS Identity and Access Management Função (IAM) com [políticas AWSAWS Systems Manager gerenciadas](#) anexadas — Isso é para o documento Systems Manager Automation.
- [AWS Systems Manager documentos](#) — Isso instala e configura o CloudWatch agente (substituindo qualquer CloudWatch configuração existente).
- AWS Systems Manager Associação [do State Manager](#) — Isso permite que os documentos do Systems Manager sejam executados em todas as instâncias da sua conta.

 Important

A execução desse modelo substitui qualquer CloudWatch configuração existente nas instâncias.

Depois, faça o seguinte:

1. Faça login no Console de gerenciamento da AWS e abra o [CloudFormation console](#).
2. No painel de navegação, escolha Pilhas.
3. escolha Criar pilha e, em seguida, escolha Com recursos existentes (importar recursos).
4. Escolha Próximo.
5. Para Origem do modelo, selecione Fazer upload de um arquivo de modelo.
6. Escolha o arquivo e depois faça o upload do arquivo `ComputeOptimize.yaml`.
7. Escolha Próximo.
8. Na página Especificar detalhes da pilha, em Nome da pilha, insira um nome para sua pilha e escolha Avançar.
9. Na página Identificar recursos, insira os valores do identificador dos recursos que você está importando.
- 10 Escolha Importar recursos.
- 11 Depois que a pilha for implantada, escolha a guia Saídas para encontrar a chave, o valor e a descrição da sua associação.

Monitorar o progresso da associação

1. Depois que a implantação da CloudFormation pilha for concluída, abra o [console do Systems Manager](#).
2. No painel de navegação, na seção Gerenciamento de nós, escolha Gerenciador de Estados.
3. Na página Associações, escolha o ID de associação da sua associação.
4. Escolha a guia Execution history (Histórico de execução).
5. Na coluna ID de execução, escolha o ID de execução da sua associação. O status deve ser Success.

Veja as métricas em CloudWatch

Recomendamos que você espere pelo menos cinco minutos para que as métricas sejam preenchidas. CloudWatch

1. Abra o [console do CloudWatch](#).
2. No painel de navegação, expanda a seção Métricas e escolha Todas as métricas.
3. Confirme se as métricas aparecem abaixo do CWAgentnamespace.

Note

Para aplicar as configurações a qualquer nova instância, execute novamente a associação.

Aplicar as recomendações do Compute Optimizer

Considere um exemplo que se concentra em fazer as alterações de dimensionamento correto em uma única conta e região. Neste exemplo, o Compute Optimizer está habilitado no nível da organização em todas as contas. Lembre-se de que o dimensionamento correto é um processo disruptivo que, na maioria dos casos, é executado com precisão pelos proprietários da aplicação durante uma janela de manutenção programada de várias semanas.

Se você navegar até o Compute Optimizer de dentro da conta gerencial de uma organização (conforme mostrado nas etapas a seguir), poderá escolher a conta que deseja investigar. Neste exemplo, há seis instâncias em execução em uma única conta na região us-east-1. Todas as seis instâncias estão superprovisionadas. A meta é redimensionar as instâncias com base nas recomendações do Compute Optimizer.

Identificar instâncias superprovisionadas e exportar os detalhes das recomendações

1. Faça login Console de gerenciamento da AWS e abra o console do [Compute Optimizer](#).
2. No painel de navegação, escolha Painel.
3. Na caixa de pesquisa na página do Painel, insira Region=US East (N. Virginia). Em seguida, insira Findings=Over-provisioned. Esses filtros permitem que você veja todas as instâncias superprovisionadas na região us-east-1.
4. Para analisar as recomendações detalhadas das instâncias superprovisionadas do EC2, role para baixo até o cartão de instâncias do EC2 e escolha Visualizar recomendações.
5. Escolha Exportar e salve o arquivo para uso futuro.
6. Em bucket do S3, insira o nome do bucket do Amazon S3 que você deseja que seja o destino do arquivo de exportação.

Note

Para salvar recomendações para análise futura, você deve ter um bucket do S3 disponível para que o Compute Optimizer grave em cada região. Para obter mais informações, consulte [Amazon S3 bucket policy for AWS Compute Optimizer](#) na documentação do Compute Optimizer.

7. Na seção Exportar filtros, marque a caixa de seleção Incluir recomendações para todas as contas de membros na organização.
8. Em Tipo de recurso, escolha Instâncias do EC2.
9. Na seção Colunas a serem incluídas, marque a caixa de seleção Selecionar tudo.
10. Escolha Exportar.

Escolher instâncias com base nas recomendações

As recomendações de instância são baseadas nas métricas de performance coletadas e analisadas pelo Compute Optimizer. É essencial estar ciente das workloads em execução na instância para garantir que você escolha a melhor instância. Este exemplo pressupõe que você possa escolher entre a última geração de instâncias [R6i](#), [R5](#) e [T3](#) do Amazon EC2. As instâncias T3 são expansíveis e têm menos recursos de largura de banda da rede. As instâncias R5 e R6 têm o mesmo custo por hora e são quase idênticas. No entanto, a instância R6 tem uma maior capacidade de largura de banda da rede, apresenta a última geração de processadores Intel e oferece a mesma pegada computacional da R5. Neste exemplo, a R6 é a melhor opção para redimensionar.

1. No [console do Compute Optimizer](#), escolha Recomendações para instâncias do EC2 na barra de navegação. Esta página mostra uma comparação do tipo de instância atual com as opções recomendadas para substituí-la.
2. Para obter o ID da instância que você deseja dimensionar corretamente, abra o [console do Amazon S3](#) na conta gerencial do AWS Organizations.
3. No painel de navegação, escolha Buckets e selecione o bucket que você está usando para armazenar seus resultados exportados.
4. Na guia Objetos, selecione seu arquivo de exportação na lista de objetos e escolha Baixar.
5. Para extrair as informações das instâncias do arquivo, você pode usar o botão Texto em colunas na guia Dados no Microsoft Excel.

Note

IDs As instâncias são representadas como Amazon Resource Names (ARNs). Certifique-se de definir o delimitador como “/” e extrair o ID da instância. Como alternativa, você pode escrever um script ou usar um ambiente de desenvolvimento integrado (IDE) para cortar o ARN.


6. No Excel, filtre a coluna de descoberta para exibir somente as instâncias de OVER_PROVISIONED. Estas são as instâncias que você está visando para o dimensionamento correto.
7. Salve a instância IDs em um editor de texto para facilitar o acesso posterior.

Marcar instâncias para obter o dimensionamento correto

Marcar suas workloads é uma ferramenta poderosa para organizar seus recursos na AWS. As tags permitem que você obtenha uma visibilidade refinada dos custos e facilitem o estorno. Para obter mais informações sobre estratégias e métodos para adicionar tags aos AWS recursos, consulte o AWS whitepaper [Best Practices for AWS Tagging](#) Resources. Neste exemplo, você pode usar o [Editor de tags da AWS](#) para fazer ajustes nas marcações nas instâncias superprovisionadas que você deseja redimensionar durante uma janela de manutenção. Você também pode usar essa tag para visualizar os custos antes e depois da alteração.

1. Faça login Console de gerenciamento da AWS e abra o [AWS Resource Groups console](#) da conta que contém as instâncias destinadas ao redimensionamento.
2. Na barra de navegação, na seção Marcação, escolha Editor de tags.

3. Em Regiões, selecione sua região de destino.
4. Para Tipos de recursos, escolha `AWS::EC2::Instance`.
5. Escolha Recursos de pesquisa.
6. Na página Resultados da pesquisa de recursos, selecione todas as instâncias que você deseja dimensionar corretamente e escolha Gerenciar tags dos recursos selecionados.
7. Escolha Adicionar Tag.
8. Em Chave da tag, insira Rightsizing. Em Valor da tag, insira enabled. Depois escolha Revisar e aplicar alterações da tag.

 Note

Você pode incluir metadados adicionais, como Team ou Business Unit, para ajudar na filtragem posterior no Cost Explorer.

Depois de criar e aplicar tags definidas pelo usuário aos recursos, pode levar até 24 horas para que as tags apareçam na página de tags de alocação de custos para ativação. Depois de selecionar as tags para ativação, pode levar até 24 horas para que as tags sejam ativadas.

Para usuários avançados, você pode usar o [AWS CloudShell](#) da conta e da região de destino para marcar várias instâncias. Por exemplo:

```
bash
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="type-m5"
# Get a list of instance IDs
INSTANCE_IDS=$(aws ec2 describe-instances --query
  "Reservations[].Instances[].InstanceId" --output text)
# Loop through each instance ID and add the tag
for INSTANCE_ID in $INSTANCE_IDS; do
  aws ec2 create-tags --resources $INSTANCE_ID --tags Key=$TAG_KEY,Value=$TAG_VALUE
done
```

Ative a etiqueta de alocação de custos para funcionar com ferramentas de AWS faturamento

Recomendamos ativar a tag de alocação de custos definida pelo usuário. Isso permite que a tag Rightsizing seja reconhecida e filtrável nas ferramentas de AWS cobrança (por exemplo, Cost Explorer e). AWS Cost and Usage Report Se você não habilitar essa funcionalidade, a opção de filtragem de tags e os dados não estarão disponíveis. Para obter informações sobre o uso de tags de alocação de custos, consulte [Activating user-defined cost allocation tags](#) na documentação do Gerenciamento de Faturamento e Custos da AWS .

1. Faça login no Console de gerenciamento da AWS e abra o [AWS Billing console](#).
2. No painel de navegação, na seção Faturamento, escolha Cost Tags de alocação de custo.
3. Na guia Tags de alocação de custos definidas pelo usuário, insira Rightsizing.
4. Selecione a chave da tag Rightsizing e, em seguida, escolha Ativar.

Após 24 horas, a tag deve aparecer no Cost Explorer.

Implementar recomendações de dimensionamento correto com o Systems Manager Automation

O redimensionamento é um cenário que exige que uma instância seja interrompida e inicializada. Nesse cenário, você pode precisar lidar com essa interrupção em uma janela de manutenção e exigir que diferentes equipes lidem com seus próprios redimensionamentos. Antes de alterar um tipo de instância, revise as [considerações sobre os tipos de instância compatíveis](#) na documentação do Amazon EC2.

As etapas de exemplo nesta seção implementam recomendações de dimensionamento correto por conta e região usando um documento do Systems Manager Automation chamado [AWS-ResizeInstance](#). Essa abordagem é típica da maioria das organizações, pois elas exigem tipos de instância diferentes para finalidades distintas. Você também pode usar o mesmo documento de automação AWS-ResizeInstance para direcionar implantações de uma ou várias contas.

1. Faça login no Console de gerenciamento da AWS e abra o [console do Systems Manager](#).
2. No painel de navegação, na seção Recursos compartilhados, escolha Documentos.
3. Na barra de pesquisa, insira AWS- ResizeInstance e escolha AWS- nos resultados ResizeInstance da pesquisa.

4. Escolha `Execute automation`.
5. Na página `Executar runbook de automação`, escolha `Execução simples`.
6. Na seção `Parâmetros de entrada`, insira `InstanceIdInstanceType`. Deixe os demais valores padrão.
7. Escolha `Executar` e aguarde até que a automação siga as etapas para alterar o tipo de instância.

Considerar métodos alternativos de redimensionamento

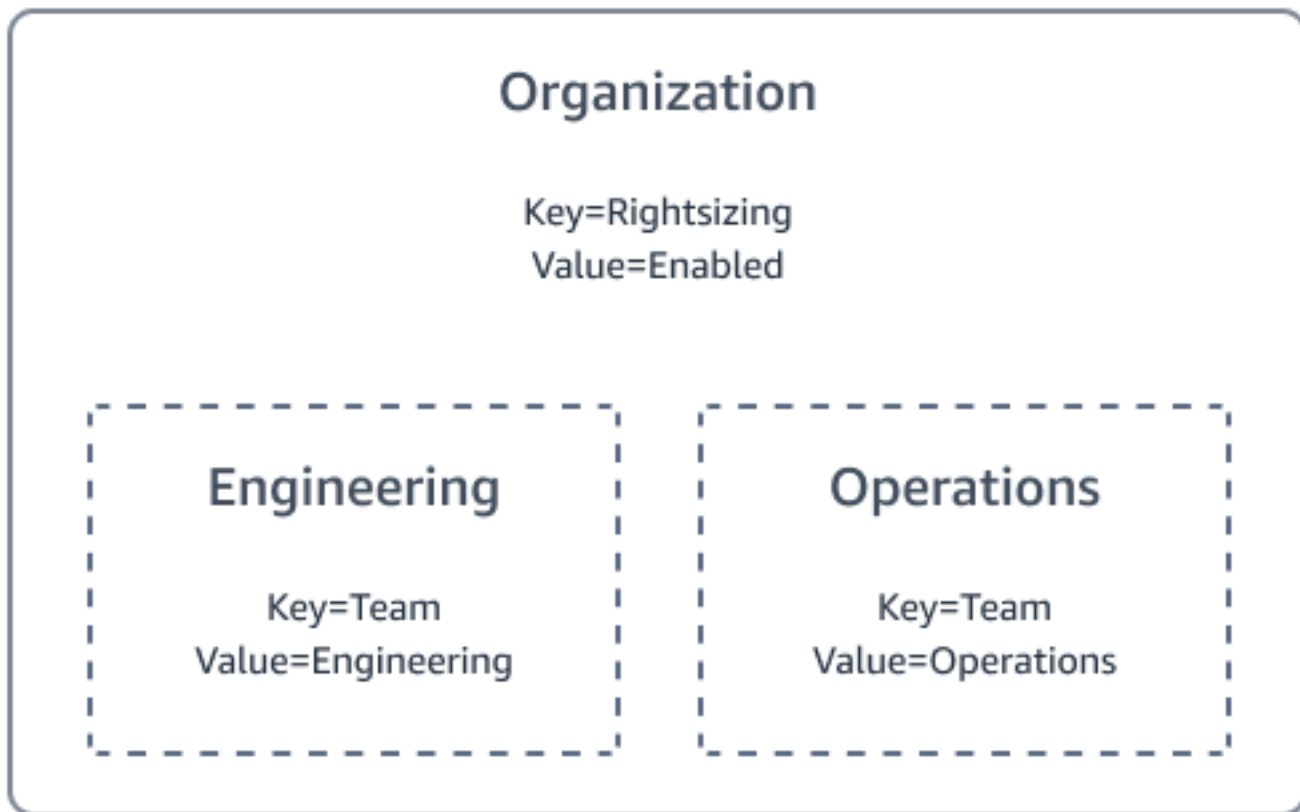
Se você estiver usando um modelo de inicialização para implantar suas instâncias, poderá atualizar esse modelo com o tipo de instância do tamanho certo e, em seguida, realizar uma atualização da instância para substituir as instâncias pela versão do tamanho certo.

Se você planeja usar o processo de dimensionamento correto em várias contas e regiões, deve criar um documento personalizado do `Systems Manager Automation`. Este documento permite alimentar várias instâncias como parâmetro e direcionar instâncias para o mesmo tipo de instância de destino (por exemplo, todas as instâncias em transição para `t3a.medium`, independentemente do tipo de instância de origem).

Revisar os custos antes e depois no Cost Explorer

Depois de dimensionar corretamente seus recursos, você pode usar o `Cost Explorer` para mostrar os custos antes e depois usando a tag `Rightsizing`. Lembre-se de que você pode usar [tags de recursos](#) para controlar os custos. Ao usar várias camadas de tags, você pode obter visibilidade granular de seus custos. No exemplo abordado neste guia, a tag `Rightsizing` é usada para aplicar uma tag genérica a todas as instâncias de destino. Em seguida, uma tag `team` é usada para organizar ainda mais os recursos. A próxima etapa é introduzir tags de aplicação para mostrar ainda mais o impacto nos custos de operar uma aplicação específica.

O diagrama a seguir mostra a estrutura de tags de uma organização.



Considere o exemplo de uma empresa que dimensiona corretamente os servidores web de produção de propriedade da equipe de operações. No Cost Explorer, a tag Rightsizing é definida como enabled, e a tag Team está definida como operations. Neste exemplo, o esforço de dimensionamento correto reduz os custos operacionais de 0,89 centavos para 0,28 centavos por hora. Supondo 744 horas por mês, o custo anual antes do dimensionamento correto é de USD 7.945,92. Após o dimensionamento correto, o custo anual cai para USD 2.499,84. Isso se traduz em uma redução de 68,5% nos custos anuais da workload. Imagine o impacto dessa redução em uma grande organização. Lembre-se de que isso é feito em um ambiente de exemplo e as instâncias estão, em sua maioria, ociosas. Em um ambiente de produção, você pode ver economias entre 10 e 35%.

Agora, considere o impacto do dimensionamento correto do bastion host de produção de propriedade da equipe de engenharia. No Cost Explorer, a tag Rightsizing é definida como enabled, e a tag Team está definida como engineering. Neste exemplo, o esforço de dimensionamento correto reduz os custos operacionais de 0,75 centavos para 0,44 centavos por hora. Supondo 744 horas por mês, o custo anual antes do dimensionamento correto é de USD 6.696,00. Após o dimensionamento correto, o custo anual cai para USD 3.928,32.

Se você usar várias tags, poderá filtrar os dados até obter os detalhes minuciosos de custo. Neste exemplo, a tag Team reduz o ruído para que você possa visualizar o impacto em nível de equipe. Como a tag Rightsizing está habilitada, você também pode filtrar por qualquer instância que tenha essa tag com o valor enabled ou nenhum valor presente. Isso pode fornecer uma visão global de seus esforços de dimensionamento correto, especialmente quando visualizados na conta gerencial (pagador) no nível do Cost Explorer. Essa visualização permite que você veja todas as contas e instâncias.

Considere um exemplo no nível de conta única em que a tag Rightsizing está definida como enabled.

Os custos operacionais caem de USD 1,64 por hora para USD 0,72 por hora. Supondo 744 horas por mês, o custo anual antes do dimensionamento correto é de USD 14.641,92. Após o dimensionamento correto, o custo anual cai para USD 6.428,16. Isso se traduz em uma redução de 56% nos custos de computação dessa conta.

Antes de embarcar em sua jornada do dimensionamento correto, considere o seguinte:

- AWS oferece muitas opções para redução de custos. Isso inclui o [AWS OLA](#), onde AWS analisa suas instâncias locais antes de migrar para o. AWS O AWS OLA também fornece recomendações corretas de dimensionamento e orientação de licenciamento.
- Conclua todo o dimensionamento correto antes de adquirir os [Savings Plans](#). Isso pode ajudar você a evitar aquisições excessivas em seu compromisso com os Savings Plans.

Recomendações

Recomendamos as próximas etapas abaixo:

1. Analise seu cenário atual e considere a conversão de volumes gp2 do Amazon EBS em volumes gp3.
2. Revise os [Savings Plans](#).

Recursos adicionais do

- [AWS Compute Optimizer](#) (documentação da AWS)
- [Melhores práticas para a marcação de AWS recursos](#) (AWS whitepapers)
- [Como coletar dados de AWS Compute Optimizer e AWS Trusted Advisor entre seu AWS Organizations](#) (YouTube)

- [Optimizing performance and reducing licensing costs: Leveraging AWS Compute Optimizer for Amazon EC2 SQL Server instances](#) (Microsoft Workloads no Blog da AWS)

Selecionar o tipo de instância certo para workloads do Windows

Visão geral do

Uma distinção significativa entre workloads operando na nuvem em comparação com ambientes on-premises é a prática de superprovisionamento. Ao comprar hardware físico para uso on-premises, você faz uma despesa de capital que deve durar por um período predeterminado, normalmente de três a cinco anos. Para acomodar o crescimento previsto durante a vida útil do hardware, o hardware é adquirido com mais recursos do que sua workload exige atualmente. Consequentemente, o hardware físico geralmente é superprovisionado, muito além das necessidades de sua workload real.

A tecnologia de máquina virtual (VM) surgiu como um meio eficaz de utilizar recursos de hardware excedentes. Os administradores provisionaram em excesso VMs com v CPUs e RAM, permitindo que o hipervisor gerenciasse o uso de recursos físicos entre servidores ocupados e ociosos alocando recursos não utilizados para cada VM. Durante o gerenciamento VMs, os recursos de vCPU e RAM alocados para cada VM funcionavam mais como governadores de recursos do que como indicadores do uso real. A superalocação de recursos da VM pode facilmente exceder três vezes os recursos computacionais disponíveis.

[O Amazon Elastic Compute Cloud \(Amazon EC2\)](#) evita o VMs provisionamento excessivo no hardware subjacente, pois é desnecessário. A computação em nuvem é uma despesa operacional, não uma despesa de capital, e você paga apenas pelo que usa. Se sua workload exigir mais recursos no futuro, provisione-os quando você realmente precisar deles, em vez de fazê-lo preventivamente.

Há centenas de opções para escolher os [tipos certos de instâncias do Amazon EC2](#). Se você planeja migrar uma carga de trabalho do Windows para a nuvem, AWS oferece um [AWS OLA](#) para ajudá-lo a entender melhor sua carga de trabalho atual e fornecer um exemplo de seu desempenho em AWS. A análise do AWS OLA visa combinar o tipo e o tamanho de instância EC2 adequados ao seu uso real no local.

Se você já tem workloads em execução no Amazon EC2 e busca estratégias de otimização de custos, esta seção do guia ajuda a identificar diferenças entre as instâncias do Amazon EC2 e sua aplicabilidade às workloads típicas do Windows.

Recomendações de otimização de custos

Para otimizar os custos dos tipos de instâncias do EC2, recomendamos:

- Escolher a família certa de instâncias para sua workload
- Entender as variações de preços entre as arquiteturas de processadores
- Entender as diferenças de preço/performance entre as gerações do EC2
- Migrar para instâncias mais novas
- Usar instâncias expansíveis

Escolher a família certa de instâncias para sua workload

É importante escolher a família de instâncias certa para sua workload.

As instâncias do Amazon EC2 são divididas nos seguintes grupos:

- Uso geral
- Otimizadas para computação
- Otimizado para memória
- Computação acelerada
- Otimizada para armazenamento
- Otimizadas para HPC

A maioria das workloads do Windows se encaixam nas seguintes categorias:

- Uso geral
- Otimizadas para computação
- Otimizado para memória

Para simplificar ainda mais, considere uma instância de linha de base do EC2 em cada categoria:

- Otimizada para computação: C6i
- Uso geral: M6i
- Otimizada para memória: R6i

A geração anterior de instâncias do EC2 apresentava pequenas diferenças nos tipos de processadores. Por exemplo, as instâncias otimizadas para computação C5 têm processadores mais rápidos do que as instâncias de uso geral M5 ou as instâncias otimizadas para memória R5. Todas as instâncias de última geração do EC2 (C6i, M6i, R6i, C6a, M6a e R6a) usam o mesmo processador em todas as famílias de instâncias. Como o processador é consistente na última geração de instâncias, a diferença de preço entre as famílias de instâncias agora depende mais da quantidade de RAM. Quanto mais RAM uma instância tiver, mais cara ela será.

O exemplo a seguir ilustra o preço por hora de uma instância de 4 vCPUs baseada em Intel em execução na região us-east-1.

Instância	v CPUs	RAM	Custo por hora
c6i.xlarge	4	8	0,17 US\$
m6i.xlarge	4	16	0,19 US\$
r6i.xlarge	4	32	\$0,25

Note

Os preços são baseados nos preços por hora sob demanda na região us-east-1.

Instâncias expansíveis

Embora seja uma prática recomendada na computação em nuvem desativar recursos computacionais não utilizados para evitar cobranças, nem todas as workloads podem ser desativadas e ativadas sempre que forem necessárias. Algumas workloads permanecem ociosas por longos períodos, mas devem estar acessíveis 24 horas por dia.

As instâncias expansíveis (T3) oferecem uma maneira de manter workloads com picos ou de baixa utilização on-line o dia todo, mantendo os custos de computação baixos. As instâncias expansíveis do EC2 têm uma quantidade máxima de recursos de vCPU que a instância pode usar por breves períodos. Essas instâncias empregam um sistema baseado em [créditos de expansão de CPU](#). Esses créditos são acumulados durante os períodos de ociosidade ao longo do dia. As instâncias intermitentes oferecem vCPU-to-RAM proporções variáveis, o que as torna alternativas para

instâncias otimizadas para computação em alguns casos e para outras instâncias de uso geral em outros.

O exemplo a seguir ilustra o preço por hora de uma instância T3 (ou seja, instância expansível) em execução na região us-east-1.

Instância	v CPUs	RAM (GB)	Custo por hora
t3.nano	2	0,5	\$0,0052
t3.micro	2	1	\$0,0104
t3.small	2	2	\$0,0208
t3.medium	2	4	\$0,0416
t3.large	2	8	\$0,0832
t3.xlarge	4	16	0,164 US\$
t3.2xlarge	8	32	\$0,328

Note

Os preços são baseados nos preços por hora sob demanda na região us-east-1.

Entender as variações de preços entre as arquiteturas de processadores

Os processadores [Intel](#) têm sido o padrão para instâncias do EC2 desde o início. Gerações anteriores de instâncias do EC2, como C5, M5 e R5, não indicam a Intel como a arquitetura do processador (já que era o padrão). As novas gerações de instâncias do EC2, como C6i, M6i e R6i, incluem um “i” para indicar o uso de um processador Intel.

A mudança na anotação da arquitetura do processador se deve à introdução de opções adicionais de processador. O processador mais comparável ao Intel é o [AMD](#) (indicado com um “a”). Os processadores AMD EPYC usam a mesma arquitetura x86 e oferecem performance semelhante aos processadores Intel, mas a um preço mais baixo. Conforme demonstrado nos exemplos de preços a

seguir, as instâncias EC2 do AMD oferecem um desconto de aproximadamente 10% nos custos de computação em comparação com as do Intel.

Instância Intel	Custo por hora	Instância do AMD	Preço	% da diferença
c6i.xlarge	0,17 US\$	c6a.xlarge	\$0,153	10%
m6i.xlarge	\$0,192	m6a.xlarge	\$0,1728	10%
r6i.xlarge	\$0,252	r6a.xlarge	\$0,268	10%

Note

Os preços são baseados nos preços por hora sob demanda na região us-east-1.

A terceira principal opção de arquitetura de processador são os [processadores AWS Graviton](#) (indicados com um “g”) em instâncias do EC2. Projetados por AWS, os processadores Graviton oferecem a melhor relação preço/desempenho no Amazon EC2. Os processadores Graviton atuais não são apenas 20% mais baratos do que seus equivalentes Intel, mas também oferecem um aumento de performance de 20% ou mais. Espera-se que a próxima geração de processadores Graviton amplie ainda mais essa diferença de performance, com testes mostrando um aumento adicional de 25% na performance.

O Windows Server não pode ser executado nos processadores Graviton, que são baseados na arquitetura ARM. Na verdade, o Windows Server opera somente em processadores x86. Embora você não possa obter um aumento de 40% na relação preço/performance usando instâncias baseadas em Graviton para Windows Server, você ainda pode usar processadores Graviton com workloads específicas da Microsoft. Por exemplo, [versões mais recentes do .NET podem ser executadas no Linux](#). Isso significa que essas workloads podem usar processadores ARM e se beneficiar de instâncias Graviton EC2 mais rápidas e acessíveis.

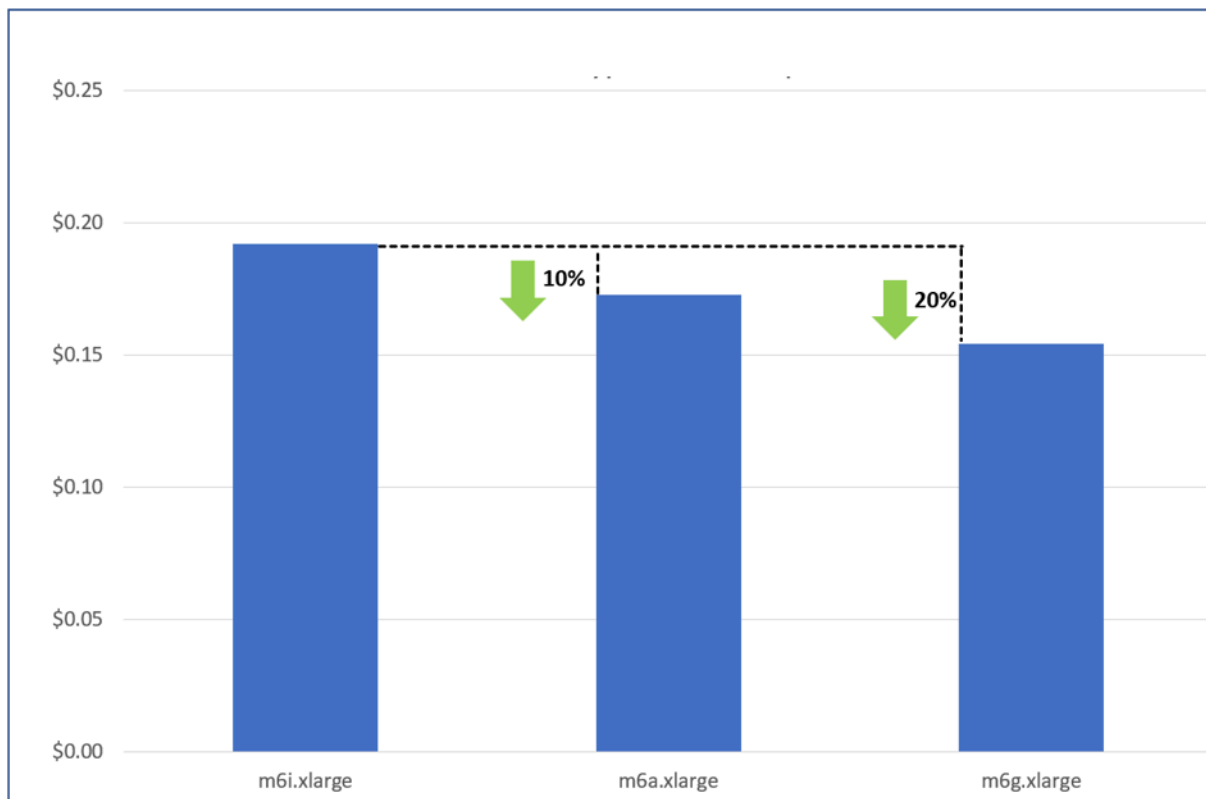
O exemplo a seguir ilustra o preço por hora de uma instância Graviton em execução na região us-east-1.

Instância Intel	Custo por hora	Instâncias Graviton	Custo por hora	% da diferença
c6i.xlarge	0,17 US\$	c6g.xlarge	0,136 US\$	20%
m6i.xlarge	\$0,192	m6g.xlarge	\$0,154	20%
r6i.xlarge	\$0,252	r6g.xlarge	\$0,2016	20%

Note

Os preços são baseados nos preços por hora sob demanda na região us-east-1.

O gráfico a seguir compara os preços das instâncias da série M.



Entenda as diferenças da relação preço/performance entre as gerações do EC2

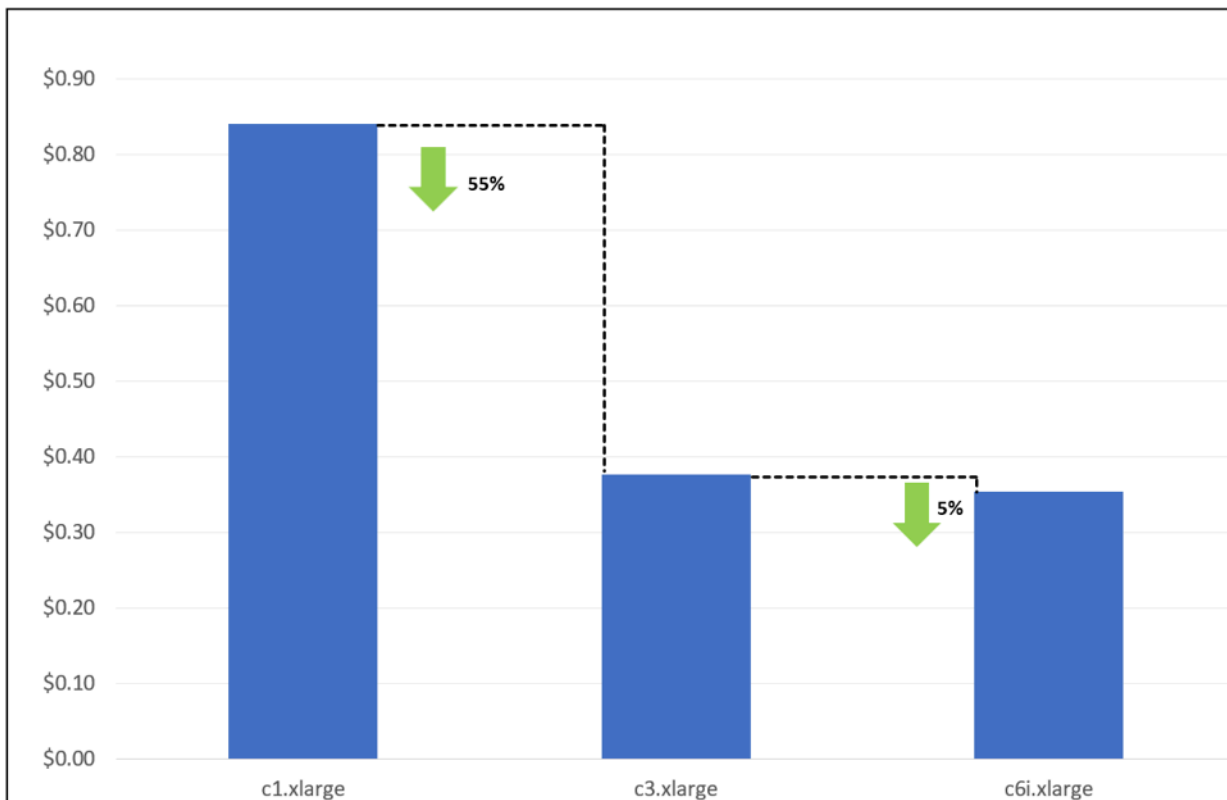
Uma das características mais consistentes do Amazon EC2 é que cada nova geração oferece uma melhor relação preço/performance do que sua antecessora. Como mostra a tabela a seguir, o preço das instâncias da nova geração do EC2 diminui a cada versão subsequente.

Instância otimizada para computação	Custo por hora	Instância de uso geral	Custo por hora	Instância otimizada para memória	Custo por hora
C1.xlarge	\$0,52	M1.xlarge	\$0,35	r1.xlarge	n/a
C3.xlarge	\$0,21	M3.xlarge	\$0,266	r3.xlarge	\$0,333
C5.xlarge	0,17 US\$	M5.xlarge	\$0,192	r5.xlarge	\$0,252

Note

Os preços são baseados nos preços por hora sob demanda na região us-east-1.

O gráfico a seguir compara os custos das diferentes gerações de instâncias da série C.



No entanto, a 6.^a geração de instâncias tem o mesmo preço da 5.^a geração, conforme mostra a tabela a seguir.

Instância otimizada para computação	Custo por hora	Instância de uso geral	Custo por hora	Instância otimizada para memória	Custo por hora
C5.xlarge	0,17 US\$	M5.xlarge	\$0,192	r5.xlarge	\$0,252
C6i.xlarge	0,17 US\$	M6i.xlarge	\$0,192	r6i.xlarge	\$0,252

Note

Os preços são baseados nos preços por hora sob demanda na região us-east-1.

Apesar de ter o mesmo custo, a nova geração oferece uma relação preço/performance superior devido a processadores mais rápidos, maior throughput de rede e maior throughput e IOPS do Amazon Elastic Block Store (Amazon EBS).

Uma das melhorias mais significativas na relação preço/performance é o aprimoramento da [instância X2i](#). Essa geração da instância oferece uma relação preço/performance até 55% maior do que a geração anterior. Como mostra a tabela a seguir, a x2iedn demonstra melhorias em todos os aspectos de performance (tudo pelo mesmo preço da geração anterior).

Instância	Custo por hora	v CPUs	RAM	Velocidade do processador	Armazenamento de instâncias	Redes	Throughput do Amazon EBS	IOPS do EBS
x1e.2xlarge	\$1,66	8	244	2.3 GHz	SSD DE 237 GB	10 Gbps	125 Mb/s	7400
x1iedn.2xlarge	\$1,66	8	256	3.5 GHz	SSD de 240 GB NVMe	25 Gbps	2500 MB/s	65000

Note

Os preços são baseados nos preços por hora sob demanda na região us-east-1.

Cenários de exemplo

Considere o exemplo de uma empresa de analytics que rastreia veículos de entrega e deseja melhorar a performance do SQL Server. Depois que um SME em MACO analisa os gargalos de performance dessa empresa, ela faz a transição de instâncias x1e.2xlarge para instâncias x2iedn.xlarge. O novo tamanho da instância é menor, mas os aprimoramentos nas instâncias x2 permitem maior performance e otimização do SQL Server por meio do uso de extensões de grupos de buffers. Isso permite que a empresa faça o downgrade da edição SQL Server Enterprise para a edição SQL Server Standard. Também permite que a empresa reduza seu licenciamento do SQL Server de 8 v CPUs para 4 v. CPUs


Antes da otimização:

Servidor	Instância do EC2	Edição do SQL Server	Custo mensal
Cutucar DB1	x1e.2xlarge	Enterprise	\$3.918,64
Cutucar DB2	x1e.2xlarge	Enterprise	\$3.918,64
Total			\$7.837,28

Após a otimização:

Servidor	Instância do EC2	Edição do SQL Server	Custo mensal
Cutucar DB1	x2iedn.xlarge	Standard	\$1.215,00
Cutucar DB2	x2iedn.xlarge	Standard	\$1.215,00
Total			\$2.430,00

Ao todo, a mudança de instâncias x1e.2xlarge para instâncias x2iedn.xlarge permite que a empresa, no cenário de exemplo, economize USD 5.407 por mês em seus servidores de banco de dados de produção. Isso reduz o custo total da workload em 69%.

 Note

Os preços são baseados nos preços por hora sob demanda na região us-east-1.

Migrar para instâncias mais novas

As gerações mais antigas do Amazon EC2 são executadas no hipervisor Xen, enquanto as gerações mais novas operam no [AWS Nitro System](#). O Nitro System fornece quase todos os recursos de computação e memória do hardware host para suas instâncias. Isso resulta em melhor performance geral. Há considerações especiais ao [migrar de instâncias baseadas em Xen para instâncias baseadas em Nitro](#). Por exemplo, [AWS o Windows AMIs](#) é configurado com configurações e personalizações padrão usadas pela mídia de instalação da Microsoft. As personalizações incluem

drivers e configurações compatíveis com os tipos de instância de última geração, que são ([instâncias criadas no Nitro System](#)).

Se você estiver iniciando instâncias do Windows personalizado AMIs ou do Windows AMIs fornecido pela Amazon que foram criadas antes de agosto de 2018, recomendamos que você conclua as etapas de [Migrar para os tipos de instância de última geração](#) na documentação do Amazon EC2.

Usar instâncias expansíveis

Embora as instâncias expansíveis sejam uma boa maneira de economizar nos custos de computação, recomendamos que você as evite nos seguintes cenários:

- As [especificações mínimas do Windows Server](#) com a Experiência de desktop exigem 2 GB de RAM. Evite usar instâncias t3.micro ou t3.nano com o Windows Server porque elas não têm a quantidade mínima de RAM.
- Se a sua workload tiver picos, mas não permanecer ociosa por tempo suficiente para acumular créditos de expansão, usar instâncias normais do EC2 é mais eficiente do que usar instâncias expansíveis. Recomendamos [monitorar seus créditos de CPU](#) para verificar essa questão.
- Recomendamos evitar usar instâncias expansíveis com o SQL Server na maioria dos cenários. O licenciamento do SQL Server é baseado no número de v CPUs atribuído a uma instância. Se o SQL Server ficar ocioso a maior parte do dia, você pagará por licenças SQL que não está utilizando totalmente. Nesses cenários, recomendamos que você consolide várias instâncias do SQL Server em um servidor maior.

Próximas etapas

Recomendamos que você execute as próximas etapas abaixo para otimizar seus custos para instâncias Amazon EC2 Windows:

- Use a instância de última geração do EC2 para obter a melhor relação preço/performance.
- Use instâncias do EC2 com processadores AMD para uma redução de 10% nos custos de computação.
- Maximize a utilização dos recursos escolhendo um tipo de instância do EC2 que corresponda à sua workload.

A tabela a seguir mostra exemplos de pontos de partida comuns para as workloads do Windows. Opções adicionais estão disponíveis, como volumes de armazenamento de instâncias para

aprimorar as cargas de trabalho do SQL Server ou instâncias do EC2 com proporções muito maiores vCPU-to-RAM. Recomendamos que você teste suas workloads minuciosamente e use ferramentas de monitoramento, como o AWS Compute Optimizer , para ajudar a fazer os ajustes necessários.

Workload	Típico	Opcional
Active Directory	T3, M6i	R6i
Servidores de arquivos	T3, M6i	C6i
Servidores da web	T3, C6i	M6i, R6i
SQL Server	R6i	x2iedn, X2iezn

Se você precisar alterar o tipo de instância do EC2, o processo normalmente envolve apenas uma simples reinicialização do servidor. Para obter mais informações, consulte [Change the instance type](#) na documentação do Amazon EC2.

Antes de alterar o tipo de instância, recomendamos que você considere o seguinte:

- É necessário interromper suas instâncias baseadas no Amazon EBS para poder alterar o tipo de instância. Certifique-se de planejar um tempo de inatividade enquanto a instância estiver parada. Interromper a instância e alterar o tipo de instância pode levar alguns minutos, e o tempo necessário para iniciar a instância pode variar dependendo dos scripts de startup da aplicação. Para obter mais informações, consulte [Início e interrupção de instâncias](#) na documentação do Amazon EC2.
- Quando você interrompe e inicia uma instância, AWS move a instância para um novo hardware. Se sua instância tiver um IPv4 endereço público, AWS liberará o endereço e fornecerá à instância um novo IPv4 endereço público. Se você precisar de um IPv4 endereço público que não mude, use um [endereço IP elástico](#).
- Você não poderá alterar o tipo de instância se [hibernation](#) estiver habilitado nela.
- Você não pode alterar o tipo de instância de uma [instância spot](#).
- Se sua instância estiver em um grupo do Auto Scaling, o Amazon EC2 Auto Scaling marcará a instância interrompida como não íntegra e poderá encerrá-la e inicializar uma instância substituta. Para evitar isso, é possível suspender os processos de escalabilidade para o grupo enquanto estiver alterando o tipo de instância. Para obter mais informações, consulte [Suspend and resume a process for an Auto Scaling group](#) na documentação do Amazon EC2 Auto Scaling.

- Quando você altera o tipo de instância de uma instância com volumes de armazenamento de NVMe instâncias, a instância atualizada pode ter volumes adicionais de armazenamento de instâncias, porque todos os volumes de armazenamento de NVMe instâncias estão disponíveis mesmo que não estejam especificados na Amazon Machine Image (AMI) ou no mapeamento de dispositivos de blocos de instâncias. Caso contrário, a instância atualizada tem o mesmo número de volumes de armazenamento de instância que você especificou ao iniciar a instância original.

Recursos adicionais do

- [Tipos de instância do Amazon EC2 \(documentação\)](#) AWS
- [Avaliação de otimização e licenciamento da AWS](#) (documentação da AWS)

Trazer licenças para workloads do Windows e SQL Server

Visão geral do

Se você tiver investimentos significativos em cargas de trabalho da Microsoft e contratos de licenciamento corporativo existentes, poderá escolher entre várias AWS opções para oferecer suporte a essas cargas de trabalho, incluindo as opções de [licença incluída \(fornecida por AWS\)](#) e [Bring Your Own License \(BYOL\)](#). Você pode usar os [hosts dedicados do Amazon EC2](#) para aproveitar totalmente os contratos de licenciamento existentes da Microsoft e trazer o Windows Server para a AWS. Isso pode fazer você economizar até 50% nos custos de instâncias do Amazon EC2. Como as licenças do Windows representam aproximadamente metade dos custos da instância, colocar o Windows Server AWS em hosts dedicados pode resultar em economias substanciais. Como o Windows Server não pode ser colocado em [locação padrão \(compartilhada\)](#), os hosts dedicados são a escolha ideal se você quiser usar suas licenças existentes para o Windows Server.

AWS

Os hosts dedicados não servem apenas para instâncias BYOL do Windows Server. Eles também oferecem a flexibilidade de adequar seu licenciamento on-premises às workloads existentes do SQL Server. Os hosts dedicados expõem os núcleos físicos do servidor subjacente e permitem que você licencie o SQL Server no nível do núcleo físico. Isso não é possível na [locação padrão \(compartilhada\)](#) em que o licenciamento do SQL Server é baseado no número de virtuais CPUs alocados para a instância. Esse recurso permite que você licencie cargas de trabalho do SQL Server de AWS forma consistente com sua estratégia de licenciamento local. Consequentemente, você pode economizar até 50% nos custos de licenciamento do SQL Server em comparação com

a locação padrão (compartilhada), além da economia nos custos de instâncias, usando licenças qualificadas do Windows. Para obter mais informações sobre esse cenário, consulte a seção [Compreender o licenciamento do SQL Server](#) deste guia.

Hosts dedicados do Amazon EC2

Um host dedicado do Amazon EC2 é essencialmente o mesmo host EC2 AWS usado para executar suas ofertas de computação do EC2. A diferença é que esses hosts são totalmente dedicados a um único cliente e fornecem acesso exclusivo à infraestrutura física subjacente. Você pode usar hosts dedicados para executar suas instâncias em hardware totalmente dedicado ao seu uso, em vez de compartilhar recursos com outros clientes da AWS. Isso lhe dá maior controle sobre os recursos da nuvem e permite que você reduza custos trazendo suas próprias licenças de software, como Windows Server e SQL Server, para a AWS.

Lembre-se do seguinte:

- Um host dedicado é um servidor físico totalmente dedicado a um único cliente. Você tem visibilidade dos soquetes e dos núcleos físicos do host dedicado para que possa cumprir os requisitos de conformidade de licenciamento, como contratos de licenciamento de software por soquete, por núcleo ou por VM.
- Hosts dedicados que são compatíveis com vários tamanhos de instâncias da mesma família de instâncias são conhecidos como hosts dedicados heterogêneos. Essas [famílias de instâncias](#) incluem T3, A1, C5, M5, R5, C5n, R5n e M5n. Em contraste, outras famílias de instâncias são compatíveis apenas com um tamanho de instância no mesmo host dedicado. Eles são chamados de hosts dedicados homogêneos.
- Os hosts dedicados são cobrados por host. Isso significa que você é cobrado por host dedicado, independentemente de quantas instâncias estejam sendo executadas nele. Os preços de hosts dedicados varia de acordo com a família de instâncias, a região e a opção de pagamento selecionada. Você pode escolher a configuração ideal para sua workload a fim de alcançar os resultados desejados de performance e custo.

Este diagrama ilustra as diferenças entre instâncias de locação compartilhada e hosts dedicados.



Hosts dedicados homogêneos

Considere um cenário em que um host dedicado M6i é usado. Os hosts dedicados M6i e R6i têm dois soquetes, 64 núcleos físicos e são compatíveis com tipos de instância do mesmo tamanho. Eles são chamados de hosts dedicados homogêneos. Isso significa que o número de instâncias que você pode inicializar em um único host dedicado M6i depende do tamanho da instância.

Por exemplo:

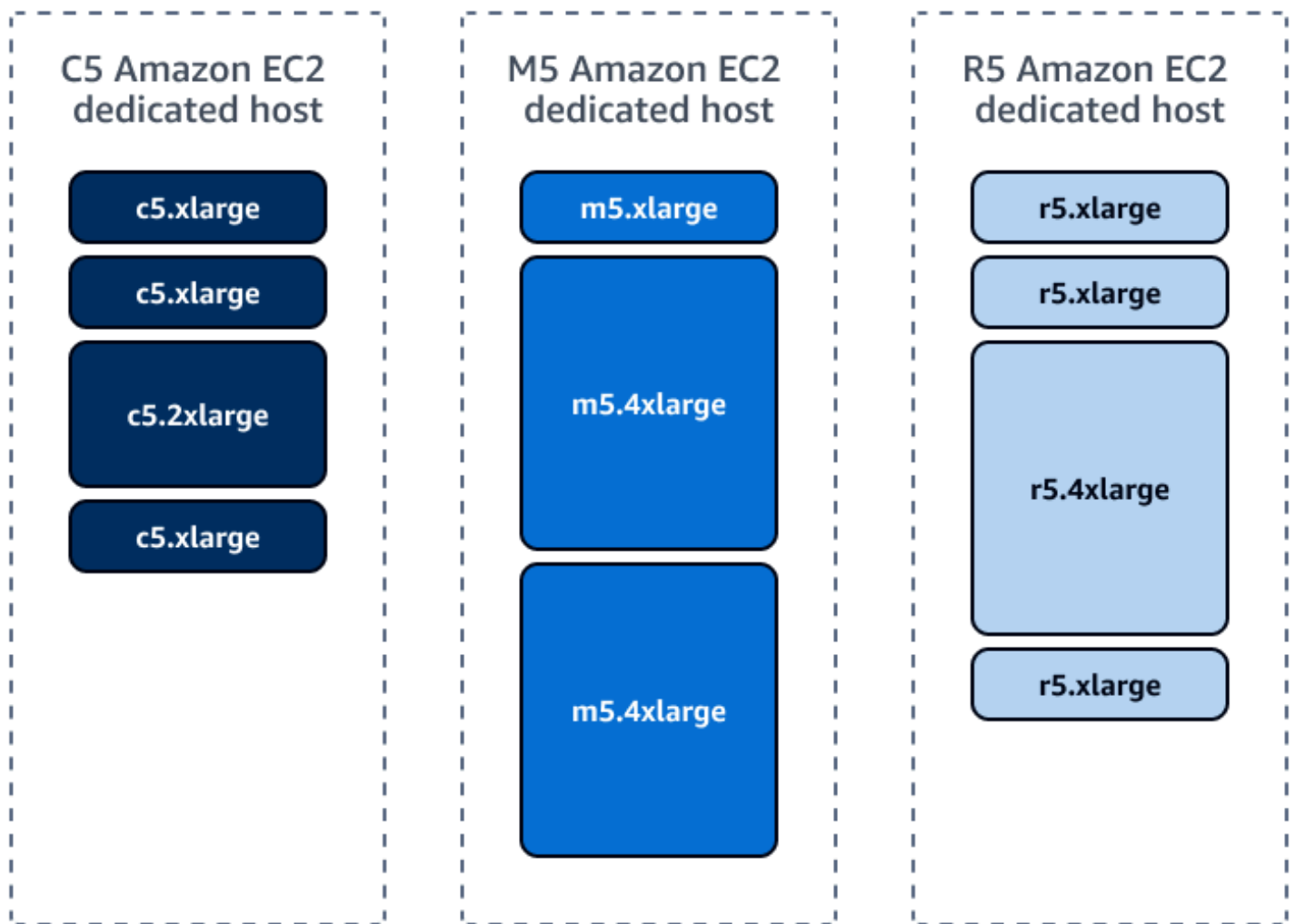
- No caso do xlarge (4 vCPUs), você pode executar no máximo 32 instâncias m6i.xlarge nesse host dedicado.
- No caso de 8xlarge (32 vCPUs), você pode executar no máximo 4 instâncias m6i.8xlarge nesse host dedicado.
- No caso do metal (128 vCPUs), você pode iniciar no máximo 1 instância m6i.metal nesse host dedicado.

O diagrama a seguir mostra as opções de host dedicado para instâncias M6.



Hosts dedicados heterogêneos

Os hosts dedicados compatíveis com vários tamanhos de instância no mesmo host são chamados de hosts dedicados heterogêneos do Amazon EC2. O diagrama a seguir mostra um exemplo de hosts dedicados C5, M5 e R5 com vários tamanhos de instâncias, como 2xlarge, xlarge e 4xlarge.



Gerenciamento de hosts dedicados

Recomendamos considerar o seguinte em relação ao gerenciamento de hosts dedicados do Amazon EC2:

- Para aproveitar ao máximo os hosts dedicados, você pode [compartilhar um único host entre várias contas em sua organização](#). O compartilhamento de hosts permite a otimização de recursos e pode resultar em economia de custos ao usar todos os slots disponíveis no host. Ao compartilhar um host dedicado entre unidades de negócios, você pode centralizar sua infraestrutura de TI e melhorar a utilização dos recursos, mantendo a separação entre as workloads. Se você faz parte de uma organização AWS Organizations e o compartilhamento está habilitado em sua organização, os consumidores em sua organização recebem automaticamente acesso ao Host dedicado compartilhado. Caso contrário, os consumidores receberão um convite para participar

do compartilhamento de recursos e acesso ao Host dedicado compartilhado depois de aceitar o convite.

- Você pode executar o Windows Server 2022 em hosts dedicados sob o modelo de licença incluída, já que o Windows Server 2019 é a versão mais recente em que há a disponibilização de BYOL. Se você quiser usar o Windows Server 2022 em hosts dedicados, deverá usar instâncias com licença incluída do Windows Server 2022.
- [AWS License Manager](#) é uma solução abrangente para gerenciar licenças de software de vários fornecedores em ambientes locais AWS e internos. Se você [usa o License Manager](#), é possível obter maior visibilidade e controle sobre como as licenças de software são usadas, resultando em economia de custos e maior conformidade. Você pode usar o License Manager para definir regras para emular seus termos de licenciamento exclusivos. Isso permite que você aplique essas regras e evite o uso indevido da licença. Isso pode reduzir o risco de não conformidade e melhorar os processos de gerenciamento de licenças.
- É possível usar o License Manager para automatizar o posicionamento, a liberação e a recuperação de hosts usando [grupos de recursos de hosts](#). Isso pode aumentar a produtividade e reduzir a sobrecarga de gerenciamento. O License Manager também fornece uma visão centralizada do uso de licenças em ambientes locais AWS e em todos os ambientes com base nas regras de licenciamento, facilitando o gerenciamento de compras incrementais de licenciamento, conformidade e auditorias de fornecedores em toda a organização. Além disso, o License Manager se integra com AWS Organizations e AWS Resource Access Manager (AWS RAM) para compartilhar configurações de licença entre contas e regiões. Isso permite que você crie relatórios para todo o seu ambiente com base em uma programação e gerencie as regras de licenciamento de forma centralizada em uma Conta da AWS. Em última análise, isso pode melhorar a governança e reduzir a complexidade.
- Ao projetar a alta disponibilidade para hosts dedicados em uma única região, certifique-se de ter alocado no mínimo dois hosts dedicados em no mínimo duas zonas de disponibilidade para workloads críticas de produção. Para obter mais informações, consulte a implantação de referência de [hosts dedicados do Amazon EC2 para Microsoft Windows na AWS](#).
- Para cada família de instâncias de hosts dedicados, há um limite no número de instâncias que podem ser executadas para cada tamanho de instância. Para obter mais informações, consulte [Tabela de configuração de hosts dedicados](#) na documentação do Amazon EC2.

AWS opções de licenciamento

As licenças são classificadas nas seguintes categorias principais:

- **Licença incluída:** esta opção de licenciamento permite que você compre e use licenças sob demanda, pagando somente pelo que usa. É ideal para casos de uso em que você busca flexibilidade no uso do licenciamento e deseja evitar custos iniciais. Você pode escolher entre uma variedade de produtos Windows Server, SQL Server e outros produtos da Microsoft.
- **Produtos BYOL com Mobilidade de Licenças:** se você já tem licenças e deseja usá-las na nuvem, esta opção permite que você traga suas próprias licenças para a nuvem por meio do [programa Mobilidade de Licenças da Microsoft](#). Produtos com Mobilidade de Licenças, como o SQL Server com o Software Assurance (SA), podem ser trazidos para locação compartilhada ou dedicada. Isso reduz os custos da AWS instância.
- **Produtos BYOL sem mobilidade de licenças** — Para produtos da Microsoft, como o Windows Server, que não possuem mobilidade de licenças, AWS fornece opções dedicadas para o uso desses produtos na nuvem. Além disso, os hosts dedicados permitem o licenciamento no nível do núcleo físico, possibilitando economizar 50% ou mais nas licenças necessárias para executar suas workloads. Os hosts dedicados são uma excelente opção para workloads estáveis e previsíveis que são executadas na maior parte do tempo.

Portabilidade de licenças do Windows Server

Trazer suas próprias licenças do Windows é uma das estratégias mais eficazes para otimização de licenças, pois permite que você aproveite os investimentos existentes e reduza suas AWS despesas. Cenários específicos de BYOL excluem a exigência de SA ou Mobilidade de Licenças, mas a infraestrutura dedicada do Amazon EC2 é sempre necessária. Para se qualificar, você deve ter comprado licenças perpétuas antes de 1.º de outubro de 2019 ou tê-las adicionado como um true-up sob um contrato corporativo ativo que estava em vigor antes dessa data. Nesses cenários específicos de BYOL, você pode atualizar somente licenças para versões disponíveis antes de 1.º de outubro de 2019. Por exemplo, se você cancelou o SA em 2017, você tem o direito de implantar somente até o Windows Server 2016, não o 2019. No entanto, 2019 é a última versão elegível para BYOL. AWS Para obter mais informações, consulte [Licenciamento — Windows Server](#) na AWS documentação.

A obtenção de licenças pode afetar significativamente o custo de execução das workloads da Microsoft na AWS. Ao trazer suas próprias licenças, você não precisa pagar custos adicionais de licenciamento pelas instâncias em execução na nuvem, o que pode levar a uma economia considerável.

A tabela a seguir mostra o custo mensal sob demanda da execução de uma única instância c5.xlarge, 24 horas por dia, 7 dias por semana, em várias configurações.

Configuração	Custo mensal (USD)
Windows Server + edição SQL Server Enterprise	USD 1.353,00 (LI)
Windows Server + edição SQL Server Standard	USD 609,00 (LI)
Apenas Windows Server	USD 259,00 (LI)
Somente computação (Linux)	\$127,00

Você pode usar suas licenças existentes para reduzir os custos de licenciamento e economizar dinheiro em sua fatura geral AWS .

Para se qualificar para BYOL em hosts dedicados do Amazon EC2, você deve trazer suas próprias licenças de software, como do Windows Server e SQL Server. O BYOL permite que você use suas licenças existentes AWS e pode resultar em economia de custos. Para trazer suas próprias licenças, você deve ter os direitos de licença do provedor do software e também fornecer a mídia ou a imagem de instalação do software. A mídia ou a imagem de instalação podem ser usadas para inicializar instâncias em hosts dedicados. Para saber mais sobre a criação de uma AMI BYOL, consulte [Como criar o Windows Server Bring-Your-Own-License AMIs localmente com o VM Import/Export](#) no blog Microsoft Workloads on. AWS

Note

Um tipo de licença definido como Auto equivale a uma [opção de licença incluída da AWS](#). Essa opção pode resultar em gastos indesejados sob demanda. Você precisará mudar os [tipos de licenciamento](#).

Cenário de otimização de custos

O dimensionamento correto e a otimização de licenças são um componente essencial da otimização de custos na AWS. Se você implementar as estratégias certas, poderá reduzir os custos de licenciamento, manter a conformidade e obter o melhor valor possível do seu investimento em licenciamento usando os hosts dedicados do Amazon EC2 e a opção BYOL.

Esta seção aborda os seguintes cenários de exemplo:

- Economia de custos com os hosts dedicados T3
- Comparação da locação compartilhada com hosts dedicados com BYOL do SQL Server
- Implantações altamente disponíveis do SQL Server

Economia de custos com os hosts dedicados T3

Os hosts dedicados T3 diferem de outros hosts dedicados do Amazon EC2 que tradicionalmente fornecem recursos fixos de CPU. Os hosts dedicados T3, por outro lado, oferecem suporte a instâncias expansíveis que são capazes de compartilhar recursos da CPU, fornecer performance básica da CPU e expandir quando necessário. O compartilhamento de recursos de CPU, também conhecido como excesso de inscrições, é o que permite que um único host dedicado T3 ofereça suporte a até quatro vezes mais instâncias do que hosts dedicados de uso geral comparáveis.

Os hosts dedicados T3 geram um TCO menor ao oferecer maior densidade de instâncias do que qualquer outro host dedicado do Amazon EC2. As instâncias T3 com capacidade de intermitência permitem consolidar um número maior de instâncias com utilização low-to-moderate média da CPU em menos hosts do que nunca. Os hosts dedicados T3 também oferecem tamanhos de instância menores em um número maior de combinações de vCPU e memória do que outros hosts dedicados do Amazon EC2. Instâncias menores podem contribuir para reduzir o TCO e ajudar a fornecer taxas de consolidação equivalentes ou superiores às dos hosts on-premises.

Os hosts dedicados T3 são mais adequados para executar software BYOL com utilização de low-to-moderate CPU e licenças de software elegíveis por soquete, por núcleo ou por VM, incluindo Microsoft Windows desktop, Windows Server, SQL Server e bancos de dados Oracle.

Use hosts dedicados T3 para reduzir as licenças do Windows Server Datacenter (por núcleo)

Em ambientes locais, você está aproveitando o fato de poder facilmente substituir sua assinatura física CPUs em VMware hosts e alcançar altos níveis de consolidação.

Considere o exemplo a seguir. No momento, você está usando VMware hosts de 10x36 núcleos e 384 GB de RAM em um ambiente local. Além disso, cada host está executando máquinas virtuais Windows Server de 96x2 vCPU e 4 GB de RAM com baixa média de utilização da CPU.

Agora você pode alcançar níveis muito mais altos de consolidação movendo suas máquinas virtuais para hosts dedicados T3, que têm o dobro da quantidade de RAM em comparação com seus hosts

locais VMware atuais. Você pode executar o mesmo número de servidores em hosts dedicados T3 com uma redução de 50% no custo do host. Isso pode ajudar você a reduzir os custos de licenciamento do Windows Server em 33%. A tabela a seguir destaca a economia do uso de hosts dedicados T3.

	Hosts locais VMware	Hosts dedicados T3	Poupança
Servidores físicos	10	5	
Núcleos físicos por host	36	48	
RAM por host (GB)	384	768	
2 vCPU, 4 GB de RAM VMs por host	96	192	
Número total de VMs	960	960	
Total de licenças do Windows Server Datacenter (por núcleo) = (Número de servidores * Contagem de núcleos físicos)	$10 * 36 = 360$	$5 * 48 = 240$	33%

Comparação da locação compartilhada com hosts dedicados com BYOL do SQL Server

Considere um exemplo prático para demonstrar o valor dos hosts dedicados do Amazon EC2. Nesse cenário, uma organização executa uma workload do SQL Server em um ambiente on-premises com 240 núcleos e deseja implantar a mesma workload de forma econômica na AWS. Se essa organização trazer suas próprias licenças (BYOL), ela continuará pagando pelo SA, e a redução do número de núcleos afetará diretamente seus custos.

O diagrama a seguir compara a AWS economia entre direitos da Microsoft e do SQL Server.

Microsoft entitlements (Enterprise Agreements)		SQL Server savings with AWS	
	Number of cores	AWS shared vCPUs	AWS BYOL/Dedicated Hosts cores
SQL Server Enterprise edition	208	120	96
SQL Server Standard edition	32	20	-
Total SA cost	\$341,000	\$197,418	\$151,355

Ao dimensionar corretamente as instâncias na locação AWS compartilhada, você pode reduzir as licenças do SQL Server para 140 núcleos. Isso resulta em custos de SA de USD 197.000.

Os hosts dedicados do Amazon EC2 permitem que você licencie o SQL Server no nível do núcleo físico. Isso não é possível na locação compartilhada em que o licenciamento do SQL Server é baseado no número de v CPUs alocado para a instância. Conseqüentemente, ao usar dois hosts dedicados R5 com 48 núcleos cada, você só precisa cobrir 96 núcleos em vez dos 140 v CPUs necessários na locação compartilhada. Ao implantar hosts dedicados R5 e licenciar as workloads no nível físico, você pode reduzir o número necessário de licenças da edição SQL Server Enterprise para 96 núcleos. Isso significa que você pode implantar até 192 núcleos (considerando o hyper-threading) de workloads do SQL Server, sem deixar de cumprir os requisitos de licenciamento e obter economias de custo significativas.

Nesse caso, a organização paga aproximadamente USD 341 mil anualmente em custos de SA. Após o dimensionamento correto da locação compartilhada, eles reduzem os custos para USD 197.000 com 140 vCPUs. Os hosts dedicados do Amazon EC2 reduzem ainda mais os custos para USD 151.000 (uma redução de aproximadamente 56%).

Implantações altamente disponíveis do SQL Server

Este exemplo analisa como o custo pode influenciar uma implantação do SQL Server na AWS com várias considerações de licenciamento. Suponha que uma organização precise implantar seis servidores SQL Server Enterprise AWS para oferecer suporte a três aplicativos. Esses servidores exigem alta disponibilidade e têm 16 v CPUs e 256 GB de RAM cada. Confira os detalhes do seguinte cenário:

- Servidor: SQL Server
- Edição do sistema operacional: Windows Server Datacenter 2019
- Edição do SQL Server: SQL Server Enterprise 2019
- vCPU: 16
- Memória (GB): 256

- Quantidade: 6

Para otimizar os custos AWS sem sacrificar o desempenho, recomendamos que você dimensione corretamente as instâncias com base na utilização de CPU, memória, rede e disco (IOPS/BW). Depois de dimensionar corretamente as cargas de trabalho, coloque-as no tipo de instância x2iedn.4xlarge, que oferece 16 v. CPUs. No entanto, esse tipo de instância também inclui o dobro da memória necessária para as cargas de trabalho. É possível otimizar ainda mais.

Cenário 1

Uma organização implanta seis servidores SQL Server Enterprise em locação AWS compartilhada usando a opção de licença incluída para Windows e SQL Server. Com essa opção, o custo das licenças do Windows e do SQL Server é incorporado ao preço das instâncias. Confira os detalhes do seguinte cenário:

- Locação compartilhada (instância): x2iedn.4xlarge
- Custo por hora (USD): USD 10,0705
- Custo por mês por unidade (USD): USD 7.351,47
- Número de servidores: 6
- CPU: 16
- Memória: 512
- Custo por mês para 6 servidores: USD 44.108

Cenário 2

Uma organização tem SA e BYOL para SQL Server em locação compartilhada. Isso significa que a organização usa a opção de licença incluída para Windows, mas fornece suas próprias licenças do SQL Server com base no número de v CPUs alocado para a instância. Como a organização tem seis servidores SQL Server Enterprise com 16 v CPUs cada, CPUs é necessário um total de 96 v. Confira os detalhes do seguinte cenário:

- Locação compartilhada (instância): x2iedn.4xlarge
- Custo por hora (USD): USD 4,0705
- Custo por mês por unidade (USD): USD 2.971,47
- Número de servidores: 6
- CPU: 16

- Memória: 512
- Núcleos BYOL: 96
- Custo por mês para 6 servidores: USD 17.828

Ao trazer suas próprias licenças do SQL Server com o SA, a organização nesse cenário pode obter economia de custos em comparação com o uso da opção de licença incluída para o SQL Server. A economia precisa de custos depende dos preços e dos termos específicos do contrato de licenciamento. Nesse cenário, AWS os custos diminuem em USD 26.280 por mês ao trazer licenças do SQL Server Enterprise para o. AWS

Cenário 3

Uma organização tem BYOL para o Windows e o SQL Server em hosts dedicados do Amazon EC2. Isso significa que a organização atribuirá licenças no nível do núcleo físico, permitindo que licenciem somente os núcleos físicos do host. O licenciamento no nível do núcleo físico permite que você implante o número máximo de instâncias sem afetar as licenças necessárias. Esse modelo de licenciamento é comumente usado com o Windows Server Datacenter e a edição SQL Server Enterprise.

Esse cenário usa dois hosts dedicados X2iezn do Amazon EC2. Cada host tem 24 núcleos físicos e 48 CPUs v. Isso fornece capacidade adequada para os seis servidores SQL Server Enterprise com 16 v CPUs e 256 GB de RAM cada. Confira os detalhes do seguinte cenário:

- Número de hosts dedicados: 2
- Família de instâncias: x2iezn
- Custo por hora (USD): USD 11,009
- Custo mensal por unidade (USD): USD 8.036
- Núcleo físico: 48
- vCPU disponível: 96
- Licenças de núcleo necessárias do Windows Server: 24
- Licenças necessárias para os núcleos do SQL Server Enterprise: 24
- Custo mensal: 16.073

O custo total de dois hosts dedicados da família X2iezn do Amazon EC2 é de USD 16.073 por mês. Para obter mais informações sobre preços, consulte a AWS Calculadora de Preços [estimativa](#) para

esse cenário. Nesse cenário, a organização pode economizar USD 1.755,65 por mês trazendo suas licenças do Windows. Se ela usar hosts dedicados do Amazon EC2, também poderá reduzir o número de licenças necessárias do SQL Server. Em locação compartilhada, eles precisarão de 96 licenças do SQL Server Enterprise para cobrir os seis servidores SQL Server Enterprise com 16 vCPUs cada. No entanto, usando os hosts dedicados do Amazon EC2 e o licenciamento no nível do núcleo físico, ela pode reduzir o número de licenças necessárias para 48 núcleos.

Os detalhes a seguir comparam os custos do exemplo 3 e mostram quanto você pode economizar implantando workloads em hosts dedicados do Amazon EC2 com a opção BYOL em comparação com outros cenários.

- Servidor on-premises: SQL Server
- vCPU: 16
- Memória: 256
- Número de servidores: 6
- Custo mensal do cenário 1: Windows (LI) + SQL Server Enterprise (LI): USD 44.108
- Custo mensal do cenário 2: Windows (LI) + SQL Server Enterprise (BYOL): USD 17.828
- Custo mensal do cenário 3: Windows (LI) + SQL Server Enterprise (BYOL) no host dedicado Amazon EC2: USD 16.073

Note

O custo é baseado em preços sob demanda. Você pode reduzir ainda mais os custos usando os Savings Plans ou as Instâncias Reservadas Dedicadas. Essas opções oferecem um modelo de preços flexível com economias de custo significativas em comparação com os preços sob demanda. Com esses planos, você pode se comprometer com um prazo de um ou três anos. Para obter mais informações, consulte a seção [Otimizar gastos com o Windows no Amazon EC2](#) deste guia.

Considere as seguintes opções de pagamento para os hosts dedicados do Amazon EC2:

- [Hosts dedicados](#) (documentação do Amazon EC2)
- [Reservas de hosts dedicados](#) (documentação do Amazon EC2)
- [Savings Plans](#) (documentação do Amazon EC2)

A [AWS Calculadora de Preços](#) agora oferece suporte a preços de hosts dedicados. Isso pode ajudar você a escolher o host dedicado subjacente apropriado.

Recomendações de otimização de custos

Recomendamos que você execute as etapas abaixo para otimizar seus custos usando o AWS Cost Explorer:

1. [Ative o Cost Explorer](#).
2. Use o Cost Explorer para [visualizar e analisar os custos e o uso](#) de suas implantações de hosts dedicados do Amazon EC2.
3. Valide que você está executando o BYOL. Você pode exibir os detalhes da plataforma e os valores de operação de uso a seguir nas páginas de instâncias ou AMIs no console do Amazon EC2 ou na resposta retornada pelo comando `describe-images` ou `describe-instances`.
 - Detalhes da plataforma: Windows, Operação de uso ::0002 RunInstances (Licença incluída)
 - Detalhes da plataforma: Windows BYOL, Operação de uso: :0800 RunInstances

Recursos adicionais do

- [Eligible license types for license type conversion](#) (documentação da AWS License Manager)
- [AWS License Manager e workshop dedicado para anfitriões](#) (AWS License Manager Workshop)
- [Hosts dedicados do Amazon EC2 \(documentação FAQs da AWS\)](#)
- [Como criar o Windows Server Bring-Your-Own-License AMIs localmente com o VM Import/Export](#) (Microsoft Workloads no blog) AWS
- [VM Import/Export \(documentação\)AWS](#)
- [Amazon Web Services e Microsoft: perguntas frequentes](#) (AWS documentação)
- [Conversões de tipo de licença no License Manager](#) (AWS License Manager documentação)
- [Implantação de SQL Server de alta disponibilidade em hosts dedicados do Amazon EC2](#) (blog de operações e migrações AWS na nuvem)

Otimizar gastos com o Windows no Amazon EC2

Visão geral do

Uma das principais preocupações sobre a migração de servidores para lá AWS são os custos de infraestrutura. É verdade que uma das vantagens da nuvem é pagar pelos recursos sob demanda, mas há workloads de produção que precisam estar disponíveis 24 horas por dia, 7 dias por semana, 365 dias por ano. Os [Savings Plans](#) foram projetados para economizar dinheiro em seu AWS uso estável em todas as instâncias do EC2, e. AWS Lambda AWS Fargate

Os Savings Plans oferecem um modelo de preços flexível e podem ajudá-lo a reduzir os preços do uso do Amazon EC2, Fargate, Lambda e SageMaker Amazon AI em troca do compromisso com uma quantidade consistente de uso (por exemplo, \$10/hora). Você se compromete com uma quantidade consistente de gastos de computação por hora em um ou três anos e, em troca, recebe um desconto por esse uso.

Você pode escolher entre três opções de pagamento diferentes com os Savings Plans:

- A opção Sem pagamento inicial não exige nenhum pagamento inicial, e seu compromisso é cobrado apenas mensalmente.
- A opção Pagamento inicial parcial oferece preços mais baixos nos Savings Plans. Você paga antecipadamente pelo menos metade do seu compromisso, e o restante é cobrado mensalmente.
- A opção Pagamento inicial total oferece os preços mais baixos e todo o seu compromisso é cobrado em um único pagamento.

É possível rastrear a validade do seus Savings Plans e o próximo Savings Plans em fila no AWS Cost Explorer. Você pode usar os alertas dos Savings Plans para receber e-mails 1, 7, 30 ou 60 dias antes da data de expiração do plano, ou quando um compromisso estiver na fila para aquisição. Essas notificações também alertam você sobre a data de expiração. Você pode enviar notificações para até dez destinatários de e-mail.

Conceitos básicos do Savings Plans

Todo tipo de uso de computação tem uma taxa sob demanda e uma taxa de Savings Plans. Se você se comprometer com USD 10/hora de uso de computação, obterá preços do Savings Plans em todo o uso de até USD 10 à taxa do Savings Plans. Qualquer uso além do compromisso de gasto computacional é cobrado de acordo com as taxas sob demanda. Você pode começar a utilizar os Savings Plans usando o Cost Explorer no Console de gerenciamento da AWS.

Você pode facilmente se comprometer com os Savings Plans usando as recomendações fornecidas no [Cost Explorer](#) para obter a maior economia. O compromisso por hora recomendado é baseado em seu histórico de uso sob demanda e em sua escolha de tipo de plano, período de vigência e opção de pagamento. A opção de Savings Plans é aplicada primeiro à conta que comprou o plano e, em seguida, é compartilhada com outras contas da família de faturamento consolidado.

Note

A opção de compartilhamento do Savings Plans AWS Organizations está ativada por padrão. Você pode recusar essa opção no AWS Billing console da conta do pagador. Você pode visitar sua página de [Recomendações](#) para ver os Savings Plans que AWS recomendam ajudá-lo a economizar no uso qualificado. As recomendações podem ser atualizadas a qualquer momento para facilitar a aquisição dos Savings Plans ideais.

Savings Plans para computação

Os Savings Plans para computação oferecem a maior flexibilidade e ajudam a reduzir seus custos. Esses planos se aplicam ao uso de instâncias do EC2, independentemente da família, do tamanho, da zona de disponibilidade, da região, do sistema operacional ou da localização da instância. Eles também se aplicam ao uso do Fargate e do Lambda. Por exemplo, com os Savings Plans para computação, você pode mudar de instâncias C4 para M5, transferir uma workload da UE (Irlanda) para a UE (Londres) ou migrar uma workload do EC2 para o Fargate ou Lambda a qualquer momento. Você continua pagando automaticamente o preço dos Savings Plans.

Savings Plans para instâncias do EC2

Os Savings Plans para instâncias do EC2 oferecem os maiores descontos em troca do compromisso com o uso de famílias de instâncias individuais em uma região (por exemplo, compromisso com um nível consistente de uso do M5 no Norte da Virgínia). Isso oferece descontos automaticamente no preço sob demanda da família de instâncias selecionada nessa região, independente da zona de disponibilidade, do tamanho, do sistema operacional ou da localização. O Savings Plans para instâncias do EC2 permite que você altere seu uso entre instâncias dentro de uma família nessa região. Por exemplo, você pode mudar de c5.xlarge executando Windows para c5.2xlarge executando Linux e se beneficiar automaticamente dos preços dos Savings Plans.

Os Savings Plans para instâncias do EC2 e computação se aplicam às instâncias do EC2 que fazem parte dos clusters do Amazon EMR, Amazon Elastic Kubernetes Service (Amazon EKS) e Amazon

Elastic Container Service (Amazon ECS). As cobranças do Amazon EMR, Amazon EKS e Amazon ECS não são cobertas pelos Savings Plans, mas as instâncias do EC2 subjacentes sim. Os Savings Plans para instâncias do EC2 são aplicados antes dos Savings Plans para computação porque estes têm uso mais amplo.

Note

Você não pode alterar uma opção de Savings Plans facilmente depois de assumir um compromisso. Recomendamos que você planeje cuidadosamente antes de se comprometer com qualquer uma das opções de Savings Plans. Os Savings Plans oferecem preços mais baixos em comparação com os preços sob demanda em troca de um compromisso, e não podem ser cancelados durante o período.

Exemplo de compromisso por hora

Se você adquirir uma opção de Savings Plans, você assume um compromisso financeiro por hora para a vigência do plano. Se você se comprometer com USD 10/hora de uso de computação, o preço do Savings Plans será aplicado automaticamente a todo o uso, até USD 10 dólares por hora. Qualquer uso além do compromisso é cobrado de acordo com as taxas sob demanda. Você pode usar a ferramenta de recomendações de aquisição de Savings Plans no Cost Explorer para obter compromissos recomendados que possam maximizar sua economia. O compromisso financeiro por hora de um plano específico não pode ser modificado durante a vigência do plano. Se você quiser um compromisso maior após analisar o uso, poderá adquirir uma opção de Savings Plans adicional para cobrir o uso excessivo.

Vantagens dos Savings Plans

Em comparação com as Instâncias Reservadas, os Savings Plans oferecem um modelo de preços mais flexível que pode economizar dinheiro enquanto você aproveita a seleção mais ampla de opções de computação oferecidas pelos Savings Plans. Os Savings Plans oferecem descontos, mesmo quando suas necessidades de computação mudam. Isso pode ajudar você a acompanhar seu ambiente dinâmico em constante mudança sem incorrer em nenhuma sobrecarga adicional de gerenciamento. Confira algumas outras vantagens de usar os Savings Plans:

- Fácil de usar: receba descontos automáticos em troca de um compromisso financeiro.
- Flexibilidade: um compromisso único que se aplica a vários tipos de uso.
- Economia potencial: há várias maneiras de economizar. Considere os seguintes exemplos:

- 60% de economia nas workloads do Windows Server usando Savings Plans para computação ([d2.8xlarge, 3 anos, tudo adiantado, windows, locação compartilhada, us-east-2](#))
- 73% de economia nas workloads do Windows Server usando Savings Plans para instâncias do EC2 ([d2.8xlarge, 3 anos, tudo adiantado, windows, locação compartilhada, us-east-2](#))
- Economia de 28 a 41% em tipos de instâncias não exóticas ([família t3, 3 anos, tudo adiantado, Windows, locação compartilhada, us-east-2](#))
- Economia média de 25 a 40% para servidores Windows

Note

Os Savings Plans para instâncias do EC2 oferecem um desconto maior do que os Savings Plans para computação devido à flexibilidade reduzida. Você se compromete com o uso por um preço com desconto.

Todo tipo de uso de computação tem uma taxa sob demanda e uma taxa de Savings Plans. A tabela a seguir mostra os Savings Plans e as taxas sob demanda para cada tipo de sistema operacional. Você paga as tarifas dos Savings Plans referentes ao seu compromisso de uso, e qualquer uso além do compromisso é cobrado de acordo com as tarifas regulares sob demanda.

Nome da instância	Taxa de Savings Plans	Economia sob demanda	Tarifa sob demanda	Sistema operacional	Região	Opção de pagamento	Período de vigência
x2iedn.xlarge	\$0,32	61%	\$0,83	Linux	Leste dos EUA (Norte da Virgínia)	Sem taxas iniciais	3
x2iedn.xlarge	\$2,01	50%	\$1,02	Windows	Leste dos EUA (Norte da Virgínia)	Sem taxas iniciais	3

Nome da instância	Taxa de Savings Plans	Economia sob demanda	Tarifa sob demanda	Sistema operacional	Região	Opção de pagamento	Período de vigência
x2iedn.xlarge	\$1,02	20%	\$2,52	Windows com licença incluída + edição SQL Server Enterprise	Leste dos EUA (Norte da Virgínia)	Sem taxas iniciais	3
x2iedn.xlarge	\$0,32	61%	\$0,83	BYOL	Leste dos EUA (Norte da Virgínia)	Sem taxas iniciais	3

Os Savings Plans incluem o sistema operacional e têm um desconto separado para BYOL. Todos eles estão detalhados na [calculadora dos Savings Plans para computação](#).

Modelo de preços de instâncias reservadas

AWS tem outro modelo de preços baseado no compromisso conhecido como Instâncias Reservadas. Esse modelo pode ser problemático se sua computação mudar depois que você já se comprometeu, fazendo com que as Instâncias Reservadas não sejam usadas. Os Savings Plans foram projetados para oferecer reduções de custo semelhantes às [Instâncias Reservadas Padrão e Conversíveis](#), mas com muito mais flexibilidade. Os Savings Plans para computação oferecem preços mais baixos no uso de instâncias do Amazon EC2, independente da família, tamanho, sistema operacional, localização ou região. Eles também permitem a máxima flexibilidade.

A tabela a seguir ajudará você a escolher entre Savings Plans ou instâncias Reservadas.

	Reserved Instance	Savings Plans para instâncias do EC2	Savings Plans para computação
Desconto médio de 1 ano	Até 38%	Até 29%	Até 29%
Desconto médio de 3 anos	Até 58%	Até 73%	Até 60%
Família de instâncias	Fixo	Fixo	Flexível
Tamanho da instância	Fixo (não Linux)	Flexível	Flexível
Geografia	1 região	1 região	Flexível
Sistema operacional	Fixo	Flexível	Flexível
Serviço	Amazon EC2 ou Amazon RDS	Amazon EC2	Amazon EC2, Fargate, Lambda
Opções de pagamento	Tudo, parcial, sem pagamento inicial	Tudo, parcial, sem pagamento inicial	Tudo, parcial, sem pagamento inicial
Limites de instâncias	20 por zona de disponibilidade	Sem limite	Sem limite

Note

Os Savings Plans funcionam oferecendo um desconto com base em um compromisso financeiro por hora. O compromisso financeiro por hora não pode ser cancelado ou alterado durante a vigência do seu plano, mas você pode adquirir Savings Plans adicionais para cobrir o uso adicional. Isso permite que você mantenha um compromisso por hora consistente à medida que sua frota cresce.

Você pode usar ferramentas como o [AWS Cost Explorer](#) ou [Painéis de inteligência da Nuvem AWS](#) para monitorar seu compromisso. O Cost Explorer fornece uma diretriz meta de cobertura que pode ajudar sua organização a planejar sua estratégia de cobertura dos Savings Plans. Se 75% da sua

workload estiver estável, então 75% é uma boa meta. Isso deixa 25% dos gastos sob demanda/variáveis com base em workloads dinâmicas. Se precisar aumentar essa cobertura para 85%, você pode adquirir outro compromisso de Savings Plans para aumentar o compromisso financeiro por hora.

Note

Recomendamos que você compre Savings Plans em vez de Instâncias Reservadas, mas os dois modelos de compromisso podem funcionar juntos se você já comprou Instâncias Reservadas.

Considere um exemplo em que você comprou uma Instância Reservada, mas quer começar a experimentar uma opção de Savings Plans. Há uma lógica para essa combinação a ser aplicada ao seu faturamento final. Confira uma hierarquia que você pode aplicar às suas Contas da AWS:

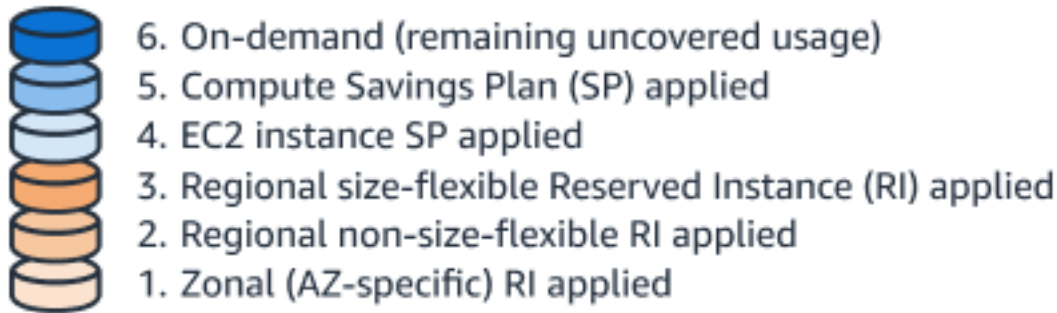
1. A Instância Reservada Zonal se aplica à conta que a possui. Se uma Instância Reservada tiver horas restantes, ela se aplicará ao resto da organização.
2. As Instâncias Reservadas Regionais de tamanho não flexível para Windows se aplicam ao uso correspondente na conta proprietária. Tudo que remanescer será transferido para o restante da organização.
3. As Instâncias Reservadas Regionais com tamanho flexível se aplicam à conta proprietária (primeiro a menor instância dentro da família, e avançando para as instâncias maiores) e depois para o restante da organização.
4. As Instâncias Reservadas Regionais se aplicam a qualquer reserva de capacidade sob demanda não utilizada.
5. Os Savings Plans para instâncias do EC2 se aplicam à conta que os adquiriu.
6. Os Savings Plans para computação se aplicam à conta que os adquiriu.

Note

Os descontos começam com o uso que resulta no maior desconto, e depois reduzem para o menor desconto. As instâncias do Windows tradicionalmente têm um potencial de desconto menor do que o Linux para os tipos de instâncias mais comuns (por exemplo, T3, M6 e

C5). Isso significa que as instâncias do Linux se beneficiam mais do que as instâncias do Windows na maioria dos casos.

O grafo a seguir mostra o preço depois de dividir as Instâncias Reservadas dos Savings Plans. Os Savings Plans para instâncias do EC2 e para computação se aplicam primeiro às instâncias em execução e depois às reservas de capacidade sob demanda não utilizadas.



Cenário de otimização de custos

Esta seção aborda cenários de otimização de custos para hosts dedicados do Amazon EC2 e instâncias do Amazon EC2 que usam um modelo de cobrança com licença incluída.

Hosts dedicados do Amazon EC2

Considere um cenário em que você migrará suas workloads on-premises do Windows para a AWS. Seu data center tem os seguintes servidores:

- Dois servidores com 16 vCPU e 128 GB de RAM
- Dois servidores com 32 vCPU e 164 GB de RAM
- Um servidor com 8 vCPU e 64 GB de RAM
- 16 servidores com vCPU e 32 GB de RAM

Além disso, suponha que você possa trazer sua própria licença AWS porque tem licenças suficientes para trazer. A tabela a seguir mostra as instâncias do servidor que você pode usar na AWS.

Tipo de instância	CPU	RAM	Valor
r5.4xlarge	16	128	2

Tipo de instância	CPU	RAM	Valor
r5.8xlarge	32	256	2
r5.2xlarge	8	64	1
r5.xlarge	4	32	16
			21

Uma análise mostra que essas 21 máquinas virtuais podem ser distribuídas em dois hosts dedicados com um host da família de instâncias R5. A tabela a seguir mostra o custo desses dois hosts dedicados.

Cenário de host dedicado sob demanda	Pagamento inicial	1 mês	1 ano	3 anos	AWS Calculadora de Preços
Sob demanda	Nenhum	\$10.123	\$121.475	\$364.392	AWS Calculadora de Preços estimativa
Savings Plans de 1 ano	Nenhum	\$7.447	\$89.362	–	AWS Calculadora de Preços estimativa
Savings Plans de 3 anos	Nenhum	\$5.476	\$65.712	\$197.128	AWS Calculadora de Preços estimativa
Savings Plans de 3 anos com	\$84.438	\$2.755	\$117.499	\$183.618	AWS Calculadora de Preços estimativa

Cenário de host dedicado sob demanda	Pagamento inicial	1 mês	1 ano	3 anos	AWS Calculadora de Preços
pagamento inicial					

Se você tiver os servidores para os quais deseja migrar AWS, o preço final de um Savings Plans de 1 ano é de \$89.362 em vez de \$121.475 para um preço sob demanda. Isso representa um desconto de 26,5% após um ano. Se você está pensando em ficar em casa AWS por um período mais longo, pode escolher os Savings Plans de 3 anos para obter uma economia ainda maior. Ao final de três anos, você pagará USD 197.128 em vez de USD 364.392. Isso resulta em uma economia de 46% do valor total após três anos.

Instâncias do Amazon EC2 com licenças incluídas

Considere um cenário em que você migrará um único aplicativo de três camadas para AWS e desejará usar as licenças fornecidas pelo. AWS Além disso, suponha que sua aplicação funcione com os seguintes servidores:

- Dois servidores web com dois v CPUs e 4 GB de RAM
- Dois servidores de aplicativos com oito V CPUs e 16 GB de RAM
- Dois servidores de bancos de dados com 16 v CPUs e 64 GB de RAM (usando a edição SQL Server Standard)

A tabela a seguir mostra as instâncias do servidor que você pode usar na AWS.

Tipo de instância	CPU	RAM	Valor
c5.large	2	4	2
c5.2xlarge	8	16	2
r5.2xlarge	8	64	2
			6 servidores

A tabela a seguir mostra o custo desses servidores em AWS.

Licença incluída por AWS	Pagamento inicial	1 mês	1 ano	3 anos	AWS Calculadora de Preços
Sob demanda	Nenhum	\$3.912	\$46.950	\$140.849	AWS Calculadora de Preços estimativa
Savings Plans de 1 ano	Nenhum	\$3.466	\$41.952		AWS Calculadora de Preços estimativa
Savings Plans de 3 anos sem pagamento inicial	Nenhum	\$3.189	\$38.264	\$114.804	AWS Calculadora de Preços estimativa
Savings Plans de 3 anos com pagamento inicial	\$112.110	Nenhum	Nenhum	Nenhum	AWS Calculadora de Preços estimativa

Se quiser executar esses servidores em ambientes de produção (24 horas por dia, 7 dias por semana) com preços sob demanda, você terá um custo mensal de USD 3.912. Pagar esse custo mensal equivale a USD 46.950 após um ano, e um total de USD 140.849 após três anos.

Se você escolher o Savings Plans de 1 ano sem pagamento inicial, o custo mensal diminuirá para USD 3.466. No final do primeiro ano, você paga USD 41.952. Isso equivale a um desconto total de 11%. Se você escolher os Savings Plans de 3 anos sem pagamento inicial, o custo mensal diminuirá para USD 3.189. Ao final de três anos, você pagará USD 114.804. Isso proporciona uma economia de 18,5%.

Recomendações de otimização de custos

Ambos os cenários ajudam você a economizar dinheiro ao planejar e prever suas workloads na AWS. É importante reconhecer que o desconto no segundo cenário é menor em comparação com o primeiro cenário. No segundo cenário, o preço da licença está incluído no preço do servidor em nuvem. AWS não oferece desconto no preço da licença, mas você sempre pode trazer suas licenças (em cenários específicos) e AWS sempre garantir o melhor compute/instance preço.

Recomendamos que você faça o seguinte para controlar seus AWS gastos com recursos de computação e instância:

- Acessar recomendações
- Personalizar as recomendações de acordo com suas necessidades
- Analisar o compromisso por hora

Acessar recomendações

Você pode usar o [console do Amazon EC2](#) para acessar as recomendações para seus Savings Plans. Você também pode baixar as recomendações dos Savings Plans no formato CSV. Para obter mais informações, consulte [Monitoring your Savings Plans](#) na documentação dos Savings Plans.

Personalizar as recomendações de acordo com suas necessidades

Abra o [console do Amazon EC2](#), expanda a seção Instâncias e escolha Savings Plans. Esta página mostra os preços de instâncias e computação antes e depois de fazer uma recomendação. Você também pode ajustar os seguintes fatores para sua recomendação:

- Vigência: por exemplo, de 1 a 3 anos
- Opção de pagamento: por exemplo, com pagamento inicial, pagamento parcial ou sem pagamento inicial
- Histórico: por exemplo, os últimos 7, 30 ou 60 dias

Analisar o compromisso por hora

Usando o mesmo exemplo, suponha que você tenha uma instância em execução 24 horas por dia, 7 dias por semana. A recomendação é usar uma opção de Savings Plans. De acordo com o tamanho, você tem um preço sob demanda de USD 120/hora. Você tem a opção de assumir o compromisso de USD 90/hora, mas isso pode variar de acordo com a região, instância e opção de

aquisição. Neste exemplo, você pode economizar 25% em comparação com o custo sob demanda. Você também pode monitorar sua utilização e cobertura, se estiverem abaixo do limite definido, e configurar um alerta quando o orçamento terminar.

Analisar recomendações

Recomendamos que você analise cuidadosamente as recomendações dos Savings Plans. A AWS não mudará nada sem sua permissão. São apenas recomendações e cabe a você aplicá-las ou não.

Adquirir um plano

Abra o [console do Amazon EC2](#), expanda a seção Instâncias e escolha Savings Plans. Em seguida, escolha Adquirir Savings Plans. Com base em seus requisitos, você pode selecionar as seguintes opções: prazo, região, família de instâncias, compromisso por hora, opção de pagamento e até mesmo data de início. Você pode escolher entre Compute Savings Plans, EC2 Instance Savings Plans e AI Savings Plans SageMaker. Para obter mais informações, consulte [Purchasing Savings Plans](#) na documentação dos Savings Plans.

Obter um relatório de utilização

Depois de adquirir um Savings Plans, você pode obter um relatório de utilização. O relatório ajuda você a verificar sua utilização, ver se o plano adquirido é suficiente para cobrir e maximizar o desconto e cancelar ou adicionar novos descontos. Esse relatório pode ser exportado para outros formatos, como CSV. Para obter mais informações, consulte [Using the utilization report](#) na documentação do Savings Plans.

Seguir as práticas recomendadas de aquisição

Recomendamos que você siga estas práticas recomendadas antes de adquirir Savings Plans:

- Use o [AWS Trusted Advisor](#) para remover recursos ociosos do EC2.
- Faça qualquer dimensionamento correto antes adquirir Savings Plans.
- Estabeleça uma taxa por hora que você mantenha de forma consistente por 30 a 60 dias.
- Adquira um compromisso para cobrir a taxa por hora consistente com a qual sua organização se sinta confortável. Considere as flutuações na demanda ou na temporada.
- Escolha uma revisão trimestral do orçamento dos Savings Plans para manter uma taxa consistente (por exemplo, meta de cobertura de 70% para a cobertura dos Savings Plans). Se a taxa cair abaixo da cobertura desejada, adquira uma opção de Savings Plans adicional como true-up para atingir sua meta de cobertura.

Recursos adicionais do

- [Savings Plans para instâncias reservadas do Amazon EC2](#) (AWS whitepapers)
- [Entendendo como os Savings Plans se aplicam ao seu AWS uso](#) (documentação do Savings Plans)
- [Anúncio do faturamento por segundo para instâncias EC2 do Windows Server e do SQL Server](#) (documentação)AWS
- [AWS Série de otimização de custos: Vídeo da Savings Plans | Amazon Web Services](#) () YouTube

Monitore os custos usando AWS ferramentas

Visão geral do

A visibilidade dos custos é um fator-chave para otimizar os custos. AWS tem várias ferramentas que você pode usar para visualizar custos e criar alertas em reação a esses custos. Isso inclui ferramentas que ajudam você a monitorar e relatar seus gastos. AWS Budgets Esta seção aborda formas específicas de monitorar os gastos do Windows na AWS , para que você possa acompanhar e reagir de acordo com seus requisitos de orçamento. Isso inclui adicionar as tags necessárias aos recursos do Windows EC2. Essas tags permitem que você monitore adequadamente o Windows EC2 e outros serviços da Microsoft usando o AWS Budgets.

Ao monitorar os gastos e criar alertas com AWS ferramentas, você pode estar mais informado sobre os gastos atuais, os gastos projetados e as anomalias dos gastos. Se você usar [Savings Plans](#) para ajudar a reduzir o preço de sua instância do EC2 por hora, recomendamos que confira a utilização geral e a cobertura dos Savings Plans. Isso pode ajudar a garantir que você esteja economizando de modo permanente. Você pode usar AWS Cost Explorer para visualizar o inventário do Savings Plans e obter recomendações para Savings Plans adicionais com base no uso anterior. Você também pode monitorar gastos específicos usando o [AWS Budgets](#) e configurando o [AWS Cost Anomaly Detection](#).

Recomendações de otimização de custos

Recomendamos que você execute as próximas etapas a seguir para otimizar seus custos usando AWS Budgets o Cost Explorer e a detecção de anomalias:

- Marcar recursos do Windows EC2
- Configure alertas usando AWS Budgets

- Habilitar a Detecção de Anomalias em Custos
- Obter uma análise de gastos em tempo real
- Conferir os gastos com a licença incluída do Windows usando o Cost Explorer

Marcar recursos do Windows EC2

Para monitorar seus AWS gastos com eficiência, você deve estabelecer uma [estratégia de marcação](#) para as cargas de trabalho que deseja monitorar. Isso é importante para que você possa agrupar recursos categoricamente e receber notificações sobre gastos específicos, em vez de gastos de uso geral. Você pode usar recursos de marcação que não apenas ajudam no custo, mas também podem ser usados para outros fins, como o [AWS Systems Manager Automation](#). Além disso, recomendamos que você implemente algum gerenciamento para as [tags necessárias](#).

Para monitorar seus gastos no AWS Budgets Cost Explorer e no Cost Anomaly Detection, você deve garantir que as tags adequadas estejam em vigor. Você pode usar tags para configurar um orçamento específico para itens que correspondam a essas tags, para que você receba alertas quando os gastos aumentarem.

Por exemplo, você pode usar uma tag simples como Key=OS Value=Windows. Isso reúne todas as suas instâncias do Windows em um grupo no qual você pode monitorar os gastos. Você também pode usar tags para outros itens, como o Systems Manager. Depois de criar uma tag, você deve ativá-la para o controle de custos. Considere adicionar uma [AWS Config regra que monitore as tags](#) anexadas a determinados recursos. AWS Config pode alertá-lo se houver recursos em execução que não contenham as tags apropriadas, que fornecem uma representação precisa de seus gastos com o Windows EC2.


Depois de inserir as tags, você pode criar um orçamento personalizado no AWS Billing. Isso fornece visibilidade de seus gastos com o Windows EC2. Você pode definir um orçamento diário ou mensal.

Configure alertas usando AWS Budgets

Neste cenário de exemplo, você cria um orçamento diário para o Windows EC2. É um orçamento recorrente que usa a opção de ajuste automático para monitorar seus gastos e ajustar o orçamento adequadamente. Se você tiver um ambiente estático, poderá usar um orçamento fixo. Certifique-se de escolher um intervalo de tempo de linha de base (por exemplo, 30 dias).

1. Faça login no Console de gerenciamento da AWS e abra o [AWS Cost Management console](#).
2. No painel de navegação, selecione Orçamentos.

3. Na parte superior da página, escolha Criar orçamentos.
4. Em Configuração de orçamento, escolha Personalizar (avançado).
5. Em Tipos de orçamento, escolha Orçamento de custo. Em seguida, escolha Próximo.
6. Em Detalhes, em Nome do orçamento, insira o nome do orçamento. Por exemplo, Gastos com o Windows EC2.
7. Em Definir valor do orçamento, em Período, escolha Diariamente.
8. Em Tipo de renovação de orçamento, escolha Orçamento recorrente para um orçamento que é redefinido após o período do orçamento.
9. Em Data de início, escolha a data de início ou o período para começar a rastrear o valor do orçamento.
10. Em Método de orçamento, escolha Ajuste automático (novo).
11. Em Intervalo de tempo de linha de base, escolha Intervalo personalizado e insira 30 dias.
12. Escolha Próximo.
13. Na seção Escopo do orçamento, selecione Filtrar dimensões de AWS custo específicas. É aqui que as tags são usadas para criar as dimensões adequadas. AWS Budgets não suporta o tipo de plataforma como uma opção em seus filtros. Por esse motivo, você deve aplicar as tags do SO.
14. Escolha Adicionar filtro e depois selecione a opção Tag em Dimensões.
15. Escolha a tag do SO, e depois escolha o valor do Windows para criar um orçamento para a tag.
16. Escolha Próximo.
17. Na página Configurar alertas, escolha Adicionar um limite de alerta. Aqui você configura dois alertas: um para um limite de 50%, e outro para um limite de 100%. Se o alerta de limite de 50% for violado antes da metade do mês, ele vai gerar um aviso. Dessa forma, você pode verificar se seus gastos estão acima do esperado e reagir antes de chegar ao final do mês.
18. Em Limite, insira 50 e selecione % do valor orçado.
19. Em Gatilho, escolha Real.
20. Em Destinatários do e-mail, insira um endereço de e-mail. Adicione outro alerta para um limite de 100.

 Note

Este exemplo usa uma notificação por e-mail para o alerta, mas você também pode usar outras abordagens, como o [Slack](#).

Habilitar a Detecção de Anomalias em Custos

Você pode usar suas tags de custo para configurar alertas de gastos que sejam uma anomalia. Por exemplo, você pode usar o [AWS Cost Anomaly Detection](#) para criar monitores para seus gastos e ser alertado quando o sistema detectar gastos anormais em sua conta.

Para configurar um monitor e alertas para as tags Key=OS e Value=Windows que você criou anteriormente, faça o seguinte:

1. Faça login no Console de gerenciamento da AWS e abra o [AWS Cost Management console](#).
2. No painel de navegação, escolha Detecção de Anomalias em Custos.
3. Escolha a guia Monitores de custo e, em seguida, Criar monitor.
4. Na Etapa 1, escolha a Tag de alocação de custos como seu tipo de monitor.
5. Em Chave da tag de alocação de custos, escolha Gastos com o Windows EC2.
6. Em Valor da tag de alocação de custos, escolha Windows.
7. Em Nomear seu monitor, insira Gastos com o Windows EC2.
8. Escolha Próximo.
9. Para criar uma assinatura para os alertas, selecione Criar uma nova assinatura. Se você tiver assinaturas existentes, selecione Escolher uma assinatura existente.
10. Em Nome da assinatura, insira Anomalia de gastos do Windows EC2.
11. Em Frequência de alertas, escolha Resumos diários.
12. Em Destinatários do alerta, insira um endereço de e-mail.
13. Escolha Adicionar limite. Em Limite, insira 10 e selecione a porcentagem acima da velocidade esperada.
14. Escolha Criar monitor.

Obter uma visualização em tempo real dos gastos

Um alerta é uma ferramenta útil para monitorar seus gastos com o Windows EC2, mas você deve usar o Cost Explorer se quiser uma visualização em tempo real dos gastos. Assista a este vídeo para saber como o Cost Explorer permite que você analise e reduza seus custos do EC2. Para obter mais informações, assista ao vídeo [AWS Supports You | Understanding and Reducing Your EC2 Costs](#) em YouTube.

Visualizar os gastos com a licença incluída do Windows

Você pode ver os gastos de instâncias EC2 Windows em sua conta usando o Cost Explorer. Para conferir os gastos com a licença incluída do Windows, você deve definir os seguintes [filtros](#) corretos no Cost Explorer:

- Em Plataforma, escolha Windows (Amazon VPC). Para operação de API, escolha:0002RunInstance. Esse é o AWS Billing código para instâncias do Windows EC2 incluídas na licença.
- Se você quiser ver seus gastos com a instância BYOL, RunInstance altere:0002 para:0800. RunInstance Este é o código de cobrança do Windows EC2 BYOL.

Com essa visibilidade no Cost Explorer, você pode filtrar rapidamente seus custos para exatamente o que você está gastando no Windows EC2. Se você quiser se aprofundar ainda mais em seus AWS gastos, use AWS Cost and Usage Report para filtrar os gastos no nível de instância individual. Você também pode gerar relatórios que podem ser visualizados no Amazon Quick e criar painéis personalizados.

Para obter mais informações, assista ao [vídeo AWS Apoia você - Visualizando seus relatórios de custo e uso](#) em YouTube.

Recursos adicionais do

- [Configurando as tags necessárias com AWS Config](#) (AWS Config documentação)
- [AWS Budgets Tutorial - Configurar alertas para AWS Billing | Amazon Web Services](#) (YouTube)
- [AWS Cost and Usage Report Biblioteca de consultas](#) (AWS Well-Architected Labs)

SQL Server

Os clientes executam cargas de trabalho da Microsoft AWS há mais de 15 anos, mais do que qualquer outro provedor de nuvem. Isso ocorre principalmente porque AWS tem a maior experiência com aplicativos da Microsoft na nuvem e oferece a melhor plataforma para Windows Server e Microsoft SQL Server nas seguintes áreas:

- Maior performance e confiabilidade
- Melhores serviços de segurança e identidade
- Mais suporte à migração
- Os recursos mais abrangentes e profundos
- Menor custo total de propriedade (TCO)
- Opções de licenciamento flexíveis

AWS oferece suporte a tudo o que é necessário para criar e executar aplicativos do Windows que dependem do SQL Server, incluindo Active Directory, .NET, SQL Server, área de trabalho do Windows como serviço e todas as versões compatíveis do Windows Server. Com experiência comprovada, AWS pode ajudá-lo a levantar e mudar, refatorar ou até mesmo modernizar suas cargas de trabalho do Windows com facilidade.

Esta seção do guia abrange os seguintes tópicos:

- [Escolher uma solução de alta disponibilidade e recuperação de desastres](#)
- [Compreender o licenciamento do SQL Server](#)
- [Selecionar a instância certa do EC2 para as workloads do SQL Server](#)
- [Consolidar instâncias](#)
- [Comparar edições do SQL Server](#)
- [Avaliar a edição SQL Server Developer](#)
- [Avaliar o SQL Server no Linux](#)
- [Otimizar as estratégias de backup do SQL Server.](#)
- [Modernizar bancos de dados do SQL Server.](#)
- [Otimizar o armazenamento para o SQL Server](#)
- [Otimizar o licenciamento do SQL Server usando o Compute Optimizer](#)
- [Otimizar o dimensionamento do SQL Server usando o Compute Optimizer](#)

- [Analise Trusted Advisor as recomendações para cargas de trabalho do SQL Server](#)

Escolher uma solução de alta disponibilidade e recuperação de desastres

Visão geral do

Recomendamos que você projete uma arquitetura para sua implantação do SQL Server na AWS que atenda às suas necessidades de negócios e, ao mesmo tempo, atenda aos [objetivos de recuperação de desastres \(DR\)](#), incluindo seu objetivo de tempo de recuperação (RTO) e objetivo de ponto de recuperação (RPO). As soluções a seguir podem ajudar você a projetar a arquitetura certa para o SQL Server no Amazon Elastic Compute Cloud (Amazon EC2) e, ao mesmo tempo, a otimizar os custos para suas workloads do SQL Server.

- Grupos de disponibilidade do SQL Server Always On — os grupos de disponibilidade do SQL Server Always On fornecem alta disponibilidade e recuperação de desastres (HA/DR) solutions for SQL Server databases. An availability group consists of a set of user databases that fail over together. Always On availability groups also provide redundancy at the database level, but don't require shared storage—each replica has its own local storage. You can deploy this feature as an HA/DRsolução). Para obter mais informações, consulte [What is an Always On availability group?](#) na documentação da Microsoft.
- Instâncias de cluster de failover (FCI) do SQL Server Always On — O SQL Server Always On FCIs usa o Windows Server Failover Clustering (WSFC) para fornecer HA no nível da instância do SQL Server. FCIs exigem armazenamento compartilhado para hospedar bancos de dados. Você pode usar o armazenamento em bloco compartilhado ou o armazenamento compartilhado de arquivos. Por exemplo, você pode usar o Amazon FSx para Windows File Server ou o Amazon FSx para NetApp ONTAP como uma solução de armazenamento compartilhado com várias zonas de disponibilidade. Para obter mais informações, consulte [Always On Failover Cluster Instances \(SQL Server\)](#) na documentação da Microsoft.
- SIOS DataKeeper — O SIOS DataKeeper pode ajudá-lo a atender aos requisitos de HA e DR ao habilitar uma FCI do SQL Server que abrange as zonas de disponibilidade e. Regiões da AWS O SIOS DataKeeper cria uma SAN virtual em cluster usando volumes locais do Amazon Elastic Block Store (Amazon EBS) e usa a replicação síncrona entre as zonas de disponibilidade para HA, enquanto usa a replicação assíncrona entre regiões e para recuperação de desastres. Para obter mais informações, consulte [High Availability Protection for Windows Applications](#) na documentação do SIOS.

- Grupos de disponibilidade distribuídos: são um tipo especial de grupo de disponibilidade que se estende por dois grupos de disponibilidade Always On separados. Um grupo de disponibilidade pode residir em duas regiões separadas (por exemplo, us-east-1 e us-west-1). Você pode considerar um grupo de disponibilidade distribuído como um grupo de disponibilidade de grupos de disponibilidade, pois os grupos de disponibilidade Always On subjacentes estão configurados em dois clusters diferentes do WSFC. A edição SQL Server Enterprise é necessária para implantar grupos de disponibilidade distribuídos. Para obter mais informações, consulte [Distributed availability groups](#) na documentação da Microsoft.
- Envio de logs: você pode implementar o envio de logs para proteger seus bancos de dados em várias regiões, no caso raro de uma região ser afetada e ficar indisponível. Dependendo da transação e da frequência de envio de logs, você pode obter RPO e RTO em questão de minutos. Para obter mais informações, consulte [About Log Shipping \(SQL Server\)](#) na documentação da Microsoft.
- AWS Elastic Disaster Recovery— O Elastic Disaster Recovery é um aplicativo de software como serviço (SaaS) que gerencia a replicação de servidores de qualquer infraestrutura AWS para fins de DR. Você também pode usar o Elastic Disaster Recovery para replicar o SQL Server em todas as regiões. O Elastic Disaster Recovery é uma solução baseada em agente que replica máquinas virtuais inteiras, incluindo o sistema operacional, todas as aplicações instaladas e todos os bancos de dados em uma área de preparação. Para obter mais informações, consulte [What is Elastic Disaster Recovery?](#) na documentação do Elastic Disaster Recovery.
- AWS Database Migration Service (AWS DMS) — AWS DMS suporta a migração ao vivo de dados de e para AWS, incluindo uma região diferente. Você pode usar esse recurso para configurar uma instância separada do SQL Server em outra região para servir como banco de dados de recuperação de desastres. Para obter mais informações, consulte [O que é AWS Database Migration Service?](#) na AWS DMS documentação.

Grupos de disponibilidade Always On do SQL Server

Se você estiver usando a edição SQL Server Enterprise apenas para um [grupo de disponibilidade Always On](#) de alta disponibilidade, poderá fazer o downgrade para a edição SQL Server Standard aproveitando os grupos de disponibilidade básicos. Você pode reduzir os custos de 65 a 75% usando grupos de disponibilidade básicos em vez de grupos de disponibilidade Always On.

Note

Para obter informações adicionais sobre diferenças de custo entre diferentes edições do SQL Server, consulte a seção [Comparar edições do SQL Server](#) deste guia.

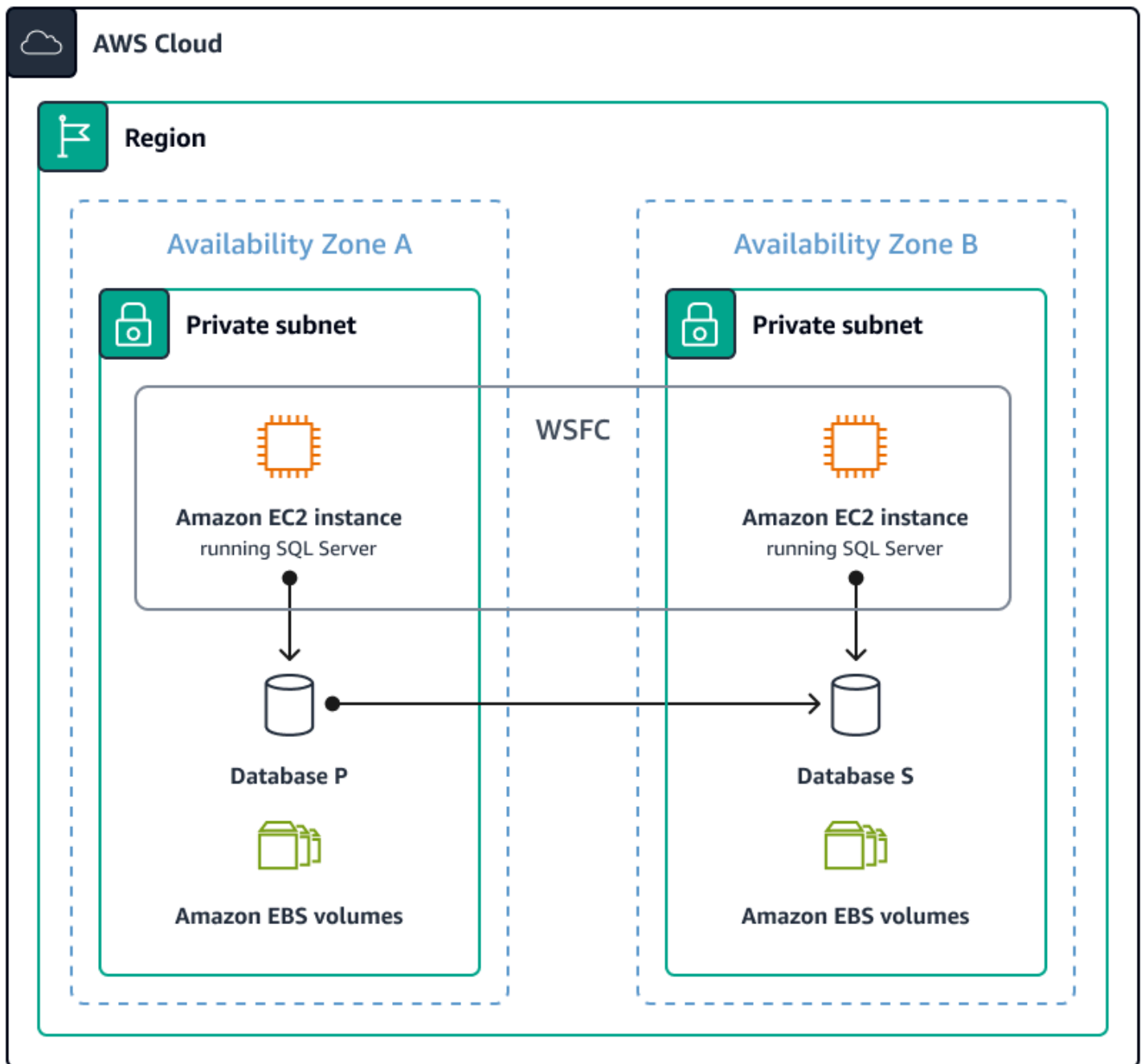
Recursos

- Disponível na edição SQL Server Standard
- Limite de duas réplicas (primária e secundária)
- Sem acesso de leitura na réplica secundária
- Sem verificações de integridade em réplicas secundárias

Limitações

- Suporte somente para um banco de dados de disponibilidade por grupo de disponibilidade
- Grupos de disponibilidade básica não podem fazer parte de um grupo de disponibilidade distribuído


O diagrama a seguir mostra um exemplo de arquitetura para uma solução de cluster de failover do Windows Server.



Instâncias de cluster de failover Always On do SQL Server

Você pode usar instâncias de cluster de failover (FCIs) para garantir operações contínuas do banco de dados, minimizando o tempo de inatividade e reduzindo o risco de perda de dados. FCIs oferece uma solução confiável se você estiver buscando alta disponibilidade para seu banco de dados SQL Server sem uma configuração de réplica de leitura.

Ao contrário dos grupos de disponibilidade, FCIs pode fornecer uma solução de failover confiável sem exigir a edição SQL Server Enterprise. Em vez disso, FCIs exige somente o licenciamento da edição SQL Server Standard. Você pode usar FCIs para reduzir os custos de licenciamento do SQL Server em 65 a 75 por cento.

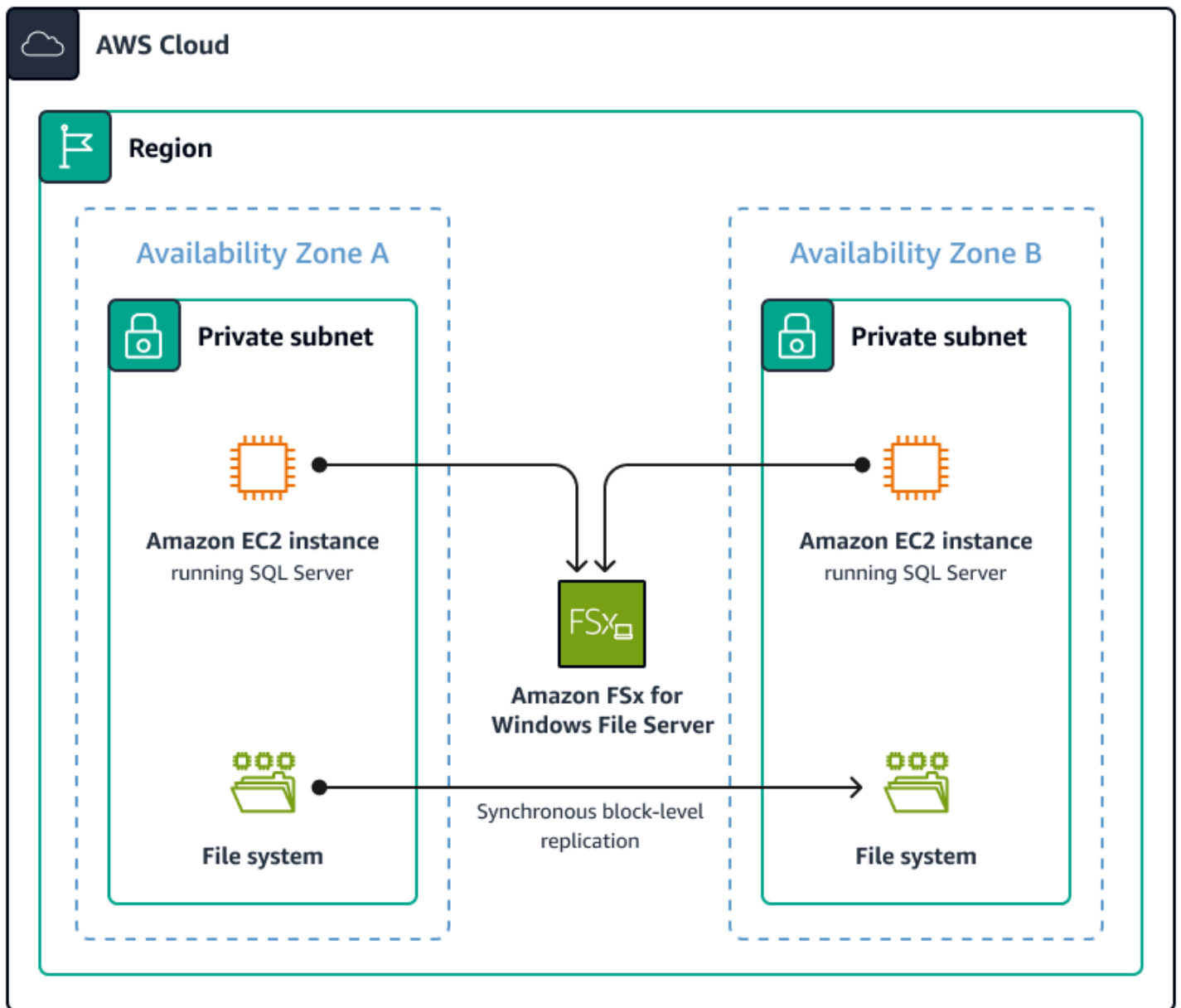
 Note

Para obter informações adicionais sobre as diferenças de custo entre as edições do SQL Server, consulte a seção [Comparar edições do SQL Server](#) deste guia.

Considere o seguinte:

- O Amazon FSx para Windows File Server oferece uma solução poderosa para atender aos seus requisitos de armazenamento compartilhado FCI do SQL Server. Você pode usar o FSx Windows File Server para evitar a necessidade de comprar uma licença para uma solução de replicação de armazenamento e gerenciar o armazenamento compartilhado por conta própria. Isso pode resultar em uma economia de custos significativa de 30 a 40%. Para obter mais informações, consulte a publicação [Simplifique suas implantações de alta disponibilidade do Microsoft SQL Server usando o Amazon FSx para Windows File Server](#) no blog AWS de armazenamento.
- Com o [resumo das vantagens do Software Assurance](#) (PDF disponível para download) e o modelo traga a sua própria licença (BYOL), você pode aproveitar as vantagens do failover passivo, desde que o servidor secundário seja passivo. Isso resulta em economia de custos para o licenciamento do SQL porque você não precisa fornecer licenças para o nó passivo do cluster.

O diagrama a seguir mostra um exemplo de arquitetura para um SQL Server FCI usando o FSx Windows File Server.




SIOS DataKeeper

Recomendamos que você considere os requisitos de armazenamento compartilhado se estiver planejando implantar o SQL Server FCIs no AWS. As instalações on-premises tradicionais geralmente usam uma rede de área de armazenamento (SAN) para atender aos requisitos de armazenamento compartilhado, mas esta não é uma opção viável na AWS. O Amazon FSx para Windows File Server é a solução de armazenamento recomendada para o SQL Server FCI on AWS, mas tem limitações que impedem a adição de servidores de cluster em diferentes Regiões da AWS.

Você pode usar o [SIOS DataKeeper](#) para criar uma FCI do SQL Server que cubra zonas e regiões de disponibilidade e reduza os custos em 58—71 por cento. O SIOS DataKeeper pode ajudar você a obter os benefícios de alta disponibilidade do FCI. Isso torna o SIOS DataKeeper uma solução econômica e confiável para organizações.

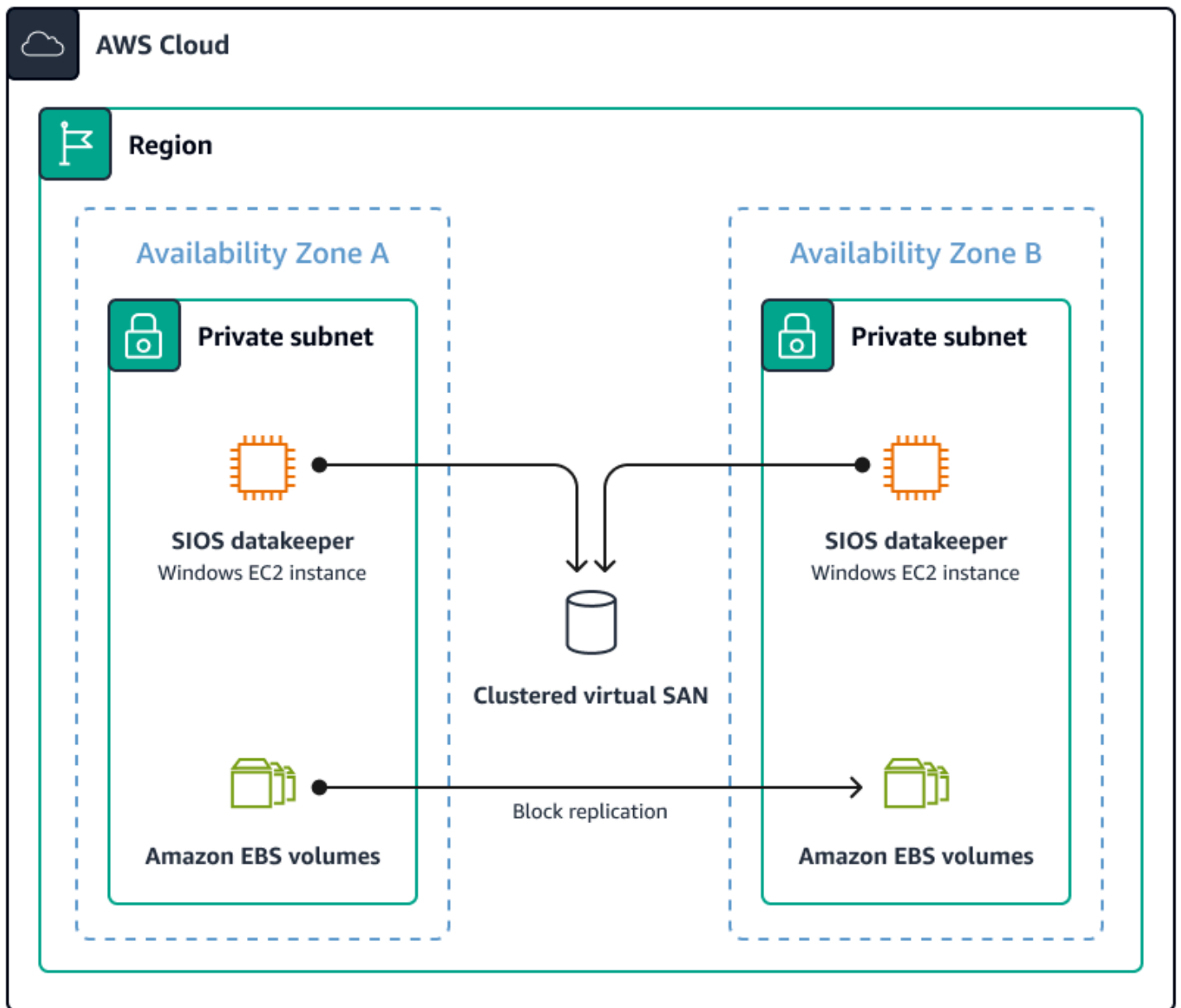
Considere os seguintes benefícios adicionais do uso do SIOS DataKeeper:

- O SIOS DataKeeper cria uma SAN virtual em cluster usando volumes locais do EBS e usa replicação síncrona entre zonas de disponibilidade para alta disponibilidade. Para recuperação de desastres, o SIOS DataKeeper usa replicação assíncrona entre regiões.
- O SIOS DataKeeper fornece recursos de clustering de classe empresarial usando a edição SQL Server Standard. Isso reduz os custos de licenciamento do SQL Server entre 65 e 75% em comparação com a implementação de alta disponibilidade com grupos de disponibilidade Always On do SQL Server que usam a edição SQL Server Enterprise. Com o SIOS DataKeeper, você pode criar um ambiente SQL Server altamente disponível, flexível e econômico que atenda às necessidades da sua organização.

 Note

Para obter informações adicionais sobre as diferenças de custo entre as edições do SQL Server, consulte a seção [Comparar edições do SQL Server](#) deste guia.

O diagrama a seguir mostra um exemplo de arquitetura para uma FCI do SQL Server usando uma solução de SAN virtual em cluster.

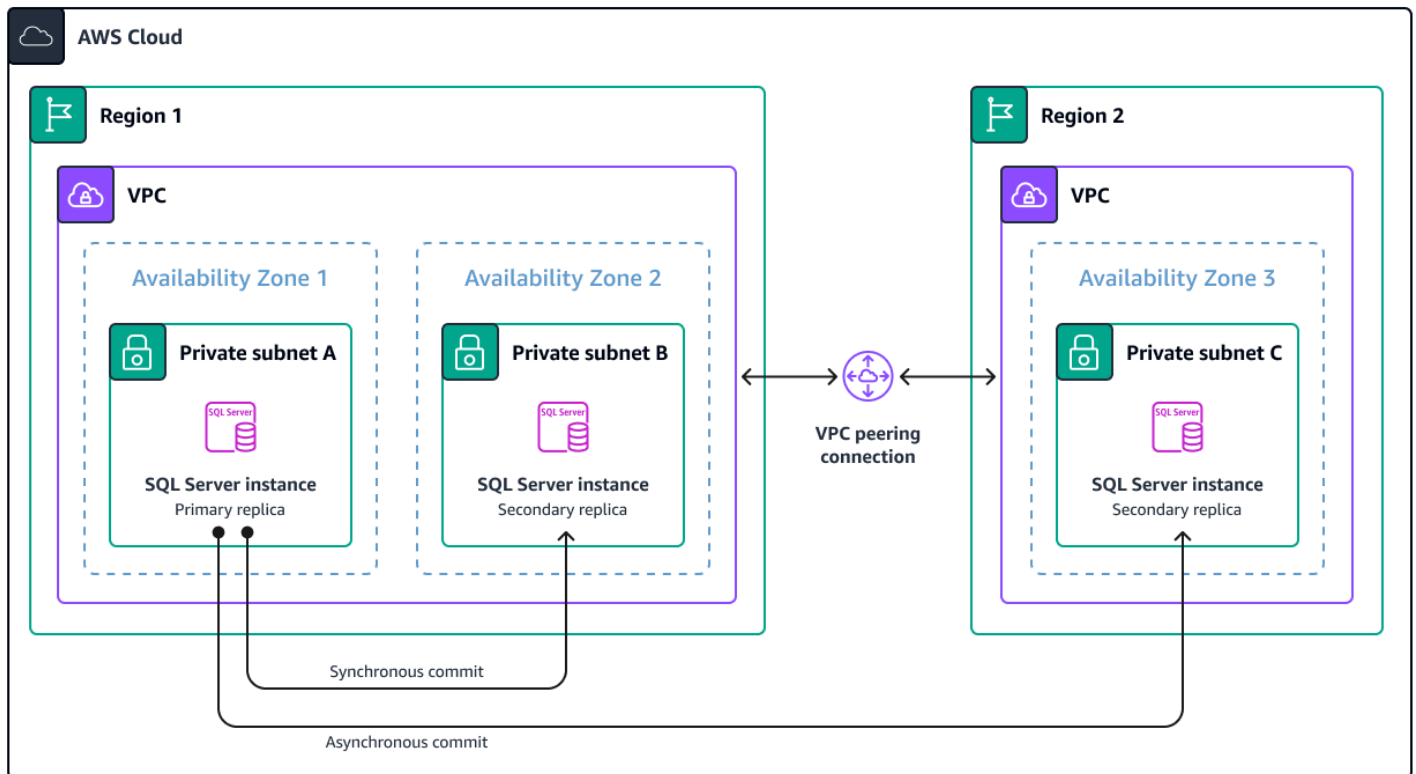


Grupos de Disponibilidade Always On

Você pode usar grupos de disponibilidade Always On para fins de alta disponibilidade e recuperação de desastres. Você pode obter alta disponibilidade implantando o SQL Server em duas zonas de disponibilidade em uma região. Você pode obter a recuperação de desastres estendendo os grupos de disponibilidade em todas as regiões.

O diagrama a seguir mostra um exemplo de arquitetura para uma solução baseada em grupos de disponibilidade Always On. As réplicas na Região 1 do diagrama estão usando um commit síncrono,

que fornece um failover automático do grupo de disponibilidade. A réplica na Região 2 está usando um commit assíncrono, que exigirá um failover manual do grupo de disponibilidade.



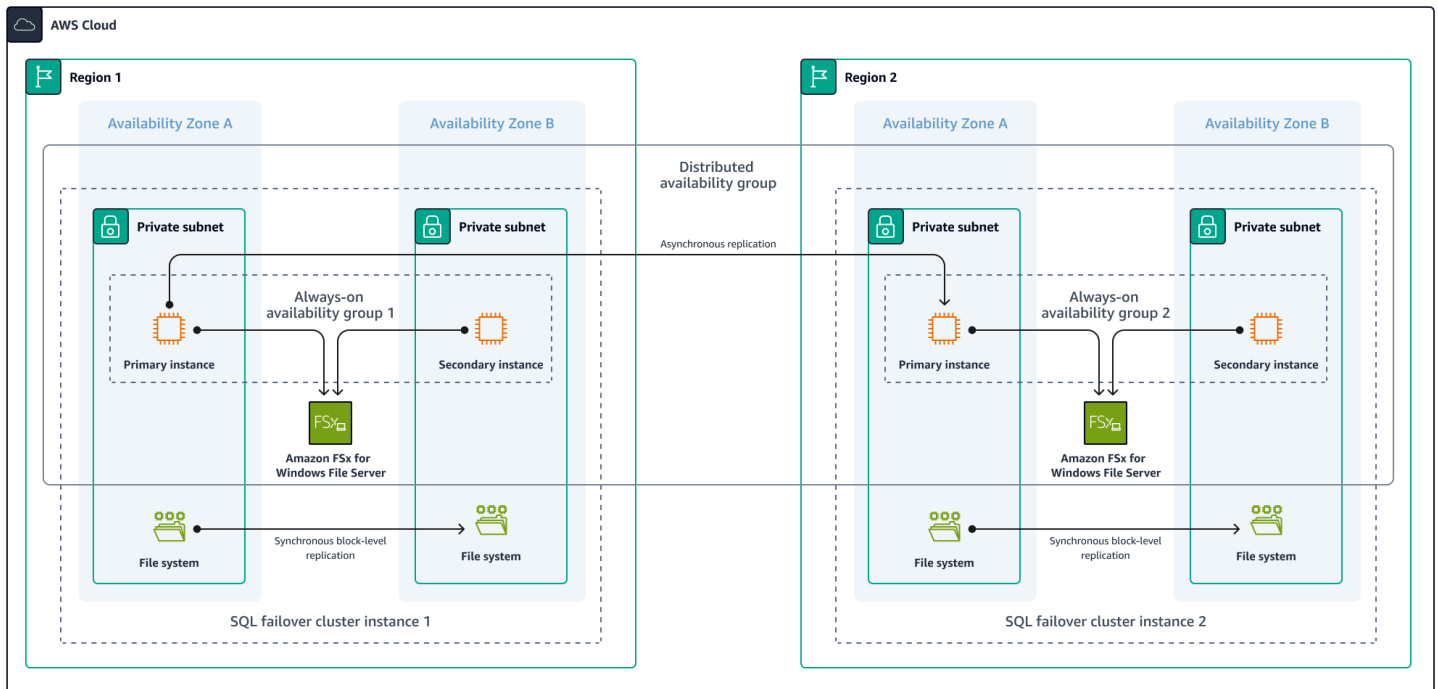
Grupos de disponibilidade distribuídos

Para implantações essenciais do SQL Server em que você não pode comprometer a confiabilidade ou a recuperação de desastres, recomendamos uma abordagem multirregional. Distribuir seus grupos de disponibilidade em várias regiões é a solução mais resiliente para manter a continuidade dos negócios e minimizar o tempo de inatividade.

Essa arquitetura aproveita ao máximo os recursos do Amazon FSx para Windows File Server, incluindo armazenamento compartilhado, replicação síncrona em nível de bloco e SQL Server. Esses recursos possibilitam a criação de um ambiente SQL Server altamente disponível que abrange várias zonas de disponibilidade. Ao replicar essa configuração em outra região, você obtém um sistema totalmente redundante que pode lidar até mesmo com as interrupções mais graves. O que diferencia essa solução é o nível de flexibilidade e segurança que ela oferece. A arquitetura independente de domínio de grupos de disponibilidade distribuídos permite que os servidores de cluster Windows subjacentes se juntem a diferentes domínios do Active Directory, enquanto a autenticação baseada em certificado garante a máxima proteção para seus ambientes SQL Server, e fornece requisitos de RTO e RPO elevados para uma estratégia de DR multirregional. Para obter

informações sobre como criar uma arquitetura multirregional, consulte [Notas de campo: Criando uma arquitetura multirregional para o SQL Server usando FCI e grupos de disponibilidade distribuídos no AWS blog](#) de arquitetura.

O diagrama a seguir mostra um exemplo de arquitetura para uma solução multirregional usando grupos de disponibilidade distribuídos.



Envio de logs


O envio de logs é um método comprovado, confiável e econômico para proteger seus bancos de dados em todas as regiões no caso de uma interrupção inesperada. As organizações usam o envio de logs para proteger seus dados há décadas.

Se você implementar o envio de registros em AWS, poderá obter RPO e RTO em minutos, dependendo da frequência das transações e dos trabalhos de envio de registros. No caso improvável de uma região ficar inacessível, o envio de logs manterá seus dados seguros e recuperáveis.

Considere as seguintes vantagens adicionais de usar o envio de logs:

- Reduza custos e atenda aos requisitos de seus negócios usando o envio de logs para resiliência de recuperação de desastres em todas as regiões. O envio de logs reduz seu TCO porque você só precisa das licenças da edição SQL Server Standard ou SQL Server Web.

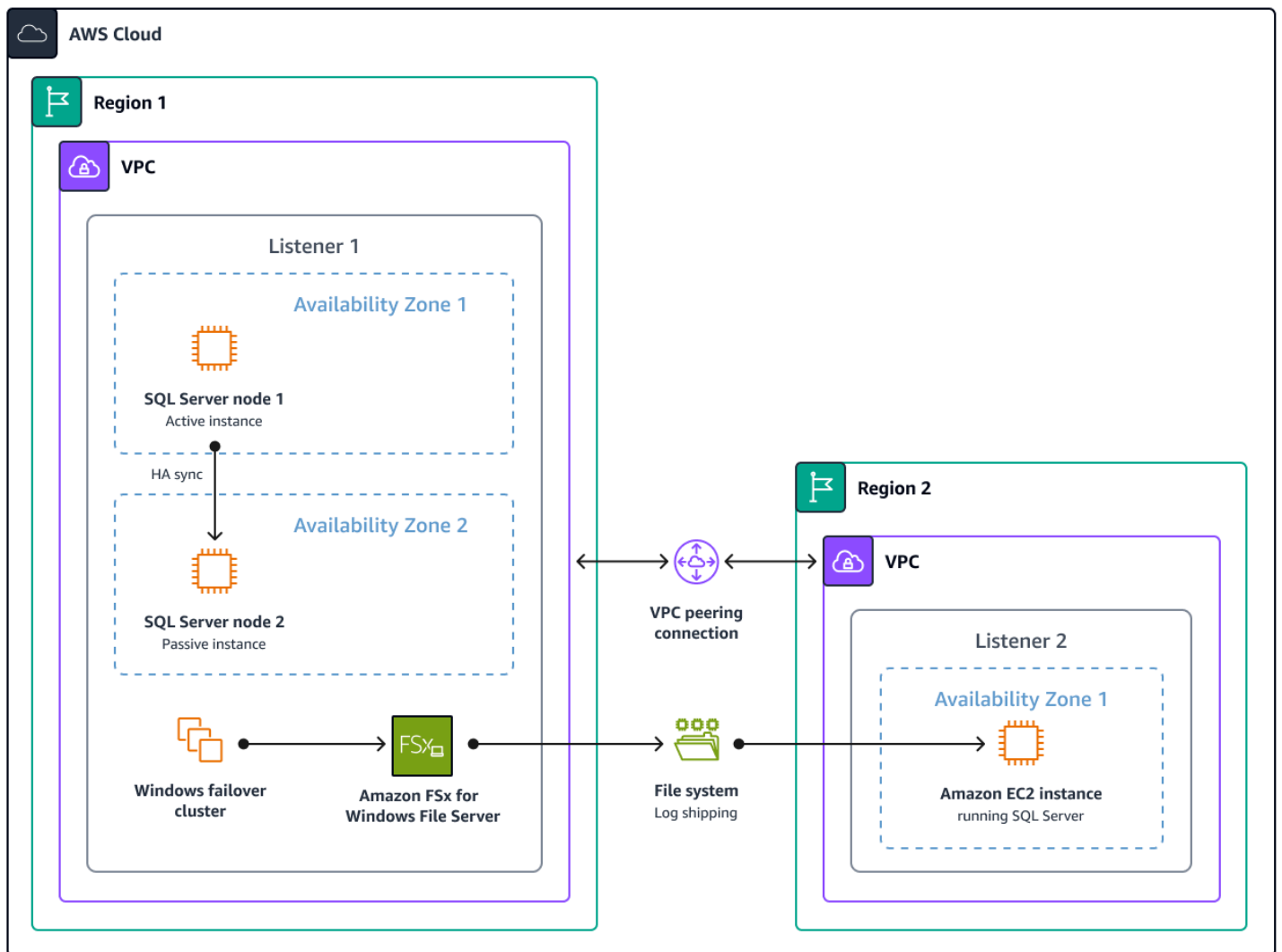
- Remova os custos de licenciamento de um recovery/passive servidor de desastres usando o envio de registros com o [Software Assurance](#) ativo. Somente o primary/active SQL Server precisa ser licenciado quando você usa o envio de registros com o Software Assurance.
- Reduza os custos de licenciamento do SQL Server em 65 a 75% eliminando a necessidade da edição SQL Server Enterprise para configurar grupos de disponibilidade distribuídos entre as regiões. Você pode fazer isso usando o SQL Server Standard Edition e o SQL Server FCIs combinados com o envio de registros para atender aos seus requisitos de recuperação de desastres.

 Note

Para obter informações adicionais sobre as diferenças de custo entre as edições do SQL Server, consulte a seção [Comparar edições do SQL Server](#) deste guia.

Para obter mais informações, consulte [Estender o SQL Server DR usando o envio de registros para a configuração do SQL Server FCI com Amazon FSx para Windows](#) no blog de AWS arquitetura.

O diagrama a seguir mostra um exemplo de arquitetura para uma solução de envio de logs.

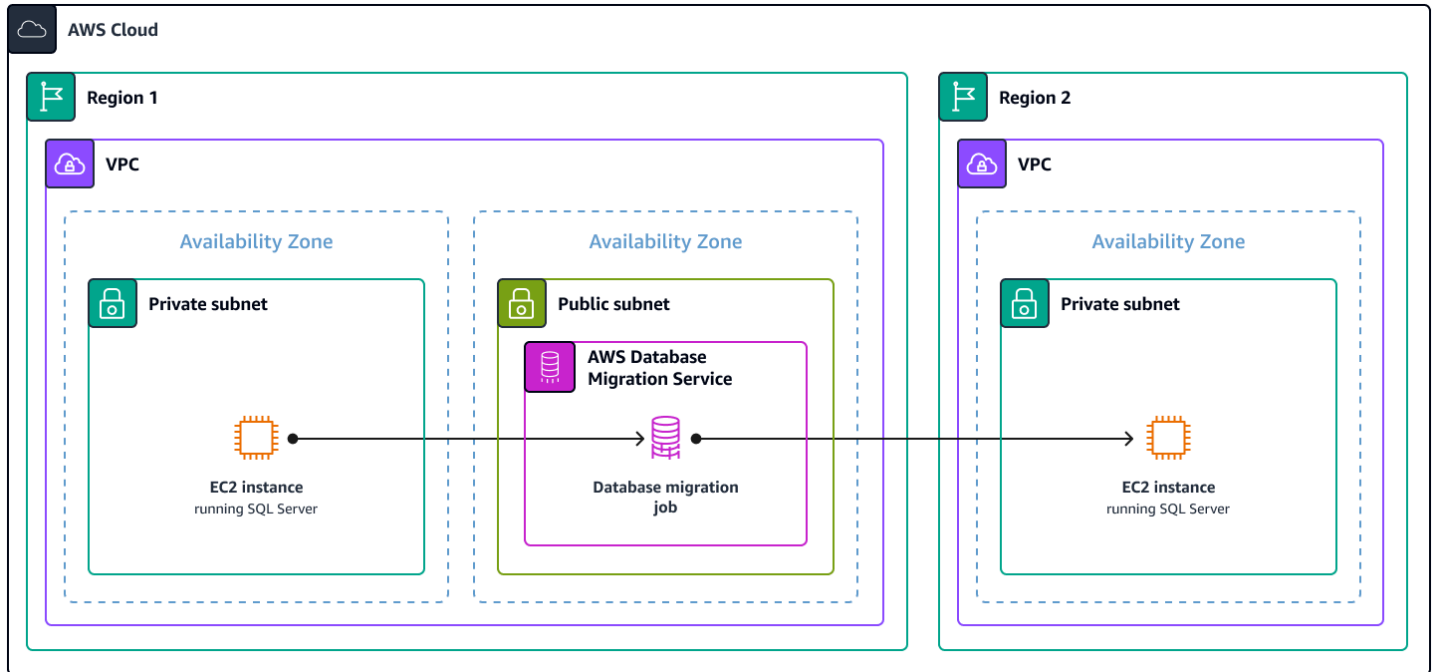


AWS Database Migration Service

Você pode usar AWS Database Migration Service (AWS DMS) para projetar uma HA/DR solução com base nas necessidades do seu aplicativo. AWS DMS permite que você copie dados facilmente para um banco de dados secundário do SQL Server na mesma região (HA) ou entre regiões (DR). Essa abordagem é tecnicamente sólida e permite que você maximize seu investimento na infraestrutura da AWS enquanto otimiza o uso de recursos.

AWS DMS é um serviço econômico. Você é cobrado somente pelos recursos de CPU usados durante o processo de transferência e por qualquer armazenamento adicional de logs. Isso significa que você pode se beneficiar dessa solução sem incorrer em custos adicionais significativos. Você pode usar AWS DMS para garantir que seus dados estejam disponíveis e acessíveis, minimizando os custos associados ao licenciamento e ao uso de recursos.

O diagrama a seguir mostra um exemplo de arquitetura para uma solução baseada no AWS DMS.



AWS Elastic Disaster Recovery

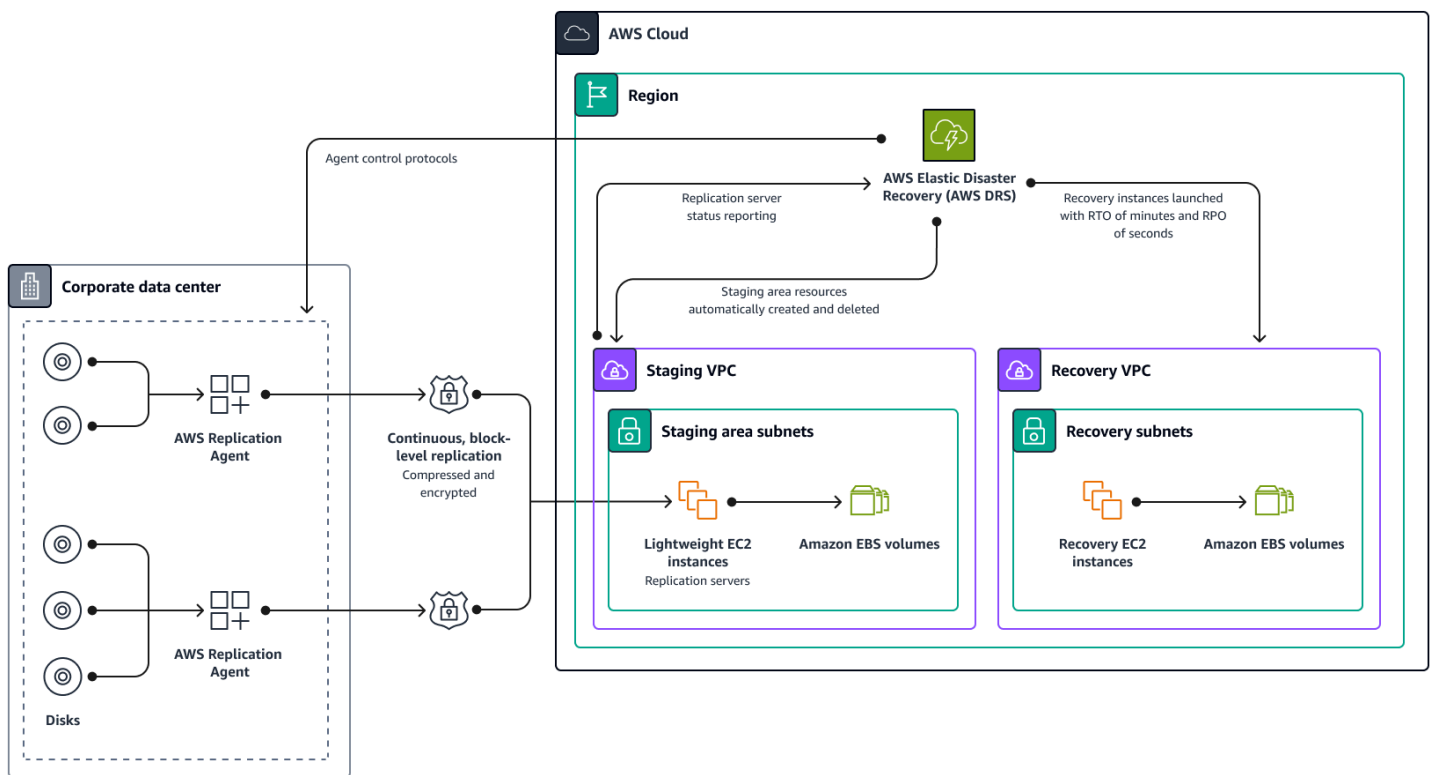
Algumas organizações devem garantir que todas as aplicações comerciais essenciais tenham um plano de recuperação de desastres em vigor. No passado, muitas dessas organizações fizeram investimentos substanciais em soluções tradicionais de recuperação de desastres, que exigem que você pré-crie e mantenha toda uma infraestrutura duplicada. Essa abordagem é cara, demorada e difícil de escalar.

Agora, você pode usar AWS Elastic Disaster Recovery para eliminar a necessidade de pré-construir uma infraestrutura de recuperação de desastres. As máquinas de recuperação de desastres não são iniciadas no Elastic Disaster Recovery até que sejam necessárias, então você paga somente pelo que usar quando precisar. Isso significa que você pode reduzir significativamente seus custos de licenciamento de software e computação de alta performance.

Além disso, a área de preparação da solução de recuperação de desastres contém volumes de baixo custo do Amazon Elastic Block Store (Amazon EBS). Os volumes do EBS reduzem ainda mais o custo do provisionamento de recursos duplicados. Isso permite que você reduza seus custos gerais de recuperação de desastres e, ao mesmo tempo, mantenha uma solução de recuperação de desastres robusta e confiável que atenda aos requisitos de sua empresa. Você pode usar o Elastic Disaster Recovery para se concentrar em suas principais atividades comerciais, enquanto AWS cuida da infraestrutura subjacente da sua solução de recuperação de desastres.

Para o SQL Server, você pode usar o Elastic Disaster Recovery como uma opção econômica de recuperação de desastres. O licenciamento do nó passivo em uma arquitetura SQL Server altamente disponível e tolerante a falhas está incluso se você usar o Software Assurance ativo. No entanto, você ainda está pagando pelos custos de computação para que o servidor passivo esteja on-line. Com o Elastic Disaster Recovery, o servidor primário pode se replicar para o ambiente de DR sem a necessidade de manter o Software Assurance ativo e sem ter que pagar pelos custos computacionais da recuperação de desastres. Essa combinação de economias pode reduzir seus custos de recuperação de desastres do SQL Server em 50% ou mais.

O diagrama a seguir mostra um exemplo de arquitetura para uma solução baseada no Elastic Disaster Recovery.



Para obter mais informações, consulte [Como configurar a alta disponibilidade do SQL Server no site de DR que foi restaurado usando AWS Elastic Disaster Recovery](#) o Microsoft Workloads on AWS Blog.

Comparação de custos

A tabela a seguir compara os custos das HA/DR soluções abordadas nesta seção. As seguintes suposições foram feitas para fins dessa comparação:

- Tipo de instância: r5d.xlarge
- Tipo de licença: licença incluída para Windows e SQL Server
- Região: us-east-1

Solução	Alta disponibilidade	Recuperação de desastres	Enterprise Edition	Standard Edition	Custo
Envio de logs	Não	Sim	Sim	Sim	Edição SQL Server Enterprise: USD 32.674,80 (2 nós) Edição SQL Server Standard: USD 14.804,40 (2 nós)
Grupos de Disponibilidade Always On	Sim	Sim	Sim	Sim, mas grupos de disponibilidade básica (2 nós)	Edição SQL Server Enterprise: USD 32.674,80 (2 nós) Edição SQL Server Standard: USD 14.804,40 (2 nós)
Sempre ligado FCIs	Sim	Não	Sim	Sim (2 nós)	Edição SQL Server Standard: USD 14.804,40

Solução	Alta disponibilidade	Recuperação de desastres	Enterprise Edition	Standard Edition	Custo
Grupos de disponibilidade distribuídos	Sim	Sim	Sim	Não	Edição SQL Server Enterprise: USD 65.349,60 (4 nós)

Solução	Alta disponibilidade	Recuperação de desastres	Enterprise Edition	Standard Edition	Custo
Elastic Disaster Recovery	Não	Sim	Sim	Sim	<p>Aproximadamente USD 107,48/mês para replicação de 1 instância e 1 TB de armazenamento</p> <p>Observação: o Elastic Disaster Recovery é cobrado por hora, por servidor de replicação. O custo é o mesmo, independentemente do número de discos, do tamanho do armazenamento, do número de testes ou de ativações de recuperação ou da região para a qual</p>

Solução	Alta disponibilidade	Recuperação de desastres	Enterprise Edition	Standard Edition	Custo
					<p>you are replicating.</p>
SIOS DataKeeper	Sim	Sim	Sim	Sim	<p>Grupos de disponibilidade Always On com o Software Assurance (2 nós, 24 núcleos): USD 213.480</p> <p>Cluster SQL Server de 2 nós em execução na edição SQL Server Standard com SIOS DataKeeper e Software Assurance : \$61.530 (2 nós)</p>
AWS DMS	Não	Sim	Sim	Sim	<p>USD 745,38/ mês para instância r5.xlarge e 1 TB de armazenamento</p>

Recomendações de otimização de custos

Recomendamos que você execute as próximas etapas a seguir para escolher uma HA/DR solução que atenda aos requisitos da sua organização:

- Consulte a seção [Selecionar a instância certa do EC2 para as workloads do SQL Server](#) deste guia.
- Determine os requisitos de IOPS e throughput de suas workloads executando contadores de performance durante picos de workload:
 - IOPS = disco reads/sec + gravações em disco/segundo
 - Taxa de transferência = leitura de bytes/sec disco+bytes/seg de gravação em disco
- Use os seguintes tipos de volume de armazenamento para obter melhor performance e economia de custos:
 - NVMe armazenamento de instâncias tempdb e extensão do buffer pool
 - volumes io2 para arquivos de banco de dados
- Use o [AWS Trusted Advisor](#) para obter recomendações sobre a otimização de custos do SQL Server no Amazon EC2. Você não precisa instalar um agente para Trusted Advisor fazer verificações de otimização do SQL Server. Trusted Advisor inspeciona suas configurações de instância incluídas na licença do Amazon EC2 SQL Server, como CPUs virtual (CPUsv), versão e edição. Em seguida, Trusted Advisor faz recomendações com base nas melhores práticas.
- Use tanto AWS Compute Optimizer para a instância do Amazon EC2 quanto para as recomendações de dimensionamento correto do Amazon EBS.
- Use [AWS Calculadora de Preços](#) para projetar sua HA/DR estratégia para estimativas de custos.
- Para determinar se o downgrade da edição SQL Server Enterprise para a edição SQL Server Standard é uma opção possível, use a visualização de gerenciamento dinâmico [sys dm_db_persisted_sku_features](#) para identificar recursos específicos da edição que estejam ativos no banco de dados atual.

Note

Side-by-side migrações são necessárias para alterações na edição do SQL Server ao usar instâncias EC2 com licença incluída.

- Realize testes semestrais ou anuais de recuperação de desastres para melhor arquitetar um projeto que possa recuperar o banco de dados com o RTO e o RPO definidos. Essa verificação também pode auxiliar na identificação de pontos fracos da arquitetura.

Recursos adicionais do

- [Simplifique suas implantações de alta disponibilidade do Microsoft SQL Server usando o Amazon FSx para Windows File Server](#) (AWS Storage Blog)
- [Notas de campo: Criando uma arquitetura multirregional para o SQL Server usando FCI e grupos de disponibilidade distribuídos](#) (blog de AWS arquitetura)
- [Arquitete uma recuperação de desastres para o SQL Server em AWS: Parte 1](#) (Blog do AWS banco de dados)
- [Alta disponibilidade do Microsoft SQL com Amazon FSx para Windows](#) (YouTube)
- [Maximizing Microsoft SQL Server Performance with Amazon EBS](#) (blog AWS Storage)
- [Comparando seus padrões de armazenamento local com os serviços AWS de AWS armazenamento](#) (Storage Blog)
- [Planejando substituir um NAS de data center pelo Amazon FSx File Gateway](#) (AWS Storage Blog)
- [Otimizando o custo de suas implantações de alta disponibilidade do SQL Server no AWS](#) (AWS Storage Blog)
- [How to set up disaster recovery for SQL Server Always On Availability Groups using AWS Elastic Disaster Recovery](#) (Microsoft Workloads on AWS)
- [Como configurar a alta disponibilidade do SQL Server no local de DR que foi restaurado usando AWS Elastic Disaster Recovery](#) (Microsoft Workloads on AWS)

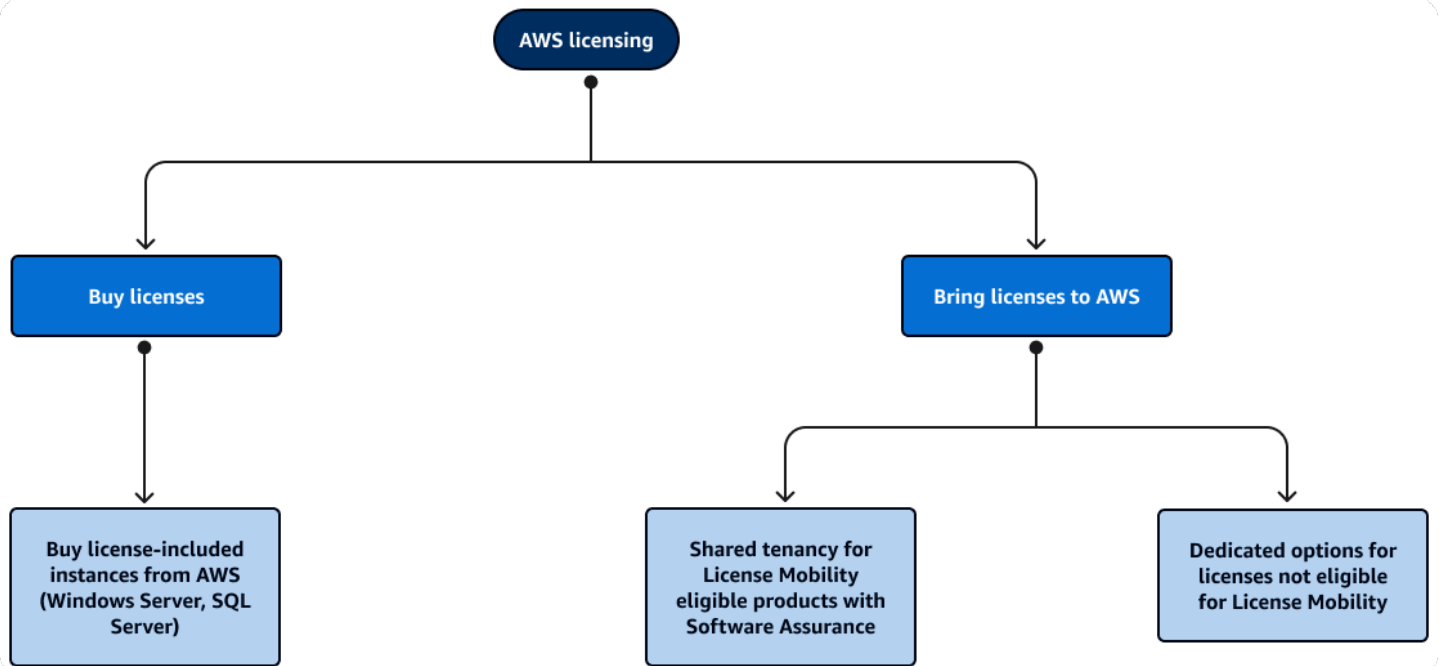
Compreender o licenciamento do SQL Server

Visão geral do

À medida que mais e mais empresas transferem suas workloads para a nuvem, otimizar os custos nas plataformas de nuvem tornou-se uma prioridade máxima. O licenciamento é um dos custos mais significativos associados à execução de cargas de trabalho da Microsoft em AWS. Esta seção explica como otimizar os custos AWS otimizando o licenciamento da Microsoft para o SQL Server.

AWS opções de licenciamento

AWS oferece uma variedade de opções flexíveis de otimização de custos para licenciamento. Essas opções de licenciamento foram projetadas para ajudar você a reduzir custos, manter a conformidade e atender às suas necessidades comerciais.



AWS categoriza as licenças em três tipos principais:

1. Licença incluída: esta opção de licenciamento permite que você compre e use licenças sob demanda, pagando somente pelo que usa. A opção de licença incluída é ideal para cenários em que você precisa de flexibilidade no uso do licenciamento e deseja evitar custos iniciais. Você pode escolher entre uma variedade de produtos Windows Server, SQL Server e outros produtos da Microsoft.
2. Produtos Bring Your Own License (BYOL) com mobilidade de licenças — Essa opção de licenciamento foi projetada para cenários em que você já tem licenças existentes e deseja usá-las na nuvem. AWS permite que os clientes tragam suas próprias licenças para a nuvem por meio do programa [License Mobility](#) da Microsoft. Você pode levar produtos que tenham mobilidade de licenças, como o SQL Server com Software Assurance (SA), para locação compartilhada ou dedicada para reduzir seus custos de AWS instância.
3. Produtos BYOL sem mobilidade de licenças — Para produtos da Microsoft que não têm mobilidade de licenças, como o Windows Server, AWS oferece opções dedicadas para usar esses produtos na nuvem. Além disso, os hosts dedicados oferecem a oportunidade de licenciamento

no nível básico físico. Isso pode economizar 50% ou mais nas licenças necessárias para executar suas workloads. Os hosts dedicados são uma ótima opção para workloads estáveis e previsíveis em execução na maior parte do tempo.

Impacto nos custos de trazer licenças

Trazer licenças pode ter um impacto significativo no custo de execução das workloads da Microsoft na AWS. Se você trazer suas próprias licenças, não precisará pagar por custos adicionais de licenciamento pelas instâncias em execução na nuvem. Isso pode gerar uma economia significativa.

A comparação a seguir mostra o custo mensal sob demanda da execução de uma única instância c5.xlarge 24 horas por dia, 7 dias por semana:

- Windows Server + Edição SQL Server Enterprise: USD 1353/mês (licença incluída)
- Windows Server + Edição SQL Server Standard: USD 609/mês (licença incluída)
- Somente Windows Server: USD 259/mês (licença incluída)
- Somente computação (Linux): USD 127/mês

Em última análise, trazer suas próprias licenças pode ter um impacto significativo no custo de execução das workloads da Microsoft na AWS. Se você usar suas licenças existentes, poderá reduzir os custos de licenciamento e economizar dinheiro em sua fatura geral AWS .

Otimização de licença

Uma Avaliação AWS de Otimização e Licenciamento (AWS OLA) pode ajudá-lo a otimizar seu licenciamento reduzindo os custos de computação e licenciamento. AWS O OLA foi projetado para avaliar seus requisitos de licenciamento para cargas de trabalho em execução AWS ou para cargas de trabalho planejadas para migração. AWS O OLA fornece recomendações sobre como otimizar o uso da licença.

Uma das principais estratégias para otimizar o uso do licenciamento é o [dimensionamento correto das instâncias](#). O dimensionamento correto envolve selecionar o tipo de instância certo para sua workload com base nos requisitos de CPU, memória e armazenamento. Ao escolher o tamanho adequado da instância, você pode garantir que está usando os recursos de maneira econômica. Isso pode gerar uma economia significativa.

Com o licenciamento de software da Microsoft, o número de núcleos em que o software é executado é um fator crítico na determinação dos custos de licenciamento. Por exemplo, as licenças do

Windows Server e do SQL Server geralmente são licenciadas com base no número de núcleos. Ao dimensionar corretamente as instâncias, você pode reduzir o número de núcleos em que o software da Microsoft é executado e, por sua vez, reduzir o custo da instância e o número de licenças necessárias.

Recomendações de otimização de custos

A otimização de licenças é um componente essencial da otimização de custos na AWS. Ao implementar as estratégias certas, você pode reduzir os custos de licenciamento, manter a conformidade e obter o melhor valor possível de seu investimento em licenciamento. Esta seção descreve várias estratégias para a otimização de licenças.

Traga suas licenças qualificadas do Windows Server

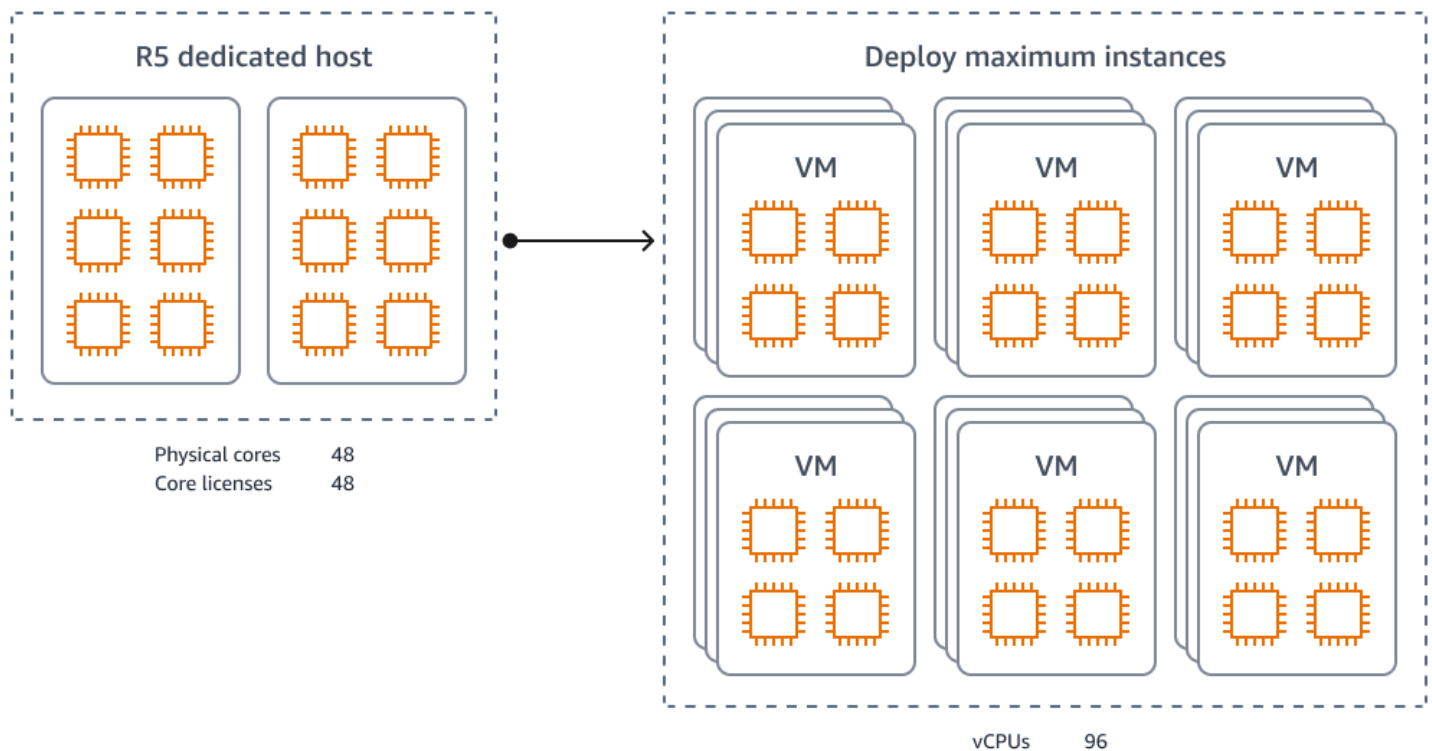
Trazer suas próprias licenças do Windows Server é uma das estratégias mais eficazes para a otimização de licenças. Essa estratégia permite que você aproveite seus investimentos existentes para reduzir seus AWS gastos.

Por exemplo, você pode implantar o Windows Server 2019 e versões anteriores em [hosts dedicados do Amazon EC2](#) se tiver comprado as licenças antes de 10/01/2019, ou se tiver comprado as licenças como true-ups no sob contratos corporativos ativos assinados antes dessa data. Essa regra se baseia em uma alteração que a Microsoft fez em 2019 em seus termos e condições de licenciamento para produtos sem mobilidade de licenças, como o Windows Server, quando implantados em [fornecedores listados](#) (por exemplo, Alibaba ou Google Cloud AWS). De acordo com os novos termos, você não pode trazer suas próprias licenças do Windows Server, AWS mas deve usar instâncias com licença incluída. No entanto, se você comprou licenças perpétuas antes dessa data, ainda poderá implantar essas licenças do Windows Server nos hosts dedicados do Amazon EC2.

Licenças de nível físico

O licenciamento no nível do núcleo físico permite licenciar apenas os núcleos físicos de um host para que você possa implantar um número máximo de instâncias sem afetar o número de licenças necessárias. Normalmente, isso é feito usando o Windows Server Datacenter e a edição SQL Server Enterprise.

Como exemplo, considere um host dedicado R5 com 48 núcleos, o que se traduz em 96 v. CPUs. Se você usa a edição Windows Server Datacenter, você só precisa de 48 licenças. Isso permite que você implante uma combinação de instâncias com até 96 vCPUs, como mostra o diagrama a seguir.



Essa abordagem pode ser especialmente econômica se você tiver workloads suficientes para maximizar o número de instâncias que podem ser executadas em um host. Ao licenciar no nível do núcleo físico, você pode evitar custos adicionais de licenciamento para cada instância e obter o melhor valor possível para seu investimento em licenciamento.

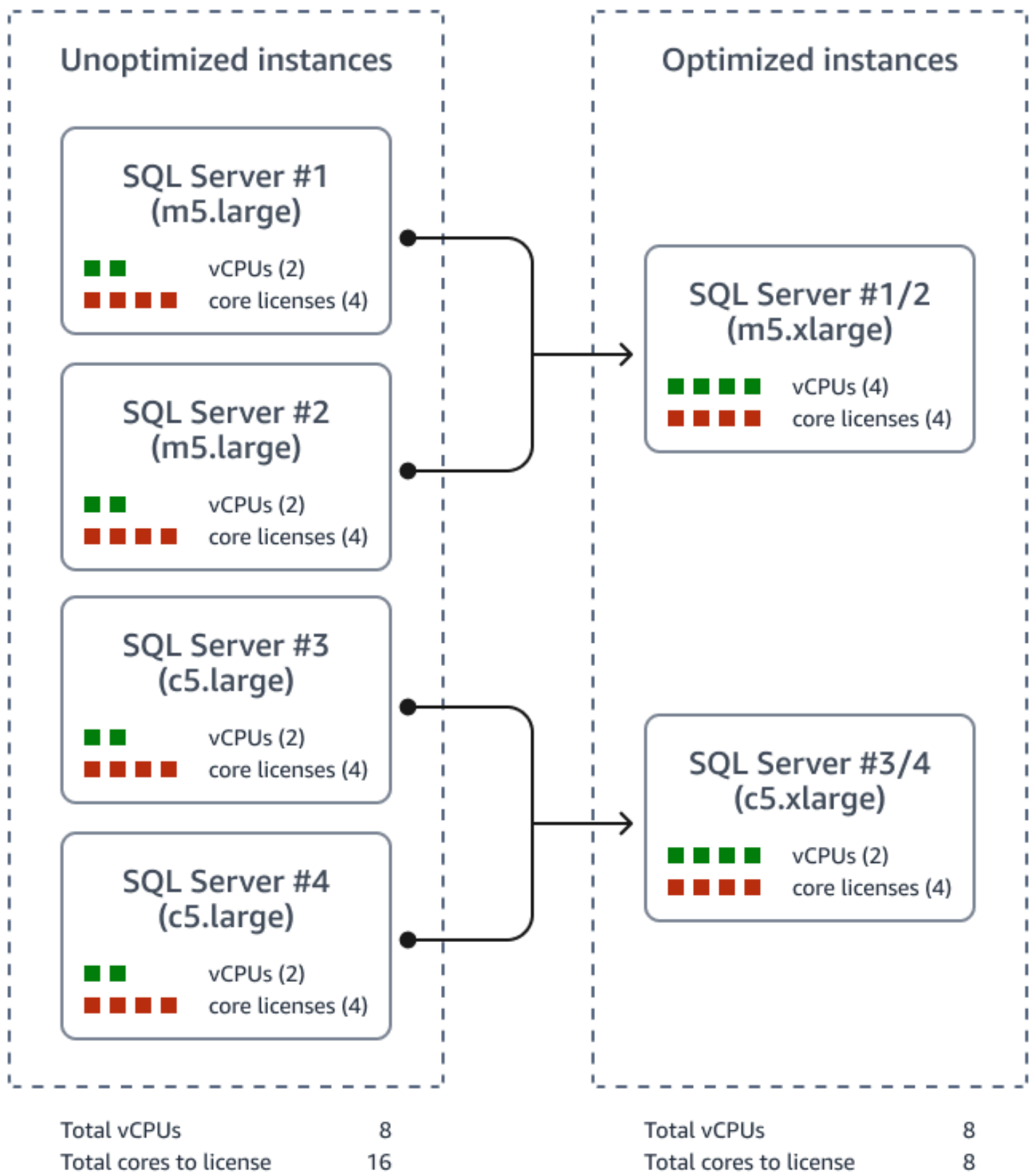
Licença no nível do núcleo físico do SQL Server

Na locação compartilhada, o licenciamento do SQL Server é baseado no número de v CPUs alocado para a instância. Por outro lado, com hosts dedicados, você pode licenciar a edição SQL Server Enterprise no nível do núcleo físico ou no nível da vCPU.

Assim como no exemplo anterior do host dedicado R5, se você licenciar a edição SQL Server Enterprise no nível do núcleo físico, precisará apenas de 48 licenças da edição SQL Server Enterprise para licenciar os hosts. Por outro lado, na locação compartilhada, em que a única opção é licenciar por vCPU, você deve ter 96 licenças da edição SQL Server Enterprise para a mesma workload. Portanto, os hosts dedicados podem economizar até 50% nos custos de licenciamento do SQL Server em comparação com a locação compartilhada. Isso se soma à economia nos custos de instâncias ao trazer licenças qualificadas do Windows.

Consolidar instâncias do SQL Server

A [consolidação do SQL Server](#) é o processo de combinar várias instâncias do SQL Server em um servidor. O SQL Server exige um mínimo de quatro licenças principais por instância, mesmo que a instância tenha apenas duas v. CPUs. Isso significa que a execução do SQL Server em servidores com menos de quatro núcleos pode fazer com que você licencie demais essas instâncias e use mais licenças do que o necessário.



Por exemplo, consolidar duas instâncias com dois v CPUs cada em uma única instância com quatro v CPUs pode reduzir a exigência de licenciamento em 50%. Isso ocorre porque são necessárias apenas quatro licenças de núcleo, em vez de oito.

Para obter mais informações sobre consolidação, consulte a seção de [Consolidação do SQL Server](#) deste guia.

Fazer downgrade das edições do SQL Server

[Alterar as edições do SQL Server](#) pode ser uma estratégia fundamental para otimizar o uso do licenciamento e reduzir custos. A edição Enterprise do SQL Server é consideravelmente mais cara do que a edição Standard, portanto, realizar o downgrade pode resultar em economias de custo significativas.

A criptografia transparente de dados (TDE) e os grupos de disponibilidade Always On são dois recursos conhecidos na edição SQL Server Enterprise. No entanto, existem alternativas econômicas para esses recursos que você pode considerar se não precisar do conjunto completo de recursos da edição SQL Server Enterprise. Por exemplo, você pode obter a TDE na edição SQL Server Standard a partir do SQL Server 2019. No lugar dos grupos de disponibilidade Always On, você pode usar o cluster de failover com armazenamento compartilhado ativado para o Windows File Server FSx para obter alta disponibilidade com a edição SQL Server Standard.

Ao fazer o downgrade da edição SQL Server Enterprise para a edição SQL Server Standard, você pode reduzir significativamente seus custos de licenciamento. Para obter mais informações, consulte a postagem [Otimizando o custo de suas implantações de alta disponibilidade do SQL Server no AWS](#) blog AWS de armazenamento.

Além de reduzir os custos de licenciamento, o downgrade das edições do SQL Server pode ajudar a reduzir seus gastos com o Software Assurance e evitar futuros true-ups. Se você devolver licenças não utilizadas para a reserva, poderá evitar custos adicionais de licenciamento e obter o melhor valor possível de seu investimento em licenciamento.

É importante avaliar cuidadosamente suas workloads do SQL Server e determinar quais recursos são essenciais para suas necessidades de negócios. Para obter mais informações, consulte [Avaliando seu ambiente](#) na Orientação AWS Prescritiva e determine se seu banco de dados Microsoft SQL Server usa recursos específicos da edição SQL Server Enterprise.

Se você escolher a edição certa do SQL Server e usar alternativas aos recursos da edição SQL Server Enterprise, poderá obter economias de custo significativas enquanto mantém a conformidade

e atende às suas necessidades de negócios. Para obter mais informações sobre as opções de downgrade, consulte a seção [Comparar edições do SQL Server](#) deste guia.

Usar a edição SQL Server Developer em ambientes de não produção

Em ambientes de não produção, você pode implantar edições licenciáveis do SQL Server, como a edição Enterprise ou Standard, usando assinaturas do MSDN em ambientes on-premises. No entanto, a assinatura do MSDN não tem a Mobilidade de Licenças. Portanto, se você migrar para AWS, não poderá trazer essas licenças. Você deverá usar a edição SQL Server Developer.

A edição SQL Server Developer é uma edição completa do SQL Server que está disponível gratuitamente. Essa edição está disponível para as versões 2016 e posteriores do SQL Server. Você pode baixá-la no site da Microsoft. A edição SQL Server Developer foi projetada para ser usada em todos os ambientes que não sejam de produção, como desenvolvimento, teste e preparação, desde que não esteja conectada a dados de produção ativos.

Se você usa a edição SQL Server Developer em ambientes que não são de produção, você pode evitar custos adicionais de licenciamento. Para obter mais informações, consulte a seção [Avaliar a edição SQL Server Developer](#) deste guia.

Otimizar a CPU para workloads do SQL Server

Em alguns casos, pode ser necessário escolher um tipo de instância com CPUs mais do que o necessário para sua carga de trabalho devido a outros fatores, como RAM ou limites de rede. No entanto, AWS fornece uma solução para ajudá-lo a otimizar seus custos de licenciamento nessas situações.

Você pode, como a maioria dos clientes que trazem licenças básicas do SQL Server, desativar o hyperthreading ou desativar as CPUs na instância do EC2 para limitar o número de CPUs disponíveis para o host. CPUs Essa opção permite que você aproveite outros recursos da instância, como RAM, enquanto ainda economiza no custo da compra de licenças extras.

Por exemplo, se você implantar uma instância r5.4xlarge porque sua carga de trabalho exige 128 GB de memória, mas você só precisa de oito núcleos do SQL Server, poderá desativar o hyperthreading de uma instância com apenas oito ativos. CPUs Ao fazer isso, você pode economizar 50% nas licenças necessárias do SQL Server, pois só precisará licenciar os oito núcleos que estão sendo usados ativamente.

Tipo de instância	Total v CPUs	vCPU ativa com recurso de otimização CPUs	Economia de licenças do SQL Server
r5.4xlarge	16	8	50%
r5.12xlarge	48	8	83%

O recurso Optimize CPU pode ser configurado durante a configuração de inicialização do Amazon EC2 ou modificando uma instância existente. Ela também pode ser aplicada tanto às instâncias BYOL quanto às instâncias Amazon EC2 com licença incluída. Essa flexibilidade ajuda você a dimensionar corretamente sua CPU de acordo com as necessidades de sua carga de trabalho, além de reduzir Windows Server as licenças. SQL Server Para instâncias do Amazon EC2 incluídas na licença, a CPUs redução proporciona uma economia instantânea nos custos de licenciamento.

Se você dimensionar corretamente suas instâncias, poderá garantir que está usando os tipos de instância mais econômicos para suas workloads. À medida AWS que introduz novos tipos de instância, é importante avaliar se essas novas instâncias podem atender aos requisitos de carga de trabalho com menos núcleos.

Recursos adicionais do

- [Amazon Web Services e Microsoft: perguntas frequentes](#) (AWS documentação)

Selecionar a instância certa do EC2 para as workloads do SQL Server

Important

Antes de ler esta seção, recomendamos que você leia primeiro as seções [Compreender o licenciamento do SQL Server](#) e [Selecionar o tipo de instância certo para workloads do Windows](#) deste guia.

Visão geral do

O Microsoft SQL Server está sendo executado nas instâncias do Amazon Elastic Compute Cloud (Amazon EC2) há mais de 15 anos. AWS pegou essa experiência e a usou para ajudar a desenvolver instâncias do Amazon EC2 para se adequar às cargas de trabalho do SQL Server, executadas desde especificações mínimas até clusters multirregionais de alto desempenho.

A escolha da instância certa do EC2 para o SQL Server depende muito da sua workload. Entender como o SQL Server é licenciado, como ele usa a memória e como os recursos do SQL Server se alinham às ofertas do Amazon EC2 pode ajudar a orientar você até a melhor instância do EC2 para sua aplicação.

Esta seção aborda uma variedade de workloads do SQL Server e como elas podem ser combinadas com determinadas instâncias do EC2 para reduzir ao mínimo seus custos de licenciamento e computação.


Comparação de custos

O Amazon EC2 permite que você traga a sua própria licença (BYOL) ou pague conforme o uso com o licenciamento do Windows Server e do SQL Server. Para pay-as-you-go licenciamento, os custos de licenciamento das licenças do Windows Server e do SQL Server são incorporados ao custo por hora da instância do EC2. Por exemplo, você pode ter preços diferentes AMIs com preços diferentes. O preço da AMI depende da edição do SQL Server em que a AMI é executada.

Os preços do Windows Server e do SQL Server não são discriminados. Você não encontrará preços discriminados em ferramentas como a [AWS Calculadora de Preços](#). Se você selecionar diferentes combinações de ofertas com licença incluída, os custos de licenciamento poderão ser deduzidos, conforme mostra a tabela a seguir.

Instância do EC2	AMI	Preço de computação	Preço de licença do Windows	Preço da licença do SQL	Preço total
r5.xlarge	Linux (preços de computação)	\$183,96	-	-	\$183,96
r5.xlarge	Linux + SQL Developer	\$183,96	\$0	\$0	\$183,96

Instância do EC2	AMI	Preço de computação	Preço de licença do Windows	Preço da licença do SQL	Preço total
r5.xlarge	Windows Server (LI)	\$183,96	\$134,32	-	\$318,28
r5.xlarge	Windows + SQL Developer	\$183,96	\$134,32	\$0	\$318,28
r5.xlarge	Windows + SQL Web (LI)	\$183,96	\$134,32	\$49,64	\$367,92
r5.xlarge	Windows + SQL Standard (LI)	\$183,96	\$134,32	\$350,4	\$668,68
r5.xlarge	Windows + SQL Enterprise (LI)	\$183,96	\$134,32	\$1095	\$1413,28

 Note

Os preços na tabela anterior são baseados nos preços sob demanda na região us-east-1.

O método mais econômico para executar o SQL Server é permanecer em uma edição de nível inferior até que você precise de um recurso de uma edição de nível superior. Para obter mais informações, consulte a seção [Comparar edições do SQL Server](#) deste guia. A atualização da edição SQL Server Web para a edição SQL Server Standard é sete vezes o custo de licenciamento do SQL Server e mais de três vezes o custo da mudança da edição Standard para a edição Enterprise. A disparidade nos custos de licenciamento é um fator importante a ser considerado, e será analisada no restante desta seção.

Cenário de otimização de custos

Considere um exemplo de cenário em que uma empresa de analytics que rastreia veículos de entrega está buscando melhorar a performance do SQL Server. Depois que um especialista em MACO analisa os gargalos de performance da empresa, ela faz a transição das instâncias x1e.2xlarge para as instâncias x2iedn.xlarge. Embora o tamanho da instância seja menor, os aprimoramentos nas instâncias x2 melhoram a performance e a otimização do SQL Server usando extensões do grupo de buffers. Isso permitiu que a empresa fizesse o downgrade da edição SQL Server Enterprise para a edição SQL Server Standard e reduzisse seu licenciamento do SQL Server de 8 v CPUs para 4 v CPUs

Antes da otimização:

Servidor	Instância do EC2	Edição do SQL Server	Custo mensal
Cutucar DB1	x1e.2xlarge	Enterprise	\$3.918,64
Cutucar DB2	x1e.2xlarge	Enterprise	\$3.918,64
Total			\$7.837,28

Após a otimização:

Servidor	Instância do EC2	Edição do SQL Server	Custo mensal
Cutucar DB1	x2iedn.xlarge	Standard	\$1.215,00
Cutucar DB2	x2iedn.xlarge	Standard	\$1.215,00
Total			\$2.430,00

As mudanças combinadas de instâncias x1e.2xlarge para instâncias x2iedn.xlarge permitiram que o cliente do exemplo economizasse USD 5.407 por mês em seus servidores de banco de dados de produção. Isso reduziu o custo total da workload em 69%.

Note

Os preços na tabela anterior são baseados nos preços sob demanda na região us-east-1.

Recomendações de otimização de custos

Instâncias otimizadas para memória

Um dos aspectos mais importantes do SQL Server é entender sua dependência da memória. O SQL Server tenta usar toda a RAM disponível que não está sendo usada pelo sistema operacional (até 2 TB para uma instalação padrão). Ele faz isso por motivos de performance. Trabalhar com dados na memória é muito mais eficiente do que ter que extrair dados constantemente do disco, fazer alterações e depois gravá-los de volta no disco. Em vez disso, o SQL Server tenta carregar o máximo possível de dados dos bancos de dados anexados e os mantém na RAM. As alterações feitas nos dados acontecem na memória e são posteriormente gravadas no disco.

Note

Para obter uma explicação detalhada de como o SQL Server grava alterações, consulte [Writing Pages](#) na documentação da Microsoft.

Como o SQL Server funciona melhor com grandes quantidades de RAM, geralmente recomendamos começar com os tipos de instâncias [otimizadas para memória do Amazon EC2](#). As instâncias otimizadas para memória são versáteis e oferecem uma variedade de opções diferentes. A família R tem uma vCPU-to-RAM proporção de 1 para 8 e tem opções para processadores Intel, processadores AMD, redes aprimoradas, desempenho aprimorado do EBS, armazenamento de instâncias e velocidade aprimorada do processador. Para cargas de trabalho com muita memória, há também uma família X que combina muitas das mesmas opções e estende a vCPU-to-RAM proporção para 1 a 32. Devido à versatilidade das instâncias otimizadas para memória, você pode aplicá-las a workloads do SQL Server de todas as formas e tamanhos.

Cargas de trabalho abaixo do mínimo de recursos (menos de 4 v) CPUs

Embora alguns casos de uso funcionem bem com instâncias expansíveis (T3), recomendamos que você geralmente evite usá-las para workloads do SQL Server. O licenciamento do SQL Server é baseado no número de v CPUs atribuído a uma instância. Se o SQL Server ficar ocioso a maior parte

do dia e estiver adquirindo créditos de expansão, você pagará pelas licenças do SQL que não está utilizando totalmente. Além disso, o SQL Server tem um requisito mínimo de licença de 4 núcleos por servidor. Isso significa que, se você tem uma carga de trabalho do SQL Server que não exige 4 vCPUs de potência computacional, você está pagando uma licença do SQL Server que não está usando. Nesses cenários, seria melhor [consolidar várias instâncias do SQL Server](#) em um servidor maior.

Workloads usando recursos mínimos (menos de 64 GB de RAM)

Muitas workloads do SQL Server com menos de 64 GB de RAM não priorizam a alta performance ou a alta disponibilidade. Para esses tipos de workloads, a edição SQL Server Web pode ser uma boa opção se a aplicação estiver coberta pelas restrições de licenciamento da Microsoft.

Important

A edição SQL Server Web tem um caso de uso restrito com base nos termos de licenciamento da Microsoft. A edição SQL Server Web pode ser usada apenas para fornecer suporte a páginas da Web, sites, aplicativos da Web e serviços da Web públicos e acessíveis pela Internet. Ele não pode ser usado para oferecer suporte a line-of-business aplicativos (por exemplo, gerenciamento de relacionamento com clientes, gerenciamento de recursos corporativos e outros aplicativos similares).

O SQL Server Web Edition pode ser expandido até 32 vCPUs e 64 GB de RAM e é 86% mais barato do que o SQL Server Standard Edition. Para workloads com poucos recursos, usar uma instância otimizada para memória AMD, como a r6a, que tem um preço de computação 10% mais barato do que sua equivalente da Intel, também é uma boa maneira de reduzir ao mínimo os custos de computação e licenciamento do SQL.

Workloads com recursos médios (menos de 128 GB de RAM)

A edição SQL Server Standard é usada na maioria das workloads do SQL Server com até 128 GB de RAM. A edição SQL Server Standard é 65 a 75% mais barata do que a edição SQL Server Enterprise e pode ser expandida para até 48 vCPUs e 128 GB de RAM. Como a limitação de 128 GB de RAM geralmente é atingida antes da limitação de 48 vCPUs, este é o foco da maioria dos clientes que desejam evitar a atualização para a edição SQL Server Enterprise.

O SQL Server tem um recurso chamado [extensão do grupo de buffers](#). Esse recurso permite que o SQL Server use uma parte de um disco para atuar como uma extensão da RAM. A extensão do

buffer pool funciona bem quando combinada com armazenamento ultrarrápido, como o NVMe SSDs usado no armazenamento de instâncias do [Amazon EC2](#). As instâncias do Amazon EC2 que contêm armazenamento de instâncias são indicadas com um “d” no nome da instância (por exemplo, r5d, r6id e x2iedn).

As extensões do grupo de buffers não substituem a RAM normal. No entanto, se você precisar de mais de 128 GB de RAM, poderá usar extensões de grupo de buffers com instâncias do EC2, como r6id.4xlarge e x2iedn.xlarge, para retardar a atualização para o licenciamento da edição Enterprise.

Workloads de alta performance (mais de 128 GB de RAM)

As workloads do SQL Server que exigem alta performance são um desafio para a otimização de custos devido à sua dependência de muitos recursos. No entanto, entender as diferenças nas instâncias do EC2 pode impedir que você faça a escolha errada.

A tabela a seguir mostra uma variedade de instâncias do EC2 otimizadas para memória e seus limites de performance.

	r5b	r6idn	r7iz	x2iedn	x2iezn
Processador	3.1 GHz	3.5 GHz	3.9 GHz	3.5 GHz	4.5 GHz
	Processador Intel Xeon de segunda geração	Processador Intel Xeon de terceira geração	Processador Intel Xeon escalável de quarta geração	Processador Intel Xeon de terceira geração	Processador Intel Xeon de segunda geração
Proporção CPU:RAM	1:8	1:8	1:8	1:32	1:32
Máximo de vCPU	96	128	128	128	48
RAM máxima	768 GB	1.024 GB	1.024 GB	4.096 GB	1.536 GB
Armazenamento de instância	–	NVMe SSD (4x 1.900 GB)	–	NVMe SSD (2x 1.900 GB)	–

	r5b	r6idn	r7iz	x2iedn	x2iezn
io2 Block Express	Compatível	Compatível	Compatível	Compatível	–
IOPS máxima do EBS	260.000	350,000	160.000	260.000	80.000
Throughput máximo do EBS	60 Gbps	80 Gbps	40 Gbps	80 Gbps	19 Gbps
Largura de banda da rede máxima	25 Gbps	200 Gbps	50 Gbps	100 Gbps	100 Gbps

Cada instância é usada para uma finalidade diferente. Compreender sua workload do SQL Server pode ajudar você a escolher o tipo de instância mais adequado para você.

Detalhes sobre atributos:

- r5b: o atributo “b” em r5b significa que esse tipo de instância está focado na alta performance do EBS. Na quinta geração de instâncias otimizadas para memória, a r5b foi a escolha de preferência. Foi o primeiro tipo de instância a utilizar volumes io2 Block Express e a atingir a IOPS máxima de armazenamento de 260 mil. O tipo de instância r5b ainda é uma alternativa econômica para as necessidades de alta performance do EBS.
- r6idn: a sexta geração de instâncias otimizadas para memória ofereceu melhorias consideráveis em relação à geração anterior. Os aprimoramentos de performance do EBS da r5b são levados um passo adiante com a r6idn, aumentando a IOPS máxima para 350 mil. A r6idn também tem um volume de armazenamento de instâncias para extensões de grupos de buffers e tempdb para aumentar ainda mais a performance do SQL Server.
- x2iedn: a x2iedn é semelhante à r6idn. Ele oferece níveis semelhantes de EBS aprimorado, rede aprimorada e armazenamento de instâncias NVMe SSD, mas com uma vCPU-to-RAM proporção de 1:32 para altas cargas de trabalho de memória e baixa quantidade de CPU (menores custos de licenciamento do SQL Server).
- x2iezn: o atributo “z” em x2iezn indica que esse tipo de instância está focado na alta performance do processador. O processador Cascade Lake tem uma frequência turbo para todos os núcleos de

até 4,5 GHz. Recomendamos que você use essa instância do EC2, juntamente com uma vCPU-to-RAM proporção de 1:32, em um cenário em que você deseja manter a quantidade de vCPU baixa. Isso, por sua vez, pode manter baixos os custos de licenciamento do SQL Server.

- **r7iz**: o atributo “z” em r7iz indica que esse tipo de instância está focado na alta performance do processador. O processador Sapphire Rapids tem uma frequência turbo de todos os núcleos de até 3,9. GHz Como as instâncias x2iezn, o r7iz prioriza o desempenho do processador de alta frequência, mas com uma proporção de 1:8. vCPU-to-RAM

Recursos adicionais do

- [Instâncias do Amazon EC2 de uso geral \(documentação\)](#) AWS
- [Comparison tool](#) (Vantage)
- [Licenciamento — SQL Server](#) (AWS documentação)

Consolidar instâncias

Esta seção se concentra na técnica de otimização de custos de combinar várias instâncias do SQL Server no mesmo servidor para minimizar os custos de licenciamento e maximizar a utilização dos recursos.

Visão geral do

A criação de uma instância faz parte do processo de instalação do mecanismo de banco de dados do SQL Server. A instância do SQL Server é uma instalação completa, contendo seus próprios arquivos de servidor, logins de segurança e bancos de dados do sistema (master, model, msdb e tempdb). Como uma instância tem todos os seus próprios arquivos e serviços, você pode instalar várias instâncias do SQL Server no mesmo sistema operacional sem que as instâncias interfiram umas nas outras. No entanto, como as instâncias estão todas instaladas no mesmo servidor, todas compartilham os mesmos recursos de hardware, como computação, memória e rede.

É comum usar apenas uma única instância do SQL Server por servidor em ambientes de produção para que uma instância “ocupada” não use demais os recursos de hardware compartilhados. Dar a cada instância do SQL Server seu próprio sistema operacional, com seus próprios recursos, é um limite melhor do que confiar na governança de recursos. Isso é especialmente verdadeiro para workloads de alta performance do SQL Server que exigem grandes quantidades de recursos de RAM e CPU.

No entanto, nem todas as workloads do SQL Server usam uma grande quantidade de recursos. Por exemplo, algumas organizações atribuem a cada um de seus clientes sua própria instância dedicada do SQL Server para fins de conformidade ou segurança. Para clientes menores ou clientes que normalmente não estão ativos, isso significa executar as instâncias do SQL Server com recursos mínimos.

Conforme observado no [guia Microsoft SQL Server 2019: Licenciamento](#), cada servidor que executa o SQL Server deve contabilizar um mínimo de quatro licenças de CPU. Isso significa que, mesmo que você execute um servidor com apenas dois vCPUs, ainda deverá licenciar o SQL Server para quatro CPUs v. Com base nos [preços públicos do SQL Server da Microsoft](#), essa é uma diferença de \$3.945 se você usar a edição SQL Server Standard. Para organizações que executam vários servidores com instâncias únicas do SQL Server usando recursos mínimos, o custo combinado de ter que licenciar recursos não utilizados pode ser substancial.

Cenário de otimização de custos

Esta seção explora um exemplo de cenário que compara a diferença entre executar quatro servidores Windows Server, cada um com uma única instância do SQL Server, com um único servidor Windows Server maior executando várias instâncias do SQL Server simultaneamente.

Se cada instância do SQL Server precisar apenas de dois V CPUs e 8 GB de RAM, o custo total por servidor será de 7.890 USD para a licença do SQL Server, além de um custo de computação por hora de 0,096 USD.

Instância do EC2	v CPUs	RAM	Preço	v CPUs para licenciar	Custo total de licenciamento do SQL Server
m6i.large	2	8	0,096	4	\$7.890

Expandindo isso para quatro servidores, o custo total é de USD 31.560 para a licença do SQL Server com um custo de computação por hora de USD 0,384.

Instância do EC2	v CPUs	RAM	Preço	v CPUs para licenciar	Custo total de licenciamento do SQL Server
4x m6i.large	2	32	0,384	16	\$31.560

Se você combinar todas as quatro instâncias do SQL Server em uma única instância do EC2, a quantidade total de recursos computacionais e computação permanecerá a mesma. No entanto, ao remover os custos desnecessários de licenciamento do SQL Server, você pode reduzir o custo total de execução da workload em USD 15.780.

Instância do EC2	v CPUs	RAM	Preço	v CPUs para licenciar	Custo total de licenciamento do SQL Server
m6i.2xlarge	8	32	0,384	8	\$15.780

Note

Nas tabelas anteriores, os custos computacionais mostram preços por hora sob demanda para servidores do Amazon EC2 executando o Windows Server na região us-east-1. Os custos de licenciamento da edição SQL Server Standard referem-se aos [preços públicos do SQL Server da Microsoft](#).

Recomendações de otimização de custos

Se você está pensando em consolidar instâncias do SQL Server, a maior preocupação é o consumo de recursos para cada uma das instâncias que você deseja consolidar. É importante obter métricas de performance por longos períodos para entender melhor os padrões das workloads em cada servidor. Algumas ferramentas comuns para monitoramento do consumo de recursos são [Amazon CloudWatch](#), [Windows Performance Monitor](#) (perfmon) e as [ferramentas de monitoramento nativas](#) do SQL Server.

Recomendamos que você considere as seguintes questões ao analisar se suas workloads do SQL Server poderiam ser combinadas para usar os mesmos recursos do servidor sem que elas interfiram umas nas outras:

- Quais recursos (CPU, memória e largura de banda da rede) são consumidos durante seu estado estável?
- Quais recursos (CPU, memória e largura de banda da rede) são consumidos durante os picos?
- Com que frequência os picos ocorrem? Os picos são consistentes?
- Os picos de recursos de um servidor coincidem com os picos de recursos de outro servidor?
- Quais são as IOPS e o throughput de armazenamento usados pelo SQL Server?

Se você quiser seguir em frente com um plano para combinar instâncias do SQL Server, consulte a publicação [Run multiple instances of SQL Server on one Amazon EC2 instance](#) no blog AWS Cloud Operations & Migrations. Essa publicação fornece instruções sobre como fazer as alterações de configuração no SQL Server para adicionar outras instâncias. Antes de começar, considere as pequenas diferenças quando várias instâncias são instaladas no mesmo servidor:

- A instância padrão do banco de dados SQL Server é nomeada MSSQLSERVER e usa a porta 1433.
- Cada instância adicional instalada no mesmo servidor é uma instância de banco de dados “nomeada”.
- Cada instância nomeada tem um nome de instância e uma porta exclusivos.
- O [navegador do SQL Server](#) deve ser executado para coordenar o tráfego para as instâncias nomeadas.
- Cada instância pode usar locais separados para arquivos de dados do banco de dados e logs separados.
- As [configurações de memória máxima do servidor](#) do SQL Server devem ser definidas de acordo com as necessidades de performance de cada instância, com seu total combinado também deixando memória suficiente para o sistema operacional subjacente.
- Você pode usar os recursos [nativos de backup e restauração](#) do SQL Server ou o [AWS DMS](#) para migração ou consolidação.

Recursos adicionais do

- [Ficha técnica de licenciamento do SQL Server](#) (blog de operações e migrações AWS na nuvem)

- Publicação no blog sobre [configuração de várias instâncias do SQL Server \(blog de operações e migrações AWS na nuvem\)](#)

Comparar edições do SQL Server

Visão geral do

O licenciamento do Microsoft SQL Server é uma das maiores despesas de um ambiente de workload do Windows. Os custos de licenciamento do SQL Server podem facilmente ir além dos custos de computação para executar a workload. Se você escolher a edição errada, poderá pagar por recursos que não está usando ou de que nem precisa. Esta seção compara as seguintes edições do SQL Server, incluindo seus recursos e custos relativos:

- Enterprise: a edição SQL Server Enterprise fornece recursos de data center com alta performance, virtualização ilimitada e várias ferramentas de business intelligence (BI).
- Standard: a edição SQL Server Standard fornece gerenciamento básico de dados e business intelligence para organizações e departamentos menores.
- Web — A edição Web do SQL Server é adequada para empresas que são hospedeiras na web ou provedores de valor agregado na web (VAPs). Esta edição oferece um baixo custo total de propriedade e fornece recursos de escalabilidade e capacidade de gerenciamento para propriedades da web de pequena a grande escala.

Important

Você pode usar a edição SQL Server Web para fornecer suporte a páginas da web, sites e serviços e aplicações web públicos e acessíveis pela internet. Você não pode usar o SQL Server Web Edition para oferecer suporte a line-of-business aplicativos (como aplicativos de gerenciamento de relacionamento com clientes ou aplicativos de gerenciamento de recursos corporativos).

- Developer: a edição SQL Server Developer inclui todas as funcionalidades da edição Enterprise, mas é destinada apenas para fins de desenvolvimento.
- Express: a edição SQL Server Express é um banco de dados gratuito e pode ser usado para aprender ou criar aplicações de área de trabalho. Você pode atualizar a edição Express para outras edições.

Note

A edição SQL Server Evaluation está disponível por um período de teste de 180 dias.

Impacto do custo

Você pode comprar licenças do SQL Server de um revendedor autorizado da Microsoft e trazê-las para a AWS com o Software Assurance. Como alternativa, você pode usar licenças do SQL Server com um pay-as-you-go modelo que inclua a licença do Amazon EC2. AMIs

Se você comprar licenças do SQL Server de revendedores autorizados da Microsoft, as licenças de núcleo serão vendidas em pacotes de dois, e você deverá licenciar no mínimo quatro núcleos por servidor. A tabela a seguir mostra uma comparação de custo entre as edições Enterprise e Standard.

Versão	Edição SQL Server Enterprise (pacote de 2 núcleos)	Edição SQL Server Standard (pacote de 2 núcleos)	Poupança
2022	\$15.123	\$3.945	74%
2019	\$13.748	\$3.586	74%

Note

Os preços na tabela anterior são baseados nos preços públicos da Microsoft para o [SQL Server 2022](#) e o [SQL Server 2019](#).

A comparação de custos a seguir mostra a hospedagem de diferentes edições do SQL Server com o Amazon EC2 incluído na licença. AMIs Nessa comparação, o SQL Server está hospedado na r6i.xlarge (4 vCPUs) na região us-east-1.

Instância	Custo de computação	Custo de licença do Windows	Custo de licença do SQL Server	Total
R6i.xlarge (Linux)	\$183,96	–	–	\$183,96
R6i.xlarge + Windows	\$183,96	\$134,32	–	\$318,28
R6i.xlarge + edição SQL Server Web	\$183,96	\$134,32	\$49,35	\$367,63
R6i.xlarge + edição SQL Server Standard	\$183,96	\$134,32	\$350,4	\$668,68
R6i.xlarge + edição SQL Server Enterprise	\$183,96	\$134,32	\$1.095	\$1.413,28

Você pode economizar até 95% nos custos de licenciamento do SQL Server selecionando a edição certa do SQL Server para sua workload. A tabela a seguir compara o custo das licenças do SQL Server em instâncias r6i.xlarge.

Edição	% de economia
Standard versus Enterprise	68%
Web versus Standard	86%
Web versus Enterprise	95%

Na maioria dos cenários, as organizações mudam da edição Enterprise para a edição Standard, mas há alguns casos em que é possível mudar da edição Standard ou Enterprise para a edição Web.

Recomendações de otimização de custos

Você pode escolher a melhor edição para sua workload com base em limites de escalabilidade, alta disponibilidade, performance e segurança. A tabela a seguir mostra os recursos que são compatíveis com as edições do SQL Server. Isso pode ajudar você a decidir qual edição usar. Essa comparação se aplica ao [SQL Server 2016 SP1 e versões posteriores](#).

Limites de escala

A tabela a seguir compara os limites de escalabilidade das diferentes edições do SQL Server.

Recurso	Enterprise Edition	Standard Edition	Edição Web	Edição Express
Capacidade máxima de computação usada por uma única instância do SQL Server Database Engine, do SQL Server Analysis Services (SSAS) ou do SQL Server Reporting Services (SSRS)	Sistema operacional máximo	Limitado a menos de 4 soquetes ou 24 núcleos	Limitado a menos de 4 soquetes ou 16 núcleos	Limitado a menos de 4 soquetes ou 4 núcleos
Memória máxima para grupos de buffers por instância do SQL Server Database Engine	Sistema operacional máximo	128 GB	64 GB	1.410 MB
Capacidade máxima para a extensão do	32 vezes a memória máxima configurada	4 vezes a memória máxima configurada	N/D	N/D

Recurso	Enterprise Edition	Standard Edition	Edição Web	Edição Express
grupo de buffers por instância do SQL Server Database Engine				
Tamanho máximo do banco de dados relacional	524 LIBRAS	524 LIBRAS	524 LIBRAS	10 GB
Memória máxima para caches do Columnstore ou dados otimizados para memória	Sistema operacional máximo	32 GB	16 GB	352 MB

Se seu aplicativo exigir menos de 16 núcleos (32 vCPUs) e 64 GB de RAM, você poderá começar a avaliar a partir da edição Web do SQL Server. Se sua workload exigir mais de 64 GB de memória ou outras opções de alta disponibilidade, você deverá atualizar para a edição SQL Server Standard.

Você pode usar a edição SQL Server Web para oferecer suporte a páginas da web, sites e serviços e aplicações web públicos e acessíveis pela internet, mas não pode usar a edição SQL Server Web para oferecer suporte a aplicações de linha de negócios. Para obter mais informações sobre casos de uso da edição SQL Server Web, entre em contato com o [Microsoft Licensing Support](#) ou com seu revendedor autorizado da Microsoft.

Você pode usar a edição SQL Server Standard para cargas de trabalho de até 24 núcleos (48 vCPUs) e 128 GB de memória. No entanto, você pode usar [extensões de grupos de buffers](#) para permitir que a edição SQL Server Standard utilize o [armazenamento de instâncias locais](#), como as presentes nas instâncias r6id do EC2. Isso estende a memória até o tamanho de quatro vezes a configuração máxima de memória. Essa combinação de recursos poderá retardar a atualização do servidor para a edição Enterprise quando os requisitos de memória começarem a aumentar.

Você pode identificar a utilização da memória localizando as páginas dos bancos de dados no grupo de buffers e nos contadores de [expectativa de vida da página](#). A expectativa de vida da página indica

quanto tempo a página fica na memória antes de ser transferida de volta para o disco. Esse valor padrão do contador é de 300. Se as páginas permanecerem na memória por horas ou dias, há uma chance de reduzir a memória alocada.

Alta disponibilidade

A tabela a seguir compara os recursos de alta disponibilidade das diferentes edições do SQL Server.

Recurso	Enterprise Edition	Standard Edition	Edição Web	Edição Express
Suporte básico do servidor 1	Sim	Sim	Sim	Sim
Envio de logs	Sim	Sim	Sim	Não
Espelhamento de banco de dados	Sim	Modo de segurança COMPLETO	Somente como testemunha	Somente como testemunha
Compactação de backup	Sim	Sim	Não	Não
Instâncias de cluster de failover Always On	16 nós	2 nós	Não	Não
Grupos de Disponibilidade Always On	Até 8 réplicas secundárias, incluindo 2 réplicas secundárias síncronas	Não	Não	Não
Grupos básicos de disponibilidade	Não	2 nós	Não	Não

Recurso	Enterprise Edition	Standard Edition	Edição Web	Edição Express
Restauração on-line de páginas e arquivos	Sim	Não	Não	Não
Indexação on-line	Sim	Não	Não	Não
Alteração do esquema on-line	Sim	Não	Não	Não
Recuperação rápida	Sim	Não	Não	Não
Backups espelhados	Sim	Não	Não	Não
Adição dinâmica de memória e CPU	Sim	Não	Não	Não
Backup criptografado	Sim	Sim	Não	Não
Backup híbrido para o Microsoft Azure (backup para URL)	Sim	Sim	Não	Não
Servidor de failover para recuperação de desastres	Sim	Sim	Não	Não
Servidores de failover para alta disponibilidade	Sim	Sim	Não	Não

Outros recursos comuns

A tabela a seguir compara os recursos mais comuns das diferentes edições do SQL Server. Para obter uma lista extensa de recursos, consulte [Editions and supported features of SQL Server 2019](#) na documentação da Microsoft.

Recurso	Enterprise Edition	Standard Edition	Edição Web	Edição Express
(Performance) Gerenciador de recursos	Sim	Não	Não	Não
(Segurança) Transparent Database Encryption (TDE)	Sim	Sim	Não	Não
(Segurança) Gerenciamento extensível de chaves (EKM)	Sim	Não	Não	Não
(Replicação) Publicação da Oracle	Sim	Não	Não	Não
(Replicação) Replicação transacional ponto a ponto	Sim	Não	Não	Não
Captura de dados de alteração	Sim	Sim	Não	Não

Edição SQL Server Developer

Todas as workloads que não sejam de produção, como ambientes de desenvolvimento, controle de qualidade, testes, preparação e UAT, podem usar a edição SQL Server Developer para economizar 100% nos custos de licenciamento do SQL Server. Depois de [baixar o SQL Server](#), você pode instalar a edição SQL Server Developer em instâncias do EC2 usando locação compartilhada. A infraestrutura dedicada não é necessária na edição SQL Server Developer. Para obter mais informações, consulte a recomendação deste guia para a [edição SQL Server Developer](#).

Troca de edições

Para workloads existentes, mudar de uma edição para outra requer testes extensivos. É uma prática recomendada verificar as workloads em execução nas edições Enterprise ou Standard para ver se os recursos específicos da edição são usados e se há alguma solução alternativa para esses recursos. Por exemplo, se você quiser ver se seus bancos de dados estão usando algum recurso de nível corporativo, você pode executar [visualizações de gerenciamento dinâmico \(DMV\)](#) em todos os bancos de dados, conforme mostra o comando do exemplo a seguir.

```
SELECT feature_name FROM sys.dm_db_persisted_sku_features; GO
```

Há alguns recursos da edição Enterprise que não podem ser capturados no T-SQL, como a reindexação on-line como parte dos trabalhos de manutenção do SQL. Eles devem ser verificados manualmente.

Considerações sobre a migração

A forma como você licencia o SQL Server determinará suas opções de troca de edições. As AMIs, incluindo as AMIs do SQL Server, têm o custo de licenciamento incluído no preço da instância do EC2. O custo do licenciamento está vinculado à AMI. Você pode usar [códigos de faturamento da AWS](#) para verificar a versão do SQL Server incluída na sua AMI. Para instâncias da AWS com licença incluída, mudar a edição do SQL Server dentro do sistema operacional não vai alterar o faturamento associado à AMI. Você deve migrar seus bancos de dados para uma nova instância do EC2 com uma AMI executando a nova edição do SQL Server.

Se você estiver trazendo sua própria licença, terá mais flexibilidade. Geralmente, ainda é recomendável migrar para outra instância do EC2 que esteja executando a nova versão. Isso permite um failback fácil se algo não sair conforme o planejado. No entanto, se você precisar usar o servidor existente, ainda poderá fazer uma side-by-side instalação do SQL Server e migrar os bancos de dados entre as instâncias. Para obter etapas mais detalhadas sobre rebaixamentos de side-by-side edições, consulte [Atualização e rebaixamento de edições no SQL Server](#) no MSSQLTips site.

Recursos adicionais do

- [Editions and supported features of SQL Server 2022](#) (Microsoft Learn)
- [sys.dm_db_persisted_sku_features \(Transact-SQL\)](#) (Microsoft Learn)
- [Which Version of SQL Server Should You Use?](#) (Brent Ozar Unlimited)
- [AWS Calculadora de Preços](#) (AWS)

Avaliar a edição SQL Server Developer

Visão geral do

A [edição SQL Server Developer](#) é uma edição gratuita do SQL Server que contém todos os recursos da edição Enterprise e pode ser usada em qualquer ambiente que não seja de produção. Na nuvem, onde o licenciamento do Microsoft Developer Network (MSDN) não pode ser usado, a edição SQL Server Developer é uma boa maneira de economizar custos sem precisar fornecer licenças para workloads de desenvolvimento e teste. Isso é especialmente verdadeiro para equipes que administram grandes ambientes de desenvolvimento e teste e buscam reduzir custos desnecessários.

Um ambiente de produção é definido como um ambiente que é acessado pelos usuários finais de uma aplicação (como um site da internet) e é usado para finalidades que vão além da coleta de feedback ou do teste da aceitação dessa aplicação. Outros cenários que constituem ambientes de produção incluem:

- Ambientes que se conectam a um banco de dados de produção
- Ambientes que oferecem suporte à recuperação de desastres ou ao backup para um ambiente de produção
- Ambientes que são usados para produção pelo menos em parte do tempo, como um servidor que é colocado em produção durante os períodos de pico de atividade

Para obter mais informações de licenciamento, consulte [Amazon Web Services and Microsoft: Frequently Asked Questions](#) na documentação da AWS .

Impacto do custo

Se você usa a edição SQL Server Developer para workloads que não são de produção, você pode economizar 100% dos seus custos atuais de licenciamento do SQL Server para ambientes de desenvolvimento e teste.

Versão do SQL Server	Edição SQL Server Enterprise (pacote de 2 núcleos)	Edição SQL Server Standard (pacote de 2 núcleos)	Edição SQL Server Developer
2022	\$15.123	\$3.945	Gratuito
2019	\$13.748	\$3.586	Gratuito

Note

Os preços na tabela anterior são baseados nos preços públicos da Microsoft para o [SQL Server 2022](#) e o [SQL Server 2019](#).

A tabela a seguir compara o custo de diferentes edições do SQL Server executadas com 4 v CPUs e usando preços sob demanda na região. us-east-2 Isso se aplica a cenários que dependem de instâncias com licença incluída de AWS

Instância do EC2	AMI	Preço de computação	Preço de licença do Windows	Preço da licença do SQL Server	Preço total
r5.xlarge	Linux (preços de computação)	\$183,96	–	–	\$183,96
r5.xlarge	Linux + edição SQL Server Developer	\$183,96	\$0	\$0	\$183,96

Instância do EC2	AMI	Preço de computação	Preço de licença do Windows	Preço da licença do SQL Server	Preço total
r5.xlarge	Windows Server (LI)	\$183,96	\$134,32	–	\$318,28
r5.xlarge	Windows + edição SQL Server Developer	\$183,96	\$134,32	\$0	\$318,28
r5.xlarge	Windows + edição SQL Server Web (LI)	\$183,96	\$134,32	\$49,64	\$367,92
r5.xlarge	Windows + edição SQL Server Standard (LI)	\$183,96	\$134,32	\$350,4	\$668,68
r5.xlarge	Windows + Edição SQL Server Enterprise (LI)	\$183,96	\$134,32	\$1095	\$1413,28

Cenário de otimização de custos

Depois que uma empresa de integridade de dados fez uma nova aquisição, ela queria migrar a workload recém-adquirida de sua localização atual em um provedor de hospedagem gerenciada para consolidar com suas outras workloads na Nuvem AWS. O preço inicial mostrou que a carga de trabalho do SQL Server da empresa custaria 60% AWS mais em execução do que no atual provedor de serviços gerenciados. Um SME em MACO avaliou a estimativa e descobriu que o cliente estava, na verdade, pagando pelas licenças do SQL Server no provedor de hospedagem gerenciada para seus ambientes de desenvolvimento e teste. Ao mudar as workloads de não produção para a edição

SQL Server Developer durante a migração, a empresa reduziu o licenciamento do SQL Server em 40%.

Licença do SQL Server incluída no Amazon EC2

Se você tiver o SQL Server em instâncias do EC2 que usam a [licença incluída AMIs](#), não é possível fazer uma conversão direta da edição Enterprise para a edição Developer. Os custos de licenciamento para instâncias com licença incluída estão vinculados à AMI. Mesmo que o SQL Server seja desinstalado do sistema operacional, a instância do EC2 ainda será cobrada pelos custos de licenciamento.

Para converter para a edição Developer, você deve [baixar a edição SQL Server Developer](#), instalá-la em uma nova instância do EC2 e, em seguida, migrar seus bancos de dados. Você pode migrar bancos de dados do SQL Server entre instâncias do EC2 usando diversos métodos. Para obter mais informações, consulte [SQL Server database migration methods](#) no guia Migrating Microsoft SQL Server databases to the Nuvem AWS. Você também pode usar a [solução Automated SQL Server Developer](#) para preparar a nova instância para a qual planeja migrar.

BYOL do SQL Server no Amazon EC2

Se você tiver instâncias do SQL Server que usam um BYOL, você pode escolher entre as seguintes opções de conversão ou side-by-side downgrade no local:

- Baixe a [edição SQL Server Developer](#) no site da Microsoft. Para obter instruções de instalação manual ou automatizada, consulte a publicação [Automating SQL Server Developer deployments](#) no Blog da AWS .
- Use o [backup e a restauração nativos do SQL Server](#) para migrar bancos de dados ou detach/attach bancos de dados de uma instância do SQL para outra.
- Use uma [ferramenta de automação](#) para implantações em massa.

Note

A edição SQL Server Developer é somente para ambientes de não produção.

Recursos adicionais do

- [Automatizando implantações do SQL Server Developer para implantar o SQL Server Developer Edition no EC2 \(Blog\)AWS](#)
- [SQL 2022 pricing](#) (Microsoft)
- [SQL 2019 pricing](#) (Microsoft)
- [Licensing options](#) (SQL Server no Amazon EC2)
- [AWS Calculadora de Preços](#) (documentação do SQL Server no Amazon EC2)
- [Microsoft SQL Server 2019 Licensing guide](#) (download da Microsoft)
- [SQL Server 2022 Developer edition](#) (download da Microsoft)

Avaliar o SQL Server no Linux

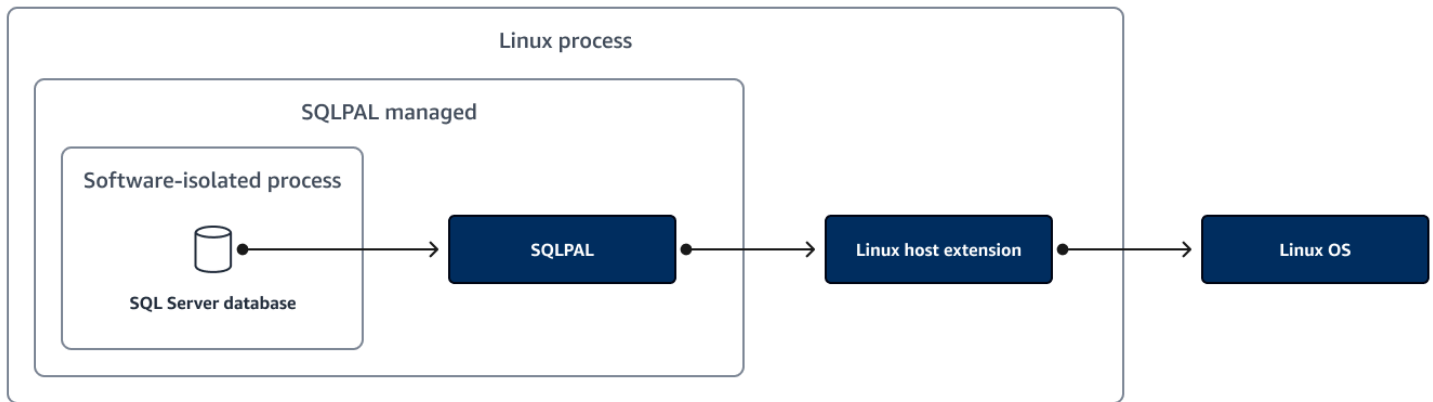
Visão geral do

Desde o SQL Server 2017, é possível instalar o SQL Server em sistemas operacionais Linux. O SQL Server no Linux está pronto para uso corporativo e oferece flexibilidade, alto desempenho, recursos de segurança, TCO reduzido, HA/DR recursos e uma ótima experiência do usuário. Você pode mudar do SQL Server no Windows Server para o SQL Server no Linux para economizar nos custos de licenciamento do Windows Server.

Para Linux, o SQL Server está disponível para implantação no Red Hat Enterprise Linux (RHEL), SUSE Linux Enterprise Server (SLES), Ubuntu e Amazon Linux 2. O mecanismo de banco de dados do SQL Server é executado da mesma forma no Windows Server e no Linux, mas há algumas alterações fundamentais em determinadas tarefas ao usar o Linux. Uma diferença fundamental entre a execução de aplicações SQL Server Always On no Linux e no Windows está relacionada ao cluster de failover. Se você implantar grupos de disponibilidade Always On em um host do Windows Server, poderá aproveitar o [Windows Server Failover Clustering \(WSFC\)](#) e o Active Directory como recursos integrados que são compatíveis com o cluster de failover. No entanto, nem o WSFC nem o Active Directory estão disponíveis para fornecer suporte ao cluster de failover no Linux. [Se você quiser iniciar o clustering de failover para o SQL Server no Linux, você pode usá-lo para simplificar AWS Launch Wizard a configuração do cluster e a instalação do SQL em instâncias Linux usando o Pacemaker. ClusterLabs](#)

O SQL Server no Windows e no Linux compartilha uma base de código comum. Ou seja, o mecanismo principal do SQL Server não precisou ter qualquer alteração para ser executado no

Linux. O SQL Server introduziu uma camada de abstração de plataforma (SQLPAL), conforme mostrado no diagrama a seguir.



O SQLPAL é responsável pela abstração de chamadas e comunicação entre o SQL Server e o sistema operacional subjacente. A extensão do host é simplesmente uma aplicação Linux nativa. As funções de baixo nível do sistema operacional são chamadas nativas para otimizar o uso de E/S, memória e CPU. Quando a extensão do host é iniciada, ela carrega e inicializa o SQLPAL, que então ativa o SQL Server. O SQLPAL inicia processos de software isolados que fornecem a conversão necessária para o restante do código. Adicionar essa nova camada à arquitetura do SQL Server significa que os mesmos recursos e benefícios básicos de nível corporativo que tornaram o SQL Server tão robusto no Windows estão disponíveis, independentemente do sistema operacional.

Impacto do custo

Para instâncias r5.2xlarge, a redução do custo de licenciamento do Windows Server é de cerca de USD 268 em cada cenário. A redução é uma porcentagem maior do custo total do servidor em comparação com o uso de edições de menor custo do SQL Server. A tabela a seguir mostra a economia de custos.

Instância	Edição	Custo mensal do SQL Server no Windows	Custo mensal do SQL Server no Linux	Poupança
r5.2xlarge	Web	\$735	\$466	37%
r5.2xlarge	Standard	\$1.337	\$1.068	20%
r5.2xlarge	Enterprise	\$2.826	\$2.558	10%

Note

A estimativa de preços na tabela anterior é baseada nos preços sob demanda na região us-east-1 e pode ser visualizada diretamente na [AWS Calculadora de Preços](#).

Considere um exemplo de cenário em que um cliente ISV no segmento de pequenas e médias empresas está procurando reduzir os custos em seu ambiente de desenvolvimento. Ele já estão usando a edição SQL Server Developer em um conjunto de servidores Windows. Ao mudar do Windows com a edição SQL Server Developer para o Linux com a edição SQL Server Developer, o cliente ISV pode economizar 33% em sua workload de desenvolvimento. A tabela a seguir mostra os custos estimados para esse cenário.

Estimativa	Custo mensal
Windows + SQL Server	\$9.307,72
Linux + SQL Server	\$6.218,36
Economia estimada de custos	USD 3.089,36 (33%)

Em outro cenário de exemplo, uma empresa migra instâncias EC2 do SQL Server incluídas na licença do Windows para o Linux. A empresa economiza um total de USD 300 mil por ano em custos de licenciamento do Windows Server, cerca de 20% de sua fatura total da AWS .

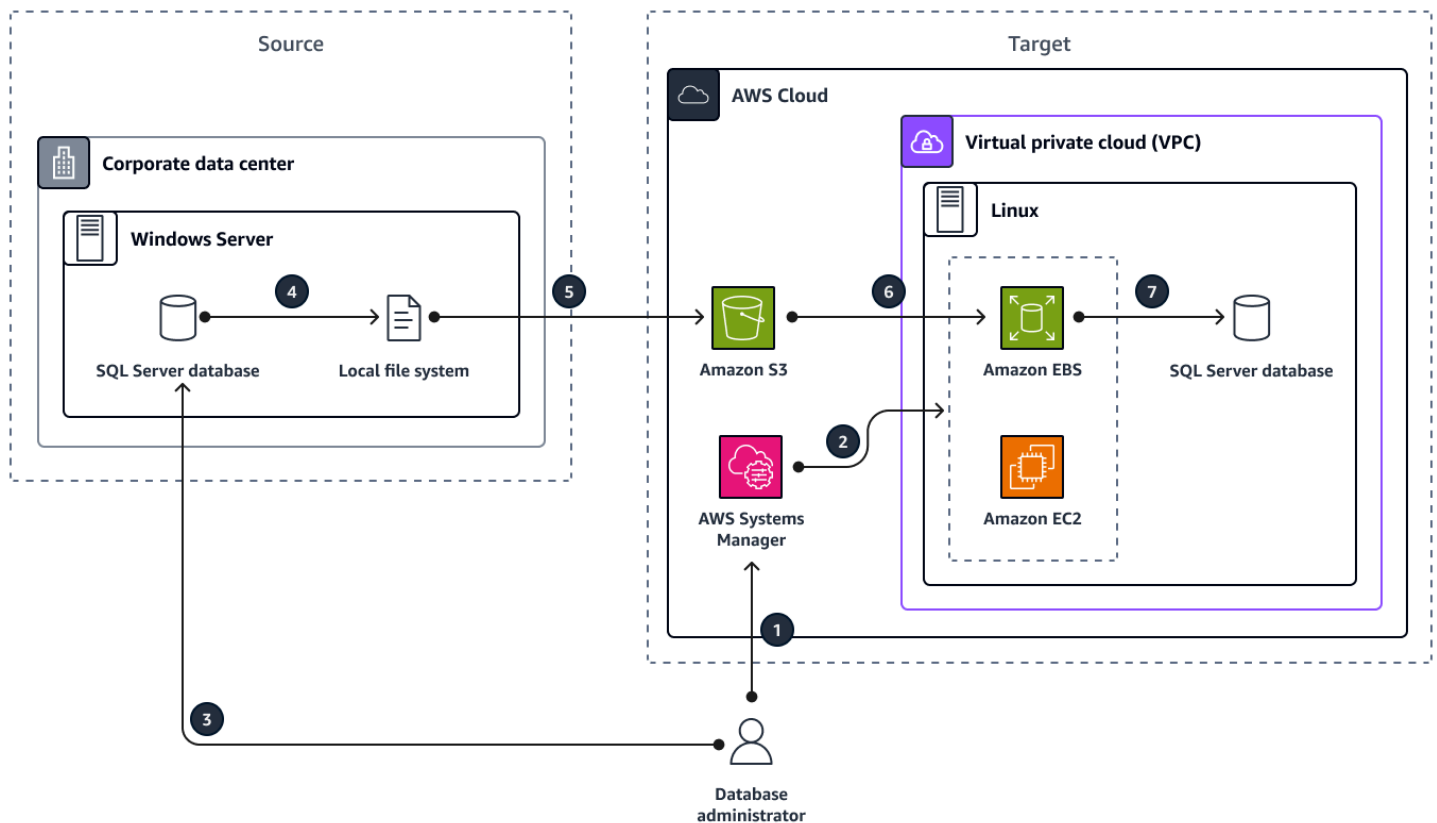
Recomendações de otimização de custos

Recomendamos que você considere o seguinte:

- O SQL Server no Linux é compatível a partir do SQL Server 2017.
- Para ajudar a fazer a mudança, você pode usar o [assistente de redefinição de plataformas de Windows para Linux para bancos de dados Microsoft SQL Server](#). O assistente de redefinição de plataformas é uma ferramenta de script que pode ajudar você a mover workloads existentes do SQL Server dos sistemas operacionais Windows para Linux verificando as incompatibilidades comuns, exportando bancos de dados do host Windows e importando os bancos de dados para uma instância do EC2 executando o Microsoft SQL Server 2017 no Ubuntu 16.04.

- Você também pode usar os recursos de [backup e restauração](#) no SQL Server para mudar do SQL Server no Windows para o Linux.
- Você pode implantar com facilidade e rapidez no SQL Server no Linux ou no Ubuntu usando o [AWS Launch Wizard](#). O assistente de inicialização pode implantar o SQL Server no Linux ou Ubuntu em cenários autônomos e de alta disponibilidade, com base nas necessidades da sua aplicação. Para obter mais informações, consulte a AWS Launch Wizard postagem [Implantando no SQL Server Always on Linux com](#) no blog Microsoft Workloads on AWS .

O diagrama a seguir mostra uma arquitetura para uma solução que usa o assistente de redefinição de plataformas de Windows para Linux para bancos de dados Microsoft SQL Server.



Recursos adicionais do

- [Overview of SQL Server on Linux](#) (Microsoft Learn)
- [Installation guide for SQL Server on Linux](#) (Microsoft Learn)
- [Implantação no SQL Server Always on Linux com AWS Launch Wizard](#) (Microsoft Workloads on AWS Blog)
- [SQL Server altamente disponível no Linux](#) (blog de código AWS aberto)

Otimizar as estratégias de backup do SQL Server.

Visão geral do

A maioria das organizações busca a solução certa para proteger seus dados no SQL Server no [Amazon EC2](#) e atender aos requisitos atuais de objetivo de ponto de recuperação (RPO), o tempo máximo aceitável desde o último backup, e de objetivo de tempo de recuperação (RTO), o atraso máximo aceitável entre a interrupção do serviço e a sua restauração. Se estiver executando o SQL Server em instâncias do EC2, há várias opções para criar backups dos seus dados e restaurá-los. As estratégias de backup para proteger os dados do SQL Server no Amazon EC2 incluem as seguintes:

- Backup em nível de servidor usando snapshots do [Amazon Elastic Block Store \(Amazon EBS\)](#) habilitados para Windows Volume Shadow Copy Service (VSS) ou o [AWS Backup](#)
- Backup em nível de banco de dados usando [backup e restauração nativos](#) no SQL Server

Se você tiver as seguintes opções de armazenamento para [database-level native backup](#):

- Um backup local com um [volume do Amazon EBS](#)
- Um backup do sistema de arquivos de rede com o [Amazon FSx para Windows File Server](#) ou o Amazon FSx para NetApp ONTAP
- Backup de rede no Amazon Simple Storage Service (Amazon S3) usando o [AWS Storage Gateway](#)
- Backup direto no Amazon S3 para SQL Server 2022

Esta seção apresenta o seguinte:

- Destaca os recursos para ajudar você a economizar espaço de armazenamento
- Compara os custos entre as diferentes opções de armazenamento de backend
- Fornece links para a documentação detalhada a fim de ajudar a implementar essas recomendações

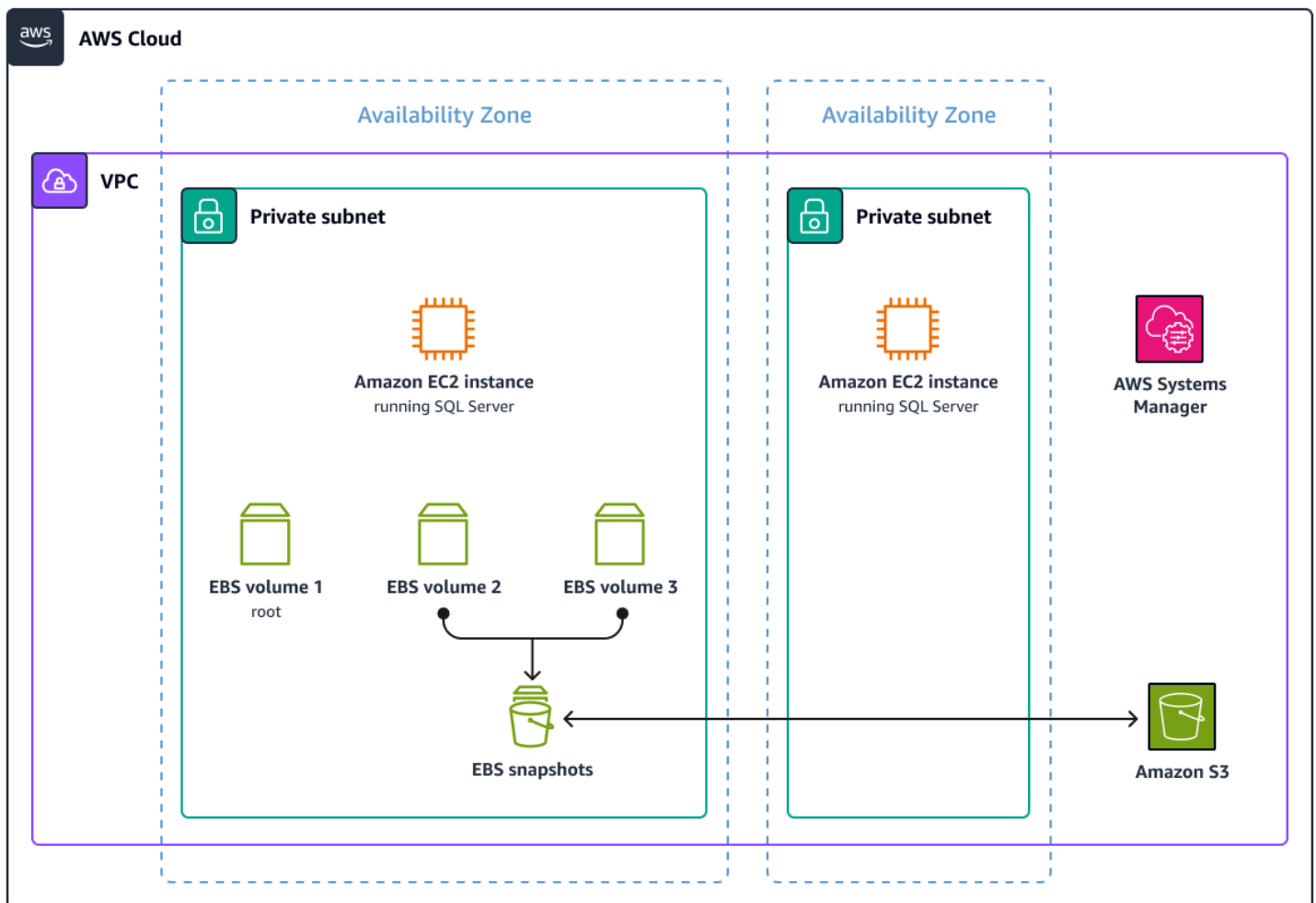
Backup no nível do servidor usando snapshots habilitados para VSS

Uma arquitetura de instantâneos habilitada para VSS usa o [comando AWS Systems Manager Executar](#) para instalar o agente VSS em suas instâncias do SQL Server. Você também pode usar o comando Executar para invocar todo o fluxo de trabalho de descarga dos buffers do sistema

operacional e do aplicativo no disco, pausar as I/O operações, tirar um point-in-time instantâneo dos volumes do EBS e, em seguida, retomar a E/S.

O Run Command cria snapshots automatizados de todos os volumes do EBS anexados a uma instância de destino. Você também tem a opção de excluir o volume raiz, pois os arquivos de banco de dados do usuário geralmente são armazenados em outros volumes. Se você fizer o stripe de vários volumes do EBS para criar um único sistema de arquivos para arquivos do SQL Server, o Amazon EBS também oferecerá suporte a snapshots em vários volumes consistentes em caso de falhas usando um único comando de API. Para obter mais informações sobre [snapshots do EBS habilitados para VSS](#) consistentes com aplicações, consulte [Create a VSS application-consistent snapshot](#) na documentação do Amazon EC2.

O diagrama a seguir mostra uma arquitetura para backup no nível de servidor usando snapshots habilitados para VSS.



Considere as seguintes vantagens de usar snapshots habilitados para VSS:

- O primeiro snapshot de uma instância de banco de dados contém os dados da instância de banco de dados completa. Os snapshots subsequentes da mesma instância de banco de dados são [incrementais](#), o que significa que somente os dados que foram alterados depois do snapshot mais recente serão salvos.
- Os instantâneos do EBS fornecem point-in-time recuperação.
- É possível [restaurar para uma nova instância do EC2 do SQL Server desde um snapshot](#).
- Se uma instância for criptografada usando o Amazon EBS, ou se um banco de dados for criptografado na instância usando TDE, essa instância ou banco de dados será restaurado automaticamente com a mesma criptografia.
- É possível copiar seus [backups automatizados entre regiões](#).
- Ao restaurar um volume do EBS de um snapshot, ele se torna imediatamente disponível para acesso pelas aplicações. Isso significa que você pode colocar o SQL Server online imediatamente após restaurar de snapshots um ou mais de seus volumes subjacentes do EBS.
- Por padrão, os volumes restaurados buscam blocos subjacentes do Amazon S3 na primeira vez que uma aplicação tenta lê-los. Isso significa que pode haver um atraso na performance após a restauração de um volume do EBS a partir de um snapshot. O volume em algum momento atingirá a performance nominal. No entanto, é possível evitar esse atraso usando [snapshots de restauração rápida \(FSR\)](#).
- Você pode usar o [gerenciamento do ciclo de vida para snapshots do EBS](#).

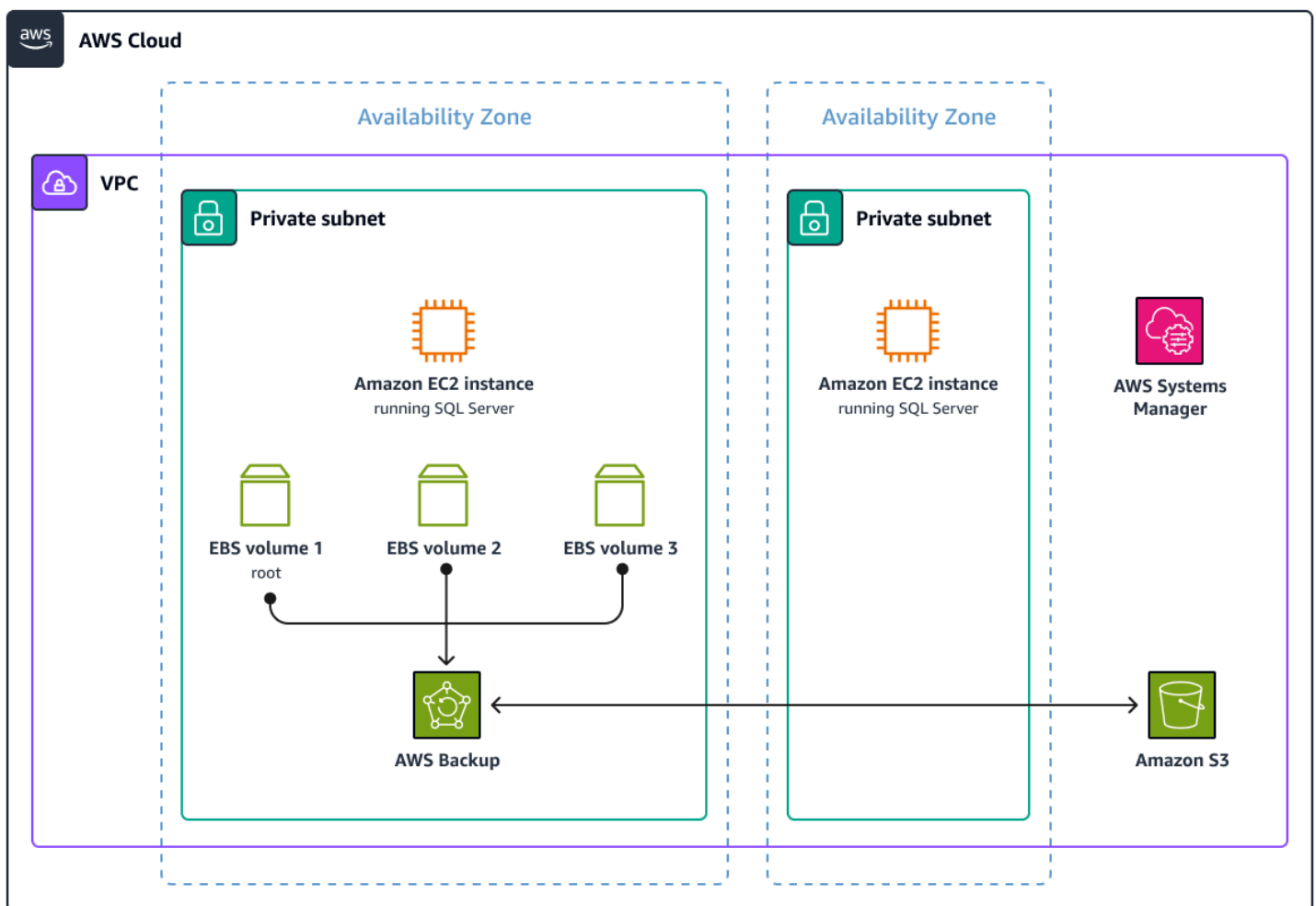
Considere as seguintes limitações do uso de snapshots habilitados para VSS:

- Você não pode realizar a point-in-time recuperação entre regiões com um instantâneo criptografado para uma instância do SQL Server.
- Não é possível criar um snapshot criptografado de uma instância não criptografada.
- Não é possível restaurar um banco de dados individual porque o snapshot é obtido no nível do volume do EBS.
- Não é possível restaurar a instância para si mesma.
- Um snapshot da instância de banco de dados deve ser criptografado usando a mesma chave do AWS Key Management Service (AWS KMS) que a instância de banco de dados.
- I/O O armazenamento é suspenso por uma fração de segundo (aproximadamente 10 milissegundos) durante o processo de backup instantâneo.

Backup do SQL Server usando AWS Backup

Você pode usar [AWS Backup](#) para centralizar e automatizar a proteção de dados em todo o mundo. Serviços da AWS oferece uma solução econômica, totalmente gerenciada e baseada em políticas que simplifica a proteção de dados em grande escala. AWS Backup também ajuda você a apoiar suas obrigações de conformidade regulatória e a cumprir suas metas de continuidade de negócios. Junto com AWS Organizations, AWS Backup permite que você implante centralmente políticas de proteção de dados (backup) para configurar, gerenciar e governar sua atividade de backup em toda a organização Contas da AWS e nos recursos.

O diagrama a seguir mostra a arquitetura de uma solução de backup e restauração para o SQL Server no EC2 usando o AWS Backup.



Considere as seguintes vantagens de fazer backup do SQL Server usando o AWS Backup:

- É possível automatizar os cronogramas de backup, o gerenciamento de retenção e o gerenciamento do ciclo de vida.

- Você pode centralizar sua estratégia de backup em toda a organização, abrangendo várias contas e Regiões da AWS
- É possível centralizar o monitoramento da sua atividade de backup e os alertas nos Serviços da AWS.
- Também é possível implementar backups entre regiões para planejar a recuperação de desastres.
- A solução oferece suporte a backups entre várias contas.
- É possível fazer backups seguros usando criptografia de backup secundária.
- Todos os backups oferecem suporte à criptografia usando chaves AWS KMS de criptografia.
- A solução funciona com o TDE.
- Você pode restaurar para um ponto de recuperação específico via console do AWS Backup .
- É possível fazer backup de uma instância do SQL Server inteira, o que inclui todos os bancos de dados do SQL Server.

Backup em nível de banco de dados

Essas abordagens usam a funcionalidade nativa de backup do Microsoft SQL Server. Você pode fazer backups de bancos de dados individuais na instância do SQL Server e restaurar um banco de dados individual.

Cada uma dessas opções para backup e restauração nativas do SQL Server também oferece suporte a:

- Compactação e backup de vários arquivos
- Backups completos, diferenciais e T-log
- Bancos de dados criptografados com TDE

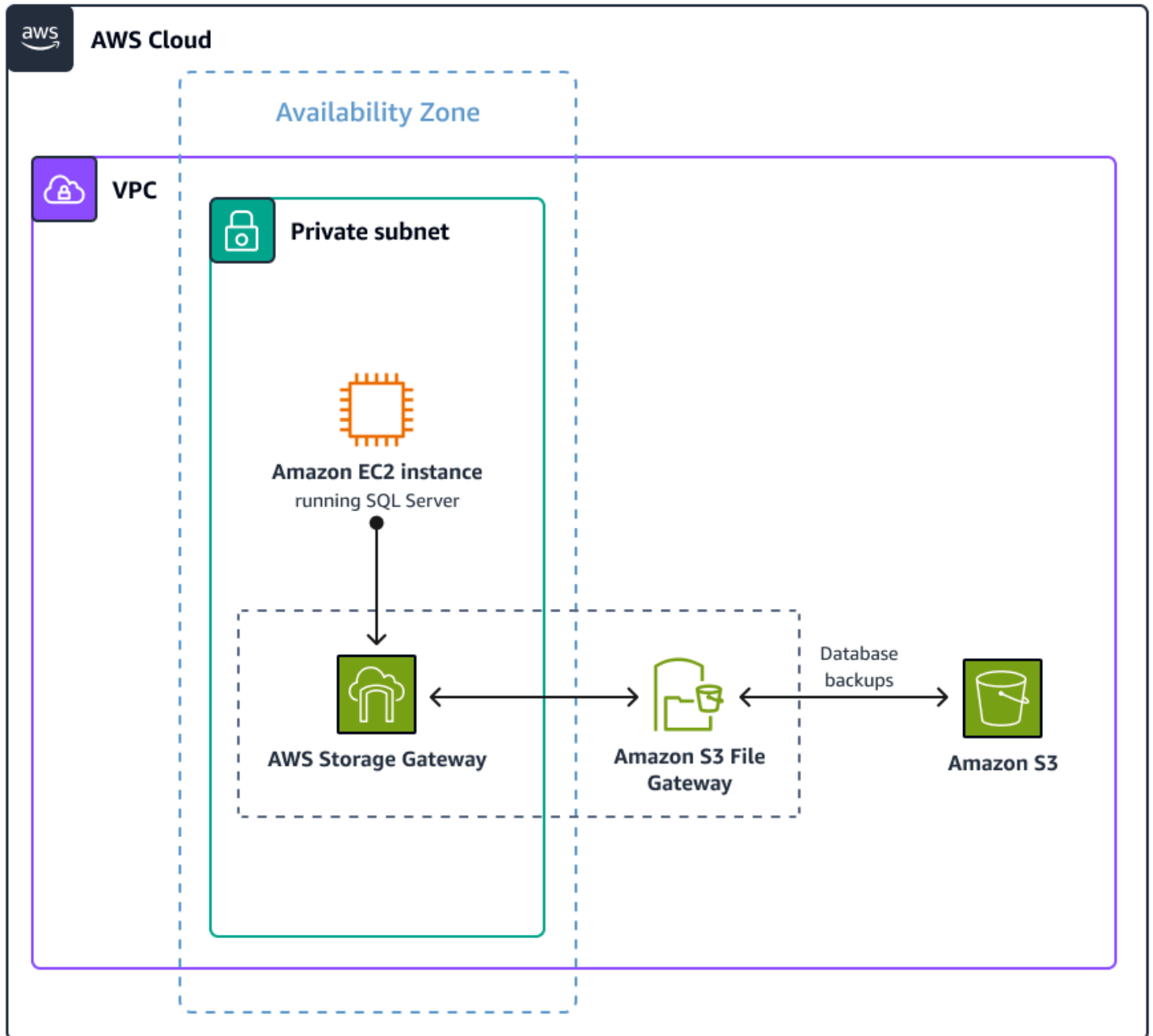
Backup e restauração nativos do SQL Server no Amazon S3

O SQL Server no Amazon EC2 é compatível com backup e restauração nativos de bancos de dados do SQL Server. É possível fazer um backup do banco de dados do SQL Server e depois restaurar o arquivo de backup em um banco de dados existente ou em uma nova instância do EC2 do SQL Server, Amazon RDS para SQL Server ou em um servidor on-premises.

O Storage Gateway é um serviço de armazenamento em nuvem híbrida que fornece às aplicações on-premises acesso a armazenamento em nuvem praticamente ilimitado. Você pode usar o Storage Gateway para fazer backup de bancos de dados do Microsoft SQL Server diretamente no

Amazon S3, reduzindo sua pegada de armazenamento on-premises e usando o Amazon S3 para armazenamento durável, escalável e econômico.

O diagrama a seguir mostra a arquitetura de uma solução nativa de backup e restauração usando o Storage Gateway e o Amazon S3.



Considere as seguintes vantagens de usar o backup nativo do SQL Server com o Storage Gateway:

- É possível mapear um gateway de armazenamento como um compartilhamento de arquivos Server Message Block (SMB) na instância do EC2 e enviar o backup para o Amazon S3.

- O backup ocorre diretamente no bucket do S3 ou por meio do cache de arquivos do Storage Gateway.
- Há suporte a backups de vários arquivos.

Considere as seguintes limitações do backup nativo usando o Storage Gateway:

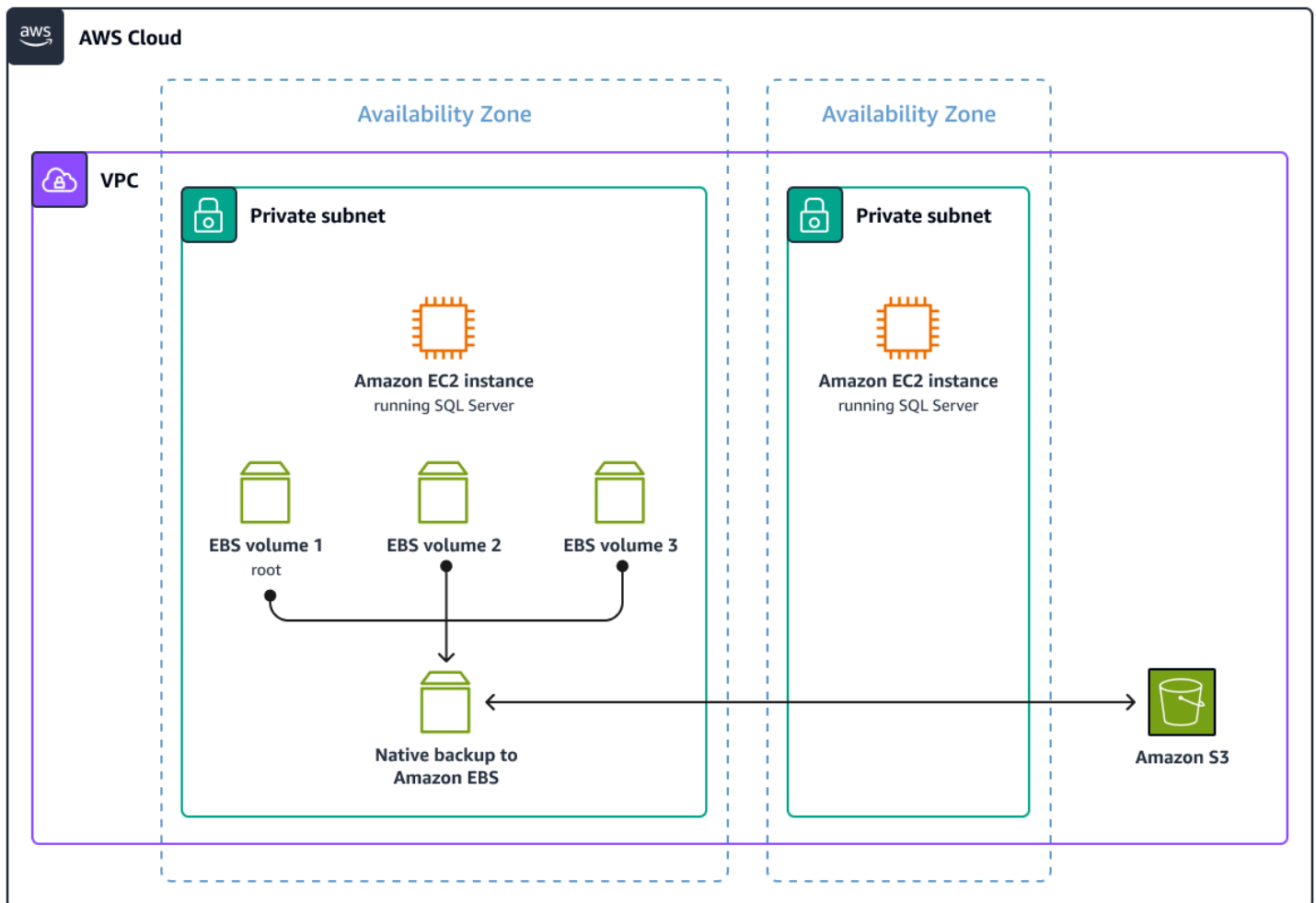
- É necessário configurar o backup e a restauração para cada banco de dados individual.
- É necessário gerenciar a [Política de ciclo de vida do Amazon S3](#) para os arquivos de backup.

Para obter mais informações sobre como configurar o Storage Gateway, consulte a publicação [Store SQL Server backups in Amazon S3 using AWS Storage Gateway](#) no Blog da AWS .

Backup nativo do SQL Server em volumes do EBS

É possível fazer um backup nativo do seu banco de dados do SQL Server e armazenar o arquivo em um volume do Amazon EBS. O Amazon EBS é um serviço de armazenamento em blocos de alta performance. Os volumes do EBS são elásticos, o que é compatível com criptografia. Eles podem ser desanexados e anexados a uma instância do EC2. Você pode fazer o backup do SQL Server em uma instância do EC2 no mesmo tipo ou em um tipo diferente de volume do EBS. Uma vantagem de fazer backup em um volume diferente do EBS é a economia de custos.

O diagrama a seguir mostra a arquitetura de um backup nativo em um volume do EBS.



Considere as seguintes vantagens de usar o backup nativo do SQL Server para volumes do EBS:

- Você pode fazer backups de bancos de dados individuais em uma instância do EC2 do SQL Server e restaurar um banco de dados individual em vez de precisar restaurar a instância completa.
- Há suporte a backups de vários arquivos.
- É possível programar trabalhos de backup usando o SQL Server Agent e o mecanismo de trabalho do SQL Server.
- Você pode obter benefícios de performance por meio de suas opções de hardware. Por exemplo, você pode usar volumes de armazenamento st1 para obter maior throughput.

Considere as seguintes limitações de usar backup nativo para volumes do EBS:

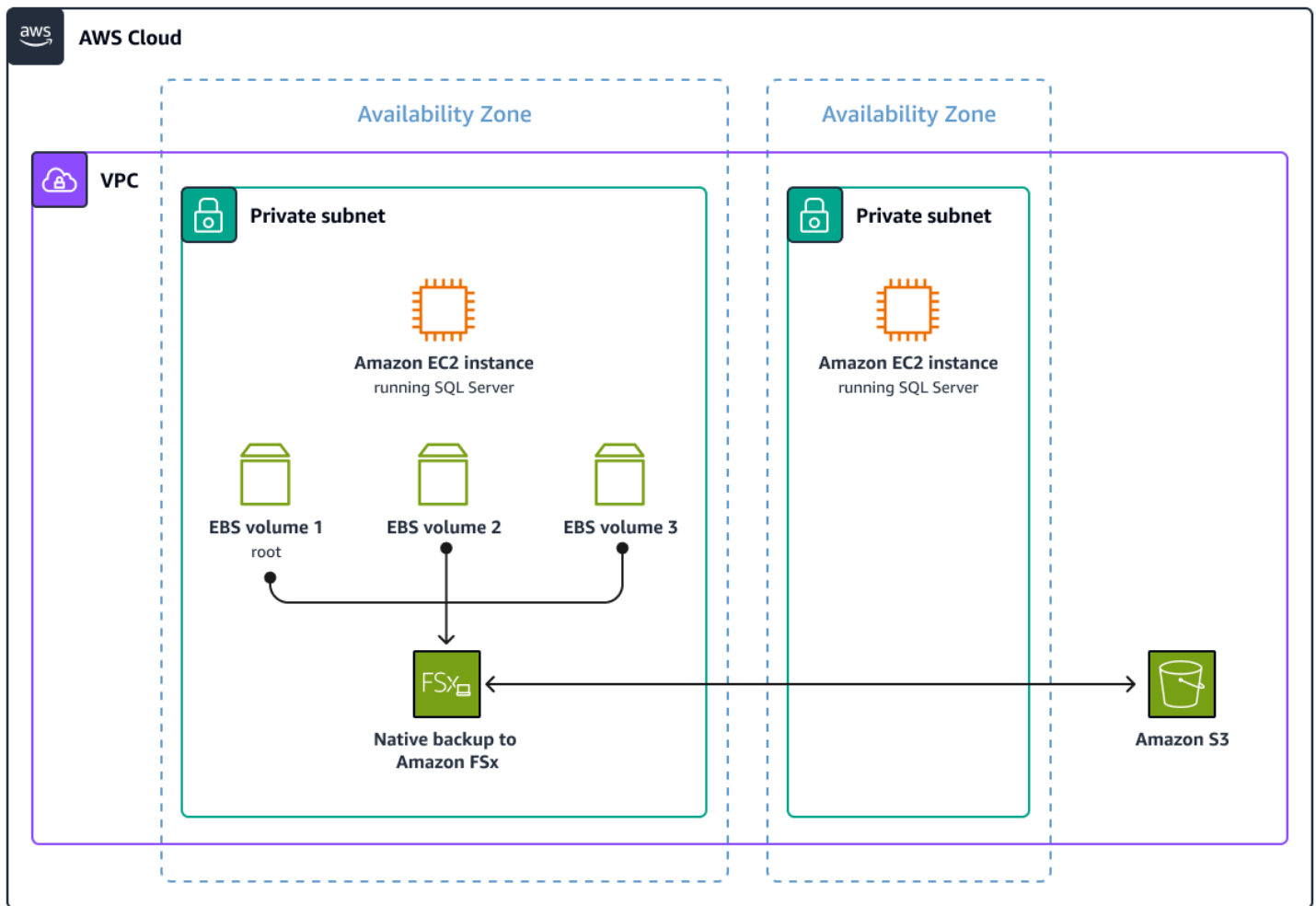
- É necessário mover manualmente os backups do volume do EBS para o Amazon S3.

- Para backups grandes, é necessário gerenciar o espaço em disco no Amazon EC2.
- Na instância do EC2, a throughput do Amazon EBS pode ser um gargalo.
- É necessário armazenamento adicional para armazenar backups no Amazon EBS.

Backup nativo do SQL Server para Amazon FSx para Windows File Server

[O Amazon FSx para Windows File Server](#) é um sistema de arquivos Windows nativo totalmente gerenciado que oferece até 64 TB de armazenamento projetado para oferecer desempenho rápido, previsível e consistente. AWS introduziu [suporte nativo para implantações de sistemas de arquivos Multi-AZ](#) no FSx Windows File Server. O suporte nativo facilita a implantação do armazenamento de arquivos do Windows na AWS com alta disponibilidade e redundância em várias zonas de disponibilidade. A AWS também introduziu suporte para [compartilhamentos de arquivos SMB de disponibilidade contínua \(CA\)](#). Você pode usar o FSx Windows File Server como armazenamento de backup para um banco de dados do SQL Server.

O diagrama a seguir mostra a arquitetura de um backup nativo do SQL Server FSx para o Windows File Server.



Considere os seguintes benefícios de usar o backup nativo do SQL Server FSx para o Windows File Server:

- Você pode fazer backup do seu banco de dados SQL Server em um compartilhamento de FSx arquivos da Amazon.
- Você pode fazer backups de bancos de dados individuais em uma instância do SQL Server e restaurar um banco de dados individual em vez de precisar restaurar a instância completa.
- Os backups multipartes são compatíveis.
- É possível programar trabalhos de backup usando o agente e o mecanismo de trabalho do SQL Server.
- As instâncias têm maior largura de banda da rede em comparação com o Amazon EBS.

Considere as seguintes limitações do uso do backup nativo do SQL Server FSx para o Windows File Server:

- Você deve mover manualmente os backups da Amazon para o Amazon S3 FSx usando AWS Backup ou AWS DataSync
- Backups grandes podem exigir uma sobrecarga adicional para o gerenciamento do espaço em disco na Amazon FSx.
- A throughput da rede da instância do EC2 pode ser um gargalo.
- É necessário armazenamento adicional para armazenar backups no FSx Windows File Server.

Backup do SQL Server na Amazon FSx para NetApp ONTAP

Os instantâneos com FSx for ONTAP são sempre consistentes com falhas, mas exigem que você desative (ou pause) seu banco I/O de dados para criar um instantâneo consistente com o aplicativo. Você pode usar NetApp SnapCenter (uma ferramenta de orquestração com plug-ins para aplicativos específicos, incluindo o SQL Server) com o ONTAP FSx para criar instantâneos consistentes com aplicativos e proteger, replicar e clonar seus bancos de dados sem custo adicional.

NetApp SnapCenter

NetApp SnapCenter é uma plataforma unificada para proteção de dados consistente com aplicativos. SnapCenter refere-se aos instantâneos como backups. Este guia adota a mesma convenção de nomenclatura. SnapCenter fornece um painel único para gerenciar backups, restaurações e clones consistentes com aplicativos. Você adiciona um SnapCenter plug-in para seu aplicativo de banco de dados específico para criar backups consistentes com o aplicativo. O SnapCenter plug-in para SQL Server fornece a seguinte funcionalidade que simplifica seu fluxo de trabalho de proteção de dados.

- Opções de backup e restauração com granularidade para backups completos e de logs
- Restauração no local e em um local alternativo

Para obter mais informações sobre SnapCenter, consulte a postagem [Proteja suas cargas de trabalho do SQL Server usando NetApp SnapCenter com a Amazon FSx for NetApp ONTAP](#) no blog AWS de armazenamento.

Otimização de custos para backups

As opções a seguir podem ajudar você a reduzir o custo de armazenamento de backups do SQL Server na AWS.

- Habilite a [compactação do SQL Server](#) durante a criação do arquivo de backup e envie o menor arquivo possível para o armazenamento. Por exemplo, uma taxa de compactação de

3:1 indica que você está economizando cerca de 66% em espaço em disco. Para consultar essas colunas, você pode usar a seguinte instrução Transact-SQL: `SELECT backup_size/compressed_backup_size FROM msdb..backupset;`

- Para backups que vão para buckets do S3, habilite a classe de armazenamento [Amazon S3 Intelligent-Tiering](#) para reduzir os custos de armazenamento em 30%.
- Para backups que vão FSx para o Windows File Server ou FSx para o ONTAP, use uma única zona de disponibilidade para economizar 50% (em comparação com o uso de várias zonas de disponibilidade). Para obter informações sobre preços, consulte Preços do [Amazon FSx para Windows File Server e Preços do Amazon FSx for NetApp ONTAP](#).
- A opção mais eficiente para o SQL Server 2022 é o backup direto no Amazon S3. Você pode economizar custos adicionais evitando o Storage Gateway.

Resultados do teste da avaliação comparativa de backups

Esta seção compara as opções a seguir, do ponto de vista de custo e performance, para um banco de dados de exemplo de 1 TB, com base nos resultados dos testes da avaliação comparativa de performance nas soluções de backup abordadas neste guia.

- Especificação de instância do EC2: r5d.8xlarge com Windows Server 2019 e a edição SQL Server 2019 Developer
- Especificação do banco de dados: 1 TB de tamanho com o TDE desabilitado

Os testes foram realizados com uma instância r5d.8xlarge e um banco de dados SQL Server de 1 TB como origem. O sistema de origem foi configurado de acordo com as práticas recomendadas, e o banco de dados de origem continha quatro arquivos de dados (250 GB cada) e um arquivo de log (50 GB) distribuídos em volumes gp3 separados. O comando BACKUP nativo do SQL Server incluiu a gravação em dez arquivos de backup usando compactação para otimizar a performance do backup e reduzir a quantidade de dados enviados pela rede e gravados no destino. Em todos os casos de teste, a performance do armazenamento foi o gargalo.

Há uma variedade quase infinita de configurações possíveis para esses tipos de teste. Esse teste se concentrou na otimização para performance, custo, escalabilidade e casos de uso reais. A tabela a seguir mostra as métricas de performance que foram capturadas para as opções de destino de backup.

Opções de backup	Nível	Duração da execução (Aprox.)	Taxa de backup	Custo em USD por mês*
Backup nativo para HDD st1 EBS local, 2 TB	Banco de dados	00:30:46 min.	554,7 Mbps	\$92,16
Backup nativo para SSD gp3 EBS local, 2 TB	Banco de dados	00:22:00 min.	512 Mbps	\$193,84
Backup nativo FSx para Windows File Server HDD, taxa de transferência de 2 TB @512 Mbps	Banco de dados	00:20:58 min.	814,0 Mbps	\$1.146
Backup nativo FSx para Windows File Server SSD, taxa de transferência de 2 TB @512 Mbps	Banco de dados	00:20:00 min.	814,0 Mbps	\$1.326
Backup nativo no Gateway de Arquivos do S3 m6i.4xlarge (16 vCPU, 64 GB) com 2 TB gp3	Banco de dados	00:23:20 min.	731,5 Mbps	\$470,42
Snapshot do EBS VSS	Volume do EBS	00:00:02 seg.	Snapshot N/A	\$51

Opções de backup	Nível	Duração da execução (Aprox.)	Taxa de backup	Custo em USD por mês*
		00:00:53 seg.		
AWS Backup (Backup da AMI)	AMI	00:00:04 seg. 00:08:00 min.	Snapshot N/A	<u>\$75</u>
Backup nativo do SQL Server diretamente no Amazon S3 (SQL Server 2022)	Banco de dados	00:12:00 min.	731,5 Mbps	<u>Primeiros 50 TB/mês, USD 0,023 por GB USD 23,55 por mês</u>
Backup nativo FSx para ONTAP (usando SnapCenter)	Banco de dados	–	–	<u>\$440,20</u>

A tabela anterior supõe o seguinte:

- Os custos de transferência de dados e do Amazon S3 não estão incluídos.
- O preço do armazenamento está incluído no preço da instância.
- Os custos são baseados na região us-east-1.
- O throughput e a IOPS aumentam em 10% com vários backups que têm uma taxa geral de alteração de 10% ao longo do mês.

Os resultados do teste mostram que a opção mais rápida é um backup nativo do banco de dados SQL Server FSx para o Windows File Server. Um backup no Storage Gateway e nos volumes do EBS anexados localmente é a opção mais econômica, mas tem uma performance mais lenta. Para backups em nível de servidor (AMI), recomendamos o uso AWS Backup para otimizar o desempenho, o custo e a capacidade de gerenciamento.

Recomendações de otimização de custos

Compreender as possíveis soluções para fazer backup do SQL Server no Amazon EC2 é fundamental para proteger seus dados, garantindo que atenda às suas necessidades de backup, e colocar um plano em prática para se recuperar de eventos críticos. As diferentes formas de fazer backup e restaurar suas instâncias e bancos de dados do SQL Server analisados nesta seção podem ajudar você a criar uma estratégia de backup e restauração que proteja seus dados e atenda aos requisitos da sua organização.

Esta seção abrange as seguintes opções de backup:

- Compactação
- Amazon S3 Intelligent-Tiering
- Zona de disponibilidade única
- Backup em URL

A orientação fornecida para cada uma dessas opções é de alto nível. Se você deseja implementar alguma dessas recomendações em sua organização, recomendamos que entre em contato com a equipe da sua conta. A equipe pode então contratar um Microsoft Specialist SA para conduzir a conversa. Você também pode entrar em contato pelo e-mail optimize-microsoft@amazon.com.

Em resumo, recomendamos o seguinte:

- Se você estiver usando o SQL Server 2022, fazer backup no Amazon S3 é a opção mais econômica.
- Se você estiver usando o SQL Server 2019 e edições anteriores do SQL Server, considere fazer backup no Storage Gateway baseado no Amazon S3 como a opção mais econômica.

Compactação

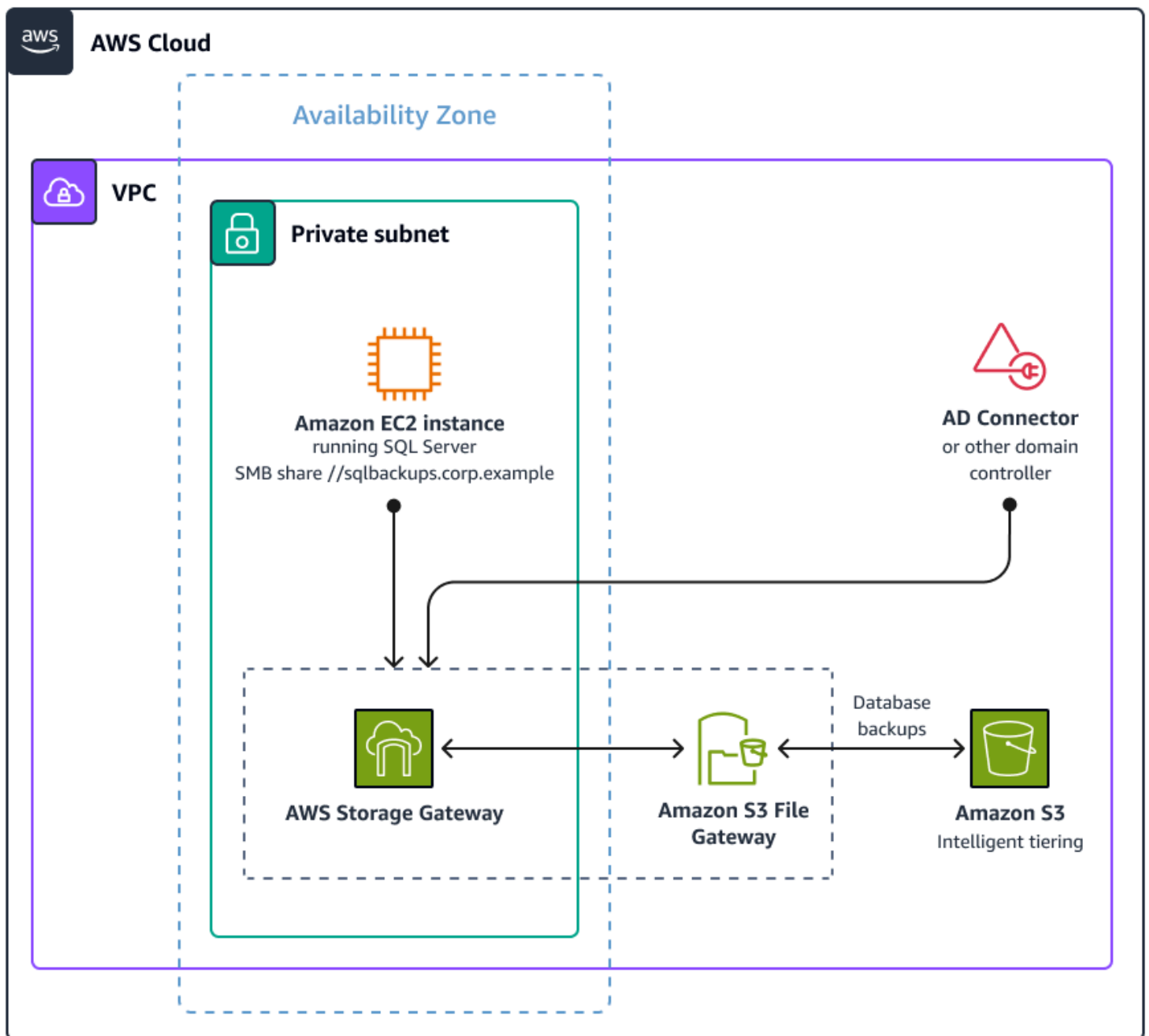
O objetivo da compactação é ter menos armazenamento consumido por cada backup, o que é benéfico para várias opções de armazenamento. Você deve habilitar a compactação para um backup do SQL Server no nível da [instância do SQL Server](#). O seguinte exemplo mostra como adicionar a palavra-chave de compactação com um banco de dados de backup:

```
BACKUP DATABASE <database_name> TO DISK WITH COMPRESSION (ALGORITHM = QAT_DEFLATE)
```

Amazon S3 Intelligent-Tiering

Para backups que vão para buckets do Amazon S3, você pode habilitar o [Amazon S3 Intelligent-Tiering](#) como sua [classe de armazenamento](#) do Gateway de Arquivos do Amazon S3. Isso pode reduzir os custos do armazenamento em até 30%. Em seguida, você monta o Gateway de Arquivos do S3 em seus servidores SQL usando um compartilhamento de arquivos SMB que pode ser integrado ao seu [domínio do Active Directory](#). Isso fornece controle de acesso para seu compartilhamento, a capacidade de aproveitar contas de serviço existentes e acesso ao Amazon S3 usando um protocolo de arquivo comum focado na Microsoft. Para contas que talvez não tenham conectividade direta com um controlador de domínio, você pode usar o [Active Directory Connector](#) para facilitar a comunicação com o Active Directory on-premises ou na nuvem. Para definir as configurações do Active Directory no gateway, você deve especificar o conector do Active Directory IPs para que o controlador de domínio faça solicitações de proxy para o Active Directory.

O diagrama a seguir mostra uma arquitetura para uma solução baseada no S3 Intelligent-Tiering.



Por padrão, os arquivos de backup gravados no bucket do S3 usam a camada Standard. Para converter os arquivos de backup da camada Standard para o S3 Intelligent-Tiering, você deve [criar uma regra de ciclo de vida](#). Você também pode usar o [Console de gerenciamento da AWS](#) para habilitar o S3 Intelligent-Tiering. Para obter mais informações, consulte [Comece a usar o Amazon S3 Intelligent-Tiering](#) na documentação da AWS .

Zona de disponibilidade única

Para criar um sistema de arquivos de zona de disponibilidade única, escolha a opção Single-AZ ao [criar um sistema de arquivos FSx para Windows File Server](#). A Amazon FSx também faz backups altamente duráveis (armazenados no Amazon S3) do seu sistema de arquivos diariamente usando o Windows Volume Shadow Copy Service e permite que você faça backups adicionais a qualquer momento. Tenha em mente alguns dos problemas com o uso de uma única zona de disponibilidade. Por exemplo, o compartilhamento de arquivos SMB ficará inacessível se uma zona de disponibilidade afetada em que o sistema de arquivos é provisionado ficar inativa por várias horas. Se você precisar de acesso aos dados, deverá restaurá-los de backups em uma zona de disponibilidade disponível na região de origem. Para obter mais informações, consulte a seção [Usar uma única zona de disponibilidade](#) deste guia.

Backup em URL

Para o SQL Server 2022, o recurso de [backup em URL](#) permite o backup direto no Amazon S3. Essa é a abordagem de backup ideal para o SQL Server 2022 em execução, AWS pois você obtém o conjunto completo de recursos do Amazon S3 na camada de armazenamento e remove o custo do AWS Storage Gateway dispositivo necessário nas versões anteriores para facilitar essa funcionalidade. Há dois custos principais a serem considerados ao implementar esse recurso: custos de transferência de dados e a classe de armazenamento S3 escolhida. Se você quiser os recursos nativos de recuperação de desastres do Amazon S3, deve considerar que a [replicação entre regiões](#) incorre em [custos de saída de dados](#) entre regiões. Para saber mais sobre como configurar essa opção, consulte a publicação [Backup SQL Server databases to Amazon S3](#) em Microsoft Workloads no Blog da AWS .

Recursos adicionais do

- [Opções de backup e restauração para SQL Server no Amazon EC2 \(orientação AWS prescritiva\)](#)
- [Point-in-time recuperação e backup contínuo para Amazon RDS com AWS Backup](#) (AWS Storage Blog)
- [Proteja suas cargas de trabalho do SQL Server usando o NetApp SnapCenter Amazon FSx for NetApp ONTAP](#) (AWS Storage Blog)
- [Introdução ao uso do Amazon S3 Intelligent-Tiering](#) (Centro de recursos para começar a usar)AWS
- [Estratégias de backup e restauração do Amazon RDS for SQL Server](#)AWS (blog do banco de dados)

- [Migre um banco de dados local do Microsoft SQL Server para o Amazon EC2 \(orientação prescritiva\)](#) AWS
- [Melhores práticas para implantar o Microsoft SQL Server no Amazon EC2](#) AWS (Whitepaper)

Modernizar bancos de dados do SQL Server.

Visão geral do

Se você está iniciando uma jornada para modernizar bancos de dados legados para escalabilidade, performance e otimização de custos, talvez esteja enfrentando desafios com bancos de dados comerciais como o SQL Server. Os bancos de dados comerciais são caros, criam dependência do fornecedor e oferecem termos de licenciamento punitivos. Esta seção fornece uma visão geral de alto nível das opções de migração e modernização do SQL Server para bancos de dados de código aberto e informações sobre como escolher a melhor opção para sua workload.

Você pode refatorar seus bancos de dados do SQL Server em bancos de dados de código aberto, como o Amazon Aurora PostgreSQL, para economizar nos custos de licenciamento do Windows e do SQL Server. Bancos de dados modernos nativos da nuvem, como o Aurora, combinam a flexibilidade e o baixo custo dos bancos de dados de código aberto com os recursos robustos e de nível corporativo dos bancos de dados comerciais. Se você tiver workloads variáveis ou workloads multioatárias, também poderá migrar para o [Aurora Sem Servidor v2](#). Isso pode reduzir os custos em até 90%, dependendo das características da workload. Além disso, AWS oferece recursos como o [Babelfish para Aurora PostgreSQL](#), ferramentas como [AWS Schema Conversion Tool \(AWS SCT\)](#) e serviços como [AWS Database Migration Service \(AWS DMS\)](#) para simplificar a migração e a modernização dos bancos de dados do SQL Server no. AWS

Ofertas de banco de dados

A migração do SQL Server no Windows para bancos de dados de código aberto como o Amazon Aurora, Amazon RDS para MySQL ou Amazon RDS para PostgreSQL pode oferecer economias de custo significativas sem comprometer a performance ou os recursos. Considere o seguinte:

- A mudança da edição SQL Server Enterprise no Amazon EC2 para o Amazon RDS para PostgreSQL ou o Amazon RDS para MySQL pode resultar em economias de até 80%.
- A mudança da edição SQL Server Enterprise no Amazon EC2 para a edição do Amazon RDS compatível com PostgreSQL ou a edição do Amazon Aurora compatível com MySQL pode resultar em economias de até 70%.

Para workloads de banco de dados tradicionais, o Amazon RDS para PostgreSQL e o Amazon RDS para MySQL atendem aos requisitos e fornecem uma solução econômica para bancos de dados relacionais. O Aurora adicionou vários recursos de disponibilidade e performance antes limitados a fornecedores comerciais caros. Os recursos de resiliência do Aurora têm um custo adicional. No entanto, em comparação com recursos semelhantes de outros fornecedores comerciais, os custos de resiliência do Aurora ainda são mais baratos do que os cobrados por softwares comerciais pelo mesmo tipo de recursos. A arquitetura do Aurora é otimizada para oferecer melhorias significativas na performance em comparação com as implantações padrão do MySQL e do PostgreSQL.

Como o Aurora é compatível com bancos de dados PostgreSQL e MySQL de código aberto, há o benefício adicional da portabilidade. Se a melhor opção for o Amazon RDS para PostgreSQL, o Amazon RDS para MySQL ou o Aurora, tudo se resume a entender os requisitos dos negócios e mapear os recursos necessários para a melhor opção.

Comparação entre o Amazon RDS e o Aurora

A tabela a seguir resume as principais diferenças entre o Amazon RDS e o Amazon Aurora.

Categoria	Amazon RDS para PostgreSQL e Amazon RDS para MySQL	Aurora PostgreSQL ou Aurora MySQL
desempenho	Boa performance	Performance 3x ou melhor
Failover	Normalmente, de 60 a 120 segundos*	Normalmente, 30 segundos
Escalabilidade	Até 5 réplicas de leitura Atraso em segundos	Até 15 réplicas de leitura Atraso em milissegundos
Armazenamento	Até 64 TB	Até 128 TB
Armazenamento de HA	Multi-AZ com um ou dois modos de espera, cada um com cópia do banco de dados	6 cópias de dados em 3 zonas de disponibilidade por padrão
Backup	Backups de logs e snapshots diários	Backup contínuo e assíncrono no Amazon S3
Inovações com o Aurora	NA	100 GB

Categoria	Amazon RDS para PostgreSQL e Amazon RDS para MySQL	Aurora PostgreSQL ou Aurora MySQL
		Clonagem rápida de banco de dados
	Réplicas de leitura com ajuste de escala automático	
	Gerenciamento de planos de consultas	
	Aurora Sem Servidor	
	Réplicas entre regiões com banco de dados global	
	Gerenciamento do cache de clusters**	
	Consulta paralela	
	Fluxos de atividades do banco de dados	

*Transações grandes podem aumentar os tempos de failover

**Disponível no Aurora PostgreSQL

A tabela a seguir mostra o custo mensal estimado dos diferentes serviços de banco de dados abordados nesta seção.

Serviço de banco de dados	Custo em USD por mês*	AWS Calculadora de Preços (requer Conta da AWS)
Amazon RDS para edição SQL Server Enterprise	\$3.750	Estimativa

Serviço de banco de dados	Custo em USD por mês*	AWS Calculadora de Preços (requer Conta da AWS)
Amazon RDS para edição SQL Server Standard	\$2.318	Estimativa
Edição SQL Server Enterprise no Amazon EC2	\$2.835	Estimativa
Edição SQL Server Standard no Amazon EC2	\$1.345	Estimativa
Amazon RDS para PostgreSQL	\$742	Estimativa
Amazon RDS para MySQL	\$712	Estimativa
Aurora PostgreSQL	\$1.032	Estimativa
Aurora MySQL	\$1.031	Estimativa

* O preço do armazenamento está incluído no preço da instância. Os custos são baseados na região us-east-1. O throughput e a IOPS são suposições. Os cálculos são para instâncias r6i.2xlarge e r6g.2xlarge.

Recomendações de otimização de custos

As migrações de banco de dados heterogêneas geralmente exigem a conversão do esquema do banco de dados do mecanismo de banco de dados de origem para o de destino e a migração de dados do banco de dados de origem para o de destino. A primeira etapa para a migração é avaliar e converter objetos de código e esquema do SQL Server no mecanismo de banco de dados de destino.

Você pode usar o [AWS Schema Conversion Tool \(AWS SCT\)](#) para avaliar e analisar a compatibilidade do banco de dados com várias opções de banco de dados de código aberto de destino, como o Amazon RDS para MySQL ou o Amazon RDS para PostgreSQL, o Aurora MySQL e o PostgreSQL. Você também pode usar a ferramenta Babelfish Compass para avaliar a compatibilidade com o Babelfish para Aurora PostgreSQL. Isso torna o Compass AWS SCT e o Compass ferramentas poderosas para entender o trabalho inicial envolvido antes de decidir sobre

uma estratégia de migração. Se você decidir continuar, o AWS SCT automatizará as alterações necessárias no esquema. A filosofia central por trás do Babelfish Compass é permitir a migração do banco de dados SQL para o Aurora sem nenhuma ou com pouquíssimas alterações. O Compass avaliará o banco de dados SQL existente para determinar se isso pode ser feito. Dessa forma, o resultado é conhecido antes que qualquer esforço seja feito na migração de dados do SQL Server para o Aurora.

AWS SCT automatiza a conversão e a migração do esquema e do código do banco de dados para o mecanismo do banco de dados de destino. Você pode usar o Babelfish para Aurora PostgreSQL para migrar seu banco de dados e a aplicação do SQL Server para o Aurora PostgreSQL sem nenhuma ou com o mínimo de alterações no esquema. Isso pode acelerar suas migrações.

Depois que o esquema for migrado, você poderá usá-lo AWS DMS para migrar os dados. AWS DMS pode realizar o carregamento total dos dados e replicar as alterações para realizar a migração com o mínimo de tempo de inatividade.

Esta seção explora as seguintes ferramentas com mais detalhes:

- AWS Schema Conversion Tool
- Babelfish for Aurora PostgreSQL
- Babelfish Compass
- AWS Database Migration Service

AWS Schema Conversion Tool

Você pode usar AWS SCT para avaliar seus bancos de dados SQL Server existentes e avaliar a compatibilidade com o Amazon RDS ou o Aurora. Para simplificar o processo de migração, você também pode usar AWS SCT para converter o esquema de um mecanismo de banco de dados para outro em uma migração de banco de dados heterogênea. Você pode usar AWS SCT para avaliar seu aplicativo e converter o código do aplicativo incorporado para aplicativos escritos em C#, C++, Java e outras linguagens. Para obter mais informações, consulte [Converting application SQL using AWS SCT](#) na documentação do AWS SCT .

AWS SCT é uma AWS ferramenta gratuita que oferece suporte a várias [fontes](#) de banco de dados. Para usá-lo AWS SCT, você o aponta para o banco de dados de origem e, em seguida, executa uma avaliação. Em seguida, o [AWS SCT](#) avalia o esquema e gera o relatório de avaliação. Os relatórios de avaliação incluem um resumo executivo, a complexidade e o esforço de migração, mecanismos

de banco de dados de destino adequados e recomendações para conversão. Para fazer o download AWS SCT, [consulte Instalação, verificação e atualização AWS SCT](#) na AWS SCT documentação.

A tabela a seguir mostra um exemplo de resumo executivo gerado pelo AWS SCT para mostrar a complexidade envolvida na alteração do banco de dados para diferentes plataformas de destino.

Plataforma de destino	Alterações automáticas ou mínimas			Ações complexas			
	Objetos de armazenamento	Objetos de código	Ações de conversão	Objetos de armazenamento		Objetos de código	
Amazon RDS para MySQL	60 (98%)	8 (35%)	42	1 (2%)	1	15 (65%)	56
Amazon Aurora Edição Compatível com MySQL	60 (98%)	8 (35%)	42	1 (2%)	1	15 (65%)	56
Amazon RDS para PostgreSQL	60 (98%)	12 (52%)	54	1 (2%)	1	11 (48%)	26
Amazon Aurora Edição Compatível com	60 (98%)	12 (52%)	54	1 (2%)	1	11 (48%)	26

PostgreSQL

L

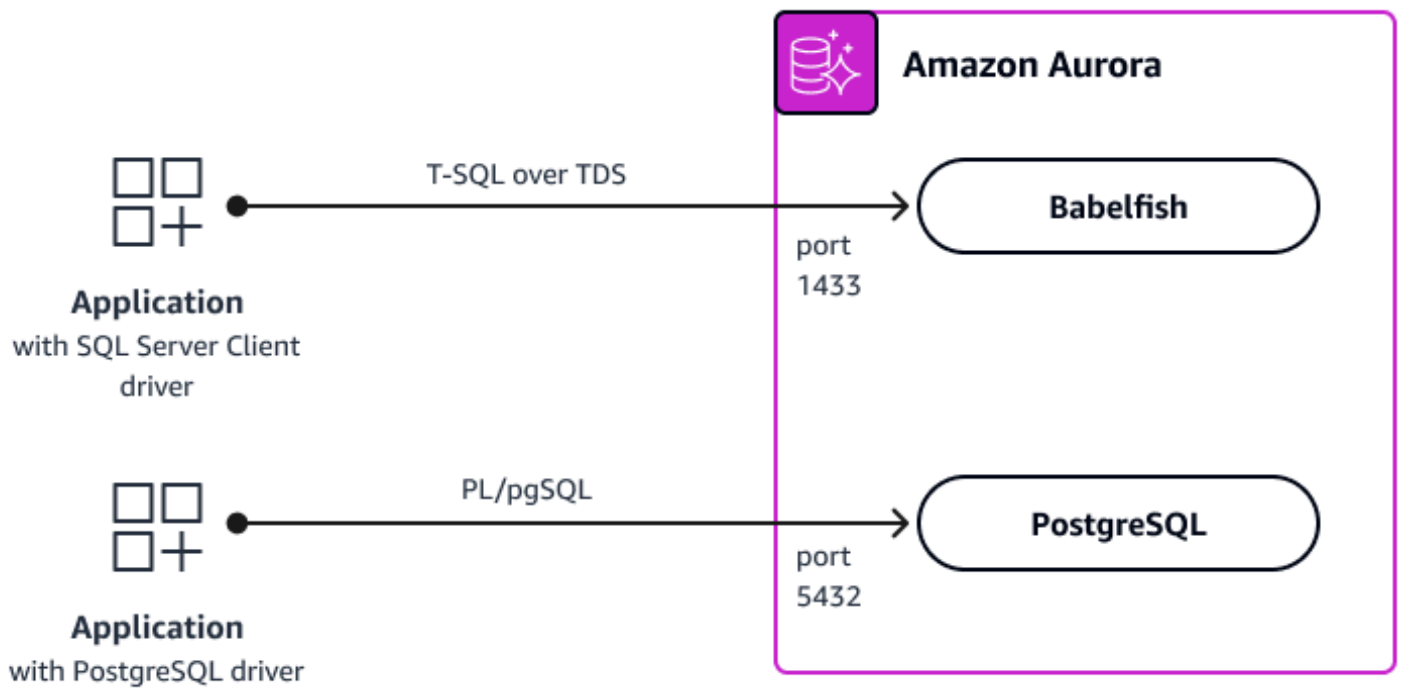
Amazon RDS para MariaDB	60 (98%)	7 (30%)	42	1 (2%)	1	16 (70%)	58
banco de dados de origem	61 (100%)	9 (39%)	124	0 (0%)	0	14 (61%)	25
AWS Glue	0 (0%)	17 (100%)	0	0 (0%)	0	0 (0%)	0
Babelfish	59 (97%)	10 (45%)	20	2 (3%)	2	12 (55%)	30

Um AWS SCT relatório também fornece detalhes sobre os elementos do esquema que não podem ser convertidos automaticamente. Você pode fechar as lacunas AWS SCT de conversão e otimizar os esquemas de destino consultando os manuais de [AWS migração](#). Há muitos playbooks de migração de banco de dados para ajudar em migrações heterogêneas.

Babelfish for Aurora PostgreSQL

O Babelfish para Aurora PostgreSQL estende o Aurora PostgreSQL com a capacidade de aceitar conexões de banco de dados de clientes do SQL Server. O Babelfish permite que as aplicações originalmente criadas para o SQL Server possam funcionar diretamente com o Aurora PostgreSQL, com poucas alterações de código e sem alterar drivers de banco de dados. O Babelfish torna o Aurora PostgreSQL bilíngue para que o Aurora PostgreSQL possa funcionar tanto com o T-SQL quanto com as linguagens PL/pgSQL. O Babelfish minimiza os esforços de migração do SQL Server para o Aurora PostgreSQL. Isso acelera as migrações, minimiza os riscos e reduz significativamente os custos de migração. Você pode continuar a usar T-SQL após as migrações, mas também há a [opção de usar ferramentas nativas do PostgreSQL](#) para desenvolvimento.

O diagrama a seguir ilustra como uma aplicação usando T-SQL se conecta à porta padrão 1433 no SQL Server e usa o tradutor Babelfish para se comunicar com o banco de dados Aurora PostgreSQL, enquanto uma aplicação usando PL/pgSQL pode se conectar direta e simultaneamente ao banco de dados Aurora PostgreSQL usando a porta padrão 5432 no Aurora PostgreSQL.



O Babelfish não é compatível com determinados recursos T-SQL do SQL Server. Por esse motivo, a Amazon fornece ferramentas de avaliação para fazer uma line-by-line análise de suas instruções SQL e determinar se alguma delas não é suportada pelo Babelfish.

Há duas opções para avaliações do Babelfish. AWS SCT pode avaliar a compatibilidade do seu banco de dados SQL Server com o Babelfish. Outra opção é a ferramenta Babelfish Compass, que é uma solução recomendada porque a ferramenta Compass é atualizada de acordo com as novas versões do Babelfish para Aurora PostgreSQL.

Babelfish Compass

O [Babelfish Compass](#) é uma ferramenta gratuita para download que se alinha à versão mais recente do Babelfish para Aurora PostgreSQL. Em contraste, AWS SCT suportará versões mais recentes do Babelfish depois de algum tempo. O [Babelfish Compass](#) é executado no esquema do banco de dados SQL Server. Você também pode extrair o esquema do banco de dados do SQL Server de origem usando ferramentas como o SQL Server Management Studio (SSMS). Em seguida, você pode executar o esquema por meio do Babelfish Compass. Isso gera o relatório detalhando a compatibilidade do esquema do SQL Server com o Babelfish e se alguma alteração é necessária antes da migração. A ferramenta Babelfish Compass também pode automatizar muitas dessas alterações e, finalmente, acelerar suas migrações.

Depois que a avaliação e as alterações forem concluídas, você poderá migrar o esquema para o Aurora PostgreSQL usando ferramentas nativas do SQL Server, como o SSMS ou sqlcmd. Para obter instruções, consulte a publicação [Migrate from SQL Server to Amazon Aurora using Babelfish](#) no blog AWS Database.

AWS Database Migration Service

Depois que o esquema for migrado, você poderá usar AWS Database Migration Service (AWS DMS) para migrar os dados AWS com o mínimo de tempo de inatividade. AWS DMS não apenas carrega os dados por completo, mas também replica as alterações da origem para o destino enquanto o sistema de origem está funcionando. Depois que os bancos de dados de origem e de destino estiverem sincronizados, a atividade de transição pode ocorrer onde o aplicativo é direcionado para o banco de dados de destino, concluindo a migração. AWS DMS atualmente só executa o carregamento total de dados com o Babelfish para um destino do Aurora PostgreSQL e não replica as alterações. Para obter mais informações, consulte [Usando o Babelfish como alvo AWS Database Migration Service](#) na AWS DMS documentação.

AWS DMS pode fazer migrações homogêneas (no mesmo mecanismo de banco de dados) e heterogêneas (em diferentes mecanismos de banco de dados). AWS DMS oferece suporte a vários mecanismos de banco de dados de origem e destino. Para obter mais informações, consulte a publicação Como [migrar seu banco de dados do SQL Server para o Amazon RDS for SQL Server AWS DMS](#) usando AWS o blog do banco de dados.

Recursos adicionais do

- [Adeus Microsoft SQL Server, Olá Babelfish \(Blog de notícias\)](#)AWS
- [Converta esquemas de banco de dados e SQL de aplicativos usando a AWS Schema Conversion Tool CLI](#)AWS (Database Blog)
- [Migre o SQL Server para o Amazon Aurora PostgreSQL usando as melhores práticas e lições aprendidas](#) em campo (blog do banco de dados)AWS
- [Valide objetos de banco de dados após a migração do Microsoft SQL Server para o Amazon RDS for PostgreSQL e o Amazon Aurora PostgreSQL](#) (blog do banco de dados)AWS

Otimizar o armazenamento para o SQL Server

Visão geral do

Esta seção se concentra nas otimizações de custos do armazenamento SSD do Amazon Elastic Block Store (Amazon EBS) para SQL Server em workloads do EC2.

Você tem uma grande variedade de opções de armazenamento para implantar e executar cargas de trabalho do SQL Server em. AWS A seleção do armazenamento certo deve ser baseada na finalidade, na arquitetura, na durabilidade, no desempenho, na capacidade e no custo. AWS os clientes que executam cargas de trabalho do SQL Server geralmente utilizam uma combinação de armazenamento Amazon EBS FSx, NVMe Amazon e Amazon Simple Storage Service (Amazon S3).

O Amazon EBS é um armazenamento anexado à rede conectado a instâncias computacionais do EC2, e é utilizado para armazenar e processar arquivos gerais do sistema operacional, da aplicação, do banco de dados e de backup. O armazenamento em unidade de estado sólido (SSD) do Amazon EBS inclui SSD de uso geral (gp2 e gp3) e SSD de IOPS provisionadas (io1, io2 e io2BX). Considere o seguinte:

- Algumas instâncias do EC2, como r5d, têm um local NVMe SSDs fisicamente conectado à instância hospedeira. Esses volumes fornecem armazenamento ao nível do bloco que é comumente usado para a extensão de grupo de buffers ou tempdb do SQL Server.
- O Amazon FSx for Windows File Server é um serviço de armazenamento de arquivos totalmente gerenciado, enquanto o Amazon FSx NetApp for ONTAP é um armazenamento compartilhado totalmente gerenciado NetApp baseado no popular sistema de arquivos ONTAP. O Amazon FSx é frequentemente usado para executar workloads do SQL Server em uma configuração de instância clusterizada de failover (FCI) do SQL Server de alta disponibilidade. Essa solução hospeda arquivos de dados e logs do SQL Server, o que reduz os requisitos de performance do EBS nas instâncias do EC2.
- O Amazon S3 é um serviço de armazenamento de objetos que oferece escalabilidade, disponibilidade de dados, segurança e performance líderes do setor. Você pode armazenar arquivos de backup nativos do SQL Server AMIs, snapshots do EBS, logs de aplicativos e muito mais no Amazon S3.

Tipos, performance e custo de armazenamento SSD para o Amazon EBS

Os custos de armazenamento SSD do Amazon EBS geralmente aumentam à medida que a durabilidade e a performance aumentam. Atualmente, o armazenamento vem em cinco tipos de volume, cada um com suas [próprias métricas de performance exclusivas](#). Para obter um resumo dos casos de uso e das características dos volumes baseados em SSD, consulte a tabela na seção [Volumes de unidades de estado sólido \(SSD\)](#) da documentação do Amazon EBS.

Você pode usar CloudWatch a Amazon para monitorar o desempenho do SSD, capturar dados de tendências e definir alarmes quando determinados limites forem atingidos. Se você estiver executando cargas de trabalho do SQL Server em AWS, considere habilitar o [monitoramento detalhado](#) e implantar [métricas CloudWatch personalizadas para capturar métricas](#) detalhadas de desempenho de volume, como latência de disco, IOPS, taxa de transferência, tamanho da fila de disco, capacidade usada versus capacidade livre e muito mais. Você pode usar essas métricas de CloudWatch desempenho para identificar armazenamento subprovisionado e superprovisionado e fornecer pontos de dados históricos para definir com precisão os requisitos de armazenamento.

Os custos de armazenamento SSD para o Amazon EBS também variam com base na capacidade alocada. A tabela abaixo mostra uma comparação dos diferentes tipos de volume. Todos os tipos de volume têm 1 TB de capacidade e configurações de performance semelhantes.

Tipo de volume	IOPS máxima (16 KiB de E/S)	Throughput máximo (E/S de 128 KiB)	Preço por 1 TB	Percentual de economia de custos
gp2	3.000	250	\$102,40	
gp3	3.000	250	\$86,92	15%
io1	16.000	500	\$1.168	
io2	16.000	500	\$1.168	
gp3	16.000	500	\$146,92	87%
io2bx	16.000	4.000	\$1.168	
gp3	16.000	1.000	\$181,92	84%

Note

As métricas de performance e custo na tabela anterior são por volume, com base em uma [estimativa](#) da AWS Calculadora de Preços. Conta da AWS É necessário um para acessar a estimativa no AWS Calculadora de Preços.

Os volumes gp3 SSD do Amazon EBS oferecem excelente performance a um baixo custo. Você pode economizar até 87% se escolher um volume gp3 em vez de volumes io1 ou io2 para cargas de trabalho que exigem menos de 16.000 IOPS e 500 taxas de transferência. MiBps

Os volumes io2 Block Express (io2BX) oferecem maior performance em relação aos volumes io2 regulares. Com 16.000 IOPS, os volumes io1 ou io2 são capazes de atingir apenas 500 MiBps taxas de transferência, enquanto os volumes IO2Bx podem ser configurados com até 4.000 taxas de transferência. MiBps Em comparação com os volumes io1 e io2, os volumes io2BX fornecem mais de quatro vezes o throughput entre 16.000 e 64.000 IOPS, exatamente pelo mesmo preço. Volumes io2 regulares podem ser convertidos em volumes io2BX anexando-os a instâncias do EC2 compatíveis com io2BX. Para obter uma lista de instâncias do EC2 compatíveis com io2BX, consulte [Provisioned IOPS SSD volumes](#) na documentação do Amazon EBS. Antes de implantar um novo armazenamento, você pode usar a [AWS Calculadora de Preços](#) para estimar seu custo mensal e entender o impacto no custo com base nas compensações entre durabilidade, performance e capacidade.

Otimização geral de custos de SSDs para o Amazon EBS

Recomendamos que você avalie o que está armazenando e certifique-se de que está usando o tipo e a classe de armazenamento corretos. Por exemplo, o Amazon S3 oferece um ótimo preço, políticas de ciclo de vida integradas e opções de replicação ideais para backups do SQL Server. O SQL Server 2022 tem a capacidade de fazer backup diretamente no Amazon S3, enquanto as versões anteriores do SQL Server dependem de backups locais nativos. Se você estiver executando versões mais antigas do SQL Server, considere fazer backup nos volumes HDD do Amazon EBS e depois copiar o backup para o Amazon S3. Essa solução pode economizar 53% em vez de usar volumes gp3 para backups.

A tabela a seguir mostra a diferença de preços de 1 TB de armazenamento no Amazon EBS gp3, Amazon EBS HDD st1 e Amazon S3.

Tipo de armazenamento	Capacity	Preço por mês
EBS gp3 500 MiBps	1 TB	\$96,92
EBS st1 burst 500 MiBps		\$46,08
S3 Standard		\$23,55
S3 Standard (acesso infrequente)		\$12,80
S3 Glacier Deep Archive		\$1,03

Note

As métricas de custo na tabela anterior são baseadas em uma [estimativa](#) na AWS Calculadora de Preços. Conta da AWS É necessário um para acessar a estimativa no AWS Calculadora de Preços.

Recomendamos que você considere o seguinte:

- Habilite o monitoramento detalhado e implante métricas CloudWatch personalizadas para capturar com precisão seus requisitos de desempenho de armazenamento.
- Atualize o armazenamento do Amazon EBS de gp2 para gp3 para reduzir custos, aumentar a flexibilidade e melhorar a performance.
- Atualize o armazenamento do Amazon EBS de io1 para io2 para maior durabilidade e flexibilidade de performance.
- Use io2BX em vez de io1 ou io2 quando possível para aumentar a durabilidade e a performance.
- Considere uma mix-and-match abordagem ao escolher o armazenamento para ajudar a reduzir os requisitos de capacidade e o custo de volumes de alto desempenho. Por exemplo, você pode usar volumes gp3 de baixo custo para seu volume raiz (sistema operacional), instalação do SQL Server, bancos de dados do sistema (excluindo o tempdb) e bancos de dados de usuários de baixa performance. Isso pode ajudar a reduzir a capacidade e o custo de um volume io2, que pode ser dedicado a bancos de dados de usuários de alta performance.

- Se você estiver hospedando bancos de dados do SQL Server em AWS, recomendamos que você use vários arquivos de dados do SQL Server por banco de dados. Isso permite a oportunidade de distribuir read/write cargas de trabalho em vários volumes, reduzindo os requisitos de desempenho e capacidade por volume e, conseqüentemente, reduzindo os custos.
- Mesmo que as workloads de produção exijam armazenamento de alta performance, como io1 ou io2/io2BX, considere os volumes gp3 para workloads que não sejam de produção para ajudar a reduzir custos.
- Acompanhe e defina a tendência da utilização do armazenamento ao longo do tempo para identificar facilmente picos de uso e custos inesperados.
- Use o [AWS Compute Optimizer](#) para obter recomendações sobre como aumentar ou reduzir a escala verticalmente dos volumes do EBS com base na utilização real.
- Use a elasticidade de AWS para ajustar as necessidades de desempenho e capacidade de seus volumes SSD para o Amazon EBS. Ao contrário dos ambientes on-premises, você não precisa superprovisionar a performance e a capacidade de armazenamento para futuras workloads. Você pode migrar suas cargas de trabalho existentes do SQL Server AWS e ajustar o desempenho ou a capacidade conforme necessário, mantendo seus bancos de dados on-line.

Recursos adicionais do

- [Amazon EBS volume types](#) (documentação do Amazon EBS)
- [Amazon Elastic Block Store \(Amazon EBS\)](#) (documentação do Amazon EBS)
- [Provisioned IOPS SSD volumes](#) (documentação do Amazon EBS)
- [Volumes de armazenamento de instâncias SSD](#) (documentação do Amazon EC2)
- [CloudWatch Métricas da Amazon para Amazon EBS](#) (documentação do Amazon EBS)
- [Specifications for Amazon EC2 storage optimized instances](#) (documentação do Amazon EC2)
- [Proteja suas cargas de trabalho do SQL Server usando o NetApp SnapCenter Amazon FSx for NetApp ONTAP](#) (AWS Storage Blog)
- [Perguntas frequentes sobre o Amazon EC2 \(páginaAWS do produto\)](#)

Otimizar o licenciamento do SQL Server usando o Compute Optimizer

Orientação sobre como otimizar licenças para o SQL Server usando o AWS Compute Optimizer

Visão geral do

O [AWS Compute Optimizer](#) pode recomendar oportunidades de otimização de licenciamento para workloads do Microsoft SQL Server no Amazon Elastic Compute Cloud (Amazon EC2). O Compute Optimizer pode fornecer recomendações automatizadas para reduzir os custos de licenciamento. As recomendações do Compute Optimizer estão listadas ao lado de cada instância do EC2 com licenças do Microsoft SQL Server. As informações fornecidas incluem oportunidades de economia recomendadas, preços sob demanda de instâncias do EC2 e preços por hora para trazer sua própria licença (BYOL). Essas informações podem ajudar você a decidir se deve fazer o downgrade da edição da sua licença.

O Compute Optimizer descobre automaticamente suas instâncias do SQL Server no Amazon EC2 por tipo de workload inferido. Para visualizar as recomendações de licenciamento, você pode selecionar as instâncias do SQL Server no Compute Optimizer e, em seguida, autenticar-se com o [CloudWatch Amazon Application Insights](#) usando suas credenciais de banco de dados somente para leitura. O Compute Optimizer analisa se você está usando algum recurso da edição SQL Server Enterprise. Se nenhum recurso da edição Enterprise estiver sendo usado, o Compute Optimizer recomenda que você faça o downgrade para a edição Standard a fim de reduzir os custos de licenciamento.

Você também pode usar o Compute Optimizer para fazer recomendações de dimensionamento para suas instâncias do Amazon EC2 que executam workloads do SQL Server. Para obter mais informações, consulte [Otimizar o dimensionamento do SQL Server usando o Compute Optimizer](#) neste guia.

Recomendações de otimização de custos

As recomendações de licença no Compute Optimizer podem ajudar você a avaliar os recursos que você está usando no Microsoft SQL Server e escolher a edição mais econômica para suas workloads. A edição SQL Server Enterprise é significativamente mais cara do que a edição Standard. Para obter mais informações, consulte [Comparar as edições do SQL Server](#) neste guia e [Preços do SQL Server 2022](#) no site da Microsoft. Investir tempo na configuração do Compute Optimizer para avaliar sua frota do SQL Server e fornecer recomendações pode reduzir de forma significativa seus custos de licenciamento.

A página Detalhes da licença fornece as seguintes informações:

- Use a tabela para comparar suas configurações de licença atuais, como edição, modelo e número de núcleos de instâncias, com as recomendações do Compute Optimizer.

- Use os grafos de utilização para analisar a quantidade de recursos da edição Enterprise que foram usados durante o período de análise.

Para obter mais informações, consulte [Viewing details of a commercial software license recommendation](#) na documentação do Compute Optimizer.

Configurar o Compute Optimizer

O Compute Optimizer analisa as licenças de software comercial usando a métrica `mssql_enterprise_features_used`. Para obter mais informações sobre essa métrica, consulte [Metrics for commercial software licenses](#).

1. Certifique-se de que você tenha as permissões apropriadas para ativar o Compute Optimizer. Para saber mais, consulte:
 - [Política de adesão ao Compute Optimizer](#)
 - [Policies to grant access to Compute Optimizer for standalone Contas da AWS](#)
 - [Políticas para conceder acesso ao Compute Optimizer para uma conta de gerenciamento de uma organização](#)
2. Anexe as funções e a política de instância necessárias para o CloudWatch Application Insights. Para obter mais instruções, consulte [Policies to enable commercial software license recommendations](#).
3. Habilite o CloudWatch Application Insights usando suas credenciais do banco de dados Microsoft SQL Server. Para obter instruções, consulte [Configurar o aplicativo para monitoramento](#) na CloudWatch documentação.

Note


Para gerar recomendações para licenças de software comercial, são necessárias pelo menos 30 horas consecutivas de dados CloudWatch métricos. Para obter mais informações, consulte os [requisitos CloudWatch métricos](#).

4. Use a consulta SQL a seguir para configurar o acesso com privilégios mínimos para o Application Insights. CloudWatch

```
GRANT VIEW SERVER STATE TO [LOGIN];  
GRANT VIEW ANY DEFINITION TO [LOGIN];
```

Isso habilita um novo serviço, o PrometheusSqlExporter SQL.

5. Na conta de gerenciamento de destino Conta da AWS ou da organização, opte pelo Compute Optimizer. Para obter instruções, consulte [Opting in your account](#).

 Note

Depois que você aceita, as descobertas e recomendações de otimização podem levar até 24 horas para serem geradas.

6. No [console do Compute Optimizer](#), escolha Licenças no painel de navegação.
7. Na coluna Descobertas, pesquise todas as instâncias que tenham a descoberta de métricas insuficientes. O Compute Optimizer retorna essa descoberta se detectar CloudWatch que o Application Insights não está ativado ou tem permissões insuficientes. Para obter mais informações, consulte [Finding reasons](#). Faça o seguinte para resolver essas descobertas:
 - a. Escolha a instância.
 - b. Adicione um segredo.
 - c. Confirme se o perfil e a política da instância estão anexados.
 - d. Escolha Habilitar recomendações de licença.
8. Na coluna Descobertas, pesquise todas as instâncias que têm a descoberta Não otimizada. O Compute Optimizer retornará essa descoberta se detectar que a infraestrutura do Amazon EC2 não está usando nenhum dos recursos de licença do Microsoft SQL Server pelos quais você está pagando. Para obter mais informações, consulte [Finding reasons](#). Faça o seguinte para resolver essas descobertas:
 - a. Escolha a instância.
 - b. Comparar a edição da licença atual com a edição recomendada.
 - c. Analise o grafo atual de utilização da licença.
 - d. Se você quiser fazer o downgrade da licença, escolha Implementar recomendação.
 - e. Analise os requisitos e siga as instruções para fazer o downgrade da licença. Se você quiser automatizar o processo, consulte Fazer o [downgrade da edição Enterprise do SQL Server usando o AWS Systems Manager Documento para reduzir custos](#) (AWS Blog).

Recursos adicionais do

- [Reduza os custos de licenciamento do Microsoft SQL Server com AWS Compute Optimizer](#) (AWS Blog)
- [O que AWS Compute Optimizer é](#) (AWS documentação)
- [Viewing commercial software license recommendations](#) (documentação da AWS)
- [Downgrade your Microsoft SQL Server edition](#) (documentação da AWS)
- [Microsoft SQL Server na AWS](#) (AWS)
- [Microsoft Licensing on AWS](#) (AWS)
- [Microsoft SQL Server 2019 Pricing](#) (Microsoft)
- [Microsoft SQL Server 2022 Pricing](#) (Microsoft)

Otimizar o dimensionamento do SQL Server usando o Compute Optimizer

Visão geral do

O [AWS Compute Optimizer](#) ajuda os administradores do banco de dados (DBAs) a descobrir as workloads do Microsoft SQL Server no Amazon Elastic Compute Cloud (Amazon EC2) e a dimensionar corretamente as instâncias do EC2 para reduzir os custos de licença em até 25%. O recurso de [tipo de carga de trabalho inferido](#) no Compute Optimizer usa aprendizado de máquina (ML) e detecta automaticamente os aplicativos que podem estar sendo executados em seus recursos. AWS O Compute Optimizer inclui suporte para o SQL Server como um tipo de workload inferida. Ao usar o recurso de tipo de workload inferida, você pode identificar oportunidades de redução de custos com base na workload específica em execução em suas instâncias do Amazon EC2.

Com esse recurso, você pode categorizar oportunidades de economia de custos de acordo com os tipos de workloads inferidas compatíveis, como o SQL Server. O Compute Optimizer pode descobrir automaticamente as instâncias EC2 do SQL Server que estão superprovisionadas. Você pode mudar para o console do EC2 para reduzir a escala da instância, o que ajuda a reduzir os custos de licenciamento e infraestrutura.

Você também pode usar o Compute Optimizer para fazer recomendações de licenciamento do SQL Server. Para obter mais informações, consulte [Otimizar o licenciamento do SQL Server usando o Compute Optimizer](#) neste guia.

Configurar o Compute Optimizer

Para obter instruções sobre como usar o Compute Optimizer com cargas de trabalho inferidas do SQL Server, [consulte Otimizando o desempenho e reduzindo os custos de licenciamento: aproveitando AWS Compute Optimizer as instâncias do SQL Server do Amazon EC2](#) (Blog).AWS Você pode optar por contas autônomas, contas que são membros de uma organização e contas gerenciais de uma organização. Para contas autônomas e de membros, a ativação do serviço habilita o Compute Optimizer somente para essa conta. Para uma conta gerencial da organização, você pode escolher se deseja habilitar o Compute Optimizer somente nessa conta ou em todas as contas de membros da organização.

O processo de aceitação do Compute Optimizer cria automaticamente AWS Identity and Access Management uma função vinculada ao serviço (IAM). Para obter mais informações, consulte [Usando funções vinculadas a serviço para AWS Compute Optimizer](#).

O Compute Optimizer analisa seus recursos com base nas métricas da CloudWatch Amazon, como CPU, E/S, rede e uso do Amazon Elastic Block Store (Amazon EBS). Para gerar recomendações, são necessárias pelo menos 30 horas consecutivas de dados CloudWatch métricos nos últimos 14 dias. Se você habilitar o recurso aprimorado de métricas de infraestrutura, ele estenderá as métricas de utilização para 93 dias. Para obter mais informações, consulte [os requisitos CloudWatch métricos e as métricas de infraestrutura aprimorada](#) na documentação do Compute Optimizer.

O Compute Optimizer fornece opções e as economias associadas a cada opção, com base na vCPU, memória, armazenamento, rede, risco e esforço de migração. Você pode usar o painel de CloudWatch métricas para analisar os dados que estão sendo usados para fazer a recomendação. Com esses dados, você pode dimensionar corretamente suas instâncias do EC2 que estão executando workloads do SQL Server. Para obter mais informações sobre como alterar o tipo de instância, consulte [Alterações no tipo de instância](#) na documentação do Amazon EC2.

Recursos adicionais do

- [AWS Compute Optimizer identifica e filtra cargas de trabalho do Microsoft SQL Server](#) (AWS)
- [Otimizando o desempenho e reduzindo os custos de licenciamento: aproveitando AWS Compute Optimizer as instâncias do SQL Server do Amazon EC2](#) (Blog)AWS
- [O que AWS Compute Optimizer é](#) (AWS documentação)

- [Visualizando recomendações de instâncias do EC2](#) (AWS documentação)

Analise Trusted Advisor as recomendações para cargas de trabalho do SQL Server

Visão geral do

O [AWS Trusted Advisor](#) fornece recomendações que ajudam você a seguir as práticas recomendadas da AWS. Ao analisar seu uso, configuração e gastos, Trusted Advisor fornece recomendações práticas para reduzir seus custos, melhorar a disponibilidade e o desempenho do sistema ou ajudar a fechar lacunas de segurança. Esta seção se concentra nas Trusted Advisor verificações que podem ajudá-lo a reduzir os custos de operação das cargas de trabalho do SQL Server no Nuvem AWS.

Recomendações de otimização de custos

Trusted Advisor fornece recomendações que ajudam você a otimizar suas cargas de trabalho do SQL Server no Amazon Elastic Compute Cloud (Amazon EC2). As verificações inspecionam suas workloads do SQL Server e listam automaticamente as instâncias que precisam de otimização. A operacionalização Trusted Advisor das recomendações pode reduzir custos e melhorar a postura de segurança da sua organização.

A seguir estão as Trusted Advisor verificações que se concentram no Microsoft SQL Server:

- [Instâncias do Amazon EC2 superprovisionadas para o Microsoft SQL Server](#): esta verificação analisa suas instâncias do Amazon EC2 que estão executando o SQL Server e gera um alerta se uma instância exceder o limite de vCPU do software do SQL Server. Por exemplo, uma instância com a edição SQL Server Standard pode usar até 48 CPUs v. Uma instância com o SQL Server Web pode usar até 32 CPUs v.

Edição	vCPU mín.	vCPU máx.
Web	4	32
Standard	4	48
Enterprise	4	Limites do SO

- [Consolidação de instâncias do Amazon EC2 para o Microsoft SQL Server](#): esta verificação analisa suas instâncias do Amazon EC2 e gera um alerta se sua instância tiver menos do que o número mínimo de licenças do SQL Server. Você pode consolidar instâncias menores do SQL Server para ajudar a reduzir custos. Se você tiver muitas instâncias pequenas de SQL Server com licença inclusa, então considere consolidar. De acordo com o [guia de licenciamento do Microsoft SQL Server 2019](#), o SQL Server exige um mínimo de quatro licenças de vCPU por instância. Se você consolidar esses bancos de dados, poderá economizar nos custos de licenciamento. Você pode tomar sua decisão com base no número de bancos de dados na instância, no tamanho máximo do banco de dados e no tamanho total dos bancos de dados. A consolidação é compatível com as edições Web, Standard e Enterprise do SQL Server. Para obter mais informações, consulte [Consolidating SQL Server Databases](#) (publicação do blog da Microsoft).

AWS não recomenda colocar grandes bancos de dados de produção em apenas um servidor. No entanto, você pode consolidar os menores usados para ambientes que não sejam de produção, como desenvolvimento, teste e preparação. Isso depende do uso atual do SQL Server. Se você tiver bancos de dados de baixo uso, poderá consolidar em um servidor.

Configure Trusted Advisor

Faça o seguinte para avaliar as verificações focadas no SQL Server Trusted Advisor.

1. Faça login no Console de gerenciamento da AWS.
2. Abra o [console do AWS Trusted Advisor](#).
3. No painel de navegação, em Recomendações, escolha Otimização de custo.
4. Na lista de verificações de otimização de custos, analise o status da consolidação de instâncias do Amazon EC2 para o Microsoft SQL Server e as instâncias do Amazon EC2 superprovisionadas para verificações do Microsoft SQL Server.
 - Símbolos de verificação verdes indicam que suas instâncias do Amazon EC2 estão configuradas de forma ideal.
 - Os símbolos de alerta laranja indicam que há oportunidades de melhoria.
5. Escolha um símbolo de verificação para ver seus detalhes e recomendações.
6. Siga as instruções fornecidas pela verificação para otimizar suas instâncias do Amazon EC2 que estão executando workloads do SQL Server.
7. Monitore suas instâncias regularmente e atualize as verificações periodicamente.

Recursos adicionais do

- [Trusted Advisor referência de verificação](#) (AWS documentação)
- [Microsoft SQL Server na AWS](#) (AWS)
- [Microsoft Licensing on AWS](#) (AWS)
- [SQL Server 2019 pricing](#) (Microsoft)
- [AWS Launch Wizard para SQL Server](#) (AWS documentação)

Contêineres

A modernização é uma jornada transformacional que oferece muitas opções, incluindo a decomposição de monólitos em microsserviços, a rearquitetura de aplicações para serem orientadas por eventos usando funções sem servidor (AWS Lambda) e o reaproveitamento de bancos de dados do SQL Server para o Amazon Aurora ou bancos de dados gerenciados criados especificamente. Os caminhos de modernização para redefinir a plataforma de aplicações .NET Framework para contêineres Linux e Windows exigem menos esforço do que outras opções de modernização. Os contêineres oferecem os seguintes benefícios:

- **Acelerar a inovação:** migrar para contêineres facilita a automação de etapas do ciclo de vida de desenvolvimento, que incluem a criação, os testes e a implantação de aplicações. Ao automatizar esses processos, as equipes de desenvolvimento e operações têm mais tempo para se concentrar na inovação.
- **Reduzir o custo total de propriedade (TCO):** migrar para contêineres também pode reduzir sua dependência das ferramentas de gerenciamento de licenças e proteção de endpoints. Como os contêineres são unidades de computação efêmeras, você pode automatizar e simplificar as tarefas de gerenciamento, como aplicação de patches, escalabilidade, backup e restauração. Isso reduz o TCO da administração e operação de workloads baseadas em contêineres. Por fim, os contêineres são mais eficientes em comparação com as máquinas virtuais porque você pode usar contêineres para maximizar o posicionamento de suas aplicações, fornecendo melhor isolamento. Isso aumenta a utilização dos recursos de infraestrutura da aplicação.
- **Melhorar a utilização dos recursos:** os contêineres são mais eficientes em comparação com as máquinas virtuais porque você pode usá-los para maximizar o posicionamento das aplicações. Isso aumenta a utilização dos recursos de infraestrutura da aplicação, fornecendo melhor isolamento.
- **Elimine a lacuna de habilidades** — AWS oferece dias de imersão para aprimorar suas equipes de desenvolvimento em tecnologia e DevOps práticas de contêineres.

Esta seção abrange os seguintes tópicos:

- [Migrar aplicações do Windows para contêineres](#)
- [Otimize os custos AWS Fargate das tarefas no Amazon ECS](#)
- [Obter visibilidade dos seus custos do Amazon EKS](#)
- [Redefinir a plataforma das aplicações do Windows com o App2Container](#)

Para obter informações sobre licenciamento, consulte a seção Licensing em [Amazon Web Services and Microsoft: Frequently Asked Questions](#), ou envie suas perguntas por e-mail para microsoft@amazon.com.

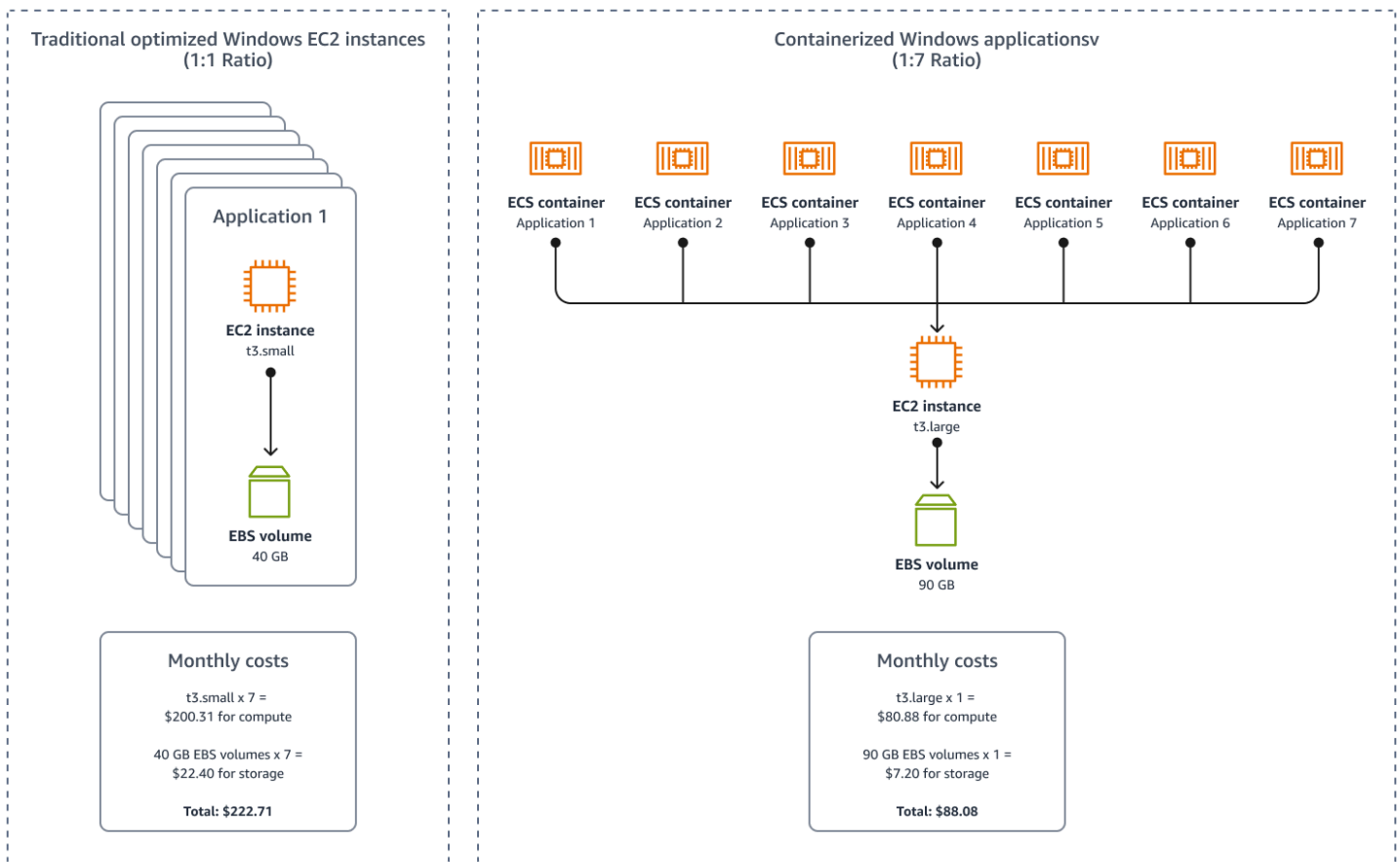
Migrar aplicações do Windows para contêineres

Visão geral do

De acordo com a [Pesquisa anual da CNCF de 2021](#), 96% das organizações estão usando ou avaliando contêineres para modernizar sua infraestrutura. Isso ocorre porque os contêineres podem ajudar sua organização a reduzir riscos, aumentar a eficiência e a velocidade operacionais e possibilitar a agilidade. Você também pode usar contêineres para reduzir o custo de execução das suas aplicações. Esta seção oferece recomendações para a execução econômica de contêineres em vários serviços de AWS contêineres, incluindo [Amazon Elastic Container Service \(Amazon ECS\)](#), [Amazon Elastic Kubernetes Service \(Amazon EKS\)](#) e [AWS Fargate](#)

Benefícios de custo

[O infográfico a seguir mostra a economia de custos que uma empresa pode obter consolidando seus aplicativos ASP.NET Framework em instâncias do Amazon Elastic Compute Cloud \(Amazon EC2\) com base AWS em uma recomendação de Avaliação de Otimização e Licenciamento \(OLA\).AWS](#) O infográfico a seguir mostra quais economias adicionais podem ser obtidas migrando aplicações para um contêiner Windows.



O AWS OLA recomendou que a empresa fizesse uma mudança para instâncias t3.small individuais. A empresa poderá obter essas economias executando sete aplicações ASP.NET em servidores on-premises, conforme mostra a análise de performance de utilização a seguir.

Server name	Storage	Operating system	On-premises CPU AVG utilization	On-premises CPU peak utilization	On-premises RAM (GB)	On-premises RAM AVG utilization (GB)	On-premises RAM peak utilization (GB)	Instance size	vCPU	RAM (GB)
1 AppServer01	60	Windows Server 2012	7.00%	17.00%	8	13.50%	17.10%	t3.small	2	2
2 AppServer02	39	Windows Server 2012	20.07%	22.00%	16	7.50%	12.40%	t3.small	2	2
3 AppServer03	39	Windows Server 2012	24.00%	25.50%	16	8.80%	11.90%	t3.small	2	2
4 AppServer04	4	Windows Server 2012	21.40%	24.00%	16	7.80%	10.70%	t3.small	2	2
5 AppServer05	40	Windows Server 2012	21.30%	23.00%	16	8.20%	12.00%	t3.small	2	2
6 AppServer06	39	Windows Server 2012	21.50%	23.50%	16	7.90%	10.90%	t3.small	2	2
7 AppServer07	39	Windows Server 2012	21.60%	22.90%	16	8.40%	11.50%	t3.small	2	2

Annotations above the table:

- Not sharable (points to Storage)
- Nearing end of life (points to Operating system)
- Approximately 15% overhead (points to On-premises CPU AVG utilization)
- The OS accounts for 1.2 GB per instance (points to On-premises RAM (GB))
- Optimized for EC2 (points to Instance size, vCPU, and RAM (GB))

Uma análise mais aprofundada revelou que a empresa pode economizar ainda mais em custos executando suas workloads em contêineres. Os contêineres reduzem a sobrecarga do sistema operacional sobre os recursos do sistema, como CPU, RAM e uso de disco (explicado na próxima seção). Nesse cenário, a empresa pode consolidar todas as sete aplicações em uma instância t3.large e ainda ter 3 GB de RAM disponíveis. A migração para contêineres pode ajudar a empresa a

obter uma economia média de 64% em computação e armazenamento usando contêineres em vez do Amazon EC2.

Recomendações de otimização de custos

A seção a seguir oferece recomendações para otimizar custos consolidando aplicações e usando contêineres.

Reduzir a pegada do Windows no Amazon EC2

Os contêineres Windows podem reduzir sua pegada do Windows no Amazon EC2, permitindo que você consolide mais aplicações em menos instâncias do EC2. Por exemplo, vamos supor que você tenha 500 aplicações ASP.NET. Se você estiver executando um núcleo por aplicação para Windows no Amazon EC2, isso equivale a 500 instâncias do Windows (t3.small). Se você presumir uma proporção de 1:7 (que pode aumentar significativamente, dependendo do tipo/tamanho da instância do EC2) para usar contêineres Windows (com t3.large), precisará apenas de aproximadamente 71 instâncias do Windows. Isso representa uma redução de 85,8% em sua pegada do Windows no Amazon EC2.

Reduzir os custos de licenciamento do Windows

Se você licenciar uma instância do Windows, não precisará licenciar contêineres em execução nessa instância. Como resultado, consolidar suas aplicações ASP.NET usando contêineres Windows pode reduzir significativamente seus custos de licenciamento do Windows.

Reduzir sua pegada de armazenamento

Toda vez que você inicializa uma nova instância do EC2, você cria e paga por um novo volume do Amazon Elastic Block Store (Amazon EBS) para hospedar o sistema operacional. À medida que isso escala, o custo aumenta com ele. Se você usar contêineres, poderá reduzir os custos de armazenamento porque todos os contêineres compartilham o mesmo sistema operacional básico. Além disso, os contêineres usam o conceito de camadas para reutilizar partes imutáveis de uma imagem de contêiner para todos os contêineres em execução baseados nessa imagem. No cenário de exemplo anterior, todos os contêineres estão executando o .NET Framework e, portanto, todos compartilham a camada intermediária e imutável do framework ASP.NET.

Migre end-of-support servidores para contêineres

O suporte para o Windows Server 2012 e o Windows Server 2012 R2 terminou em 10 de outubro de 2023. Você pode migrar suas aplicações em execução no Windows Server 2012 ou em versões

anteriores, colocando-as em contêineres para serem executadas em novos sistemas operacionais. Dessa forma, você evita executar suas aplicações em sistemas operacionais não compatíveis e, ao mesmo tempo, aproveita as eficiências de custo, a redução de riscos, a eficiência operacional, a velocidade e a agilidade que os contêineres oferecem.

Uma ressalva a ser considerada com essa abordagem é se seu aplicativo requer informações específicas APIs relacionadas à versão do sistema operacional atualmente em uso (COM Interop, por exemplo). Nesse caso, é necessário testar a migração da aplicação para uma versão mais recente do Windows. Os contêineres Windows alinham sua imagem de contêiner base (por exemplo, Windows Server 2019) com o sistema operacional do host do contêiner (por exemplo, Windows Server 2019). Testar e migrar para contêineres pode facilitar as atualizações do sistema operacional no futuro, alterando a imagem base em seu Dockerfile e implantando em um novo conjunto de hosts executando a versão mais recente do Windows.

Remover licenças e ferramentas de gerenciamento de terceiros

O gerenciamento de sua frota de servidores requer o uso de várias ferramentas de operação de sistema de terceiros para gerenciamento de patches e configuração. Isso pode tornar o gerenciamento da infraestrutura complexo e, muitas vezes, você incorre em custos de licenciamento de terceiros. Se você usa contêineres AWS, não precisa gerenciar nada no lado do sistema operacional. O runtime do contêiner gerencia os contêineres. Isso significa que o host subjacente é efêmero e pode ser facilmente substituído. Você pode executar seus contêineres sem a necessidade de gerenciar diretamente o host do contêiner. Além disso, você pode usar ferramentas gratuitas, como AWS Systems Manager Session Manager acessar facilmente o host e solucionar problemas.

Melhorar o controle e a portabilidade

Os contêineres oferecem um controle mais granular sobre os recursos do servidor, como CPU e RAM, do que sobre as instâncias do EC2. Para instâncias do EC2, você pode controlar a CPU e a RAM selecionando uma família de instâncias, um tipo de instância e as [opções de CPU](#). No entanto, com contêineres, você pode definir exatamente quanta CPU ou RAM deseja alocar para um contêiner nas definições de tarefas do ECS ou para [pods no Amazon EKS](#). Na realidade, e recomendável [especificar a memória e CPU em nível de contêiner](#) para contêineres Windows. Esse nível de granularidade traz benefícios de custo. Considere o seguinte exemplo de código:

```
json
{
  "taskDefinitionArn": "arn:aws:ecs:us-east-1:123456789012:task-definition/demo-
service:1",
```

```
"containerDefinitions": [  
  {  
    "name": "demo-service",  
    "image": "mcr.microsoft.com/dotnet/framework/samples:aspnetapp-  
windowsservercore-ltsc2019",  
    "cpu": 512,  
    "memory": 512,  
    "links": [],  
    "portMappings": [  
      {  
        "containerPort": 80,  
        "hostPort": 0,  
        "protocol": "tcp"  
      }  
    ],  
  },  
],
```

Acelerar a inovação

A migração para contêineres facilita a automação das etapas do ciclo de vida de desenvolvimento, que incluem a criação, os testes e a implantação de aplicações. Se você automatizar esses processos, proporcionará às suas equipes de desenvolvimento e operações mais tempo para se concentrar na inovação.

Reduzir o TCO

A migração para contêineres geralmente reduz a dependência das ferramentas de gerenciamento de licenças e de proteção de endpoints. Como os contêineres são unidades de computação efêmeras, você pode automatizar e simplificar as tarefas de gerenciamento, como aplicação de patches, escalabilidade, backup e restauração. Isso pode reduzir o TCO da administração e operação de workloads baseadas em contêineres. Os contêineres são mais eficientes em comparação com as máquinas virtuais porque permitem que você maximize o posicionamento de suas aplicações para que você possa aumentar a utilização dos recursos de infraestrutura da sua aplicação.

Eliminar a lacuna de habilidades

AWS oferece programas e dias de imersão para aprimorar as equipes de desenvolvimento de clientes em contêineres e DevOps tecnologia. Isso inclui consultoria e capacitação práticas.

Refatorar para o .NET 5+ e usar contêineres Linux

Embora você possa reduzir custos migrando suas aplicações .NET Framework para contêineres, é possível obter ainda mais economia de custos ao refatorar aplicações .NET legadas para alternativas nativas da nuvem na AWS.

Remover custos de licenças

Refatorar sua aplicação .NET Framework no Windows para .NET Core no Linux resulta em uma redução de custos de aproximadamente 45%.

Acessar os aprimoramentos mais recentes

Refatorar suas aplicações .NET Framework no Windows para .NET Core no Linux lhe dá acesso aos aprimoramentos mais recentes, como o Graviton2. O Graviton2 oferece um preço de performance 40% melhor do que instâncias comparáveis.

Melhorar a segurança e a performance

Refatorar sua aplicação .NET Framework no Windows para .NET Core em contêineres Linux traz melhorias na segurança e na performance. Isso ocorre porque você obtém os patches de segurança mais recentes, beneficia-se do isolamento de contêineres e tem acesso a novos recursos.

Usar contêineres Windows em vez de executar várias aplicações em uma instância do IIS

Considere as seguintes vantagens de usar contêineres Windows em vez de executar várias aplicações em uma instância EC2 Windows com o Internet Information Services (IIS):

- **Segurança:** os contêineres fornecem um nível de segurança pronto para uso que não é alcançado por meio do isolamento no nível do IIS. Se um site ou aplicação do IIS for comprometido, todos os outros sites hospedados ficarão expostos e vulneráveis. O escape de contêiner é raro e representa uma vulnerabilidade mais difícil de ser explorada do que assumir o controle de um servidor por meio de uma vulnerabilidade na web.
- **Flexibilidade:** a capacidade de executar contêineres isoladamente no processo e ter sua própria instância permite opções de rede mais granulares. Os contêineres também oferecem métodos complexos de distribuição em muitas instâncias do EC2. Você não obtém esses benefícios ao consolidar aplicações em uma única instância do IIS.

- Sobrecarga de gerenciamento: a Indicação do nome do servidor (SNI) cria uma sobrecarga que requer gerenciamento e automação. Além disso, você precisa lidar com operações típicas de gerenciamento do sistema operacional, como aplicação de patches, solução de problemas de BSOD (se o ajuste de escala automático não estiver em vigor), proteção de endpoints e assim por diante. Configurar sites do IIS de acordo com as [práticas recomendadas de segurança](#) é uma atividade demorada e contínua. Talvez você até precise configurar [níveis de confiança](#), o que também aumenta a sobrecarga de gerenciamento. Os contêineres são projetados para serem sem estado e imutáveis. Em última análise, suas implantações serão mais rápidas, mais seguras e reproduzíveis se você usar contêineres Windows.

Próximas etapas

Investir em uma infraestrutura moderna para executar suas cargas de trabalho legadas traz imensos benefícios para sua organização. AWS os serviços de contêiner facilitam o gerenciamento de sua infraestrutura subjacente, seja no local ou na nuvem, para que você possa se concentrar na inovação e nas necessidades de sua empresa. Quase 80% de todos os contêineres na nuvem funcionam AWS atualmente. AWS fornece um rico conjunto de serviços de contêiner para praticamente todos os casos de uso. Para começar, consulte [Contêineres na AWS](#).

Recursos adicionais do

- [Otimize o custo de cargas de trabalho de contêineres com provedores de capacidade ECS e instâncias spot do EC2 \(blog\)AWS](#)
- [Cost Optimization Checklist for Amazon ECS and AWS Fargate](#) (Blog da AWS)
- [Amazon EKS no AWS Graviton2 disponível ao público em geral: considerações sobre aplicativos de várias arquiteturas](#) (Blog)AWS
- [Otimização de custos para Kubernetes em AWS](#)(Blog)AWS
- [Otimizando seus custos de computação do Kubernetes com a consolidação do Karpenter](#) (Blog)AWS

Otimize os custos AWS Fargate das tarefas no Amazon ECS

Visão geral do

AWS Fargate As tarefas de dimensionamento correto são uma etapa importante para a otimização de custos. Muitas vezes, as aplicações são criadas com tamanhos arbitrários para tarefas do Fargate

e nunca são revisitadas. Isso pode causar provisionamento excessivo das tarefas do Fargate e gastos desnecessários. Esta seção mostra como usar o [AWS Compute Optimizer](#) para fornecer recomendações práticas para que você possa otimizar a CPU e a memória de tarefas para seus serviços do Amazon Elastic Container Service (Amazon ECS) executados no Fargate. O Compute Optimizer também quantifica o impacto nos custos da adoção dessas recomendações. Isso permite que você priorize seus esforços de otimização com base no tamanho da oportunidade de economia. As recomendações do Compute Optimizer fornecem configurações de CPU e memória em nível de contêiner para tarefas de redução de escala.

Benefícios de custo

O dimensionamento correto das tarefas do Amazon ECS no Fargate pode reduzir os custos em 30 a 70% para tarefas de longa execução. Sem revisar as métricas de performance da aplicação para dimensionar corretamente o tamanho da tarefa, você pode aplicar a mesma lógica usada nas instâncias de computação do EC2 ao dimensionamento de contêineres. Isso resulta em tarefas superdimensionadas do Fargate que aumentam os custos de recursos ociosos. Você pode usar o Compute Optimizer para identificar as oportunidades de dimensionamento correto de forma reativa. O ideal é que o proprietário da aplicação revise as métricas específicas de performance da aplicação e remova a sobrecarga do sistema operacional para garantir que o tamanho adequado da tarefa seja especificado. Para obter mais informações, consulte a seção [Migrar aplicações do Windows para contêineres](#) deste guia.

Recomendações de otimização de custos

Esta seção oferece recomendações para usar o Compute Optimizer para dimensionar corretamente seu Amazon ECS nas tarefas do Fargate.

Como parte do processo de otimização de custos, é recomendável fazer o seguinte:

- Habilitar o Compute Optimizer
- Utilizar os resultados do Compute Optimizer
- Marcar as tarefas para serem dimensionadas corretamente
- Ative a etiqueta de alocação de custos para funcionar com ferramentas de AWS cobrança
- Implementar recomendações de dimensionamento correto
- Revisar os custos antes e depois no Cost Explorer

Habilitar o Compute Optimizer

Você pode habilitar o [AWS Compute Optimizer](#) no nível da organização ou da conta única no AWS Organizations. A configuração de toda a organização fornece relatórios contínuos para instâncias novas e existentes em toda a sua frota para todas as contas de membros. Isso permite que o dimensionamento correto seja uma atividade recorrente em vez de uma point-in-time atividade.

Nível da organização

Para a maioria das organizações, a forma mais eficiente de usar o Compute Optimizer é no nível organizacional. Isso fornece visibilidade de várias contas e várias regiões em sua organização e centraliza os dados em uma fonte para análise. Para habilitar essa funcionalidade no nível da organização, faça o seguinte:

1. Entre na sua [conta gerencial do AWS Organizations](#) com um perfil que tenha as [permissões necessárias](#) e opte pela inclusão de todas as contas dessa organização. A organização deve ter [todos os recursos habilitados](#).
2. Depois de habilitar a conta gerencial, você pode entrar na conta, ver todas as outras contas de membros e consultar suas recomendações.

Note

É uma prática recomendada configurar uma [conta de administrador delegado](#) para o Compute Optimizer. Isso permite que você exerça o princípio privilégio mínimo, minimizando o acesso à conta gerencial do AWS Organizations e, ao mesmo tempo, fornecendo acesso ao serviço de toda a organização.

Nível de conta única

Se você tem como alvo uma conta com altos custos, mas não tem acesso a ao AWS Organizations, você ainda pode habilitar o Compute Optimizer para essa conta e região. Para saber mais sobre o processo de aceitação, consulte [Introdução ao AWS Compute Optimizer](#).

Note

As recomendações são atualizadas diariamente e podem levar até 12 horas para serem geradas. Lembre-se de que o Compute Optimizer precisa de 24 horas de métricas nos últimos 14 dias para gerar recomendações para o Amazon ECS no Fargate. Para obter mais

informações, consulte [Requirements for Amazon ECS services on Fargate](#) na documentação do Compute Optimizer.

O Compute Optimizer analisa automaticamente as seguintes métricas de utilização do Amazon e do CloudWatch Amazon ECS para seus serviços do Amazon ECS no Fargate:

- `CPUUtilization`: a porcentagem da capacidade da CPU usada no serviço.
- `MemoryUtilization`: a porcentagem de memória usada no serviço.

Utilizar os resultados do Compute Optimizer

Considere um exemplo que se concentra em fazer as alterações de dimensionamento correto em uma única conta e região. Neste exemplo, o Compute Optimizer está habilitado no nível da organização em todas as contas. Lembre-se de que o dimensionamento correto é um processo disruptivo que, na maioria dos casos, é executado com precisão pelos proprietários da aplicação durante uma janela de manutenção programada de várias semanas.

Se você navegar até o Compute Optimizer de dentro da conta gerencial de uma organização (conforme mostrado nas etapas a seguir), poderá escolher a conta que deseja investigar. Neste exemplo, uma tarefa está sendo executada em uma única conta superprovisionada em `us-east-1`. O foco é redimensionar para o tamanho recomendado para o serviço do Amazon ECS.

1. Abra o console do [Compute Optimizer](#).
2. Na página Painel, filtre por `Descobertas=superprovisionado` para ver todos os serviços do Amazon ECS no Fargate.
3. Para revisar as recomendações detalhadas para serviços superprovisionados do ECS no Fargate, role para baixo e escolha `Visualizar recomendações`.
4. Escolha `Exportar` e salve o arquivo para uso futuro.

Note

Para salvar recomendações para análise futura, você deve ter um bucket do S3 disponível para que o Compute Optimizer grave em cada região. Para obter mais informações, consulte [Amazon S3 bucket policy for AWS Compute Optimizer](#) na documentação do Compute Optimizer.

Para conferir as recomendações do Compute Optimizer, faça o seguinte:

1. No console do [Compute Optimizer](#), acesse a página de recomendações de exportação.
2. Em Destino do bucket do S3, escolha seu bucket do S3.
3. Na seção Filtros de exportação, em Tipo de recurso, escolha Serviços do ECS no Fargate.
4. Na página Recomendações para serviços do ECS no Fargate, analise um dos serviços do ECS no Fargate e confira as recomendações de CPU e memória do Compute Optimizer. Por exemplo, revise as recomendações nas seções Comparar as configurações atuais com o tamanho recomendado da tarefa e Comparar as configurações atuais com o tamanho recomendado do contêiner.

Para obter a lista de serviços do ECS para Fargate que você precisa do tamanho certo, faça o seguinte:

1. Abra o [console Amazon S3](#).
2. No painel de navegação, escolha Buckets e selecione o bucket em que você exportou os resultados.
3. Na guia Objetos, selecione seu objeto e escolha Baixar.
4. Nos resultados baixados, filtre a coluna de descobertas para mostrar somente serviços OVER_PROVISIONED do Amazon ECS no Fargate. Isso mostra os serviços do Amazon ECS que você planeja direcionar para o tamanho certo.
5. Armazene as definições de tarefas em um editor de texto para uso posterior.

Dimensionamento correto das tarefas de marcação

Marcar suas workloads é uma ferramenta poderosa para organizar seus recursos na AWS. Você pode usar tags para obter uma visibilidade detalhada dos custos e possibilitar o estorno. Há muitos métodos e estratégias para adicionar tags aos recursos da AWS para lidar com o estorno e a automação. Para obter mais informações, consulte o AWS whitepaper [Best Practices for AWS tagging](#) Resources. O exemplo a seguir usa o [AWS CloudShell](#) para marcar todas as tarefas que fazem parte de qualquer serviço do Amazon ECS na conta e Região da AWS de destino.

```
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="enabled"
```

```
# Get a list of ECS Clusters
ClustersArns=$( aws ecs list-clusters -query 'clusterArns' -output text)
for ClustersArn in $ClustersArns; do
  ServiceArns=$( aws ecs list-services -cluster $ClustersArn -query 'serviceArns' -
output text)
  for ServiceArn in $ServiceArns; do
    TasksArns=$( aws ecs list-tasks -cluster $ClustersArn -service-name $ServiceArn -
query 'taskArns' -output text)
    for TasksArn in $TasksArns; do
      aws ecs tag-resource -resource-arn $TasksArn -tags key=$TAG_KEY,value=$TAG_VALUE
    done
  done
done
```

O exemplo de código a seguir mostra como habilitar a [propagação de tags](#) para todos os serviços do Amazon ECS.

```
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="enabled"
# Get a list of ECS Clusters
ClustersArns=$(aws ecs list-clusters --query 'clusterArns' --output text)
for ClustersArn in $ClustersArns; do
  ServiceArns=$(aws ecs list-services --cluster $ClustersArn --query 'serviceArns' --
output text)
  for ServiceArn in $ServiceArns; do
    aws ecs update-service --cluster $ClustersArn --service $ServiceArn --propagate-tags
SERVICE &>/dev/null
    aws ecs tag-resource --resource-arn $ServiceArn --tags key=$TAG_KEY,value=$TAG_VALUE
  done
done
```

Ative a etiqueta de alocação de custos para funcionar com ferramentas de AWS cobrança

Recomendamos ativar a tag de alocação de custos definida pelo usuário. Isso permite que a tag Rightsizing seja reconhecida e filtrável nas ferramentas de AWS cobrança (por exemplo, e). AWS Cost Explorer AWS Cost and Usage Report Se você não habilitar essa funcionalidade, a opção de filtragem de tags e os dados não estarão disponíveis. Para obter informações sobre o uso de tags

de alocação de custos, consulte [Activating user-defined cost allocation tags](#) na documentação do Gerenciamento de Faturamento e Custos da AWS .

Depois de esperar 24 horas, você pode ver a tag no Cost Explorer antes de implementar as recomendações de dimensionamento correto na próxima seção. Para fazer isso, pesquise a tag Rightsizing no Cost Explorer.

Implementar recomendações de dimensionamento correto

O Compute Optimizer fornecerá recomendações de tamanho de tarefas ou contêineres. Para implementar recomendações de dimensionamento correto, faça o seguinte:

1. Abra o [console do Amazon ECS](#).
2. Na barra de navegação, selecione a região que contém a definição de tarefa.
3. No painel de navegação, escolha Task definitions (Definições de tarefa).
4. Na página Task definitions (Definições de tarefa), escolha a tarefa e, em seguida, escolha Create new revision (Criar nova revisão).
5. Na página Create new task definition revision (Criar nova revisão da definição de tarefa), faça as alterações. Para atualizar a recomendação de tamanho do contêiner, atualize a cpu e memory no bloco containerDefinitions em sua [definição de tarefa do ECS](#). Por exemplo:

```
"containerDefinitions": [  
  {  
    "name": "your-container-name",  
    "image": "your-image",  
    "cpu": 1024,  
    "memory": 2048,  
  }  
],
```

6. Verifique as informações e escolha Create (Criar).

Para atualizar o serviço do Amazon ECS, faça o seguinte:

1. Abra o [console do Amazon ECS](#).
2. Na página Clusters, selecione o cluster.
3. Na página Cluster overview (Visão geral do cluster), selecione o serviço e escolha Update (Atualizar).

4. Em Task definition (Definição de tarefa), escolha a família de definição de tarefa e a revisão que serão usadas.

Para operadores avançados, você pode usar CloudShell para atualizar o serviço Amazon ECS. Por exemplo:

```
bash
#!/bin/bash
# Set variables
ClustersName="workshop-cluster"
ServiceName="lab7-fargate-service"
TaskDefinition="lab7-fargate-demo:3"
# update the service
aws ecs update-service --cluster $ClustersName --service $ServiceName --task-definition
$TaskDefinition
```

Analisar os custos antes e depois

Depois de dimensionar corretamente seus recursos, você pode usar o Cost Explorer para mostrar os custos antes e depois usando a tag Rightsizing. Lembre-se de que você pode usar [tags de recursos](#) para controlar os custos. Ao usar várias camadas de tags, você pode obter visibilidade granular de seus custos. No exemplo abordado neste guia, a tag Rightsizing é usada para aplicar uma tag genérica a todas as instâncias de destino. Em seguida, uma tag de team é usada para organizar ainda mais os recursos. A próxima etapa é introduzir tags de aplicação para mostrar ainda mais o impacto nos custos de operar uma aplicação específica.

Considere um exemplo da redução de custos que pode ser alcançada usando a tag Rightsizing para um único nível de conta. Neste exemplo, os custos operacionais vão de USD 30,26 por dia para USD 7,56 por dia. Supondo 744 horas por mês, o custo anual antes do dimensionamento correto é de USD 11.044,90. Após o dimensionamento correto, o custo anual cai para USD 2.759,40. Isso se traduz em uma redução de 75% nos custos de computação dessa conta. Imagine o impacto dessa redução em uma grande organização.

Antes de embarcar em sua jornada do dimensionamento correto, considere o seguinte:

- AWS oferece muitas opções para redução de custos. Isso inclui o [AWS OLA](#), onde AWS analisa suas instâncias locais antes de migrar para o. AWS O AWS OLA também fornece recomendações de dimensionamento e orientação de licenciamento corretos.

- Conclua todo o dimensionamento correto antes de adquirir os [Savings Plans](#). Isso pode ajudar você a evitar aquisições excessivas em seu compromisso com os Savings Plans.

Próximas etapas

Recomendamos as próximas etapas abaixo:

1. Analise seu cenário atual e considere a conversão de volumes gp2 do Amazon EBS em volumes gp3.
2. Revise os [Savings Plans](#).

Recursos adicionais do

- [Introdução ao Compute AWS Optimizer](#) (documentação)
- [Melhores práticas para a marcação de AWS recursos](#) (AWS whitepapers)
- [Contêineres Windows ativados AWS](#) (AWS Workshop Studio)

Obter visibilidade dos seus custos do Amazon EKS

Visão geral do

Uma visão holística é necessária para monitorar com eficácia o custo de uma implantação do Kubernetes. O único custo fixo e conhecido é o do ambiente de gerenciamento do Amazon Elastic Kubernetes Service (Amazon EKS). Isso inclui todos os outros componentes que compõem a implantação, da computação e armazenamento à rede, sendo uma quantidade variável com base nas necessidades da sua aplicação.

Você pode usar o [Kubecost](#) para analisar o custo da sua infraestrutura do Kubernetes, desde os [namespaces](#) e [serviços](#) até os [pods](#) individuais, e depois exibir os dados em um painel. O Kubecost apresenta custos em cluster, como computação e armazenamento, e custos out-of-cluster como buckets do Amazon [Simple Storage Service \(Amazon S3\)](#) e [instâncias do Amazon Relational Database Service](#) (Amazon RDS). O Kubecost fará recomendações de dimensionamento correto com base nesses dados e exibirá alertas críticos que podem afetar o sistema. O Kubecost pode se [integrar](#) com o [AWS Cost and Usage Report](#) para mostrar economias dos [Savings Plans para computação](#), das [Instâncias Reservadas](#) e de outros programas de descontos.

Benefícios de custo

O Kubecost fornece relatórios e painéis que visualizam o custo de suas implantações do Amazon EKS. Ele permite que você faça uma análise detalhada do cluster em cada um dos vários componentes, como controladores, serviços, nós, pods e volumes. Isso lhe dá uma visão holística de suas aplicações em execução em um ambiente do Amazon EKS. Ao habilitar essa visibilidade, você pode agir de acordo com as recomendações do Kubecost ou visualizar os custos de cada aplicação em um nível granular. O dimensionamento correto de um grupo de nós do Amazon EKS oferece a mesma economia potencial que as instâncias padrão do EC2. Se você conseguir dimensionar corretamente seus contêineres e nós, poderá remover o excesso computacional do tamanho da instância necessária para executar o contêiner e do número de instâncias do EC2 necessárias no grupo do Auto Scaling.

Recomendações de otimização de custos

Para aproveitar as vantagens do Kubecost, é recomendável fazer o seguinte:

1. Implementar o Kubecost em seu ambiente
2. Obter um detalhamento granular dos custos das aplicações do Windows
3. Dimensionar corretamente nós de clusters
4. Dimensionar corretamente solicitações de contêineres
5. Gerenciar nós subutilizados
6. Remediar workloads abandonadas
7. Agir de acordo com as recomendações
8. Atualizar nós autogerenciados

Implementar o Kubecost em seu ambiente

O [Amazon EKS Finhack Workshop](#) ensina como implantar um ambiente Amazon EKS configurado para usar o Kubecost em uma conta própria. AWS Isso permite que você tenha experiência prática com a tecnologia. Se você estiver interessado em realizar esse workshop em sua organização, entre em contato com a equipe da sua conta.

Para implantar o Kubecost em seu cluster Amazon EKS usando o [Helm](#), consulte a publicação [AWS e a Kubecost colabore para fornecer monitoramento de custos para clientes do EKS no blog](#). AWS

Como alternativa, você pode consultar a [documentação oficial do Kubecost](#) para obter instruções sobre como instalar e configurar o Kubecost. Para obter informações sobre o suporte do Kubecost para nós do Windows, consulte [Windows Node Support](#) na documentação do Kubecost.

Obter um detalhamento granular dos custos das aplicações do Windows

Embora você possa obter uma redução de custo significativa usando [instâncias spot do Amazon EC2](#), você também pode se beneficiar do fato de que as workloads do Windows tendem a ser com estado. O uso de instâncias spot depende da aplicação, e recomendamos que você verifique se elas serão aplicáveis ao seu caso de uso.

Para obter uma análise granular dos custos de suas aplicações do Windows, [faça login no Kubecost](#). Na página de navegação, escolha Economia.

Dimensionar corretamente nós de clusters

No [Kubecost](#), escolha Economia na barra de navegação e depois Dimensionar corretamente o nó do cluster.

Considere um exemplo em que o Kubecost relata que o cluster está superprovisionado em termos de vCPU e RAM. A tabela a seguir mostra os detalhes e as recomendações do Kubecost.

	Atual	Recomendação: simples	Recomendação: complexa
Contagem total	USD 3.462,57 por mês	USD 137,24 por mês	USD 303,68 por mês
Contagem de nós	4	5	4
CPU	74 VCPUs	10 VCPUs	8 VCPUs
RAM	152 GB	20 GB	18 GB
Detalhamento da instância	2 c5.xlarge + 2 mais	5 t3a.medium	2 c5n.large + 1 mais

Conforme descrito na publicação do blog do Kubecost [Find an optimal set of nodes for a Kubernetes cluster](#), a opção simples utiliza um único grupo de nós, enquanto a complexa utiliza uma abordagem

de grupo de vários nós. O botão Saiba como adotar pode realizar o redimensionamento do cluster com um clique. Isso requer a instalação do [Kubecost Cluster Controller](#).

Se você estiver usando [nós autogerenciados do Windows](#) que não foram criados pelo [eksctl](#), consulte [Updating an existing self-managed node group](#). Essas instruções mostram como alterar o tipo de instância no modelo de inicialização do Amazon EC2 usado pelo [grupo do Auto Scaling](#).

Dimensionar corretamente solicitações de contêineres

No [Kubecost](#), escolha Economia na barra de navegação e acesse a página Solicitar recomendações de dimensionamento correto. Essa página mostra a [eficiência](#) dos pods, as recomendações de dimensionamento correto e a economia estimada de custos. Você pode usar o botão Personalizar para filtrar por Cluster, Nó, Namespace\Controlador e muito mais.

Como exemplo, considere que o Kubecost calculou que alguns de seus pods estão superprovisionados em termos de CPU e RAM (memória). Em seguida, ele recomenda que você se ajuste aos novos valores de CPU e RAM para obter sua economia mensal estimada. Para alterar os valores de CPU e RAM, você deve atualizar seu arquivo de [manifesto de implantação](#).

Gerenciar nós subutilizados

No [Kubecost](#), escolha Economia na barra de navegação e depois Gerenciar nós subutilizados.

Considere um exemplo em que a página mostra que um nó no cluster é subutilizado em termos de CPU e RAM (memória) e, portanto, pode ser drenado e encerrado ou redimensionado. Escolher os nós que não passam nas verificações de nós e pods fornecerá mais informações sobre por que eles não podem ser drenados.

Remediar workloads abandonadas

No [Kubecost](#), escolha Economia na barra de navegação e depois a página Workloads abandonadas. Neste exemplo, você filtra pelo namespace chamado windows. Essa página mostra os pods que não atingiram o limite de tráfego e foram considerados abandonados. Os pods precisam enviar ou receber uma certa quantidade de tráfego de rede durante o período definido.

Depois de considerar cuidadosamente que um ou mais pods foram abandonados, você pode economizar custos reduzindo a escala verticalmente do número de réplicas, excluindo a implantação, redimensionando-a para consumir menos recursos ou notificando o proprietário da aplicação de que acredita que a implantação foi abandonada.

Agir de acordo com as recomendações

Na seção Dimensionar corretamente os nós do cluster, o Kubecost analisa o uso dos nós de processamento no cluster e faz recomendações sobre o dimensionamento correto dos nós para reduzir os custos. Há dois tipos de grupos de nós que podem ser usados com o Amazon EKS: [autogerenciados](#) e [gerenciados](#).

Atualizar nós autogerenciados

Para obter informações sobre a atualização de nós autogerenciados, consulte [Atualizações de nós autogerenciados](#) na documentação do Amazon EKS. O documento afirma que os grupos de nós criados com o `eksctl` não podem ser atualizados e devem ser migrados para um novo grupo de nós com a nova configuração.

Como exemplo, suponha que você tenha um grupo de nós do Windows chamado `ng-windows-m5-2xlarge` (que usa uma instância `m5.2xlarge` do EC2) e queira migrar os pods para um [novo grupo de nós](#) chamado `ng-windows-t3-large` (que é apoiado por uma instância `t3.large` do EC2 para reduzir custos).

Para migrar para um novo grupo de nós ao usar grupos de nós implantados pelo `eksctl`, faça o seguinte:

1. Para encontrar o nó em que o pod está atualmente, execute o comando `kubectl describe pod <pod_name> -n <namespace>`.
2. Execute o comando `kubectl describe node <node_name>`. A saída mostra que o nó está sendo executado em uma instância `m5.2xlarge`. Também corresponde ao nome do grupo de nós (`ng-windows-m5-2xlarge`).
3. Para alterar a implantação para usar o grupo de nós `ng-windows-t3-large`, exclua o grupo de nós `ng-windows-m5-2xlarge` e execute `kubectl describe svc,deploy,pod -n windows`. A implantação começa imediatamente a ser reimplantada agora que seu grupo de nós foi excluído.

Note

Haverá um tempo de inatividade do serviço quando você excluir o grupo de nós.

4. Execute o comando `kubectl describe svc,deploy,pod -n windows` novamente após alguns minutos. A saída mostra que os pods estão todos no estado `Running` novamente.

5. Para mostrar que os pods agora estão sendo executados no grupo de nós `ng-windows-t3-large`, execute os comandos `kubectl describe pod <pod_name> -n <namespace>` e `kubectl describe node <node_name>` novamente.

Métodos alternativos de redimensionamento

Esse método se aplica a qualquer combinação de grupos de nós autogerenciados ou gerenciados. A publicação do blog [Seamlessly migrate workloads from EKS self-managed node group to EKS-managed node groups](#) fornece orientação sobre como migrar suas workloads de um grupo de nós com o tipo de instância superdimensionada para o grupo de nós que foi dimensionado corretamente sem nenhum tempo de inatividade.

Próximas etapas

O Kubecost facilita a visualização do custo de seus ambientes do Amazon EKS. A profunda integração do Kubecost com o Kubernetes e o AWS APIs pode ajudar você a encontrar possíveis economias de custo. Você pode vê-las como recomendações no painel de Economia do Kubecost. O Kubecost também pode implementar algumas dessas recomendações para você por meio do [recurso de controlador de clusters](#).

Recomendamos que você analise a step-by-step implantação no blog [AWS Containers e que a Kubecost colabore para fornecer monitoramento de custos para clientes do AWS EKS](#).

Recursos adicionais do

- [Amazon EKS Workshop](#) (Amazon EKS Workshop)
- [AWS e a Kubecost colaboram para fornecer monitoramento de custos para clientes da EKS](#) (Blog)AWS
- Workshop [Amazon EKS Finhack \(estúdio de AWS oficina\)](#)
- [Contêineres Windows ativados AWS](#) (AWS Workshop Studio)

Redefinir a plataforma das aplicações do Windows com o App2Container

Visão geral do

O [AWS App2Container](#) é uma ferramenta de linha de comando para migrar e modernizar aplicações web Java e .NET em contêineres. O App2Container analisa e cria um inventário de todas as aplicações em execução em bare metal, em máquinas virtuais, em instâncias do Amazon Elastic Compute Cloud (Amazon EC2) ou em outros provedores de nuvem. Você seleciona a aplicação que deseja containerizar. O App2Container empacota os artefatos e as dependências da aplicação em imagens de contêineres, configura as portas de rede e gera os artefatos de implantação do Amazon Elastic Container Service (Amazon ECS) e do Amazon Elastic Kubernetes Service (Amazon EKS), que são modelos de infraestrutura como código (IaC). O App2Container provisiona a infraestrutura de nuvem e os pipelines de CI/CD necessários para implantar a aplicação containerizada em um ambiente de produção. Para obter mais informações, consulte [How App2Container works](#) na documentação do App2Container.

Com o App2Container, você pode migrar AWS e modernizar seus aplicativos como contêineres, além de padronizar a implantação e as operações de seus aplicativos. Você pode usar o App2Container para ajudar a criar rapidamente uma prova de conceito (PoC) ou acelerar a implantação de workloads de produção em contêineres.

Existem vários pontos a considerar ao lidar com aplicações Windows. O App2Container oferece suporte à containerização de aplicações ASP.NET implantadas no Internet Information Services (IIS) da Microsoft, incluindo aplicações Windows Communication Foundation (WCF) hospedados no IIS que são executadas no Windows Server 2016, Windows Server 2019 ou Windows Server Core 2004. Para obter mais informações, consulte [Supported applications for Windows](#) na documentação do App2Container. O App2Container usa o Windows Server Core como imagem base para seus artefatos de contêiner, combinando a versão do contêiner do Windows Server Core com a versão do sistema operacional (SO) do servidor em que você executa os comandos de containerização. Essa abordagem separa a aplicação do SO subjacente para que você possa atualizá-lo sem realizar uma migração tradicional.

Se você usar uma máquina de operador para containerizar sua aplicação, a imagem base do contêiner, como o canal de serviço de longo prazo (LTSC) do Windows Server 2019, corresponderá ao SO da máquina de operador, como o Windows Server 2019. Se você estiver executando a containerização diretamente nos servidores de aplicações, a versão corresponderá ao SO do seu

servidor de aplicações. Se suas aplicações estiverem sendo executadas no Windows Server 2008 ou 2012 R2, você ainda poderá usar o App2Container configurando uma máquina de operador para as etapas de containerização e implantação. O App2Container não é compatível com aplicações executadas em sistemas operacionais clientes Windows, como o Windows 7 ou o Windows 10. O App2Container suporta estruturas Tomcat, TomEE e JBoss (modo autônomo) para processos Java. Para obter mais informações, consulte [App2Container compatibility](#).

Benefícios de custo

A containerização e a consolidação de seus aplicativos podem gerar até [60% de economia de computação](#) em comparação com um padrão de design de implantação one-application-to-one de servidor. O App2Container ajuda a agilizar o processo de containerização de aplicações. Confira abaixo algumas das vantagens de usar o App2Container para suas necessidades de modernização:

- O App2Container é oferecido sem custo adicional.
- O App2Container é compatível com várias aplicações em uma imagem de contêiner.
- Aborde a questão dos sistemas operacionais que estão chegando ao fim do suporte usando o App2Container para migrar suas aplicações .NET legadas para contêineres. Você pode migrar para um sistema operacional mais novo, evitar pagar por suporte estendido e reduzir os riscos de segurança.
- Os contêineres são um método eficiente e econômico de empacotar suas aplicações .NET. Analise as vantagens dos contêineres na [Recomendação MACO - Migrando para contêineres](#).
- A consolidação e a containerização de aplicações ajudam a reduzir sua pegada de computação, armazenamento e licenciamento usando seus recursos computacionais com mais eficiência.
- A migração para contêineres pode reduzir a sobrecarga operacional e os custos de infraestrutura, além de aumentar a portabilidade do desenvolvimento e a agilidade de implantação.

Recomendações de otimização de custos

Para obter instruções sobre como usar o App2Container, consulte [Getting started with AWS App2Container](#). Para obter mais informações sobre os comandos do App2Container, consulte [App2Container command reference](#).

Próximas etapas

O App2Container pode acelerar o processo de containerização de aplicações e implantação no Amazon EKS ou no Amazon ECS. A implantação de aplicações em contêineres reduz os custos

de computação, rede e armazenamento e também a sobrecarga operacional dos operadores de aplicações.

[Para uma experiência prática com o App2Container, consulte o Modernize with Workshop. AWS App2Container](#) Se você quiser ter uma experiência de aprendizado aprofundada, peça à sua equipe de AWS contas que configure um dia de imersão no App2Container.

Recursos adicionais do

- [Colocando em contêineres aplicativos Windows complexos de várias camadas usando AWS App2Container](#) (AWS postagem do blog)
- [Containerizando aplicativos ASP.NET legados usando AWS App2Container](#) (postagem do blog)AWS
- [Aplicativos compatíveis com o App2Container](#) (documentação)AWS
- [Modernize com o AWS App2Container Workshop](#) (AWS Workshop Studio)
- [AWS App2Container FAQs](#)(AWS site)

Armazenamento

Escolher o armazenamento certo para suas workloads da Microsoft é uma decisão crítica de arquitetura. Como parte do processo de tomada de decisão, recomendamos que você desenvolva um plano de armazenamento e determine os requisitos funcionais para suas aplicações e serviços. Este capítulo fornece uma visão geral das opções de armazenamento a seguir que podem ser consideradas no planejamento.

Seções:

- [Amazon EBS](#)
- [Amazon FSx](#)
- [AWS Storage Gateway](#)

Amazon EBS

O Amazon Elastic Block Store (Amazon EBS) é um serviço de armazenamento em bloco totalmente gerenciado que permite armazenar volumes de armazenamento ao nível do bloco persistentes que você pode usar com instâncias do Amazon Elastic Compute Cloud (Amazon EC2). Você pode aproveitar vários recursos do Amazon EBS para gerenciar e otimizar com eficácia seus recursos de armazenamento para workloads do Windows na nuvem. Por exemplo, você pode usar o Amazon EBS para provisionar a quantidade exata de IOPS e throughput de que você precisa para sua workload, selecionar entre uma variedade de tipos de volume para atender aos seus requisitos de workload e usar ferramentas para identificar e eliminar o desperdício de recursos de armazenamento. Esse controle granular da performance e do uso do armazenamento ajuda você a otimizar seus recursos de armazenamento, evitando custos desnecessários.

Esta seção abrange os seguintes tópicos:

- [Migrar volumes do Amazon EBS de gp2 para gp3](#)
- [Modificar snapshots do Amazon EBS](#)
- [Excluir volumes do Amazon EBS não anexados](#)

Migrar volumes do Amazon EBS de gp2 para gp3

Visão geral do

Uma unidade de estado sólido (SSD) é a opção de armazenamento padrão para workloads de produção e de alta performance. O Amazon EBS oferece um [volume SSD de uso geral](#) para workloads de média a alta performance. O padrão em muitos Serviços da AWS (incluindo o Amazon EC2) é o [gp2](#), a segunda geração desses volumes SSD de uso geral. A terceira geração de uso geral SSDs, chamada [gp3](#), foi lançada em dezembro de 2020.

A oferta gp3 fez melhorias significativas nos aspectos de personalização de performance em relação à geração anterior. Para volumes gp2 do Amazon EBS, a performance está estritamente associada ao tamanho do volume. Para cada 1 GB de capacidade, os volumes gp2 obtêm 3 IOPS de performance. Ou seja, um volume gp2 de 2.000 GB é capaz de 6.000 IOPS. Para volumes gp3, a performance pode ser personalizada independentemente da capacidade de armazenamento. Isso permite que até mesmo pequenos volumes de capacidade atinjam capacidades de desempenho de até 80.000 IOPS e 2.000 MB/s de taxa de transferência.

Outra grande mudança nos volumes gp3 é a performance de linha de base de IOPS. Os volumes do gp3 começam em 3.000 IOPS. Em comparação, os volumes gp2 devem atingir 1 TiB de tamanho antes de atingir a mesma capacidade de performance. Para o Windows Server, que geralmente tem uma unidade C: muito menor que 1 TiB, a atualização de gp2 para gp3 é uma melhoria significativa na performance.

Finalmente, o preço dos volumes gp3 é uma das melhorias mais significativas em comparação com os volumes gp2. Os volumes gp3 oferecem recursos de performance aprimorados a um custo 20% menor do que os volumes gp2.

Impacto do custo

Com a capacidade de escalar a performance independentemente da capacidade, é importante entender os aspectos de preços da adição de IOPS e throughput extras. Para volumes de gp2, o preço é baseado na capacidade provisionada de USD 0,10 por GiB/mês. Para volumes gp3, o preço é semelhante aos [volumes SSD de IOPS provisionadas](#) de alta performance, que têm um custo de capacidade e um custo separado para IOPS e throughput extras.

Conforme observado na tabela a seguir, os volumes gp3 têm um preço de capacidade de 0,08 USD por GiB por mês (20% mais barato que o gp2) e custos separados para IOPS de 0,005 USD por IOPS provisionado por mês acima de 3.000 e 0,04 USD por mês provisionado acima de 125 USD para taxa de transferência. MiBs MiBs

	gp3	gp2
Tamanho do volume	1 GiB — 64 TiB	1 GiB – 16 TiB
IOPS de linha de base	3.000	3 IOPS/GiB (mínimo de 100 IOPS) a um máximo de 16.000 IOPS Volumes menores que 1 TiB também podem se expandir até 3.000 IOPS.
IOPS máxima/volume	80.000	16.000
Linha de base de throughput	125 MiBs	O limite de taxa de transferência está entre 128 MiBs e 250 MiBs, dependendo do tamanho do volume
Throughput máximo/volume	2.000 MiBs	250 MiBs
Preço	USD 0,08/GiB/mês 3.000 IOPS gratuitas e USD 0,005/IOPS provisionadas por mês acima de 3.000 125 MiBs gratuitos e 0,04 MiBs USD/provisionado - mês acima de 125 MiBs	USD 0,10 USD/Gib por mês

⚠ Important

Embora os volumes gp3 tenham custos separados de capacidade e performance, os volumes gp3 serão sempre mais baratos do que os volumes gp2 se estiverem configurados nos mesmos níveis de performance.

As tabelas a seguir mostram exemplos de economias de custo que podem ser obtidas com a conversão de volumes gp2 em gp3 em várias configurações de capacidade e performance.

Exemplo de configuração do gp2

Tamanho do volume (GiB)	IOPS máx.	Rendimento () MiBs	Custo (USD/mês)
30	3000	128	\$3,00
100	3000	128	\$10,00
500	3000	250	\$50,00
1000	3000	250	\$100,00
2000	6000	250	\$200,00
6000	16000	250	\$600,00

Exemplo de configuração do gp3 (linha de base)

IOPS máx.	Rendimento () MiBs	Custo (USD/mês)	Redução de custos (em comparação com o gp2)
3000	125	\$2,40	20%
3000	125	\$8,00	20%
3000	125	\$40,00	20%
3000	125	\$80,00	20%
3000	125	\$160,00	20%
3000	125	\$480,00	20%

Exemplo de configuração do gp3 (correspondência com o gp2)

IOPS máx.	Rendimento () MiBs	Custo (USD/mês)	Redução de custos (em comparação com o gp2)
3000	128	\$2,52	16%
3000	128	\$8,12	19%
3000	250	\$45,00	10%
3000	250	\$85,00	15%
6000	250	\$180,00	10%
16000	250	\$550,00	8%

Para uma análise de custos, consulte a seção da calculadora de economia de custos da migração do gp2 para o gp3 do EBS no [recurso do Amazon EBS](#). É possível baixar a calculadora e usá-la para descobrir o quanto você pode economizar ao migrar seus volumes gp2 para gp3.

Recomendações de otimização de custos

Para obter instruções sobre como concluir o processo de migração, consulte a publicação [Migre seus volumes do Amazon EBS de gp2 para gp3 e economize até 20% em custos no blog de armazenamento](#). AWS

Recursos adicionais do

- [Migre seus volumes do Amazon EBS de gp2 para gp3 e economize até 20% em custos \(Storage Blog\)AWS](#)
- [Crie uma regra AWS Config personalizada para otimizar os tipos de volume do Amazon EBS \(blog de operações e migrações AWS na nuvem\)](#)
- [Controle seus AWS custos excluindo volumes não utilizados do Amazon EBS](#) (blog de operações e migrações AWS na nuvem)
- [Utilitário de migração do Amazon EBS](#) () GitHub
- [Finding savings from 2020 re:Invent announcements](#) (Gerenciamento financeiro da nuvem da AWS)

- [Workshop de otimização de custos](#) (AWS Well-Architected Labs)
- [Calculadora de economia de custos da migração do gp2 para o gp3](#) (download)

Modificar snapshots do Amazon EBS

Visão geral do

Excluir volumes do EBS e gerenciar a retenção e o arquivamento de snapshots é um aspecto importante para controlar os custos desde o início. Você pode fazer backup dos dados dos seus volumes do EBS no Amazon Simple Storage Service (Amazon S3) tirando snapshots. point-in-time Snapshots são backups incrementais, então eles só salvam os blocos nos dispositivos que foram alterados depois do snapshot mais recente. Isso minimiza o tempo necessário para criar o snapshot e economiza em custos de armazenamento ao não duplicar os dados. Cada snapshot contém todas as informações necessárias para restaurar seus dados (desde o momento em que o snapshot foi criado) para um volume novo do EBS.

As cobranças dos snapshots do EBS são calculadas em gigabytes por mês. Você é cobrado pelo tamanho do snapshot e por quanto tempo você o mantém. O preço varia de acordo com a camada de armazenamento. Para a [camada Standard](#), você é cobrado somente pelos blocos alterados que são armazenados. Para a camada Archive, você é cobrado por todos os blocos de snapshots armazenados. Você também é cobrado pela recuperação de snapshots da [camada Archive](#). Confira abaixo os exemplos de cenários para cada camada de armazenamento:

- Camada Standard: você tem um volume que armazena 100 GB de dados. Você é cobrado pelos 100 GB completos de dados do primeiro snapshot (snap A). No momento do próximo snapshot (snap B), você tem 105 GB de dados. Você então será cobrado apenas pelos 5 GB adicionais de armazenamento para o snap incremental B.
- Camada Archive: você arquiva o snap B. O snapshot é então movido para a camada Archive e você é cobrado pelo bloco de snapshot completo de 105 GB.

Você pode usar o [Amazon Data Lifecycle Manager](#) para ajudar você a configurar um ciclo de vida para reter e gerenciar seus snapshots dentro da programação.

Impacto do custo

As cobranças por volumes e snapshots do EBS são gerenciadas separadamente. Os snapshots do EBS são cobrados a uma taxa menor do que os volumes ativos do EBS. Quando uma instância é

encerrada, o valor do [DeleteOnTermination atributo](#) para cada volume do EBS conectado determina se o volume deve ser preservado ou excluído. Por padrão, o atributo `DeleteOnTermination` é definido como `True` para o volume raiz. Está definido como `False` para todos os outros tipos de volume. Isso cria situações em que o operador pretende excluir uma instância do EC2, mas deixa para trás volumes que foram adicionados à instância, além do volume raiz. Para obter instruções sobre como verificar volumes (e seus snapshots associados) de que você não precisa mais, consulte [View information about an Amazon EBS volume](#) na documentação do Amazon EBS.

Por padrão, quando você cria um snapshot, ele é armazenado na camada Standard para snapshots do Amazon EBS (camada padrão). Os snapshots armazenados na camada padrão são incrementais. Isso significa que apenas os blocos de volumes que foram alterados após o snapshot mais recente são salvos. O [Arquivo de Snapshots do Amazon EBS](#) é uma nova camada de armazenamento que você pode usar para armazenamento de baixo custo e longo prazo dos snapshots raramente acessados e que não requerem recuperação frequente ou rápida. A diferença de preço do padrão para o arquivamento é significativa e deve ser uma consideração importante ao configurar sua estratégia de snapshots. O Arquivo de Snapshots do Amazon EBS oferece custos de armazenamento de snapshots até 75% menores para os snapshots que você planeja armazenar por 90 dias ou mais e que não precisa acessar com frequência.

Armazenamento de snapshots do Amazon EBS	Custo
Standard	USD 0,05/GB/mês
Arquivo	USD 0,0125/GB/mês

Em ambientes menores, a economia de custos pode não ser significativa. A economia é mais significativa em grande escala, quando há várias contas e milhares de instâncias do EC2 sem snapshots TBs do EBS, mesmo quando os volumes do EBS foram excluídos.

A tabela a seguir compara as camadas Standard e de Archive por mês com apenas 50 TB de uso. Mesmo nessa escala mais baixa, ainda são milhares de dólares em economias anuais.

Armazenamento de snapshots do Amazon EBS	Custo por mês	Custo por ano
Standard 50 TB	\$312,50	\$3.750

Armazenamento de snapshots do Amazon EBS	Custo por mês	Custo por ano
Archive 50 TB	\$78,13	\$937,60
	Economia anual	\$2.812,40

Recomendações de otimização de custos

A exclusão de um snapshot pode não reduzir os custos de armazenamento de dados de sua organização. Outros snapshots podem fazer referência aos dados desse snapshot e os dados referenciados serão sempre preservados. Por exemplo, quando você executa o primeiro snapshot de um volume com 10 GiB de dados, o tamanho do snapshot também é de 10 GiB. Como os snapshots são incrementais, o segundo snapshot que você tira do mesmo volume contém apenas os blocos de dados alterados desde que o primeiro snapshot foi tirado. O segundo snapshot também faz referência aos dados no primeiro snapshot. Ou seja, se você alterar 4 GiB de dados e tirar um segundo snapshot, o tamanho do segundo snapshot será de 4 GiB. Além disso, o segundo snapshot faz referência aos 6 GiB inalterados no primeiro snapshot. Para obter mais informações, consulte [Por que meus custos de armazenamento não diminuíram depois que eu excluí um snapshot do meu volume do EBS e depois excluí o volume](#) em si? no Centro de AWS Conhecimento.

Considere o seguinte:

- Você não é cobrado por instantâneos que outra pessoa Conta da AWS possui e compartilha com sua conta. Só há cobrança quando você copia o snapshot compartilhado para sua conta. Também há cobrança pelos volumes do EBS que você cria do snapshot compartilhado.
- Se um snapshot (snap A) for referenciado por outro snapshot (snap B), então excluir o snap B pode não reduzir os custos de armazenamento. Ao excluir um snapshot, somente os dados exclusivos desse snapshot serão removidos. Os dados referenciados por outros snapshots permanecem, e há uma cobrança por esses dados referenciados. Para excluir um snapshot incremental, consulte [Incremental snapshot deletion](#) na documentação do Amazon EBS.

A limpeza de snapshots é uma prática operacional padrão ao executar suas workloads na AWS. Com o tempo, os snapshots podem resultar em cobranças caras por dados de que você não precisa.

Recursos adicionais do

- [Controle seus AWS custos excluindo volumes não utilizados do Amazon EBS](#) (blog de operações e migrações AWS na nuvem)
- [Delete an Amazon EBS snapshot](#) (documentação do Amazon EBS)
- [Workshop de otimização de custos](#) (AWS Well-Architected Labs)
- [Automatically archive Amazon EBS Snapshots with Amazon Data Lifecycle Manager](#) (blog AWS Storage)

Excluir volumes do Amazon EBS não anexados

Visão geral do

Volumes EBS não conectados (órfãos) podem gerar custos de armazenamento desnecessários em seu ambiente. É essencial incorporar a revisão e a exclusão regulares de volumes do EBS não utilizados e não utilizados como parte da higiene do ambiente. É uma prática recomendada ter um processo para analisar continuamente o uso dos volumes do EBS. Você pode usar o [AWS Compute Optimizer](#) para analisar instâncias subutilizadas. Esta seção ajuda você a identificar, gerenciar e excluir volumes do EBS que estão desanexados ou subutilizados.

Amazon EBS

O [Amazon Elastic Block Store \(Amazon EBS\)](#) é um dispositivo ao nível do bloco que oferece volumes de armazenamento para instâncias do Amazon Elastic Compute Cloud (Amazon EC2). O EBS fornece armazenamento persistente, com a flexibilidade de anexar e desconectar de instâncias do EC2. Isso significa que o ciclo de vida dos volumes do EBS persistirá mesmo se uma instância do EC2 for encerrada. O [DeleteOnTermination](#) atributo é um recurso que controla se os volumes anexados do EBS devem ser preservados ou excluídos após o encerramento da instância. Por padrão, o atributo é definido como `True` para o volume raiz, resultando na exclusão. Está definido como `False` para outros volumes, resultando na preservação.

Impacto do custo

Os volumes do EBS não anexados, também chamados de volumes não utilizados ou órfãos, incorrem nas mesmas cobranças que os volumes anexados, com base no tamanho do armazenamento provisionado e no tipo de armazenamento. Embora o custo médio das cobranças do Amazon EBS possa parecer mínimo, a USD 0,10 por GB/mês, é crucial reconhecer que o acúmulo de volumes não utilizados do EBS pode resultar em custos significativos ao longo do tempo.

Por exemplo, considere as ramificações de reter 50 volumes do EBS não utilizados, cada um provisionado com um tamanho de armazenamento de 100 GB, conforme mostra a tabela a seguir.

Quantidade de volumes de armazenamento	Tipo de volume	Tamanho	Custo mensal total
50 volumes	gp2 (USD 0,10)	100 GB	100 GB 50,00 volumes do EBS por mês USD 0,10 = USD 500,00

O cenário da tabela anterior gera uma redução de custo de aproximadamente USD 500 por mês ou USD 6.000 por ano. Esta é uma medida eficaz para a redução de custos. Certifique-se de incorporar a exclusão de volumes do EBS não conectados como uma prática regular na higiene do ambiente.

AWS

Recomendações de otimização de custos

Você pode usar AWS para automatizar facilmente a exclusão de volumes do EBS não anexados. Por exemplo, você pode usar AWS Lambda, AWS Config CloudWatch, Amazon e AWS Systems Manager para definir critérios para excluir volumes não anexados com base na idade, tags e outras especificações. Você também pode usá-los Serviços da AWS para automatizar o processo de limpeza em grande escala.

Para evitar consequências não intencionais, recomendamos que você realize a devida diligência antes de excluir volumes do EBS não anexados.

Gerenciar volumes do EBS não anexados

Recomendamos que você considere as seguintes práticas recomendadas:

- Atenda aos requisitos de conformidade: verifique se a exclusão de volumes do EBS não anexados está em conformidade com os requisitos de governança e conformidade da sua organização.
- Defina políticas de backup e retenção de dados: antes de excluir um volume do EBS não anexado, faça backup de todos os dados importantes em outro repositório de armazenamento (por exemplo, o [Amazon S3](#)). Para a retenção de dados, os [snapshots do Amazon EBS](#) são uma forma mais econômica de reter dados do que os volumes do EBS, e eles podem restaurar o volume, se

necessário, no futuro. Para obter mais informações sobre o gerenciamento eficaz de snapshots, consulte a seção [Modificar snapshots do Amazon EBS](#) deste guia.

- Verifique se há dependências: verifique se há dependências entre volumes do EBS não anexados e outros recursos da AWS. Você pode usar o [Console de gerenciamento da AWS ou uma API](#) para reunir informações descritivas sobre seus volumes do EBS, como tamanho, status e recursos associados. Esta é uma etapa importante para proteger contra a exclusão de recursos temporariamente não anexados.
- Crie uma política de retenção: estabeleça um período de retenção para volumes do EBS não anexados. Isso pode ajudá-lo a identificar o momento apropriado para excluir volumes não conectados, garantindo que seu AWS ambiente permaneça otimizado. Por exemplo, você pode criar uma EventBridge regra da [Amazon](#) para iniciar uma função Lambda de forma programada. A função Lambda pode usar o AWS SDK para identificar ativamente quaisquer volumes do EBS não conectados, aplicar um mecanismo de marcação para facilitar o rastreamento e enviar notificações quando um volume do EBS não vinculado atinge ou excede um limite definido.
- Marque volumes do EBS não anexados: [marcar](#) volumes do EBS é uma prática útil que pode ajudar na organização e identificação de volumes com base em atributos como ambiente, aplicação ou proprietário. Isso pode ser particularmente útil ao decidir quais volumes não anexados excluir, pois permite identificar rapidamente os volumes que não são mais necessários com base em suas tags.
- Garanta a exclusão segura: analisar quando um volume do EBS foi anexado pela última vez pode ajudar você a determinar se é seguro excluir o volume. Para obter mais informações, consulte [Como eu uso AWS CLI comandos para listar os anexos ou o histórico de destacamentos de um volume específico do Amazon EBS?](#) no Centro de AWS Conhecimento.
- Identificar volumes subutilizados do EBS — Identificar e remover volumes subutilizados do EBS é uma prática altamente recomendada para reduzir os custos de armazenamento e manter um ambiente otimizado. AWS Trusted Advisor e [AWS Compute Optimizer](#) pode ajudá-lo a identificar volumes subutilizados do EBS e fornecer recomendações para reduzir custos e melhorar a eficiência. Por exemplo, consulte [Configuração da automação para otimizar volumes do EBS com AWS Trusted Advisor](#) (GitHub), [painel Estabelecendo uma Trusted Advisor organização \(TAO\) \(AWS Workshop Studio\)](#) e Otimização de custos dos [volumes do Amazon EBS](#) usando (Storage Blog). AWS Compute OptimizerAWS

Automatizar a limpeza de volumes do EBS não anexados

Recomendamos que você considere as seguintes ferramentas para ajudar você a automatizar a limpeza de volumes do EBS não anexados:

- [AWS APIs \(DescribeVolumes\)](#) — Você pode filtrar e encontrar volumes EBS não anexados usando a AWS SDKs ou o AWS Command Line Interface (AWS CLI). Você pode economizar tempo e esforço automatizando esse processo com um script ou uma [função do Lambda](#) que é executada em uma programação. Um [exemplo de script](#) GitHub demonstra como isso funciona. O script usa o Lambda para analisar AWS CloudTrail registros e identificar volumes do EBS não conectados.
- [AWS Systems Manager Automação](#) — Isso permite que você automatize as tarefas rotineiras de manutenção e remediação em sua infraestrutura. Para começar, [crie um runbook de automação](#), que define uma série de etapas a serem executadas em uma ordem específica. Por exemplo, você pode criar um runbook que primeiro cria um snapshot do volume EBS não anexado e depois exclui o volume em si. Isso pode ajudar a automatizar tarefas que, de outra forma, seriam demoradas e propensas a erros se fossem feitas manualmente.
- [AWS Config](#) — Isso permite que você avalie, audite e acompanhe as mudanças em seus AWS recursos ao longo do tempo. Ao capturar as alterações de configuração, você pode usar AWS Config para avaliar a conformidade, a governança e a utilização de recursos em seu ambiente. Por exemplo, AWS Config pode identificar volumes do [EBS não utilizados](#). Além disso, você pode associar a AWS Systems Manager automação AWS Config para remediar automaticamente a exclusão de volumes não utilizados do EBS.

Recursos adicionais do

- [Exclua volumes não utilizados do Amazon Elastic Block Store \(Amazon EBS\) usando AWS Config](#) e (orientação prescritiva) AWS Systems Manager
- [Controle seus AWS custos excluindo volumes não utilizados do Amazon EBS](#) (blog de operações e migrações AWS na nuvem)
- [AWSConfigRemediation-DeleteUnusedEBSVolume](#) (Referência do runbook de AWS Systems Manager automação)

Amazon FSx

O Amazon FSx para Windows File Server é um serviço de armazenamento de arquivos totalmente gerenciado que é otimizado para cargas de trabalho do Windows. Ele fornece uma solução simples e escalável para executar suas aplicações e workloads baseadas em Windows, sem a necessidade de gerenciamento complexo da infraestrutura de armazenamento. Você pode usar o Windows File Server FSx para provisionar e acessar facilmente o armazenamento compartilhado de arquivos que ofereça suporte nativo aos aplicativos do Windows, incluindo Microsoft SQL Server SharePoint,

Microsoft e aplicativos.NET personalizados. Além disso, o File Server FSx for Windows ajuda você a gerenciar custos fornecendo opções flexíveis de preços, como cotas de armazenamento pay-as-you-go e deduplicação automática de dados para reduzir o espaço ocupado pelo armazenamento e otimizar o desempenho e o custo.

Esta seção abrange os seguintes tópicos:

- [Escolher o armazenamento adequado de arquivos SMB](#)
- [Habilite a deduplicação de dados na Amazon FSx](#)
- [Entenda a fragmentação de dados no FSx Windows File Server](#)
- [Entenda o uso do volume de HDD na Amazon FSx](#)
- [Usar uma única zona de disponibilidade](#)

Escolher o armazenamento adequado de arquivos SMB

Visão geral do

AWS oferece uma variedade de serviços de armazenamento totalmente gerenciados que oferecem os recursos avançados dos serviços de arquivos líderes do setor, ao mesmo tempo em que combinam as mais recentes inovações de AWS infraestrutura e segurança. Você pode incorporar os serviços da AWS aos fluxos de trabalho de infraestrutura como código (IaC) e integrá-los aos serviços de computação, monitoramento e proteção de dados da AWS. Para cargas de trabalho do Windows, você pode escolher entre dois serviços de arquivos totalmente gerenciados que podem ser usados para atender às necessidades do seu aplicativo: FSx para Windows File Server e Amazon FSx para NetApp ONTAP.

FSx para Windows File Server

O Amazon FSx para Windows File Server fornece armazenamento compartilhado totalmente gerenciado baseado no Windows Server e oferece uma ampla variedade de recursos administrativos, de acesso a dados e de gerenciamento de dados. FSx para Windows O File Server se integra facilmente aos ambientes Windows porque é um serviço nativo do Windows. Recomendamos usar o FSx Windows File Server para compartilhamentos de usuários e grupos, instâncias de cluster de failover Always On para SQL Server, aplicativos Windows e infraestrutura de desktop virtual (VDI). FSx para Windows, o File Server também se integra bem ao Amazon FSx File Gateway, ao Amazon Kendra, aos logs de auditoria do Amazon S3 e ao Amazon Data Firehose.

FSx para ONTAP

FSx for ONTAP é baseado no sistema de arquivos ONTAP proprietário NetApp da ONTAP. Isso exige algum nível de aprimoramento e é recomendado principalmente para usuários locais NetApp existentes. Os casos de uso típicos incluem compartilhamentos de usuários e grupos, instâncias de cluster de failover Always On para SQL Server e aplicativos Windows. FSx for ONTAP oferece suporte a vários protocolos, sistemas de arquivos maiores que 64 TB (escala PB sem um servidor de namespace DFS), clonagem, replicação, instantâneos, compactação (eficiência de armazenamento) e classificação inteligente de dados em camadas.

Impacto do custo

FSx para Windows File Server

FSx para Windows File Server foi a primeira solução de armazenamento compartilhado AWS para implantar instâncias de cluster de failover para SQL Server. Com FSx o Windows File Server, você pode iniciar instâncias de cluster de failover usando o licenciamento da edição SQL Standard. No entanto, isso impede que você confie nos grupos de disponibilidade Always On, que exigem licenças da edição SQL Server Enterprise. Ao mudar da edição SQL Server Enterprise Standard para a edição SQL Server Standard, você pode economizar de 65 a 75% em seu [licenciamento do SQL Server](#).

Você pode usar o FSx for Windows File Server para instâncias de cluster de failover para descarregar o armazenamento do I/O armazenamento típico do EBS. Ao transferir I/O para o FSx for Windows File Server, você pode reduzir as instâncias do EC2, que dependem da alta taxa de transferência e IOPS do Amazon EBS, sem afetar a taxa de transferência do armazenamento.

FSx para ONTAP

Você pode usar o FSx for ONTAP para executar seu cluster de failover da Microsoft no protocolo de blocos iSCSI e se beneficiar da inicialização instantânea de arquivos do SQL Server, do uso da replicação entre regiões, do suporte a antivírus e da clonagem. SnapMirror Se você criar várias cópias de bancos de dados para teste, a clonagem poderá fazer uma diferença significativa no consumo de espaço e na rapidez com que essas cópias do banco de dados podem ser criadas. Além disso, você pode usar NetApp SnapCenter para gerenciar a funcionalidade de backup, restauração e clonagem com suas instâncias do EC2 para SQL Server usando o FSx ONTAP. FSx for ONTAP também fornece escalonamento automático de SSD a um armazenamento em pool de capacidade de baixo custo para uma combinação de desempenho e economia.

FSx for ONTAP suporta NetApp o sistema de arquivos (ONTAP), ao contrário FSx do Windows File Server, que suporta um sistema de arquivos NTFS nativo do Windows. O tamanho mínimo FSx para o ONTAP é de 1024 GB, enquanto FSx para o Windows File Server pode começar a partir de 32 GB.

Integração com o Sistema de Arquivos Distribuído da Microsoft

FSx para Windows File Server e FSx ONTAP, integre-se ao [Sistema de Arquivos Distribuído \(DFS\)](#) da Microsoft para uma integração perfeita às implantações existentes. Tenha o seguinte em mente ao planejar sua arquitetura:

- FSx para Windows File Server e FSx para ONTAP oferecem suporte a [namespaces DFS \(DFSN\)](#) em ambos os tipos de implantação (várias zonas de disponibilidade e zonas de disponibilidade únicas).
- Somente FSx para Windows, o servidor de arquivos oferece suporte [à replicação DFS \(DFSR\)](#) e somente ao usar zonas de disponibilidade únicas.

Recomendações de otimização de custos

O desempenho do Windows File Server e FSx do ONTAP depende muito da configuração, assim como seus preços. FSx para Windows File Server, o preço depende principalmente da capacidade e do tipo de armazenamento, da capacidade de taxa de transferência, do backup e dos dados transferidos. Com FSx o ONTAP, você paga pelo armazenamento SSD, pelo IOPS do SSD, pelo uso do pool de capacidade, pela capacidade de transferência e pelo backup.

Serviço de arquivos	Custo do armazenamento de 5 TB	Configuração	Região
FSx para Windows File Server	\$982,78	Zona de disponibilidade única SSD (15.000 IOPS) 32 MBps Backup de 5 TB (sem economia de desduplicação)	Leste dos EUA (Norte da Virgínia)

Serviço de arquivos	Custo do armazenamento de 5 TB	Configuração	Região
FSx para ONTAP	\$979,28	Zona de disponibilidade única 100% SSD Nível de capacidade de leitura e gravação de 15.000 15.000 IOPS de SSD 128 MBps Backup de 5 TB (sem economia de desduplicação)	Leste dos EUA (Norte da Virgínia)

Lembre-se do seguinte:

- A desduplicação e a compactação permitem que você armazene mais dados em dispositivos físicos ao reduzir o tamanho dos dados, mas você paga pelo armazenamento em unidade de estado sólido (SSD) ou em unidade de disco rígido (HDD) provisionado.
- Você pode usar o ONTAP FSx para hierarquizar seus dados. É extremamente raro que 100% dos seus dados sejam acessados regularmente e exijam armazenamento SSD. Você pode mover dados frios e acessados com pouca frequência para uma camada de capacidade a fim de reduzir custos.
- Os preços mencionados aqui são calculados com dados 100% na camada SSD e 15.000 IOPS na camada SSD.

Backup

Por padrão, tanto FSx para o ONTAP quanto FSx para o Windows File Server armazenam seu backup totalmente gerenciado no Amazon S3. No entanto, com FSx o ONTAP, há uma opção adicional para o uso de backup SnapVault, que pode configurar os backups para residirem no

nível de capacidade. O backup com SnapVault é um mecanismo autogerenciado que é mais econômico do que a opção padrão de backup totalmente gerenciado. A opção de backup totalmente gerenciado custa USD 0,05 por GB/mês. O SnapVault backup ativado FSx para ONTAP (SSD 10:1 para armazenamento em pool de capacidade) é de 0,03221 USD (0,9x0,0219+0,1x0,125).

Lembre-se do seguinte:

- AWS os backups gerenciados oferecem granularidade de uma hora. [SnapVault](#) permite que você vá até cinco minutos.
- Você pode usar as ferramentas NetApp da (como a CLI e a API) para configurar os SnapVault relacionamentos e a replicação de instantâneos.
- Ative a política all de classificação em camadas em um SnapVault volume para usar a camada de capacidade como armazenamento para os dados de backup.
- SnapVault os destinos podem estar na mesma região Região da AWS, em várias regiões ou no local. Isso geralmente é para um destino de backup do sistema de arquivos de uma única zona de disponibilidade ou de várias zonas de disponibilidade. Em comparação, AWS Backup é apoiado pela resiliência regional do Amazon S3.

Dimensionamento correto


Você também pode economizar nos custos e tirar o máximo proveito do seu sistema de arquivos, dimensionando corretamente e evitando o provisionamento excessivo.

Para obter o dimensionamento correto, faça o seguinte:

1. Identifique suas necessidades atuais com base nos dados. Para workloads típicas do Windows, você pode usar ferramentas integradas do sistema operacional, como o [Monitor de Performance](#).
2. No Monitor de Performance, use os contadores a seguir para avaliar suas necessidades atuais de performance. O intervalo de captura é definido para um segundo, com um tamanho máximo de logs de 1.000 MB e a substituição habilitada.

```
Logman.exe create counter PerfLog-Short -o "c:\perflogs\PerfLog-Long.blg" -f bincirc  
-v mmddhhmm -max 1024 -c "\LogicalDisk(*)\*" "\Memory\*" "\.NET CLR Memory(*)\*"  
"\Cache\*" "\Network Interface(*)\*" "\Paging File(*)\*" "\PhysicalDisk(*)\*"  
"\Processor(*)\*" "\Processor Information(*)\*" "\Process(*)\*" "\Thread(*)\*"  
"\Redirector\*" "\Server\*" "\System\*" "\Server Work Queues(*)\*" "\Terminal  
Services\*" -si 00:00:01
```

3. Para iniciar a captura de logs, execute o comando `logman start PerfLog-Short`. Para interromper a captura de logs, execute o comando `logman stop PerfLog-Short`.

 Note

Você pode encontrar arquivos de logs de performance em `c:\perflogs` no servidor que executa a captura. Para obter mais informações, consulte [Windows Performance Monitor Overview](#) na documentação da Microsoft.

4. Depois de identificar a configuração correta, teste se sua estimativa está correta no sistema de FSx arquivos da Amazon usando ferramentas de estresse de disco, como o Microsoft [DISKSPD](#).
5. Se você estiver satisfeito com a performance, vá para o compartilhamento de arquivos.

Recomendamos uma abordagem conservadora da capacidade de armazenamento, pois ela só pode ter a escala aumentada verticalmente. A capacidade de throughput pode ter a escala aumentada e reduzida verticalmente, conforme necessário.

Recursos adicionais do

- [Amazon FSx para NetApp ONTAP FAQs](#) (AWS site)
- [Otimizando o desempenho do Amazon FSx para Windows File Server com novas métricas](#) (AWS Storage Blog)

Habilite a deduplicação de dados na Amazon FSx

Visão geral do

A deduplicação de dados é um recurso que permite armazenar seus dados com mais eficiência e com menos requisitos de capacidade. Ela envolve encontrar e remover a duplicação nos dados sem comprometer sua fidelidade ou integridade. A deduplicação de dados usa fragmentação e compactação de tamanho variável de subarquivos, que oferecem taxas de otimização de 2:1 para servidores de arquivos gerais e de até 20:1 para dados de virtualização. A deduplicação de dados é muito mais eficaz do que a compactação NTFS. A resiliência durante falhas de hardware é inerente à arquitetura de deduplicação, com validação completa da soma de verificação dos dados e metadados, incluindo redundância de metadados e os fragmentos de dados mais acessados.

FSx para Windows File Server oferece suporte total à deduplicação de dados. Seu uso pode resultar em uma economia média de 50 a 60% para compartilhamentos de arquivos de uso geral. Em compartilhamentos, as economias variam de 30 a 50% para documentos do usuário até 70 a 80% para conjuntos de dados de desenvolvimento de software. É importante entender que a economia de armazenamento que você pode obter com a deduplicação de dados depende da natureza do seu conjunto de dados, incluindo a quantidade de duplicação existente nos arquivos. A deduplicação não será uma boa opção se os dados armazenados forem de natureza dinâmica.

Impacto do custo

Para lidar com o crescimento do armazenamento de dados na empresa, os administradores consolidam os servidores e fazem da escalabilidade da capacidade e da otimização de dados os principais objetivos. As configurações padrão da deduplicação de dados podem proporcionar economia imediata, ou os administradores podem ajustar as configurações para obter ganhos adicionais. Por exemplo, você pode configurar a eliminação de duplicação para ser executada somente em determinados tipos de arquivo ou criar uma programação de trabalho personalizada.

Em um alto nível, a deduplicação tem três tipos de trabalhos: otimização, coleta de resíduos e depuração. Esteja ciente de que o espaço não será liberado até que você execute um trabalho de coleta de resíduos após a otimização. Você pode programar o trabalho ou executá-lo manualmente. Todas as configurações disponíveis quando você agenda um trabalho de deduplicação de dados também estão disponíveis quando você inicia um trabalho manualmente (exceto aquelas que são específicas da programação).

Mesmo com uma economia efetiva de apenas 25% com a deduplicação, há uma economia significativa FSx para o Windows File Server. Essas economias projetadas são baseadas em uma [estimativa](#) na AWS Calculadora de Preços.

Recomendações de otimização de custos

A deduplicação ativada FSx para sistemas de arquivos do Windows File Server não está habilitada por padrão. Para habilitar a deduplicação usando o [gerenciamento remoto ativado PowerShell](#), você deve executar o `Enable-FSxDedup` comando e, em seguida, usar o `Set-FSxDedupConfiguration` comando para definir a configuração. Para obter mais informações, consulte [Administração de sistemas de arquivos](#) na documentação do FSx Windows File Server.

Para habilitar a deduplicação, execute o seguinte comando:

```
PS C:\Users\Admin> Invoke-Command -ComputerName amznfsxxxxxxxxx.corp.example.com -  
ConfigurationName FSxRemoteAdmin -ScriptBlock {Enable-FsxDedup }
```

Para verificar se a configuração de deduplicação, execute o seguinte comando:

```
Invoke-Command -ComputerName amznfsxxxxxxxxx.corp.example.com -ConfigurationName  
FSxRemoteAdmin -ScriptBlock {  
Set-FsxDedupSchedule -Name "CustomOptimization" -Type Optimization -Days  
Mon,Tues,Wed,Sat -Start 09:00 -DurationHours 7  
}
```

Ao executar o PowerShell `Measure-DedupFileMetadata` cmdlet, você pode determinar quanto espaço em disco potencial pode ser recuperado em um volume se você excluir um grupo de pastas, uma única pasta ou um único arquivo e, em seguida, executar um trabalho de coleta de lixo. Especificamente, o valor `DedupDistinctSize` informa a quantidade de espaço que você vai recuperar se excluir esses arquivos. Os arquivos geralmente têm partes que são compartilhadas em outras pastas, então o mecanismo de deduplicação calcula quais partes são exclusivas e que seriam excluídas após o trabalho de coleta de resíduos.

As [programações de trabalho padrão de deduplicação de dados](#) foram projetadas para funcionar bem para as workloads recomendadas e serem o menos intrusivas possível (excluindo a tarefa de otimização prioritária habilitada para o tipo de uso de backup). Se as workloads tiverem grandes requisitos de recursos, recomendamos que você programe trabalhos executados somente durante as horas de ociosidade, ou que reduza ou aumente a quantidade de recursos do sistema que um trabalho de deduplicação de dados pode consumir.

Por padrão, a deduplicação de dados usa 25% da memória disponível. No entanto, isso pode ser aumentado usando `-memory switch`. Para trabalhos de otimização, recomendamos que você defina um intervalo de 15 a 50. Para trabalhos programados, você pode usar um maior consumo de memória. Por exemplo, com trabalhos de depuração e coleta de resíduos (que normalmente são programados para serem executados fora do horário de pico), você pode definir um maior consumo de memória (como 50).

Para obter informações adicionais sobre as configurações de deduplicação de dados, consulte [Reduzindo os custos de armazenamento com a deduplicação de dados na documentação](#) do Windows FSx File Server.

Recursos adicionais do

- [Understanding Data Deduplication](#) (documentação da Microsoft)
- [Reduzindo os custos de armazenamento com a deduplicação de dados](#) (FSx para documentação do Windows File Server)

Entenda a fragmentação de dados no FSx Windows File Server

Visão geral do

FSx para o Windows File Server, o desempenho depende da configuração. Ele se baseia principalmente no tipo de armazenamento, na capacidade de armazenamento e na configuração do throughput. A capacidade de taxa de transferência selecionada determina os recursos de desempenho disponíveis para o servidor de arquivos, incluindo I/O os limites de rede, a CPU e a memória e os I/O limites de disco impostos pelo servidor de arquivos. A capacidade de armazenamento e o tipo de armazenamento selecionados determinam os recursos de desempenho disponíveis para os volumes de armazenamento — os I/O limites de disco impostos pelos discos de armazenamento. Além do desempenho, as opções de configuração também influenciam o custo. FSx para Windows File Server, o preço depende principalmente da capacidade e do tipo de armazenamento, da capacidade de taxa de transferência, do backup e dos dados transferidos.

Se você tiver requisitos de performance e armazenamento de arquivos relativamente grandes, poderá se beneficiar da fragmentação de dados. A fragmentação de dados envolve a [divisão de seus dados de arquivos](#) em conjuntos de dados menores (fragmentos) e o armazenamento deles em diferentes sistemas de arquivos. As aplicações que acessam seus dados de várias instâncias podem alcançar altos níveis de performance lendo e gravando nesses fragmentos em paralelo. Ao mesmo tempo, você ainda pode apresentar às suas aplicações uma visão unificada sob um namespace comum. Além disso, elas também podem ajudar a escalar o armazenamento de dados de arquivos além do que cada sistema de arquivos suporta (64 TB) para grandes conjuntos de dados de arquivos, de até centenas de petabytes.

Impacto do custo

Para grandes conjuntos de dados, geralmente é mais eficaz implantar vários sistemas de arquivos pequenos FSx para Windows File Server, em vez de um grande compartilhamento de SSD para alcançar o mesmo nível de desempenho. O uso de uma combinação dos tipos FSx de armazenamento HDD e SSD do Windows File Server permite uma melhor economia de custos e permite que você combine a carga de trabalho com o melhor subsistema de disco subjacente. Nas

tabelas a seguir, você pode ver a diferença entre um único sistema de arquivos de 17 TB e compará-lo com vários sistemas de arquivos menores que aumentam a mesma capacidade.

Sistema de arquivos SSD grande com várias workloads

Nome do servidor	Custo	Configuração	Região
Servidor FSx de arquivos Amazon para Windows	USD 5.716	SSD DE 17 TB 30% de deduplicação 256 Mbps 17 TB de backup	Leste dos EUA (Norte da Virgínia)

Workload particionada usando DFSN

Nome do servidor	Custo	Configuração	Região	Compartilhar
Servidor FSx de arquivos Amazon para Windows	USD 1.024	SSD DE 2 TB 20% de deduplicação 128 Mbps 2 TB de backup multi-AZ	Leste dos EUA (Norte da Virgínia)	Compartilhamento 1
Servidor FSx de arquivos Amazon para Windows	USD 2.132	SSD DE 5 TB 30% de deduplicação 256 Mbps 5 TB de backup	Leste dos EUA (Norte da Virgínia)	Compartilhamento 2

Nome do servidor	Custo	Configuração	Região	Compartilhar
		multi-AZ		
Servidor FSx de arquivos Amazon para Windows	USD 1.036	HDD DE 10 TB 40% de deduplicação 128 Mbps 10 TB de backup multi-AZ	Leste dos EUA (Norte da Virgínia)	Compartilhamento 3
Instâncias EC2 Windows com DFSN	USD 27	t3a.medium 2 v CPUs 4 GiB de memória	Leste dos EUA (Norte da Virgínia)	Instâncias DFSN

O custo anual de um grande sistema de arquivos SSD é de USD 68.592. O custo anual de uma workload particionada é de USD 50.640. Neste exemplo, é possível obter uma economia de 26% ao combinar a workload com o armazenamento de backend apropriado. Para obter mais informações sobre a estimativa de preço, consulte a estimativa da [AWS Calculadora de Preços](#).

Recomendações de otimização de custos

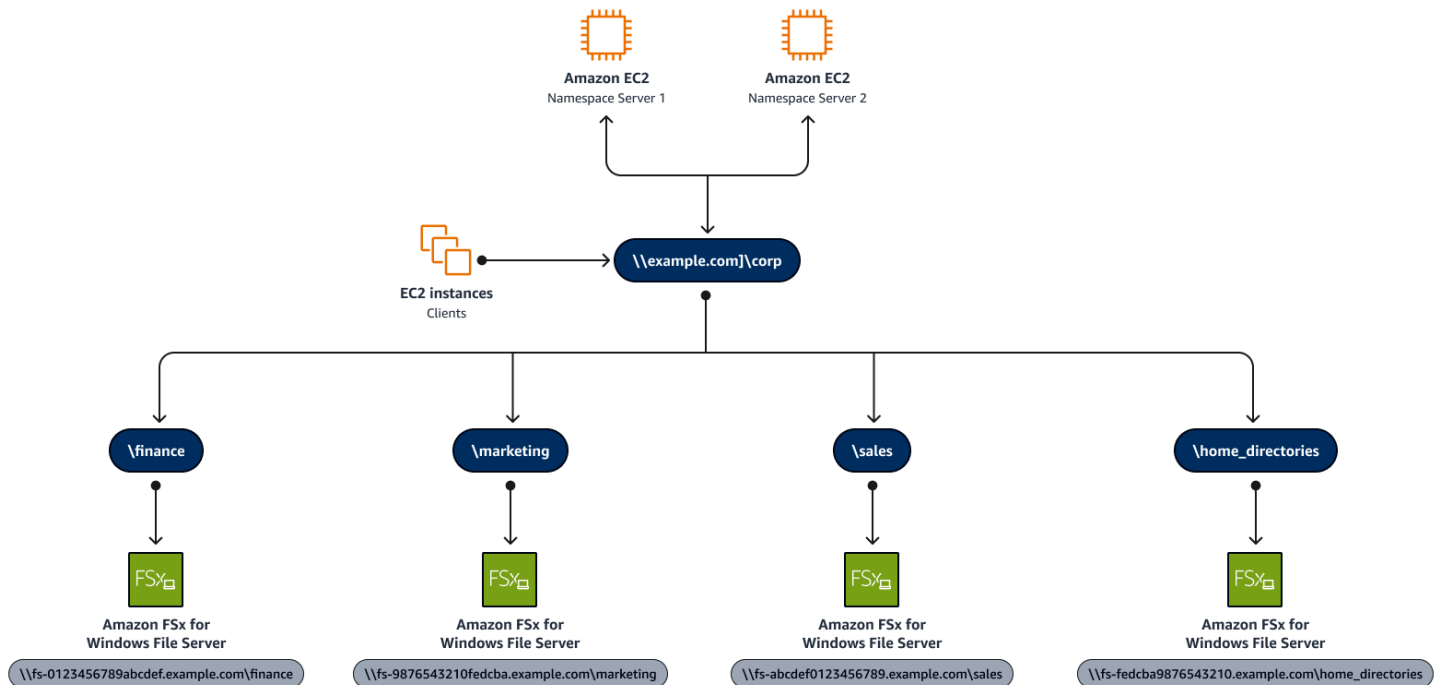
Para implantar uma solução de deduplicação de dados, você deve configurar um [Namespace Microsoft DFS](#) com base no tipo de dados, I/O tamanho e padrão de acesso. I/O Cada namespace é compatível com até 50 mil compartilhamentos de arquivos e centenas de petabytes de capacidade de armazenamento no total.

Ele funcionará de forma mais eficiente se escolher uma convenção de fragmentação que distribua a E/S uniformemente entre todos os compartilhamentos de arquivos que você planeja usar. O monitoramento de sua workload ajudará na otimização adicional ou na redução de custos. Se precisar de ajuda para avaliar as informações de desempenho do sistema de FSx arquivos da

Amazon, consulte FSx o [desempenho do Windows File Server](#) na documentação do FSx Windows File Server.

Depois de escolher uma estratégia de fragmentação, você pode agrupar os sistemas de arquivos para facilitar o acesso aos seus compartilhamentos usando namespaces DFS. Isso permite que os usuários vejam um sistema de arquivos homogêneo quando, na realidade, estão acessando uma variedade de sistemas de arquivos diferentes com casos de uso com propósito específico. É importante criar os compartilhamentos com uma convenção de nomenclatura adequada para que seus usuários finais possam decifrar facilmente para qual workload os compartilhamentos foram projetados. Também é importante rotular compartilhamentos de produção e de não produção para que os usuários finais não coloquem arquivos no sistema de arquivos errado por engano.

O diagrama a seguir mostra como um único namespace DFS pode ser usado como ponto de acesso para vários sistemas de arquivos da Amazon FSx .



Lembre-se do seguinte:

- Você pode adicionar compartilhamentos existentes FSx do Windows File Server a uma árvore DFS.
- A Amazon não FSx pode ser adicionada à raiz do caminho de compartilhamento do DFS. Você tem apenas uma subpasta.
- Você deve implantar uma instância do EC2 para atender à configuração do namespace DFS.

Para obter mais informações sobre a configuração do DFS-N, consulte [DFS Namespaces overview](#) na documentação da Microsoft. Para obter mais informações sobre o uso de namespaces DFS, consulte o vídeo [Usando namespaces DFS com Amazon FSx para Windows File Server](#) sobre YouTube

Recursos adicionais do

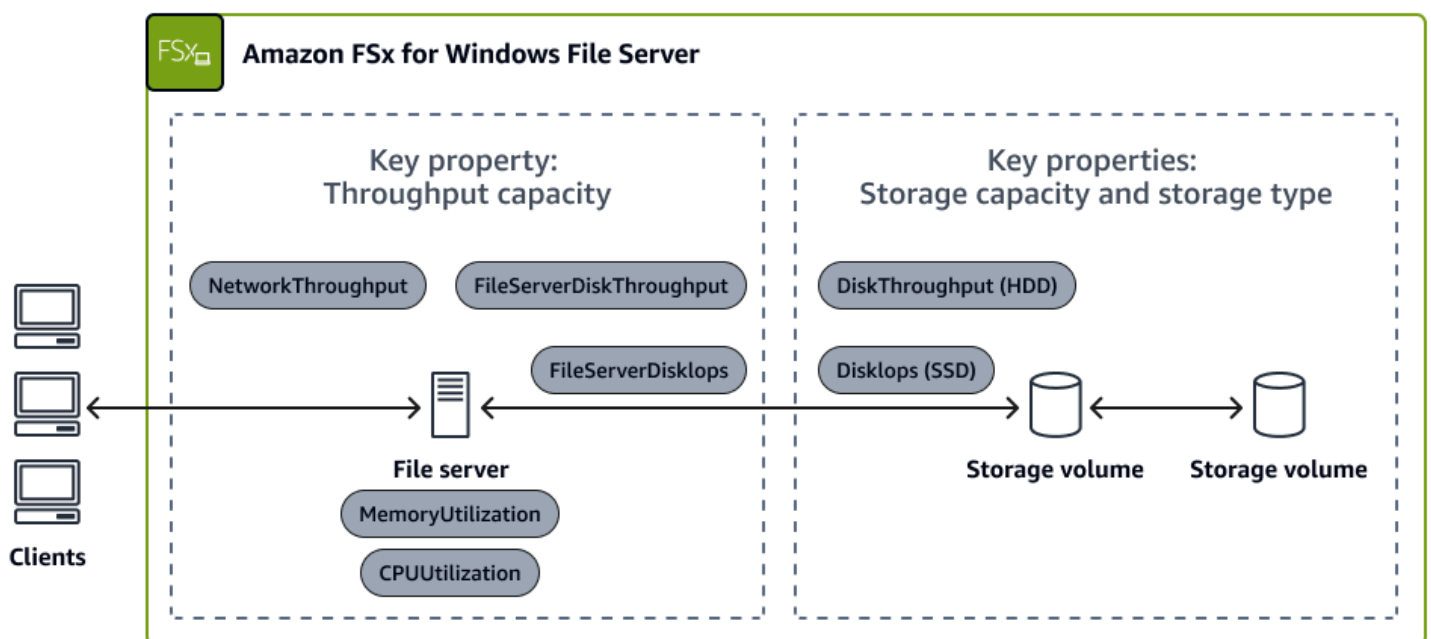
- [Agrupamento de vários sistemas de arquivos com namespaces DFS](#) (documentação da Amazon) FSx
- [Passo a passo 6: Escalando o desempenho com fragmentos](#) (documentação da Amazon) FSx
- [Usando namespaces DFS com o Amazon FSx para Windows File Server](#) (Labs)AWS

Entenda o uso do volume de HDD na Amazon FSx

Visão geral do

O Amazon FSx para Windows File Server oferece a flexibilidade de escolher a taxa de transferência independentemente da capacidade do sistema de arquivos. Duas configurações de capacidade estão disponíveis: unidade de disco rígido (HDD) e unidade de estado sólido (SSD).

O diagrama a seguir mostra o relacionamento entre as configurações de throughput e armazenamento.



Com o armazenamento baseado em HDD, você recebe uma linha de base de 12 IOPS com 80 IOPS de disco de intermitência (por IOPs TiB de armazenamento) e uma taxa de transferência de 12 linhas de base Megabytes/second com 80 de intermitência (por TiB de armazenamento). Megabytes/second Por exemplo, se seu compartilhamento for de 50 TB, você obterá $50 * 12 = 600$ como linha de base para throughput e IOPS.

O Amazon FSx para Windows File Server fornece 80 IOPS de intermitência. Os créditos de expansão são recarregados automaticamente quando sua utilização está abaixo da taxa de linha de base, e são consumidos automaticamente quando sua utilização está acima da taxa de linha de base. Por exemplo, se sua carga de trabalho estiver utilizando apenas 10 IOPS/TB por uma hora (2 IOPS/TB below your baseline rate), you can then utilize 14 IOPS/TB (2 IOPS/TB acima da linha de base) na hora seguinte, antes de ficar sem créditos intermitentes novamente.

Para operações de arquivos, o Amazon FSx for Windows File Server fornece latências consistentes de menos de um milissegundo com armazenamento SSD e latências de um dígito em milissegundos com armazenamento em HDD. Para todos os sistemas de arquivos, incluindo aqueles com armazenamento em HDD, o Amazon FSx for Windows File Server fornece um cache rápido (na memória) no servidor de arquivos, para que você possa obter alto desempenho e latências inferiores a um milissegundo para dados acessados ativamente, independentemente do tipo de armazenamento.

Quando apropriado, o uso do armazenamento HDD pode ajudar a reduzir o custo de sua capacidade geral de armazenamento e fornecer uma plataforma de armazenamento confiável para suas necessidades.

Impacto do custo

O desempenho do Amazon FSx para Windows File Server depende de três fatores: capacidade de armazenamento, tipo de armazenamento e taxa de transferência. O I/O desempenho da rede e o tamanho do cache na memória são determinados exclusivamente pela capacidade de taxa de transferência, enquanto o I/O desempenho do disco é determinado por uma combinação de capacidade de taxa de transferência, tipo de armazenamento e capacidade de armazenamento.

Embora o SSD seja recomendado para cargas de trabalho I/O intensivas, há uma variedade de cargas de trabalho cujas necessidades podem ser atendidas com as especificações de desempenho do HDD. O armazenamento em HDD foi projetado para um amplo espectro de workloads, incluindo diretórios pessoais, compartilhamentos de usuários e departamentos e sistemas de gerenciamento de conteúdo. Por exemplo, se seus usuários precisam apenas de acesso de baixa latência aos

dados que dão suporte aos projetos atuais, a maioria dos dados que você está armazenando é acessada com pouca frequência.

Você pode usar a [AWS Calculadora de Preços](#) para fornecer uma comparação de um SSD de 20 TB com um sistema de arquivos HDD em us-east-1. Como mostra a tabela a seguir, mesmo sem economia de desduplicação, a diferença de custo é significativa ao comparar sistemas de arquivos HDD com sistemas de arquivos SSD.

Configuração do sistema FSx de arquivos da Amazon	Custo mensal
SSD multi-AZ de 20 TB (us-east-1)	\$4.699,30
HDD multi-AZ de 20 TB (us-east-1)	\$542,88
Economia mensal estimada	\$4.156,42

Note

Para obter mais FSx economias sobre o Windows File Server, consulte a FSx seção [Habilitar a desduplicação de dados na Amazon](#) deste guia.

Ao identificar corretamente suas necessidades de performance, você pode selecionar o armazenamento certo para sua workload e reduzir seus custos.

Recomendações de otimização de custos

Se você decidir usar o armazenamento HDD, teste seu sistema de arquivos para garantir que ele atenda aos seus requisitos de performance. O armazenamento HDD tem um custo menor em relação ao armazenamento SSD, mas com níveis mais baixos de throughput e IOPS de disco por unidade de armazenamento. Ele pode ser adequado para compartilhamentos de usuários de uso geral e diretórios pessoais com baixos I/O requisitos, grandes sistemas de gerenciamento de conteúdo em que os dados são recuperados com pouca frequência ou conjuntos de dados com pequenos números de arquivos grandes.

O tipo de armazenamento de um sistema de arquivos existente não pode ser alterado. Para converter o tipo de armazenamento de um sistema de arquivos Amazon FSx para Windows File

Server, você deve fazer backup do seu sistema de arquivos existente e restaurá-lo em um novo sistema de arquivos com o tipo de armazenamento desejado. Se você deseja converter um sistema de arquivos SSD existente em um sistema de arquivos HDD, saiba que o HDD tem uma capacidade mínima muito maior de 2 TB.

Para restaurar um backup com um tipo de armazenamento diferente, faça o seguinte:

1. [Faça o backup de seus sistemas de arquivos.](#)
2. [Crie um novo sistema de FSx arquivos da Amazon](#) com o tipo de armazenamento HDD.
3. Restaure o backup no novo sistema de arquivos com o tipo de armazenamento desejado.
4. Verifique se o novo sistema de arquivos tem o tipo de armazenamento correto e se seus dados estão intactos.

Antes de transferir suas alterações para a produção, recomendamos que você analise o desempenho do seu sistema de FSx arquivos da Amazon e verifique se a alteração é aceitável. Para obter mais orientações, consulte a publicação [Otimizando o desempenho do Amazon FSx para Windows File Server com novas métricas](#) no blog AWS de armazenamento.

Recursos adicionais do

- [Otimizando custos com a Amazon FSx](#) (FSx documentação da Amazon)

Usar uma única zona de disponibilidade

Visão geral do

Esta seção explica quando é mais vantajoso usar uma única implementação de zona de disponibilidade do [Amazon FSx para Windows File Server](#). Ele abrange cenários em que a mudança para uma única zona de disponibilidade reduz custos e ainda permite que você use o Amazon FSx for Windows File Server como seu serviço gerenciado de armazenamento de arquivos. Recomendamos que você implemente uma única zona de disponibilidade para a Amazon FSx para cargas de trabalho de produção. Isso pode ajudar a garantir que você tenha a redundância de várias zonas de disponibilidade.

Impacto do custo

Um único sistema de arquivos de zona de disponibilidade oferece uma redução de custo de aproximadamente 40% em comparação com a implementação de várias zonas de disponibilidade.

Com um sistema de arquivos com várias zonas de disponibilidade, você paga USD 0,230 por GB/mês em SSD e USD 0,025 por GB/mês em HDD, em comparação com USD 0,130 por GB/mês para SSD e USD 0,013 por GB/mês para HDD em um único sistema de arquivos de zona de disponibilidade. Você pode ver uma comparação de custos e criar suas próprias estimativas usando a [AWS Calculadora de Preços](#).

Para um sistema de arquivos de 10 TB, isso pode ser a diferença de pagar aproximadamente USD 1.200 por mês por várias zonas de disponibilidade ou USD 680 por mês por uma única zona de disponibilidade. Este [exemplo](#) usa um sistema de arquivos do Windows File Server de 10 TB FSx com SSD. A economia estimada para a desduplicação é de 50%. No geral, uma única zona de disponibilidade tem um custo de entrada menor, mas vem com algumas advertências que serão abordadas na próxima seção.

Recomendações de otimização de custos

Implantações de zona de disponibilidade única

Para garantir que uma única zona de disponibilidade seja a opção certa, considere sua própria zona interna SLAs para os dados que estão sendo armazenados no FSx Windows File Server. Isso implica entender se você precisa fornecer SLAs aos seus clientes (internos e externos) e se os três nove de disponibilidade de uma FSx única zona de disponibilidade da Amazon ainda permitirão que você os atenda. SLAs FSx para Windows File Server com uma única zona de disponibilidade ainda tem um tempo de atividade de 99,9%. O SLA da Amazon FSx para várias zonas de disponibilidade é superior a 99,99%. Para workloads essenciais, recomendamos que você use várias zonas de disponibilidade em uma única zona de disponibilidade, mesmo com custo adicional.

As implantações de zona de disponibilidade única são ideais para workloads, como backups para bancos de dados do SQL Server. Elas podem fornecer armazenamento de baixo custo com uma camada de HDD e, ao mesmo tempo, fornecer um tempo de atividade consistente. Se você precisar de um nível mais alto de disponibilidade para uma workload de produção, como servidores SQL de alta disponibilidade ou acesso a aplicações de produção, então uma única zona de disponibilidade não é a opção certa para suas workloads. Para backups, testes de não produção e ambientes de desenvolvimento, uma implementação de zona de disponibilidade FSx única da Amazon pode reduzir seus custos operacionais.

Um caso de uso em que um sistema de arquivos de zona de disponibilidade FSx única da Amazon funciona bem é em uma situação de produção em que vários sistemas de arquivos de zona de disponibilidade FSx única da Amazon estão em uso, como armazenamento por servidor em um cluster altamente disponível do SQL Server usando grupos de disponibilidade Always On. Para

obter mais informações, consulte a postagem [Otimizando o custo de suas implantações de alta disponibilidade do SQL Server no AWS](#) blog AWS de armazenamento.

Replicação em várias regiões

Uma opção potencial para reduzir custos com um único sistema de arquivos de zona de disponibilidade (em que apenas um único sistema de arquivos de zona de disponibilidade funciona) é se você quiser aproveitar as vantagens de uma replicação multirregional com a Amazon FSx. Você pode implantar [sistemas de arquivos single-AZ](#) que sejam compatíveis com o uso com o DFS-R nativo da Microsoft. O DFS-R tem a capacidade de replicar dados automaticamente entre regiões e vários locais. Para obter mais informações sobre como configurar o DFS-R usando a FSx Amazon, consulte [Usando a replicação do sistema de arquivos distribuído da Microsoft](#) na documentação da Amazon FSx.

Outra alternativa para economia de custos em várias regiões é usar AWS Storage Gateway. Isso permite que você implemente um [Amazon FSx File Gateway](#) em outra região para acesso multirregional da Amazon FSx. Para obter mais informações, consulte a seção [AWS Storage Gateway](#) deste guia.

Se você trabalha em várias regiões, deve considerar o custo de transferência de dados para o tráfego de dados entre regiões. O tráfego que se move entre regiões incorre em uma cobrança de USD 0,02/GB. Portanto, se você tiver alterações consistentes de dados em grandes volumes, isso aumentará seu custo geral. Por [exemplo](#), 1 TB de transferência de dados equivale a aproximadamente USD 20,48.

Janela de manutenção

A janela de manutenção é uma consideração importante se você estiver usando uma única zona de disponibilidade com a Amazon FSx. Durante a janela de manutenção, o sistema de FSx arquivos da Amazon fica indisponível por aproximadamente 20 minutos, devido a correções de software de rotina para o Windows Server subjacente. Se você estiver usando o sistema de arquivos para backups noturnos, ajuste a janela de FSx manutenção da Amazon adequadamente para evitar interrupções durante o backup. Você pode ajustar a [janela de manutenção](#) depois de criar seu sistema de FSx arquivos da Amazon.

Recursos adicionais do

- [Disponibilidade e durabilidade: sistemas de arquivos Single-AZ e Multi-AZ \(documentação da Amazon FSx\)](#)
- [Preços do Amazon FSx para Windows File Server](#) (AWS site)

AWS Storage Gateway

AWS Storage Gateway é um serviço de armazenamento em nuvem híbrida que conecta ambientes locais ao armazenamento AWS em nuvem. Ele permite que você integre perfeitamente sua infraestrutura local existente AWS, permitindo que você armazene e recupere dados da nuvem e execute aplicativos em um ambiente híbrido. Para workloads do Windows, você pode usar o Storage Gateway para armazenar e acessar dados usando protocolos nativos do Windows, como SMB e NFS. Você pode usar o Storage Gateway para reduzir os custos associados à execução de cargas de trabalho do Windows AWS usando hardware e software locais como uma ponte para a nuvem. Isso permite que você aproveite a escalabilidade e a economia AWS sem precisar fazer alterações significativas em sua infraestrutura existente.

Sob a égide do Storage Gateway, você obtém o Amazon S3 File Gateway, o Amazon File Gateway, o Tape Gateway e o Volume Gateway. FSx O S3 File Gateway e FSx o File Gateway são mais comumente usados com cargas de trabalho da Microsoft.

Gateway de Arquivos do Amazon S3

[O Gateway de Arquivos do Amazon S3](#) permite que você armazene seus arquivos no Amazon S3 enquanto fornece acesso aos seus usuários usando compartilhamentos SMB tradicionais. Isso fornece uma interface de usuário familiar e ajuda a reduzir custos armazenando seus dados no Amazon S3 e aproveitando as várias camadas de armazenamento do Amazon S3. É possível implementar o Storage Gateway com o S3 Intelligent Tiering para ajudar você a mover automaticamente os arquivos do ciclo de vida para as camadas de armazenamento de menor custo, a fim de reduzir ainda mais seus custos. Recomendamos o Gateway de Arquivos do S3 para acesso com a escala aumentada verticalmente e somente leitura, leituras rápidas repetidas (do cache) e despejos de banco de dados. Geralmente, não é recomendado para gravações, edição de arquivos ou compartilhamentos departamentais de alta performance ou alta disponibilidade.

Amazon FSx File Gateway

[O Amazon FSx File Gateway](#) também pode oferecer economia de custos ao trabalhar com sistemas de arquivos Amazon FSx Windows. Você pode criar um gateway de FSx arquivos para fornecer acesso localizado a um sistema de FSx arquivos da Amazon em outra região para evitar os custos de ter dois sistemas de arquivos independentes. Isso também pode ser útil se você tiver vários servidores de arquivos on-premises e quiser consolidá-los para evitar pagar por vários dispositivos de hardware.

Impacto do custo

Gateway de Arquivos do Amazon S3

Configurar o Gateway de Arquivos do S3 é fácil porque você pode usar o assistente de inicialização do Storage Gateway. Você pode implantar o gateway em questão de minutos usando uma instância do EC2 em seu AWS ambiente. Depois que o gateway estiver configurado, você poderá definir os compartilhamentos do Storage Gateway para serem acessíveis por meio dos protocolos SMB e NFS. Para workloads típicas do Windows, você também pode usar essa configuração para aproveitar um ambiente do Active Directory e definir permissões em seus compartilhamentos de arquivos. Você pode integrar efetivamente um Storage Gateway ao seu uso normal, pois ele funcionará como um compartilhamento de arquivos típico do Windows. Arquivos e pastas são armazenados como objetos e listas de controle de acesso NTFS (ACLs) como metadados.

A tabela a seguir compara os custos de 10 TB de armazenamento com três opções de armazenamento disponíveis:

- FSx para Windows File Server
- Gateway de Arquivos do Amazon S3
- Amazon Elastic Block Store (Amazon EBS)

O preço para armazenar 10 TB de armazenamento é consideravelmente mais barato se você usar o Amazon S3, pois você pode particionar seus dados em várias camadas de uso. Na estimativa de preços, o S3 Intelligent Tiering é usado por sua flexibilidade de preços. Isso inclui 80% no S3 Standard, 10% no Infrequent Access e 10% no Amazon Glacier. Embora você possa usar o Amazon Glacier, é importante definir as regras de ciclo de vida adequadas para garantir que os arquivos transferidos para o Amazon Glacier não precisem ser acessados imediatamente. O Amazon Glacier é exclusivamente para uso de arquivamento, não para uso de acesso regular.

Sistemas de armazenamento	Custo do armazenamento de 10 TB	Região
FSx para Windows File Server (supondo uma economia de 50% na deduplicação)	USD 683,20 SSD	Leste dos EUA (Norte da Virgínia)

Sistemas de armazenamento	Custo do armazenamento de 10 TB	Região
Gateway de Arquivos do Amazon S3	USD 449.51 Intelligent Tiering	Leste dos EUA (Norte da Virgínia)
Amazon EBS	\$1.335,69 USD GP3	Leste dos EUA (Norte da Virgínia)


Considere o seguinte:

- No Amazon Glacier, você recebe I/O erros genéricos, a menos que use a [RestoreObjectAPI](#) para restaurar o objeto de volta ao Amazon S3. Recomendamos que você use uma notificação para esse I/O erro usando o Amazon CloudWatch Events. Dessa forma, sua equipe de operações pode reagir ao fato de um usuário receber esse erro em um arquivo que ele talvez precise acessar. Para obter mais informações sobre esses erros, consulte [Erro: InaccessibleStorageClass na documentação](#) do Amazon S3 File Gateway.
- Além da limitação de acesso ao Amazon Glacier, são [ACLs permitidos apenas 10 por objeto/pasta no Storage Gateway](#). Antes de decidir usar o Storage Gateway, verifique se você não precisa de mais de dez entradas da ACL.

Amazon FSx File Gateway

Semelhante a um gateway de arquivos do Amazon S3, um gateway de FSx arquivos fornece acesso a um sistema de arquivos que retém os dados a longo prazo. No Gateway de Arquivos do Amazon S3, os dados residem no Amazon S3. Para o FSx File Gateway, seus dados residem no FSx Windows File Server. Embora as opções Multi-AZ estejam disponíveis FSx para o Windows File Server, não há uma opção multirregional. Se você tem uma empresa global ou um escritório remoto, talvez seja necessário fornecer uma plataforma de armazenamento compartilhado que esteja geograficamente mais próxima do usuário final para evitar latência. Se você implantasse outro sistema de FSx arquivos da Amazon, isso adicionaria o custo de um sistema de arquivos Amazon FSx para Windows File Server totalmente novo e o armazenamento necessário. Para evitar a criação de um sistema de arquivos totalmente novo e a duplicação de custos, você pode implantar o FSx File Gateway na região secundária. Isso fornece acesso localizado aos arquivos para os usuários, ao mesmo tempo que ajuda a reduzir seus custos gerais.

Sistemas de armazenamento	Custo do armazenamento de 10 TB	Região
Servidor FSx de arquivos Amazon para Windows	USD 683,20 SSD	Leste dos EUA (Norte da Virgínia)
Amazon FSx File Gateway	USD 503,70/gateway único	Leste dos EUA (Norte da Virgínia)

 Note

Os preços na tabela anterior são baseados nos [preços do Storage Gateway](#).

Lembre-se do seguinte:

- FSx O File Gateway pode ajudar você a economizar aproximadamente 180 USD por mês (ou 2.100 USD por ano) para cargas de trabalho multirregionais.
- As taxas de transferência de dados são muito mais baixas com o FSx File Gateway, porque ele só precisa armazenar em cache os arquivos que estão sendo acessados regularmente e não uma cópia secundária completa.
- Embora você possa ter duas implantações do FSx Windows File Server em regiões diferentes e mantê-las atualizadas com AWS Backup ou AWS DataSync, nenhuma das opções está quase em tempo real.

Recomendações de otimização de custos

Gateway de Arquivos do Amazon S3

O Gateway de Arquivos do S3 oferece uma opção de baixo custo para armazenar arquivos, mas há algumas questões a serem consideradas em relação à forma como ele implementa e usa o sistema de arquivos. Por exemplo, o Gateway de Arquivos do S3 exige o uso de uma máquina virtual para executar o software Storage Gateway. Em AWS, o Storage Gateway é implantado no Amazon EC2 usando uma instância m5.xlarge, por padrão. Se quiser reduzir seus custos de armazenamento local, você pode implantar o Storage Gateway como um dispositivo virtual em plataformas de virtualização como VMware o Hyper-V.

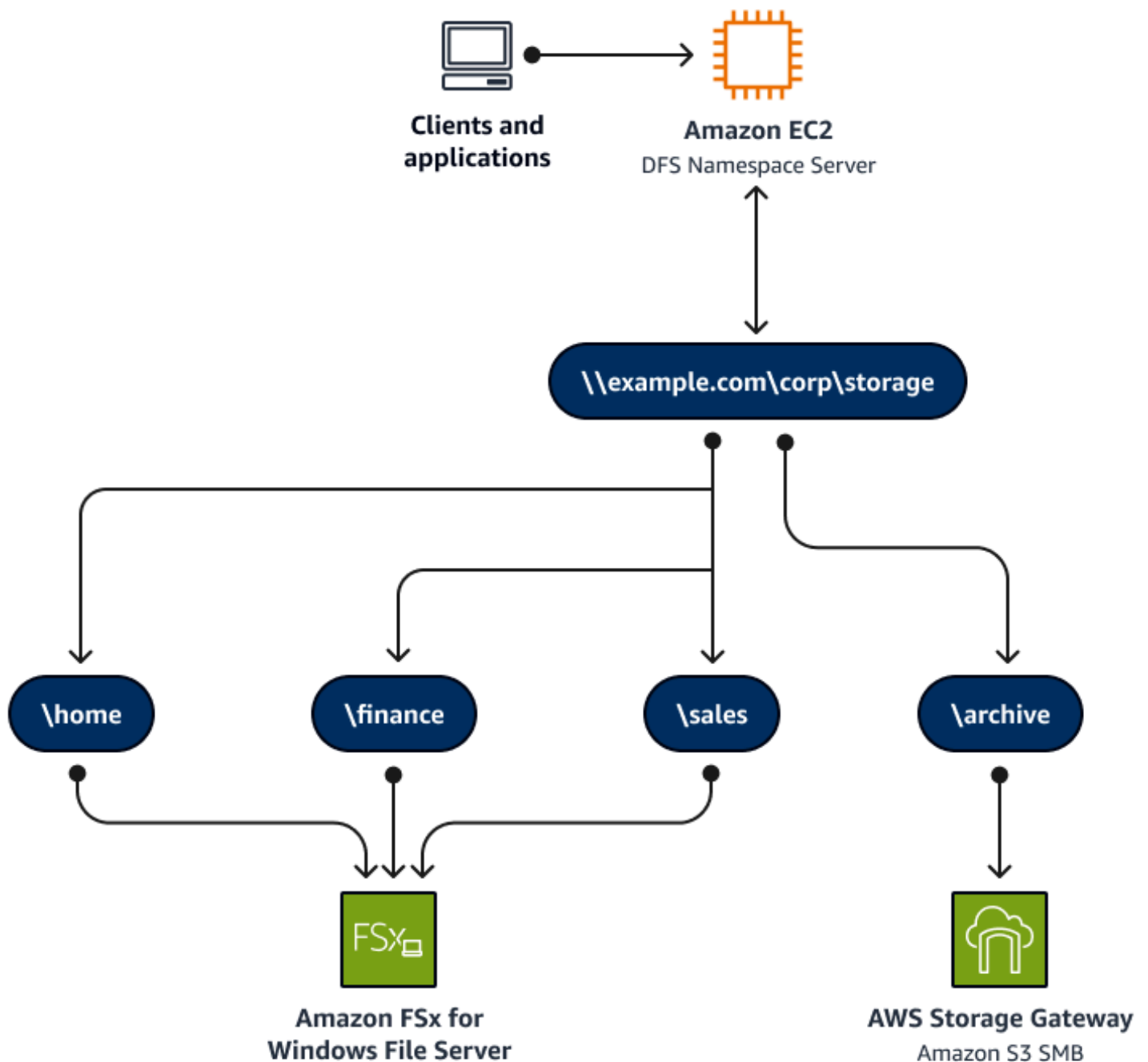
Considerações sobre alta disponibilidade

A execução do Storage Gateway é um único ponto de falha para acesso aos arquivos. Para evitar tempo de inatividade desnecessário, recomendamos que você implemente um controle de acesso rígido em que os usuários possam fazer alterações ou interromper e iniciar a instância do Storage Gateway. Além disso, para implantações em AWS, é vantajoso usar o Amazon Data Lifecycle Manager para criar instantâneos de roteamento para recuperar rapidamente sua implementação do Storage Gateway. Se você estiver executando o Storage Gateway no local usando VMware, você pode configurá-lo para [alta disponibilidade](#).

Execução de vários sistemas de arquivos

Separar suas workloads de arquivos de uso diário de suas workloads de arquivamento pode ajudar você a evitar custos de armazenamento desnecessários. O Storage Gateway tem a capacidade de ser implantado junto com um sistema de arquivos FSx para Windows File Server. Ao usar [namespaces DFS](#), você pode apresentar seu armazenamento principal de uso diário em execução no Windows File Server e seu armazenamento em execução no Amazon S3 (que é acessado FSx por meio do Storage Gateway).

O diagrama a seguir mostra como um único namespace DFS pode ser usado como ponto de acesso de frontend para diferentes opções de armazenamento de backend.



Os clientes são direcionados para uma estrutura de pastas, como `\\example.com\storage`. Esse diretório principal contém os subdiretórios. Um sistema FSx de arquivos do Windows File Server contém os compartilhamentos de arquivos acessados normalmente. Você pode usar um compartilhamento de arquivos criado no Storage Gateway para arquivar dados. Os usuários podem arquivar itens manualmente na pasta de arquivamento, ou você pode criar um processo para

automatizar a movimentação de alguns arquivos dos compartimentos normais para a pasta de arquivamento.

Considere o seguinte:

- Analise seus requisitos de armazenamento e forneça [armazenamento adequado para o cache](#).
- Adicione seu gateway à configuração do Active Directory e use o [Windows padrão ACLs para acessar os arquivos](#).

FSx Gateway de arquivos

A implantação do FSx File Gateway é semelhante à implantação do S3 File Gateway, mas é ainda mais fácil se você usar o assistente de inicialização. Para obter instruções detalhadas, consulte [Etapa 3: Criar e ativar um Amazon FSx File Gateway](#) na documentação do Amazon FSx File Gateway. Depois de implantar FSx o File Gateway em seu ambiente, você pode associá-lo aos seus sistemas de FSx arquivos existentes da Amazon e obter acesso aos seus arquivos.

O armazenamento é a principal consideração ao implantar o FSx File Gateway. O armazenamento padrão fornece 150 GB, o que é uma quantidade razoável de espaço para armazenar arquivos em cache. Criar alertas de monitoramento para pouco espaço livre pode ajudar no dimensionamento correto do armazenamento sem superalocação.

Recursos adicionais do

- [AWS Storage Gateway recursos](#) (AWS documentação)

Active Directory

O Amazon Elastic Compute Cloud (Amazon EC2) executando o Windows Server é um ambiente seguro, confiável e de alta performance para implantar aplicações e workloads com base no Windows. Você pode provisionar instâncias rapidamente e aumentar ou reduzir a escala verticalmente conforme necessário, pagando apenas pelo que usar. Os serviços do Active Directory são usados como a principal fonte de gerenciamento de identidades em ambientes do Windows Server.

Esta seção abrange os seguintes tópicos:

- [Active Directory autogerenciado no Amazon EC2](#)
- [AWS Managed Microsoft AD](#)
- [AD Connector](#)

Active Directory autogerenciado no Amazon EC2

Visão geral do

Esta seção fornece recomendações para reduzir o custo de execução do Active Directory no Amazon Elastic Compute Cloud (Amazon EC2). O foco principal é garantir que você possa dimensionar adequadamente os controladores de domínio do Active Directory e usar a flexibilidade do Nuvem AWS para ajustá-los conforme necessário ao seu ambiente. AWS podem ajudá-lo a interromper facilmente uma instância e redimensioná-la para atender às suas necessidades em constante mudança, ou a reduzir o tamanho da instância se você aumentar a escala muito rápido. Escolher o tamanho e o tipo de instância corretos pode resultar em economias significativas.

Impacto do custo

A tabela a seguir mostra a diferença entre escolher uma instância da família de instâncias expansíveis em vez de uma instância de uso geral. Essa opção pode gerar uma economia significativa para você todos os meses. O planejamento e o dimensionamento adequados da sua instância podem ajudar você a gerenciar os custos.

Tipo de instância	Número de instâncias	vCPU	Memória	Custo
t3a.medium	2	2	8	~USD 81,76/mês
m5a.large	2	2	8	~USD 259,88/mês

Para obter mais informações sobre custos, consulte a AWS Calculadora de Preços [estimativa](#).

Uma economia de USD 178,12 por mês acaba sendo uma economia de mais de USD 2.000 por ano para seus controladores de domínio. Lembre-se de que é para uma pequena pegada de apenas dois controladores de domínio em uma conta. Em grande escala, com várias contas e controladores de domínio adicionais, essa economia pode resultar em uma redução significativa de custos.

Recomendações de otimização de custos

A Microsoft fornece [recomendações de planejamento de capacidade](#) para quando você estiver implementando seu ambiente do Active Directory. Recomendamos que você leve em consideração os seguintes componentes principais ao planejar ou escalar seu ambiente do Active Directory:

- Memória
- Rede
- Armazenamento
- Processador

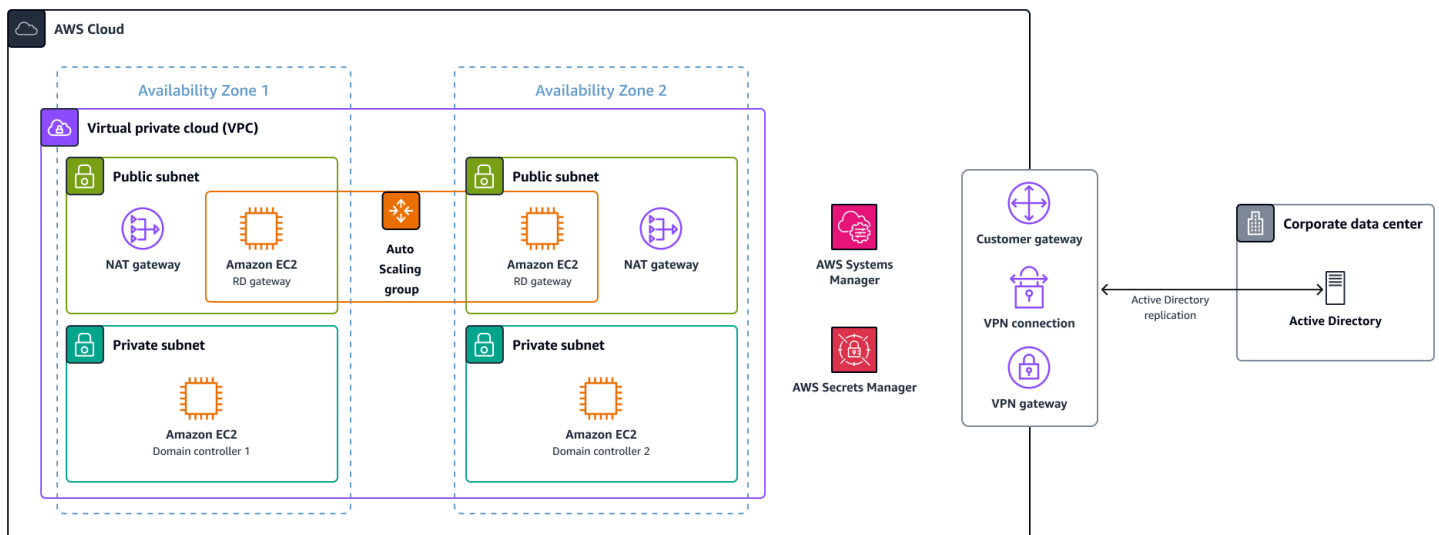
Levando em consideração esses componentes principais, você pode selecionar um tipo de instância que faça sentido para o seu ambiente do Active Directory na AWS. Esta seção aborda alguns exemplos do Active Directory para cenários AWS de implantação. Esses cenários deixam claro que não é necessário replicar seu ambiente local em AWS, se você não planeja lidar com o mesmo número de usuários e computadores que você faz em seu ambiente local.

A tabela a seguir destaca componentes importantes relacionados à vCPU, memória e disco para seu AWS espaço físico.

Componente	Estimativas
Tamanho do banco de dados/armazenamento	40 a 60 KB para cada usuário
RAM	Tamanho do banco de dados Recomendações do sistema operacional de base Aplicações de terceiros
Rede	1 GB
CPU	1.000 usuários simultâneos para cada núcleo

Cenário de implantação híbrida

O diagrama a seguir mostra uma arquitetura de exemplo para uma implantação híbrida do Active Directory.



Conforme mostra o diagrama, você normalmente tem uma pegada on-premises e, em seguida, a expande para a Nuvem AWS. Nas fases iniciais de uma migração, você normalmente não terá todos os seus usuários e servidores implementados na AWS. Por isso é importante implantar inicialmente uma pegada menor para economizar dinheiro nos esforços de migração.

Se você quiser manter uma pegada on-premises com servidores e usuários se autenticando on-premises, não precisará da mesma pegada para controladores de domínio na AWS. Seguindo as práticas recomendadas do Active Directory, você pode implementar [sites e serviços adequados do Active Directory](#) para autenticar usuários e computadores em sua pegada on-premises, ao mesmo tempo em que autentica sua pegada da AWS somente nos controladores de domínio na AWS. Isso permite que você evite sobredimensionar sua área de cobertura do Active Directory, AWS limitando o uso apenas aos AWS recursos e não a toda a sua infraestrutura local. Para obter orientação sobre como criar uma configuração híbrida, consulte [Proper placement of domain controllers and site considerations](#) na documentação da Microsoft.

Otimize para uma AWS migração por meio do dimensionamento correto

Se você estiver implantando uma nova instância do Active Directory para seus usuários ou planeja migrar totalmente AWS para sua infraestrutura do Active Directory, recomendamos que você planeje o dimensionamento de acordo com as recomendações da Microsoft para vCPU, memória e espaço em disco para a escolha de instâncias na tabela anterior.

Se esta for uma nova pegada, você pode começar aos poucos e aproveitar a capacidade de [alterar facilmente os tipos de instância](#) para redimensionar seu ambiente à medida que ele cresce na AWS. A seção [Windows no Amazon EC2](#) deste guia mostra como monitorar e analisar a utilização da CPU e da memória na AWS. Dessa forma, você sabe quando aumentar o tamanho da instância do EC2.

Se estiver migrando totalmente seu ambiente on-premises do Active Directory para a AWS, você poderá implementar os mesmos planos de dimensionamento para garantir a performance adequada. Antes de duplicar o que você tem no local AWS, recomendamos que você conclua uma análise completa do seu ambiente do Active Directory. Isso pode ajudar você a evitar o excesso de provisionamento. Certifique-se de usar o Monitor de Performance para coletar informações sobre o volume de tráfego e a utilização dos controladores de domínio existentes. Isso pode lhe dar uma compreensão do uso geral para que você possa dimensionar corretamente e, por fim, reduzir seus custos.

Otimize o Active Directory em AWS

Se você estiver executando o Active Directory AWS, é importante também monitorar continuamente a utilização e alterar os tamanhos das instâncias conforme necessário para reduzir seus gastos. Você pode usar AWS Compute Optimizer para obter informações sobre os recursos que você está usando AWS. Para obter informações sobre como usar o Compute Optimizer para dimensionar corretamente suas workloads do Windows, consulte a seção [Windows no Amazon EC2](#) deste guia. Para uma análise mais abrangente, você pode usar o Monitor de Performance para monitorar a

utilização dos controladores de domínio do Active Directory, avaliar a performance e, em seguida, redimensionar adequadamente.

Você também pode usar CloudWatch para monitorar o desempenho dos controladores de domínio. Para otimizar seus controladores de domínio (aumentando ou diminuindo a escala), você pode usar as métricas disponíveis CloudWatch para ajudá-lo a tomar as decisões certas. Você pode usar o CloudWatch agente para configurar métricas personalizadas do Monitor de Desempenho a serem enviadas para coleta de dados. Para obter instruções, consulte [Como posso usar o CloudWatch agente para visualizar métricas do Monitor de Desempenho em um servidor Windows?](#) no Centro de AWS Conhecimento.

Depois de implantar o CloudWatch agente, você pode configurar as seguintes métricas no arquivo de configuração do agente `emmetrics_collected`:

Categoria métrica	Nome da métrica
Banco de dados para instâncias (NTDSA)	% de acertos no cache do banco de dados
Latência média das leituras do banco de dados de E/S	
I/O database reads/sec	
Latência média das gravações em log de E/S	
DirectoryServices (NTDS)	Tempo de vinculação de LDAP
Operações de replicação pendentes do DRA	
Sincronizações de replicação pendentes do DRA	
DNS	Consultas recursivas/segundo
Falhas de consultas recursivas/segundo	
Consultas de TCP recebidas/segundo	
Total de consultas recebidas/segundo	
Total de respostas enviadas/segundo	

Categoria métrica	Nome da métrica
Consultas de UDP recebidas/segundo	
LogicalDisk	Tamanho médio da fila de discos
% de espaço livre	
Memória	% de bytes em uso confirmados
Tempos de vida médios do cache em espera de longo prazo	
Interface de rede	Bytes enviados/segundo
Bytes recebidos/segundo	
Largura de banda atual	
NTDS	Atraso estimado na fila de ATQ
Latência da solicitação de ATQ	
Leituras de diretório DS/segundo	
Pesquisas de diretório DS/segundo	
Gravações em diretório DS/segundo	
Sessões do cliente LDAP	
Pesquisas LDAP/segundo	
Vinculações de LDAP bem-sucedidas/segundo	
Processador	% do tempo do processador
Estatísticas abrangendo todo o sistema de segurança	Autenticações Kerberos
Autenticações NTLM	

Recursos adicionais do

- [Serviços de domínio do Active Directory em AWS: Guia de implantação da solução de parceiro](#) (AWS documentação)
- [Capacity planning for Active Directory Domain Services](#) (documentação da Microsoft)
- [Considerações de design para executar o Active Directory em instâncias do EC2](#) (AWS whitepapers)

AWS Managed Microsoft AD

Visão geral do

AWS Directory Service for Microsoft Active Directory, também conhecido como AWS Managed Microsoft AD, é alimentado por um Active Directory do Windows Server e gerenciado por AWS. Você pode usar AWS Managed Microsoft AD para migrar uma ampla variedade de aplicativos compatíveis com o Active Directory para o. Nuvem AWS AWS Managed Microsoft AD funciona com uma variedade de aplicativos e serviços nativos do Active Directory. Ele também é compatível com [aplicações e serviços gerenciados pela AWS](#). Embora não existam muitas alavancas de otimização de custos AWS Managed Microsoft AD devido ao serviço e seus mecanismos de cobrança, existem alguns princípios de design que podem ajudar você a manter os custos no mínimo.

Impacto do custo

Como AWS Managed Microsoft AD é um serviço gerenciado com base no presente SKUs, o dimensionamento é um processo relativamente simples. Atualmente, existem dois tamanhos SKUs disponíveis: edições Standard e Enterprise. Outros SKUs incluem compartilhamento de diretórios, adição de controladores de domínio adicionais (incluindo regiões adicionais) e transferência de dados entre regiões.

Recomendações de otimização de custos

Há diferenças entre a AWS Managed Microsoft AD Standard Edition e a AWS Managed Microsoft AD Enterprise Edition. A edição Enterprise tem suporte para até 500 mil objetos do Active Directory, 500 compartilhamentos de contas (limite flexível) e tem suporte multirregional. A edição Standard tem suporte para até 30 mil objetos do Active Directory, cinco compartilhamentos de conta (limite flexível de aproximadamente 25 no máximo) e não tem suporte multirregional.

Note

Os limites superiores dos objetos do Active Directory são aproximações. Seu diretório pode oferecer suporte a mais ou menos objetos, dependendo do tamanho deles e do comportamento e das necessidades de performance das suas aplicações.

As questões a serem consideradas antes de selecionar seu tipo de diretório são:

- O suporte multirregional é necessário?
- O diretório será compartilhado com mais de 25 contas?
- A quantidade de objetos do Active Directory será superior a 30 mil?

Se a resposta for sim para qualquer uma das perguntas acima, então a edição Enterprise é necessária. Se a resposta para todas as perguntas for não, recomendamos que você comece com a edição Standard.

Note

Você pode atualizar um diretório da edição Standard para a edição Enterprise, mas um diretório não pode sofrer downgrade. Implantar a edição Standard não é uma decisão sem volta. Se você deseja atualizar seu diretório para a Enterprise Edition, entre em contato com AWS.

Há um custo para cada compartilhamento quando você compartilha diretórios na edição Enterprise do AWS Managed Microsoft AD . Isso é menor do que o custo de implantar um diretório em cada conta, mas lembre-se de que os custos de compartilhamento podem aumentar se não forem verificados. Recomendamos que você compartilhe somente diretórios com contas que contenham Amazon Relational Database Service (Amazon RDS) e FSx Amazon for Windows File Server, já que somente esses serviços oferecem suporte a esse recurso. Lembre-se de que você tem a opção de integrar FSx o Windows File Server ao seu Active Directory autogerenciado, incluindo um AWS Managed Microsoft AD. Se apenas a Amazon FSx for necessária em outra conta, você poderá fazer uma FSx implantação autogerenciada da Amazon na AWS Managed Microsoft AD sem a necessidade de compartilhar o diretório.

Ao decidir quando implantar controladores de domínio adicionais, lembre-se de que o AWS Managed Microsoft AD oferece suporte a apenas duas sub-redes em zonas de disponibilidade separadas na mesma VPC. Adicionar mais controladores de domínio não permite que você adicione mais sub-redes. Para determinar se você deve adicionar outros controladores de domínio devido a problemas de desempenho, revise as [métricas de desempenho do controlador de domínio em CloudWatch](#). Isso informa se um ou todos os controladores de domínio estão sobrecarregados. Se você determinar que apenas um controlador de domínio está sobrecarregado, adicionar outros controladores de domínio não aliviará a carga e você precisará investigar as aplicações que não estão balanceando a carga nos controladores de domínio atualmente disponíveis. Se todos os controladores de domínio estiverem com uso elevado, adicionar um controlador de domínio extra poderá reduzir a carga nos controladores de domínio existentes. Para obter instruções sobre como automatizar a escalabilidade, consulte [Como automatizar a AWS Managed Microsoft AD escalabilidade com base nas métricas de utilização no blog de segurança](#). AWS

Se você estendeu seu diretório para várias regiões, recomendamos que não use os compartilhamentos NETLOGON ou SYSVOL do diretório para armazenamento de arquivos. Todos os controladores de domínio replicam o conteúdo desses compartilhamentos. Não usar os compartilhamentos para armazenamento de arquivos reduz os custos de transferência de dados ao mínimo.

Você também tem a opção de se inscrever em um Contrato Enterprise com AWS. Os contratos empresariais oferecem a você a opção de personalizar os contratos que melhor atendam às suas necessidades. Para obter mais informações, consulte [Enterprise Customers](#).

Recursos adicionais do

- [AWS Managed Microsoft AD cotas](#) (AWS Directory Service documentação)
- [AWS Directory Service Preços](#) (AWS site)
- [Active Directory Domain Services on AWS](#) (AWS Whitepapers)

AD Connector

Visão geral do

O [AD Connector](#) é um serviço de proxy que fornece uma maneira fácil de conectar seu Microsoft Active Directory local existente a [AWS aplicativos](#) compatíveis, como Amazon WorkSpaces, Amazon Quick e união de domínio perfeita para instâncias do Amazon Elastic Compute Cloud (Amazon EC2),

sem armazenar em cache nenhuma informação na nuvem. Você pode usar o AD Connector para adicionar uma conta de serviço ao seu Active Directory. O AD Connector elimina a necessidade de sincronização de diretórios ou o custo e a complexidade da hospedagem de uma infraestrutura de federação. Embora não haja muitas formas de otimizar os custos para o AD Connector devido à natureza do serviço e aos seus mecanismos de faturamento, você pode seguir as recomendações de design nesta seção para reduzir os custos ao mínimo.

Impacto do custo

O AD Connector é um serviço gerenciado baseado em predefinições SKUs. Isso torna o dimensionamento um processo simples. Há dois tamanhos SKUs disponíveis: tamanhos pequenos e grandes. Você pode usar a [AWS Calculadora de Preços](#) para estimativas de custo envolvendo o AD Connector.

Recomendações de otimização de custos

Além dos recursos de computação de backend, não há diferença entre os conectores pequenos e grandes.

As questões a serem consideradas antes de selecionar seu tipo de diretório são:

- Há um grande número (mais de 10.000) de usuários ativos usando AWS aplicativos integrados ao AD Connector?
- O usuário pertence a muitos grupos aninhados, grupos com aninhamento profundo ou grupos com aninhamento circular?

Se a resposta para as duas perguntas for não, recomendamos que você comece com o tamanho pequeno. Se você responder sim a alguma das perguntas acima, talvez seja válido considerar um tamanho grande. Você pode começar com um AD Connector pequeno e, se o diretório ficar deficiente devido à performance, você pode solicitar que ele seja atualizado para um grande.

Note

Você pode atualizar um AD Connector de pequeno para grande, mas um AD Connector não pode sofrer downgrade.

A maioria dos problemas de performance não está relacionada ao AD Connector, mas aos controladores de domínio do Active Directory on-premises que estão sobrecarregados devido ao fato

de muitos usuários pertencerem a muitos grupos aninhados, grupos com aninhamento profundo ou grupos com aninhamento circular.

Você também tem a opção de se inscrever em um Contrato Enterprise com AWS. Os contratos empresariais oferecem a você a opção de personalizar os contratos que melhor atendam às suas necessidades. Para obter mais informações, consulte [Enterprise Customers](#).

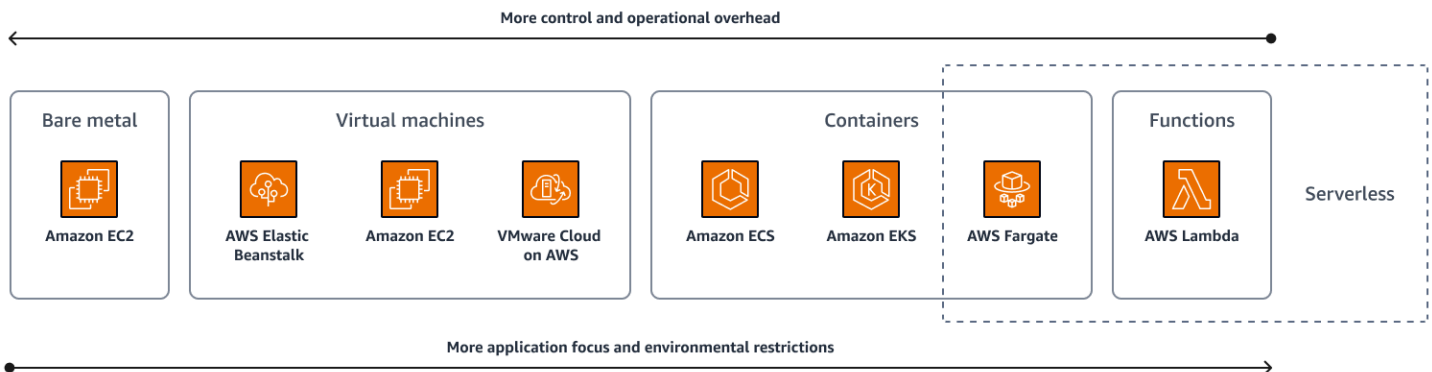
Recursos adicionais do

- [Cotas do AD Connector](#) (AWS Directory Service documentação)
- [Preços de outros tipos de diretório](#) (AWS site)
- [Active Directory Domain Services on AWS](#) (AWS Whitepapers)

.NET

Desenvolver e implantar aplicações .NET é um fator chave para ajudar você a alcançar a escala e a agilidade oferecidas pela computação em nuvem. Para muitos aplicativos .NET legados, a opção computacional mais adequada para executar aplicativos AWS é usar máquinas virtuais, seja por meio do AWS Elastic Beanstalk Amazon Elastic Compute Cloud (Amazon EC2). Também é possível executar aplicações .NET em contêineres Windows e Linux.

A introdução do .NET Core permite que você crie aplicações .NET Modern que aproveitam todas as vantagens da nuvem. Os aplicativos modernos podem usar o conjunto tradicional de opções de computação e também atingir vários tipos de ambientes sem servidor, incluindo ou. AWS Fargate AWS Lambda O .NET 6+ agora oferece hospedagem de alta performance de workloads em instâncias ARM64 do EC2, como as famílias EC2 Graviton2. Isso permite o acesso à última geração de processadores disponíveis no Amazon EC2. Isso significa que suas aplicações podem ser hospedadas em computação especializada para seu tipo de workload, como codificação de vídeo, servidores web e computação de alta performance (HPC).



Esta seção fornece recomendações para ajudar você a adaptar suas aplicações .NET para aproveitar os benefícios da nuvem com foco na eficiência de custos.

Esta seção abrange os seguintes tópicos:

- [Refatorar para o .NET Modern e migrar para o Linux](#)
- [Conteinerizar aplicações .NET](#)
- [Usar instâncias e contêineres do Graviton](#)
- [Suporte à escalabilidade dinâmica para aplicações .NET Framework estáticas](#)
- [Use o armazenamento em cache para reduzir a demanda do banco de dados](#)

- [Considerar o .NET sem servidor](#)
- [Considerar bancos de dados com propósito específico](#)

Refatorar para o .NET Modern e migrar para o Linux

Visão geral do

A modernização de aplicações .NET Framework legadas pode ajudar a melhorar a segurança, a performance e a escalabilidade. Uma forma eficaz de modernizar as aplicações .NET Framework é migrá-las para uma versão do .NET Modern (6+). Confira algumas das principais vantagens de migrar essas aplicações para o .NET de código aberto:

- Para reduzir os custos de licenciamento do Windows executando-os em um sistema operacional Linux
- Aproveitar a disponibilidade de linguagens modernas
- Obtenha performance otimizada para execução no Linux

Muitas organizações ainda estão executando versões mais antigas do .NET Framework. Isso pode representar riscos de segurança, já que as vulnerabilidades nas versões mais antigas não são mais tratadas pela Microsoft. A Microsoft encerrou o suporte para as versões recentes do .NET Framework 4.5.2, 4.6 e 4.6.1. É muito importante avaliar os riscos e as vantagens de continuar executando as versões mais antigas do framework. Para reduzir o risco e os custos, pode valer a pena investir tempo e esforço na refatoração para uma versão moderna do .NET.

Impacto do custo

Considere um tipo de instância do EC2 de uso geral (m5), que oferece um equilíbrio entre recursos de computação, memória e rede. Essas instâncias são adequadas para uma variedade de aplicações, como servidores web, bancos de dados de médio porte e repositórios de código-fonte.

Por exemplo, uma instância m5.xlarge sob demanda com 4 V CPUs e 16 GB de memória no Windows Server (licença incluída) no Leste dos EUA (Norte da Virgínia) custa 274,48 USD mensais. Os mesmos recursos em um servidor Linux custam USD 140,16 por mês. Neste exemplo, há uma redução de 49% no custo quando você migra sua aplicação do .NET Framework para uma versão moderna do .NET e executa a aplicação em um servidor Linux. Seu custo pode variar dependendo das opções (por exemplo, tipo de instância, sistema operacional, armazenamento) que você escolher ao selecionar uma [instância do EC2](#). Você pode otimizar ainda mais os custos usando [Savings](#)

[Plans](#) ou [Instâncias Reservadas](#). Para obter mais detalhes, use a [AWS Calculadora de Preços](#) para executar estimativas de custo. Para instâncias incluídas no Windows, o custo da licença é de [USD 0,046 por vCPU por hora](#), independentemente do modelo de preços.

A portabilidade dessas aplicações .NET Framework para o .NET Modern exige esforço do desenvolvedor. Você deve avaliar seus aplicativos e suas dependências para ver se eles são compatíveis com a versão da plataforma de destino. [AWS Assistente de Portabilidade para .NET](#) é uma ferramenta auxiliar que verifica aplicações .NET Framework e gera uma avaliação de compatibilidade do .NET, ajudando você a portar suas aplicações para serem compatíveis com o Linux mais rapidamente. O Assistente de Portabilidade para .NET identifica incompatibilidades com o .NET, encontra substitutos conhecidos e gera uma avaliação detalhada da compatibilidade. Depois de portar sua solução, você deve fazer alterações manuais no código para que seu projeto seja compilado com êxito com dependências. Isso reduz o esforço manual envolvido na modernização das suas aplicações para Linux. Se sua aplicação for compatível com processadores ARM, migrar para o Linux vai permitir usar instâncias do Graviton. Isso pode ajudar você a obter mais 20% em reduções adicionais de custos. Para obter mais informações, consulte [Powering .NET 5 com AWS Graviton2: Benchmarks](#) in the Compute Blog. AWS

Há outras ferramentas, como o [Kit de ferramentas da AWS para refatoração do .NET](#) e o [Assistente de Portabilidade para .NET](#), que podem ajudar a portar aplicações .NET Framework legadas para o .NET Modern.

Recomendações de otimização de custos

Para migrar aplicações .NET Framework, faça o seguinte:

1. Pré-requisitos: para usar o Assistente de Portabilidade para .NET, você deve instalar o .NET 5+ na máquina em que planeja analisar o código-fonte da aplicação. Os recursos na máquina devem ter no mínimo 1,8 velocidade de GHz processamento, 4 GB de memória e 5 Gb de espaço de armazenamento. Para obter mais informações, consulte [Prerequisites](#) na documentação do Assistente de Portabilidade para .NET.
2. Avaliação: baixe o Assistente de Portabilidade para .NET como um arquivo [executável](#) (download). Você pode baixar e instalar a ferramenta em sua máquina para iniciar a avaliação de suas aplicações. A página de avaliação contém projetos e pacotes portados APIs que são incompatíveis com o .NET moderno. Por esse motivo, você recebe erros de criação na solução após a avaliação. Você pode visualizar ou baixar as decobertas da avaliação em um arquivo CSV. Para obter mais informações, consulte [Port a solution](#) na documentação do Assistente de Portabilidade para .NET.

3. Refatoração: depois de avaliar a aplicação, você pode portar seus projetos para a versão do framework de destino. Ao portar uma solução, seus arquivos de projeto e parte do código serão modificados pelo Assistente de Portabilidade. Você pode verificar os logs para analisar as alterações em seu código-fonte. Na maioria dos casos, o código exigirá um esforço adicional para concluir a migração e os testes para deixá-lo pronto para produção. Dependendo da aplicação, algumas das alterações podem incluir o framework de entidades, a identidade e a autenticação. Para obter mais informações, consulte [Port a solution](#) na documentação do Assistente de Portabilidade para .NET.

Esta é a primeira etapa para modernizar suas aplicações em contêineres. Pode haver vários fatores comerciais e técnicos para modernizar suas aplicações .NET Framework para contêineres Linux. Um dos fatores mais importantes é reduzir o custo total de propriedade ao mudar de um sistema operacional Windows para o Linux. Isso reduz os custos de licenciamento ao migrar sua aplicação para uma versão multiplataforma do .NET e para contêineres para otimizar a utilização dos recursos.

Depois que sua aplicação for portada para o Linux, você poderá usar o [AWS App2Container](#) para containerizar sua aplicação. O App2Container usa o Amazon ECS ou o Amazon EKS como serviços de endpoint em que você pode implantar diretamente. O App2Container fornece todos os artefatos de implantação de infraestrutura como código (IaC) necessários para containerizar suas aplicações repetidamente.

Considerações e recursos adicionais

- Se você tiver aplicativos criados no VB.NET (uma estrutura legada de 2002) e quiser portá-los para o .NET 6, consulte a postagem Portar [aplicativos VB.NET legados para o .NET 6.0 com o Porting Assistant para .NET no blog](#) Microsoft Workloads on. AWS
- Se você tiver aplicações legadas no Windows Communication Foundation (WCF) e quiser executá-las no .NET Modern, você pode adotar o CoreWCF. Para obter mais informações, consulte a postagem [Modernizando aplicativos WCF legados para CoreWCF usando o Porting Assistant para .NET](#) no blog Microsoft Workloads on. AWS
- Você pode adicionar o assistente de portabilidade como uma extensão ao seu IDE do Visual Studio. Isso permite que você execute todas as tarefas necessárias para converter seu código sem precisar alternar entre o IDE e a ferramenta Assistente de Portabilidade para .NET. Para obter mais informações, consulte a postagem [Acelere a modernização do aplicativo .NET com o Porting Assistant for .NET Visual Studio IDE](#) no blog Microsoft Workloads on AWS .

- [O Assistente de Portabilidade para .NET da AWS agora é uma ferramenta de código aberto](#) com o código-fonte e os componentes de análise de compatibilidade da avaliação. Isso pode incentivar seus desenvolvedores a usar e compartilhar o conhecimento e as práticas recomendadas de portabilidade do .NET.
- Você pode portar aplicativos do .NET framework para o .NET moderno no Linux usando o AWS Toolkit for .NET Refactoring. Para obter mais informações, consulte a postagem [Acelere a modernização do .NET com o AWS Toolkit for .NET Refactoring no blog](#) Microsoft Workloads on AWS
- Você pode [acelerar a containerização e a migração de aplicativos ASP.NET Core](#) para o uso AWS App2Container

Containerizar aplicações .NET

Visão geral do

Os contêineres são uma forma leve e eficiente de empacotar e implantar aplicações de maneira consistente e reproduzível. Esta seção explica como você pode usar o AWS Fargate, um serviço de contêiner sem servidor para reduzir os custos de suas aplicações .NET e, ao mesmo tempo, fornecer uma infraestrutura escalável e confiável.

Impacto do custo

Alguns fatores que influenciam a eficácia do uso de contêineres para a redução de custos incluem o tamanho e a complexidade da aplicação, o número de aplicações que precisam ser implantadas e o nível de tráfego e demanda das aplicações. Para aplicações pequenas ou simples, os contêineres podem não oferecer reduções de custo significativas em comparação com as abordagens tradicionais de infraestrutura, pois a sobrecarga do gerenciamento dos contêineres e dos serviços associados pode, na verdade, aumentar os custos. No entanto, para aplicações maiores ou mais complexas, o uso de contêineres pode proporcionar uma redução de custos, melhorando a utilização de recursos e reduzindo o número de instâncias necessárias.

Recomendamos que você tenha em mente o seguinte ao usar contêineres para redução de custos:

- Tamanho e complexidade da aplicação: aplicações maiores e mais complexas são mais adequadas para a containerização porque tendem a exigir mais recursos e podem se beneficiar mais com a melhor utilização deles.

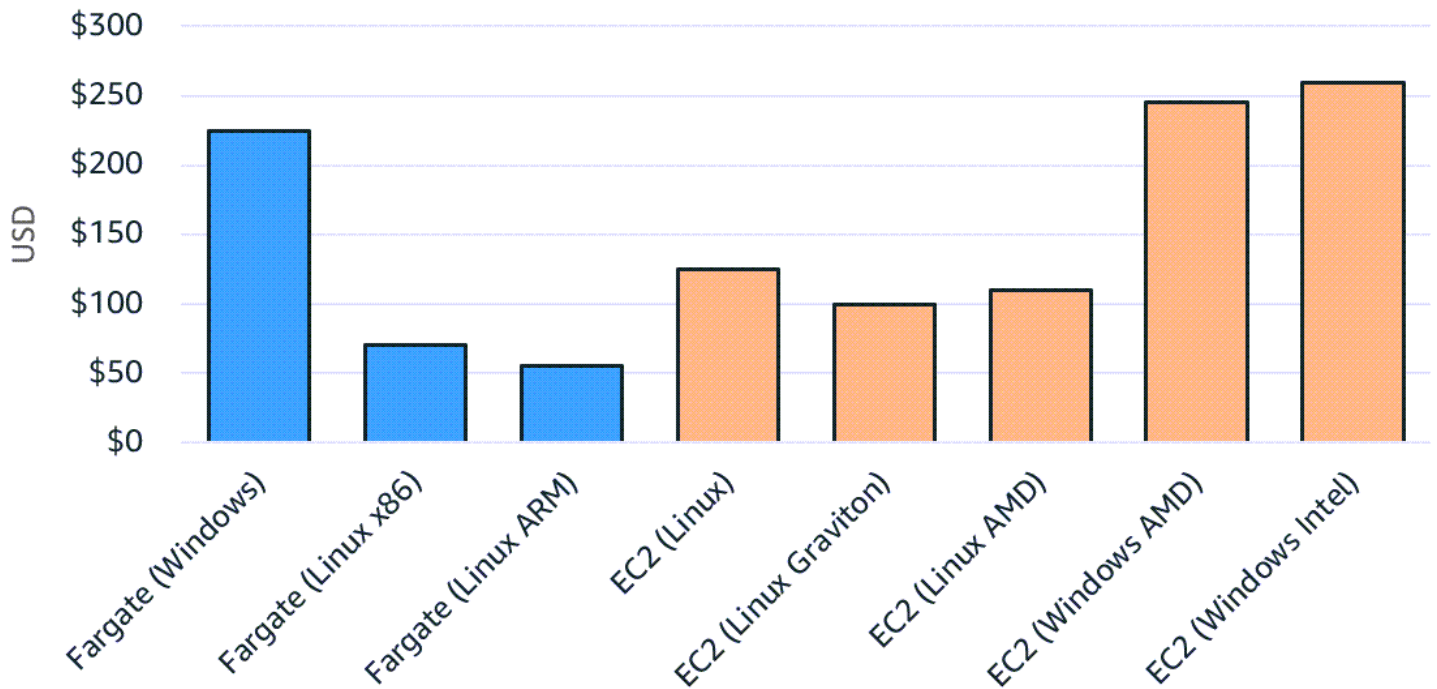
- Número de aplicações: quanto mais aplicações sua organização precisar implantar, maior será a redução de custos por meio da containerização.
- Tráfego e demanda: as aplicações que experimentam alto tráfego e demanda podem se beneficiar da escalabilidade e elasticidade que os contêineres oferecem. Isso pode resultar em uma redução de custos.

Arquiteturas e sistemas operacionais diferentes afetam os custos dos contêineres. Se você estiver usando contêineres Windows, os custos podem não diminuir devido a considerações de licenciamento. Os custos de licenciamento são menores ou não existem com os contêineres Linux. O gráfico a seguir usa uma configuração básica AWS Fargate na região Leste dos EUA (Ohio) com as seguintes configurações: 30 tarefas por mês, cada uma sendo executada por 12 horas com 4 V CPUs e 8 GB de memória alocada.

Você pode escolher entre duas plataformas de computação principais para executar seus contêineres AWS: [hosts de contêineres baseados em EC2 e](#) sem servidor ou. [AWS Fargate](#) Se você usar o Amazon Elastic Container Service (Amazon ECS) em vez do Fargate, mantenha a computação em execução (instâncias) para permitir que o mecanismo de posicionamento instancie contêineres quando necessário. Se você usar o Fargate, somente a capacidade computacional necessária será provisionada.

O gráfico a seguir mostra a diferença entre contêineres equivalentes usando o Fargate e o Amazon EC2. Devido à flexibilidade do Fargate, as tarefas de uma aplicação podem ser executadas 12 horas por dia, sem utilização fora do horário de pico. No entanto, para o Amazon ECS, você deve controlar a capacidade computacional usando um [grupo do Auto Scaling](#) de instâncias do EC2. Isso pode fazer com que a capacidade opere ininterruptamente, 24 horas por dia, o que pode, em última análise, aumentar os custos.

Monthly costs of Fargate and Amazon EC2



Recomendações de otimização de custos

Use contêineres Linux em vez de Windows

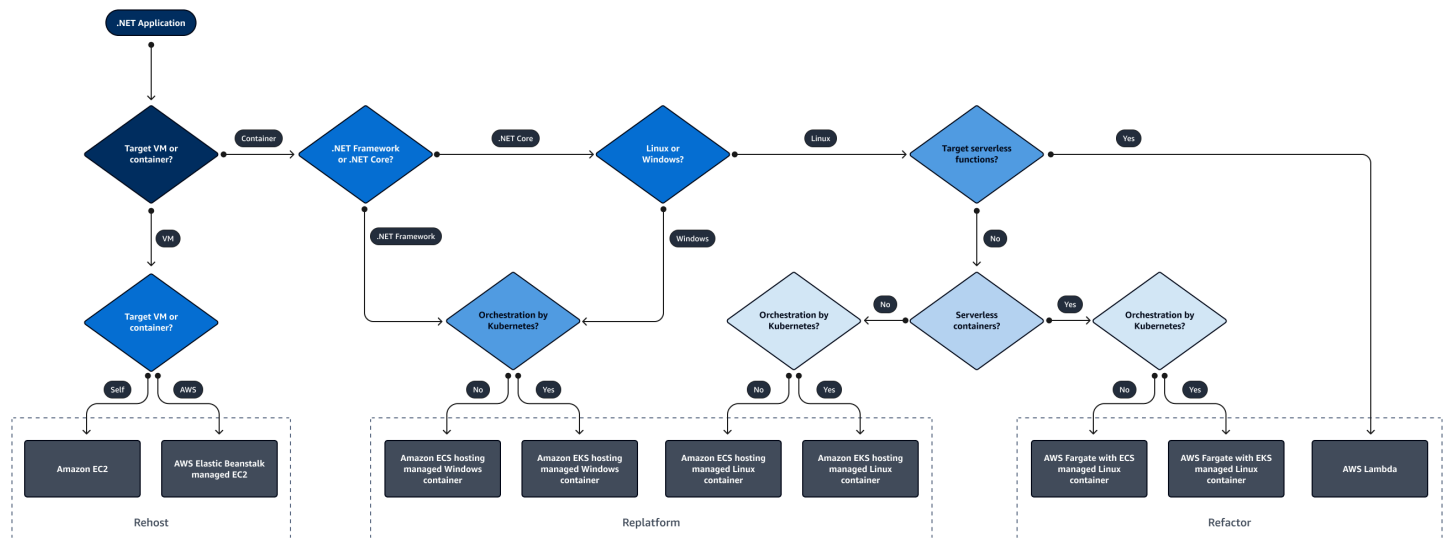
Você pode obter economias significativas se usar contêineres Linux em vez de contêineres Windows. Por exemplo, você pode obter uma economia de aproximadamente 45% nos custos de computação se executar o .NET Core em instâncias EC2 Linux em vez de executar o .NET Framework em instâncias EC2 Windows. Você pode obter uma economia adicional de 40% se usar a arquitetura ARM (AWS Graviton) em vez da x86.

Se você planeja executar contêineres baseados em Linux para aplicações .NET Framework existentes, você deve portar essas aplicações para versões modernas e multiplataforma do .NET ([como o .NET 6.0](#)) para usar contêineres Linux. Uma consideração importante é ponderar o custo da refatoração em comparação com a economia de custos obtida com a redução do custo dos contêineres Linux. Para obter mais informações sobre como portar suas aplicações para o .NET Modern, consulte [Assistente de Portabilidade para .NET](#) na documentação da AWS .

Outro benefício de migrar para o .NET Modern (ou seja, sair do .NET Framework) é que oportunidades adicionais de modernização ficam disponíveis. Por exemplo, você pode considerar a

redefinição da arquitetura de sua aplicação para uma arquitetura baseada em microsserviços que seja mais escalável, ágil e econômica.

O diagrama a seguir ilustra o processo de tomada de decisão para explorar oportunidades de modernização.



Aproveitar os Savings Plans

Os contêineres podem ajudar você a aproveitar os [Savings Plans para computação](#) para reduzir seus custos com o Fargate. O modelo de desconto flexível oferece os mesmos descontos que as Instâncias Reservadas Conversíveis. Os preços do Fargate são baseados nos recursos de vCPU e de memória usados desde o início do download da imagem do contêiner até o término da tarefa do Amazon ECS (arredondado para o segundo mais próximo). Os [Savings Plans do Fargate](#) oferecem economias de até 50% no uso do Fargate em troca do compromisso de utilizar uma quantidade específica de uso de computação (medida em dólares por hora) por um período de um ou três anos. Você pode usar o [AWS Cost Explorer](#) para ajudar você a escolher um dos Savings Plans.

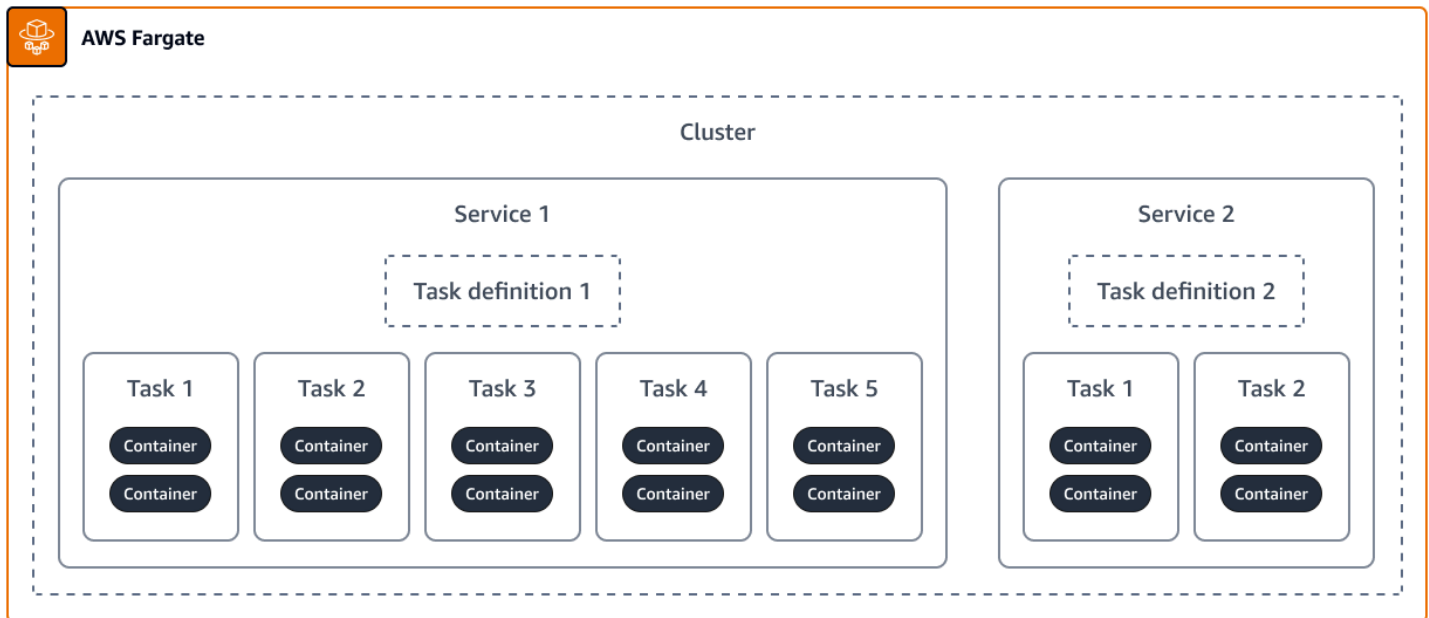
É importante entender que os Savings Plans para computação são aplicados ao uso que gera a maior economia em primeiro lugar. Por exemplo, se você estiver executando uma instância t3.medium do Linux em us-east-2 e uma instância t3.medium idêntica do Windows, a instância do Linux receberá primeiro o benefício dos Savings Plans. Isso ocorre porque a instância do Linux tem um potencial de economia de 50%, enquanto a mesma instância do Windows tem um potencial de economia de 35%. Se você tiver outros recursos elegíveis do Savings Plans em execução no seu Conta da AWS, como Amazon EC2 ou Lambda, não é necessário que seus Savings Plans sejam aplicados primeiro ao Fargate. Para obter mais informações, consulte [Entendendo como os Savings](#)

[Plans se aplicam ao seu AWS uso](#) na documentação do Savings Plans e na seção [Otimizar gastos para Windows no Amazon EC2](#) deste guia.

Tarefas do Fargate do tamanho certo

É importante garantir que as tarefas do Fargate sejam dimensionadas corretamente para alcançar o grau máximo de otimização de custos. Frequentemente, os desenvolvedores não têm todas as informações de uso necessárias ao determinar inicialmente as configurações das tarefas do Fargate usadas em suas aplicações. Isso pode levar ao provisionamento excessivo de tarefas e, em seguida, resultar em gastos desnecessários. Para evitar isso, recomendamos que você carregue as aplicações de teste em execução no Fargate para entender a performance de uma configuração de tarefa específica em diferentes cenários de uso. Você pode usar os resultados do teste de carga, a vCPU, a alocação de memória das tarefas e as políticas de ajuste de escala automático para encontrar o equilíbrio certo entre performance e custo.

O diagrama a seguir mostra como o Compute Optimizer gera recomendações para o tamanho ideal da tarefa e do contêiner.



Uma abordagem é usar uma ferramenta de teste de carga, como a descrita em [Distributed Load Testing on AWS](#), para estabelecer uma linha de base para a utilização da vCPU e da memória. Depois de executar o teste de carga para simular uma carga típica da aplicação, você pode ajustar a configuração da vCPU e da memória para a tarefa até que a utilização básica seja alcançada.

Recursos adicionais do

- [Cost Optimization Checklist for Amazon ECS and AWS Fargate](#) (publicação do blog AWS Containers)
- [Otimização teórica de custos pelo tipo de lançamento do Amazon ECS: Fargate vs EC2 AWS](#) (publicação no blog de contêineres)
- [Assistente de portabilidade para.NET](#) (AWS documentação)
- [Teste de carga distribuída em AWS](#) (Biblioteca de AWS soluções)
- [AWS Compute Optimizer lança suporte para serviços Amazon ECS em AWS Fargate\(publicação no blog AWS Cloud Financial Management\)](#)

Usar instâncias e contêineres do Graviton

Visão geral do

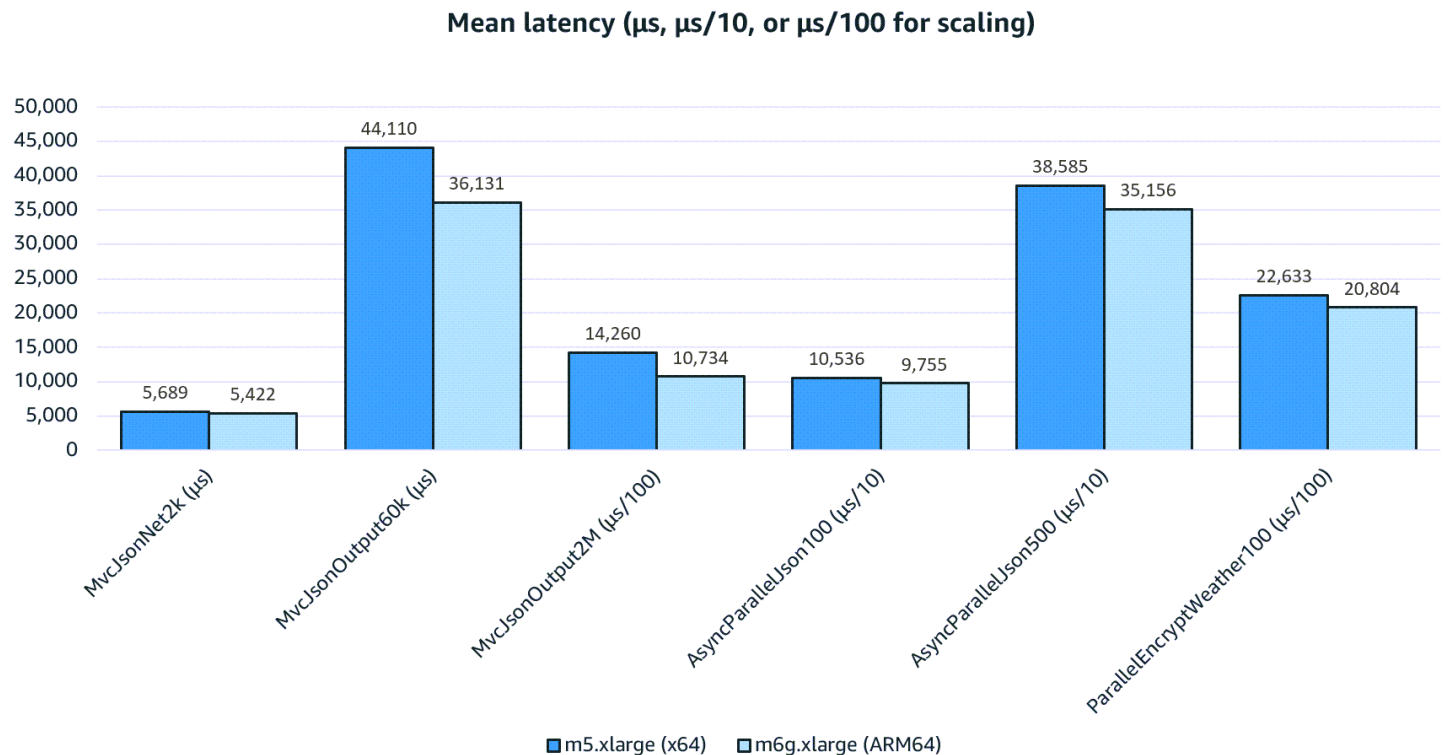
AWS As instâncias Graviton são alimentadas por processadores ARM projetados AWS para oferecer a melhor relação preço/desempenho para suas cargas de trabalho em nuvem executadas no Amazon Elastic Compute Cloud (Amazon EC2), incluindo contêineres em execução. AWS Atualmente, há três gerações do Graviton disponíveis para uso no Amazon EC2. Este guia se concentra no uso do Graviton 2 e 3 com aplicações .NET porque há uma economia significativa quando você usa as versões mais recentes do Graviton. Lembre-se de que instâncias do Graviton só executam o sistema operacional Linux. Como resultado, as instâncias do Graviton são uma boa oferta para o .NET em execução no Linux, mas não são uma opção para o sistema operacional Windows ou aplicações NET Framework legadas.

O Graviton 3 é 60% mais eficiente do que as instâncias comparáveis do EC2, com performance até 40% melhor. Este guia se concentra nos benefícios de custo do uso do Graviton, mas é importante observar que o Graviton oferece os benefícios adicionais de melhorias de performance e maior sustentabilidade ambiental.

Impacto do custo

Você pode obter até 45% de economia ao mudar para o Graviton. Depois de refatorar qualquer aplicação .NET Framework legada para uma versão do .NET Modern, você habilita a capacidade de usar instâncias do Graviton. Mudar para o Graviton é uma técnica eficaz de otimização de custos para desenvolvedores .NET.

O exemplo na tabela a seguir mostra o potencial de melhorias de performance que você pode obter ao migrar para instâncias do Graviton.



Para obter uma análise e uma explicação completas da abordagem de benchmarking usada para criar os resultados no diagrama anterior, consulte [Powering .NET 5 with AWS Graviton2: Benchmarks](#) in the Compute Blog. AWS

Uma das razões para a eficiência aprimorada é a diferença no significado de vCPU entre a x86 e o Graviton. Na arquitetura x86, uma vCPU é um núcleo lógico obtido pelo hyperthreading. No Graviton, vCPU equivale a um núcleo físico que permite que a vCPU esteja totalmente comprometida com a workload.

O resultado com o Graviton2 é uma relação preço/performance 40% melhor do que instâncias x86/x64 comparáveis. O Graviton3 oferece as seguintes vantagens em relação ao Graviton2:

- Um perfil de performance aprimorado com performance até 25% melhor
- Performance de ponto flutuante até duas vezes maior
- Performance da workload criptográfica até duas vezes mais rápida
- Performance de machine learning até três vezes melhor

Além disso, o Graviton3 é a primeira instância na nuvem a apresentar DDR5 memória.

As tabelas a seguir mostram a diferença na redução de custos entre instâncias baseadas em Graviton e instâncias equivalentes baseadas em x86.

Esta tabela mostra uma economia do Graviton de 19,20%.

Tipo de instância	Arquitetura	vCPU	Memória (GB)	Custo por hora (sob demanda)
t4g.xlarge	Arm	4	16	\$0,134
t3.xlarge	x86	4	16	0,164 US\$

Esta tabela mostra uma economia do Graviton de 14,99%.

Tipo de instância	Arquitetura	vCPU	Memória (GB)	Custo por hora (sob demanda)
c7g.4xlarge	Arm	16	32	\$0,5781
c6i.4xlarge	x86	16	32	\$0.6800

É importante testar o perfil de performance da sua aplicação ao considerar o Graviton. O Graviton não substitui práticas sólidas de desenvolvimento de software. Você pode usar os testes para verificar se está aproveitando ao máximo seus recursos computacionais subjacentes.

Recomendações de otimização de custos

Há várias maneiras de aproveitar as instâncias/processadores Graviton. Esta seção apresenta as mudanças necessárias para passar do uso de uma máquina com arquitetura x86 para instâncias do Graviton (ARM).

Alterar a configuração de runtime no Lambda

Recomendamos que você ative as configurações de tempo de execução AWS Lambda. Para obter mais informações, consulte [Modificar ambiente de runtime](#) na documentação do Lambda.

Como o .NET é uma linguagem compilada, você deve seguir um processo de criação para que isso funcione. Para obter um exemplo de como fazer isso, consulte [.NET em Graviton](#) em GitHub.

Contêineres

Para uma workload containerizada, crie uma imagem de contêiner com várias arquiteturas. Você pode fazer isso especificando várias arquiteturas no comando de criação do Docker. Por exemplo:

```
docker buildx build -t "myImageName:latest" --platform linux/amd64,linux/arm64 --push .
```

Você também pode usar uma ferramenta AWS Cloud Development Kit (AWS CDK) para ajudar a [orquestrar a construção](#). Para obter exemplos do Docker, consulte [Building Multi-Arch Images for Arm and x86 with Docker Desktops](#) na documentação do Docker.

Amazon EC2

Para migrar de x86/x64 para o ARM, direcione a arquitetura ARM na etapa de compilação. No Visual Studio, você pode criar uma ARM64 CPU. Para obter instruções, consulte [To configure a project to target Arm64 and other platforms](#) na documentação da Microsoft.

Se você estiver usando a CLI do .NET, a execução da criação em uma máquina ARM produz uma criação compatível com o Graviton. Para ver uma demonstração, assista ao [desempenho do Accelerate .NET 6 com o Arm64 no AWS Graviton2 ligado](#). Problemas de dependência resultarão em erros de tempo de compilação que podem ser resolvidos individualmente. Desde que existam bibliotecas ARM para qualquer dependência, a transição deve ser relativamente simples.

Recursos adicionais do

- [Como criar seus contêineres para ARM e economizar com instâncias Graviton e Spot no Amazon ECS \(blog\)AWS](#)
- [AWS Lambda Funções alimentadas pelo processador AWS Graviton2 — Execute suas funções no Arm e obtenha um preço/desempenho até 34% melhor \(blog\)AWS](#)
- [Migrando AWS Lambda funções para processadores AWS Graviton2 baseados em ARM \(blog\)AWS](#)
- [Crie e implante aplicativos web.NET em clusters Amazon ECS AWS Graviton 2 baseados em ARM usando \(blog\) AWS CDKAWS](#)
- [Graviton Fast Start — Um novo programa para ajudar a transferir suas cargas de trabalho para o AWS Graviton \(blog\)AWS](#)

- [Alimentando o.NET 5 com AWS Graviton2: Benchmarks](#) (blog)AWS

Suporte à escalabilidade dinâmica para aplicações .NET Framework estáticas

Visão geral do

Uma das principais vantagens de usar a nuvem para aplicações é a elasticidade, ou seja, a capacidade de aumentar ou reduzir horizontalmente a computação com base na demanda. Isso permite que você pague apenas pela capacidade computacional de que precisa, em vez de provisionar para o uso de pico. A Cyber Monday, em que os varejistas on-line podem obter rapidamente muito mais tráfego do que o normal (por exemplo, [milhares de pontos percentuais em questão de minutos](#)), é um bom exemplo de elasticidade.

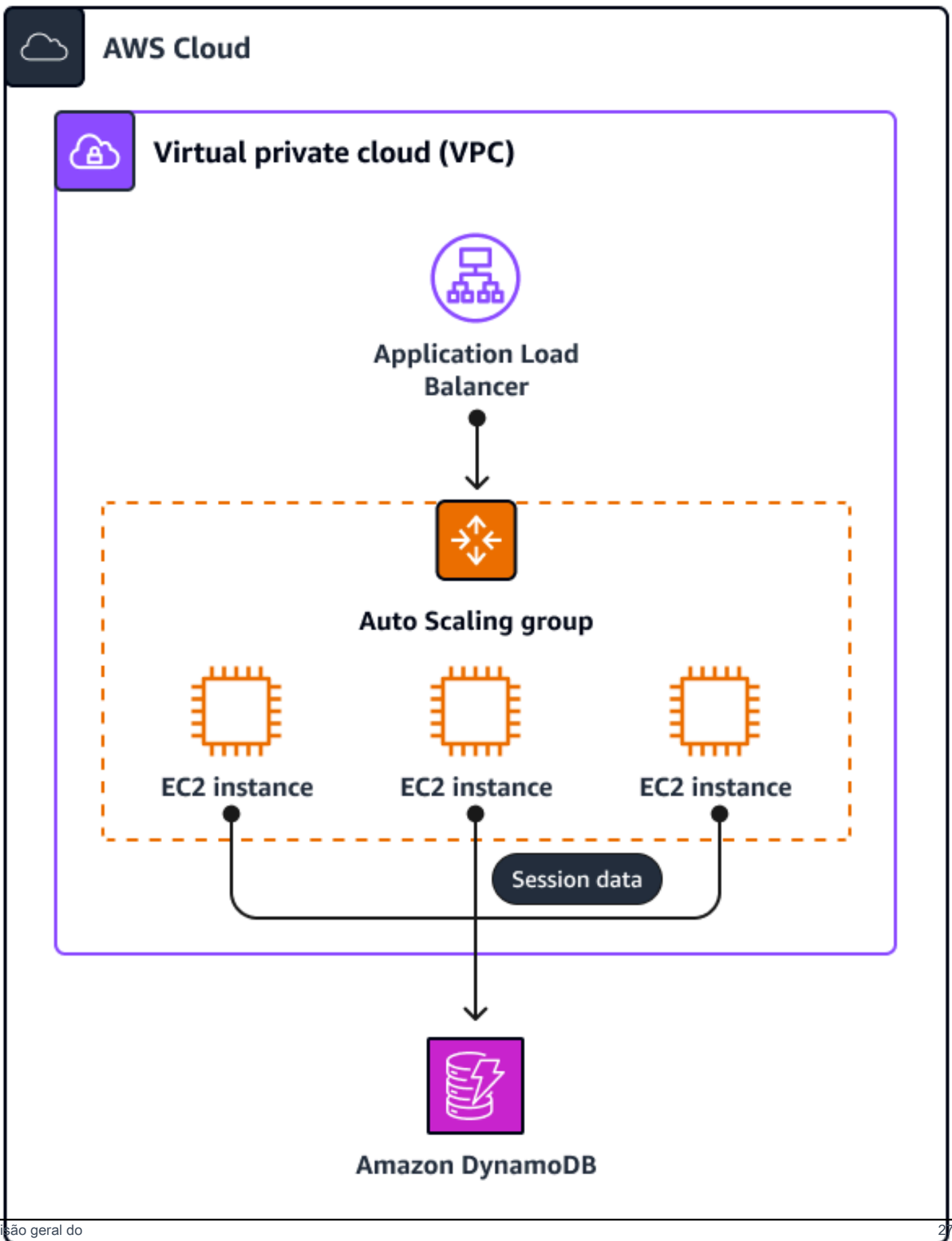
Se você estiver trazendo aplicações web .NET legadas para a nuvem (por exemplo, aplicações ASP.NET Framework executadas no IIS), a capacidade de escalar rapidamente os parques de servidores com balanceamento de carga pode ser difícil ou impossível devido à natureza estável da aplicação. Os dados da sessão do usuário são armazenados na memória da aplicação, geralmente com o [estado da sessão do ASP.NET](#) ou variáveis estáticas que contêm dados de solicitações cruzadas que devem ser persistidos. A afinidade da sessão do usuário geralmente é mantida por meio de sessões persistentes do balanceador de carga.

Isso se prova um desafio operacional. Quando é necessário aumentar a capacidade, você deve provisionar e adicionar servidores intencionalmente. Isso pode ser um processo lento. Colocar os nós fora de serviço em caso de aplicação de patches ou falhas inesperadas pode ser problemático para a experiência do usuário final, perdendo o estado de todos os usuários associados aos nós afetados. Na melhor das hipóteses, os usuários precisarão fazer login novamente.

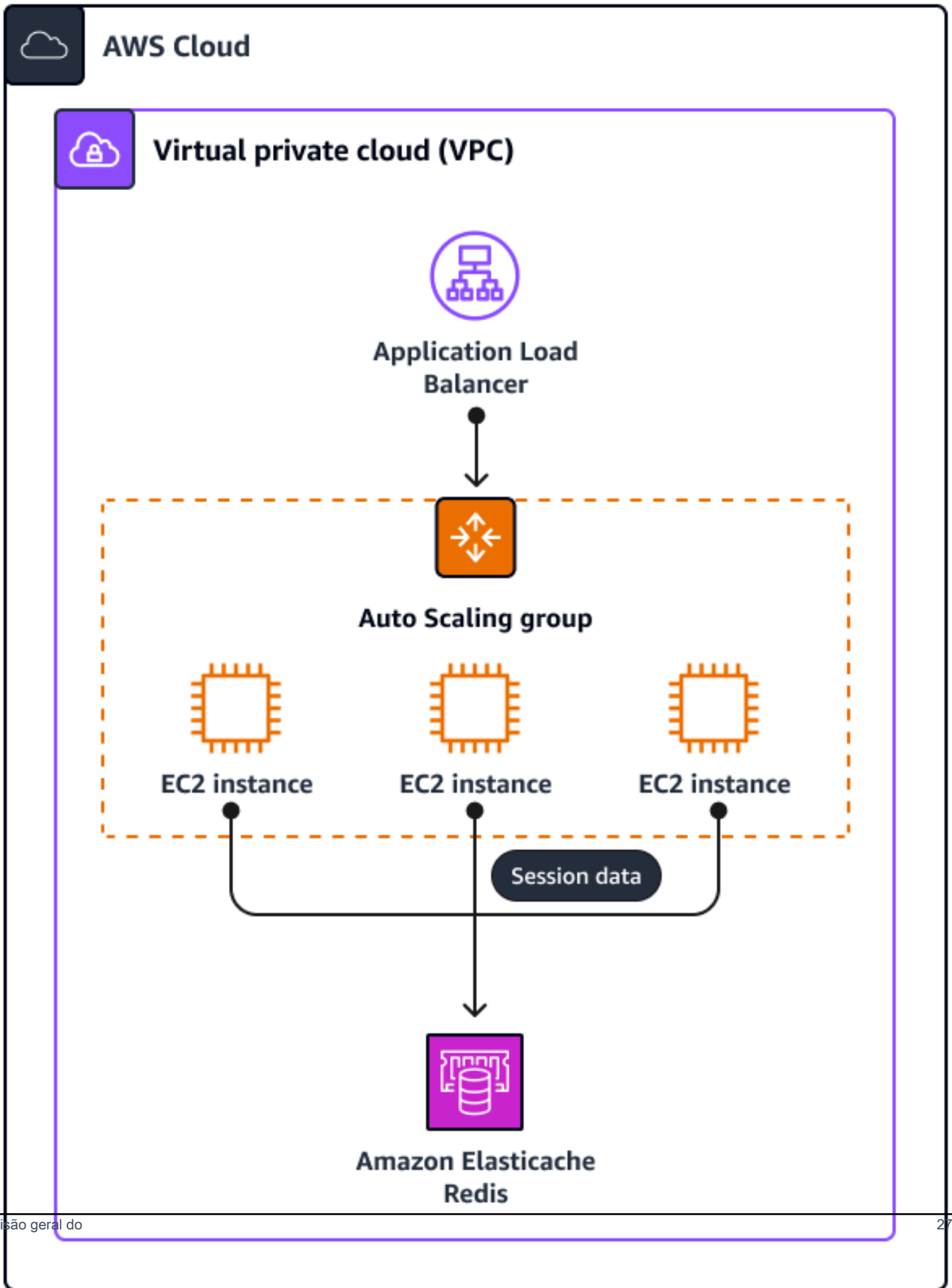
Ao centralizar o estado da sessão para aplicações ASP.NET e aplicar regras de escalonamento automático às aplicações ASP.NET legadas, você pode aproveitar a elasticidade da nuvem e, possivelmente, aproveitar a redução de custos ao executar aplicações. Por exemplo, você obtém reduções de custo por meio da escalabilidade computacional, mas também pode escolher entre os diferentes modelos de preços disponíveis, como reduzir o [uso de instâncias reservadas](#) e usar os [preços das instâncias spot da Amazon](#).

Duas técnicas comuns incluem o uso do [Amazon DynamoDB como provedor de estado de sessão](#) e o uso do [ElastiCache Amazon \(Redis OSS\)](#) como armazenamento de sessões do ASP.NET.

O diagrama a seguir mostra uma arquitetura que usa o DynamoDB como provedor de estado de sessão.



O diagrama a seguir mostra uma arquitetura que usa ElastiCache (Redis OSS) como provedor de estado de sessão.



Impacto do custo

Para determinar as vantagens do dimensionamento para uma aplicação de produção, recomendamos que você modele a demanda real. Esta seção faz as seguintes suposições para modelar uma aplicação de exemplo:

- As instâncias adicionadas e removidas da rotação são idênticas, e nenhuma variação no tamanho da instância é introduzida.
- A utilização do servidor nunca cai abaixo de dois servidores ativos a fim de manter a alta disponibilidade da aplicação.
- A quantidade de servidores é escalada linearmente com o tráfego (ou seja, o dobro do tráfego exigirá o dobro da computação).
- O tráfego é modelado ao longo de um mês em incrementos de seis horas, com variação intradiária e um pico anormal de tráfego (por exemplo, uma venda promocional) em um dia de tráfego dez vezes maior. O tráfego de fim de semana é modelado com base na utilização básica.
- O tráfego noturno é modelado com base na utilização básica, enquanto o tráfego durante a semana é modelado com uma utilização 4 vezes maior.
- O preço de instância reservada usa o modelo de um ano, sem pagamento antecipado. O preço normal durante o dia usa o modelo sob demanda, enquanto a demanda de expansão usa o preço de instância spot.

O diagrama a seguir ilustra como esse modelo aproveita a elasticidade em uma aplicação .NET em vez do provisionamento para uso de pico. Isso resulta em uma economia de aproximadamente 68%.

Comparison of cumulative costs for peak provisioning and autoscaling



Se você usa o DynamoDB como um mecanismo de armazenamento do estado da sessão, use os seguintes parâmetros:

Storage: 20GB
 Session Reads: 40 million
 Session Writes: 20 million
 Pricing Model: On demand

O custo mensal estimado desse serviço é de aproximadamente USD 35,00 por mês.

Se você usa ElastiCache (Redis OSS) como um mecanismo de armazenamento do estado da sessão, use os seguintes parâmetros:

Number of Nodes: 3
 Node size: cache.t4g.medium
 Pricing Model: 1y reserved

O custo mensal estimado desse serviço é de aproximadamente USD 91,00 por mês.

Recomendações de otimização de custos

A primeira etapa é implementar o estado da sessão em uma aplicação .NET legada. Se você estiver usando ElastiCache como seu mecanismo de armazenamento de estado, siga as orientações de [ElastiCache como um repositório de sessões do ASP.NET no blog](#) de ferramentas para AWS

desenvolvedores. Se você estiver usando o DynamoDB, siga as orientações [de O que é AWS SDK para .NET o na documentação](#). SDK para .NET

Se o aplicativo usar a InProcessão para começar, certifique-se de que todos os objetos que você planeja armazenar na sessão possam ser serializados. Para fazer isso, use o atributo `SerializableAttribute` para marcar classes cujas instâncias serão armazenadas na sessão. Por exemplo:

```
[Serializable()]
public class TestSimpleObject {
    public string SessionProperty {get;set;}
}
```

Além disso, o `.NET MachineKey` deve ser o mesmo entre todos os servidores em uso. Normalmente é o caso quando instâncias são criadas de uma Imagem de máquina da Amazon (AMI) comum. Por exemplo:

```
<machineKey
    validationKey="some long hashed value"
    decryptionKey="another long hashed value"
    validation="SHA1"/>
```

No entanto, é importante garantir que, se uma imagem base for alterada, ela será configurada com a mesma imagem de máquina `.NET` (configurável no IIS ou no nível do servidor). Para obter mais informações, consulte [SystemWebSectionGroup. MachineKey Propriedade](#) na documentação da Microsoft.

Finalmente, você deve determinar o mecanismo para adicionar servidores a um grupo do Auto Scaling em resposta a um evento de escalabilidade. Existem diversas maneiras de realizar isso. Recomendamos os seguintes métodos para implantar facilmente aplicações `.NET Framework` em uma instância do EC2 em um grupo do Auto Scaling:

- Use o [EC2 Image Builder](#) para configurar uma AMI que contenha o servidor e a aplicação totalmente configurados. Em seguida, você pode usar essa AMI para configurar o [modelo de inicialização do seu grupo do Auto Scaling](#).
- Use [AWS CodeDeploy](#) para implantar seu aplicativo. CodeDeploy permite a integração diretamente com o [Amazon EC2 Auto Scaling](#). Isso fornece uma alternativa à criação de uma nova AMI para cada versão da aplicação.

Recursos adicionais do

- [Create images with EC2 Image Builder](#) (documentação do EC2 Image Builder)
- [Implantando aplicativos Web.NET usando o AWS CodeDeploy Visual Studio Team Services](#) (blog de ferramentas para AWS desenvolvedores)

Use o armazenamento em cache para reduzir a demanda do banco de dados

Visão geral do

Você pode usar o armazenamento em cache como uma estratégia eficaz para ajudar a reduzir os custos de suas aplicações .NET. Muitas aplicações usam bancos de dados de backend, como o SQL Server, quando as aplicações exigem acesso frequente aos dados. O custo de manter esses serviços de backend para lidar com a demanda pode ser alto, mas você pode usar uma estratégia eficaz de armazenamento em cache para reduzir a carga nos bancos de dados de backend, reduzindo os requisitos de dimensionamento e escalabilidade. Isso pode ajudar a reduzir os custos e a melhorar a performance de suas aplicações.

O armazenamento em cache é uma técnica útil para economizar nos custos relacionados à leitura de workloads pesadas que usam recursos mais caros, como o SQL Server. É importante usar a técnica certa para sua workload. Por exemplo, o armazenamento em cache local não é escalável e exige que você mantenha um cache local para cada instância de uma aplicação. Você deve ponderar o impacto na performance em comparação com os possíveis custos, para que o menor custo da fonte de dados subjacente compense quaisquer custos adicionais relacionados ao mecanismo de armazenamento em cache.

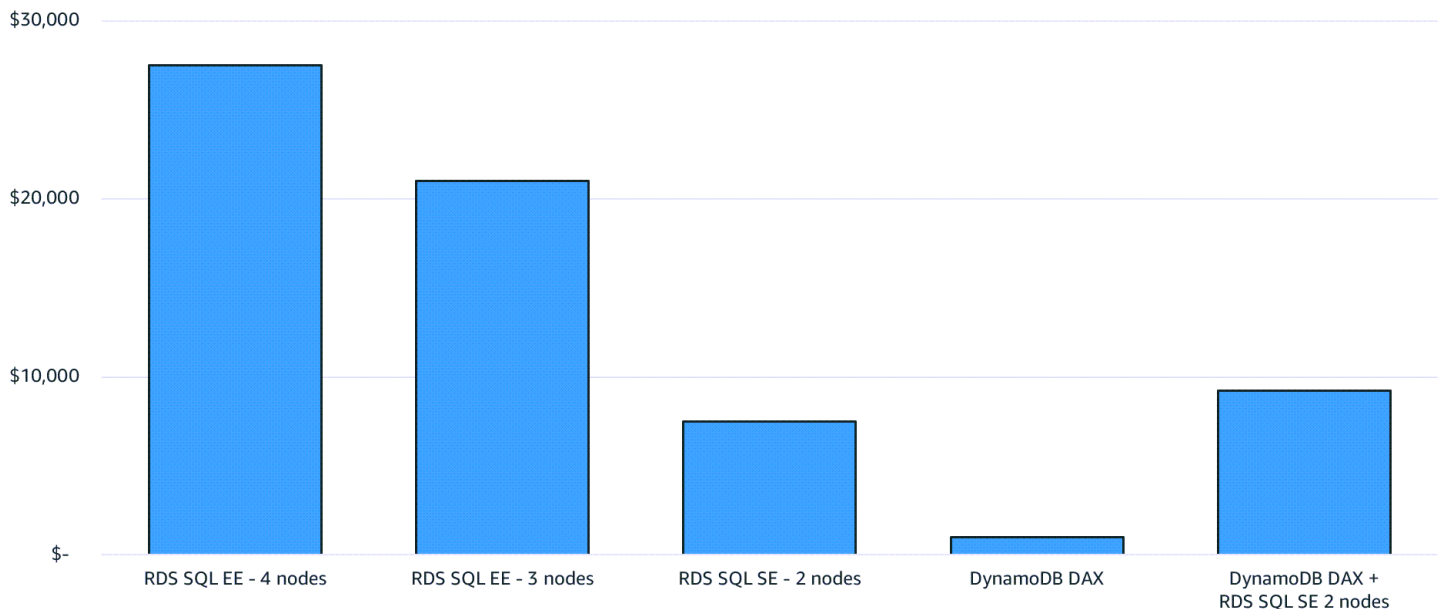
Impacto do custo

O SQL Server exige que você leve em consideração as solicitações de leitura ao dimensionar seu banco de dados. Isso pode afetar os custos, pois talvez seja necessário introduzir réplicas de leitura para acomodar a carga. Se você estiver usando réplicas de leitura, é importante entender que elas estão disponíveis somente na edição SQL Server Enterprise. Essa edição exige uma licença mais cara do que a edição SQL Server Standard.

O diagrama a seguir foi projetado para ajudar você a entender a eficácia do armazenamento em cache. Ele mostra o Amazon RDS para SQL Server com quatro nós db.m4.2xlarge executando a

edição SQL Server Enterprise. Ele é implantado em uma configuração multi-AZ com uma réplica de leitura. O tráfego de leitura exclusivo (por exemplo, consultas SELECT) é direcionado às réplicas de leitura. Em comparação, o Amazon DynamoDB usa um cluster r4.2xlarge de dois nós do DynamoDB Accelerator (DAX).

O gráfico a seguir mostra os resultados da remoção da exigência de réplicas de leitura dedicadas que lidam com alto tráfego de leitura.



Você pode obter economias de custo significativas usando o armazenamento em cache local sem réplicas de leitura ou introduzindo o DAX lado a lado com o SQL Server no Amazon RDS como uma camada de armazenamento em cache. Essa camada é descarregada do SQL Server e reduz o tamanho do SQL Server necessário para executar o banco de dados.

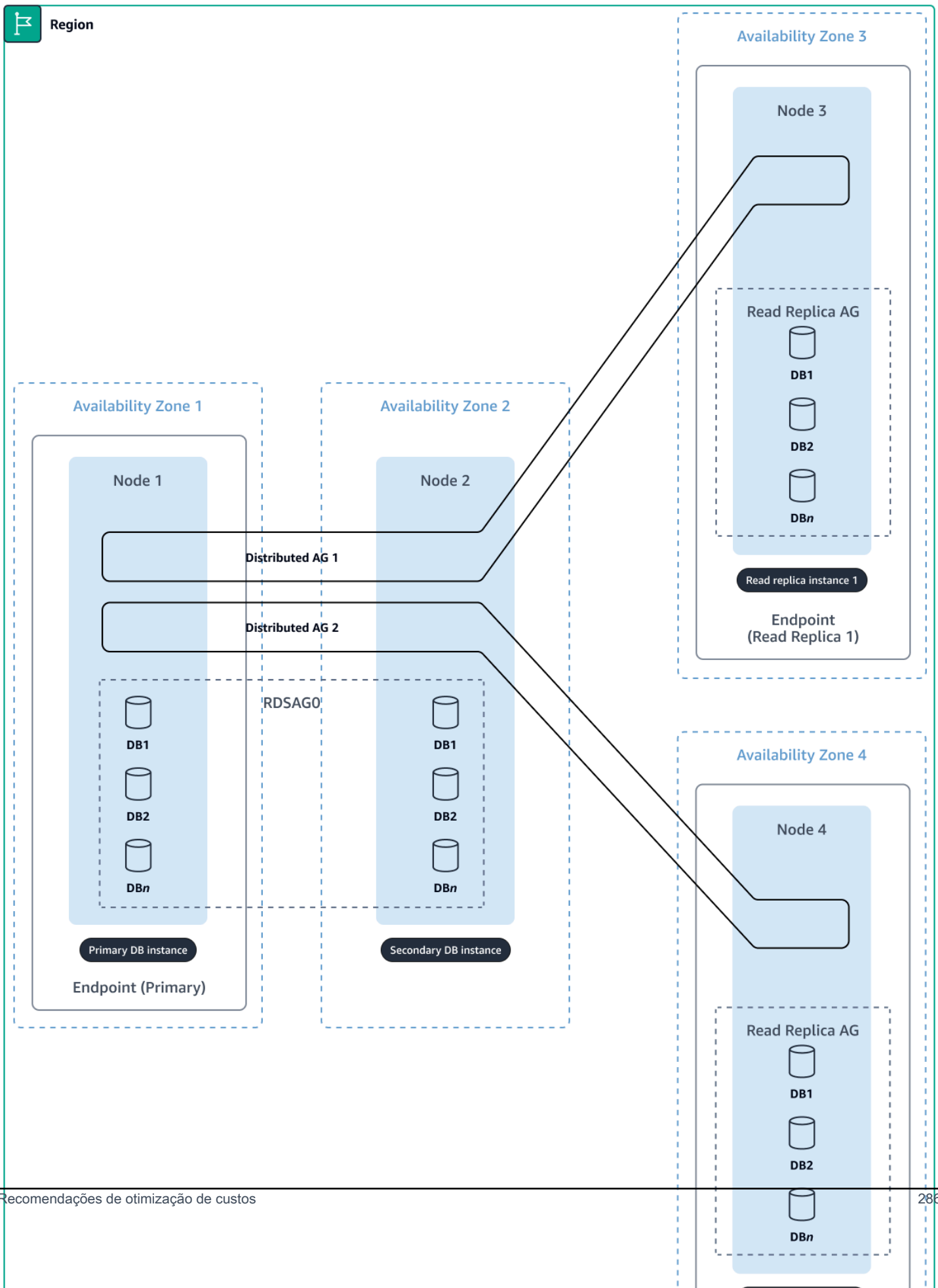
Recomendações de otimização de custos

Armazenamento em cache local

O armazenamento em cache local é uma das formas mais usadas de armazenar conteúdo em cache para aplicações hospedadas em ambientes on-premises ou na nuvem. Isso ocorre porque é relativamente fácil e intuitivo de implementar. O armazenamento em cache local envolve pegar conteúdo de um banco de dados ou de outra fonte e armazenar em cache localmente na memória ou no disco para acesso mais rápido. Essa abordagem, embora fácil de implementar, não é ideal para alguns casos de uso. Por exemplo, isso inclui casos de uso em que o conteúdo do armazenamento em cache precisa persistir ao longo do tempo, como preservar o estado da aplicação ou do usuário.

Outro caso de uso é quando o conteúdo em cache precisa ser acessado de outras instâncias da aplicação.

O diagrama abaixo ilustra um cluster do SQL Server altamente disponível com quatro nós e duas réplicas de leitura.



Com o armazenamento em cache local, talvez você precise balancear a carga do tráfego entre várias instâncias do EC2. Cada instância deve manter seu próprio cache local. Se o cache armazenar informações com estado, é necessário que haja commits regulares no banco de dados, e talvez os usuários precisem ser encaminhados para a mesma instância em cada solicitação subsequente (sessão persistente). Isso representa um desafio ao tentar escalar aplicações, pois algumas instâncias podem ser superutilizadas, enquanto outras são subutilizadas devido à distribuição desigual do tráfego.

Você pode usar o armazenamento em cache local, na memória ou usando armazenamento local, para aplicações .NET. Para fazer isso, você pode adicionar funcionalidade para armazenar objetos em disco e recuperá-los quando necessário, ou consultar dados do banco de dados e persisti-los na memória. Para realizar o armazenamento em cache local na memória e no armazenamento local de dados de um SQL Server em C#, por exemplo, você pode usar uma combinação das bibliotecas `MemoryCache` e `LiteDB`. A `MemoryCache` fornece armazenamento em cache em memória, enquanto a `LiteDB` é um banco de dados NoSQL incorporado baseado em disco que é rápido e leve.

Para realizar o armazenamento em cache em memória, use a biblioteca .NET `System.Runtime.MemoryCache`. O exemplo de código a seguir mostra como usar a classe `System.Runtime.Caching.MemoryCache` para armazenar dados em cache na memória. Essa classe fornece uma forma de armazenar dados temporariamente na memória da aplicação. Isso pode ajudar a melhorar a performance de uma aplicação ao reduzir a necessidade de buscar dados de um recurso mais caro, como um banco de dados ou uma API.

Confira como o código funciona:

1. Uma instância estática privada de `MemoryCache` denominada `_memoryCache` é criada. O cache recebe um nome (`dataCache`) para identificá-lo. Em seguida, o cache armazena e recupera os dados.
2. O método `GetData` é um método genérico que usa dois argumentos: uma chave `string` e um delegado `Func<T>` chamado `getData`. A chave é usada para identificar os dados em cache, enquanto o delegado `getData` representa a lógica de recuperação de dados que é executada quando os dados não estão presentes no cache.
3. O método primeiro verifica se os dados estão presentes no cache usando o método `_memoryCache.Contains(key)`. Se os dados estiverem no cache, o método recuperará os dados usando `_memoryCache.Get(key)` e os converterá no tipo esperado `T`.

4. Se os dados não estiverem no cache, o método chamará o delegado `getData` para buscar os dados. Em seguida, ele adiciona os dados ao cache usando `_memoryCache.Add(key, data, DateTimeOffset.Now.AddMinutes(10))`. Essa chamada especifica que a entrada do cache deve expirar após dez minutos, momento em que os dados são removidos do cache automaticamente.
5. O método `ClearCache` usa uma chave `string` como argumento e remove os dados associados a essa chave do cache usando `_memoryCache.Remove(key)`.

```
using System;
using System.Runtime.Caching;

public class InMemoryCache
{
    private static MemoryCache _memoryCache = new MemoryCache("dataCache");

    public static T GetData<T>(string key, Func<T> getData)
    {
        if (_memoryCache.Contains(key))
        {
            return (T)_memoryCache.Get(key);
        }

        T data = getData();
        _memoryCache.Add(key, data, DateTimeOffset.Now.AddMinutes(10));

        return data;
    }

    public static void ClearCache(string key)
    {
        _memoryCache.Remove(key);
    }
}
```

É possível usar o seguinte código:

```
public class Program
{
    public static void Main()
    {
        string cacheKey = "sample_data";
    }
}
```

```
Func<string> getSampleData = () =>
{
    // Replace this with your data retrieval logic
    return "Sample data";
};

string data = InMemoryCache.GetData(cacheKey, getSampleData);
Console.WriteLine("Data: " + data);
}
}
```

O exemplo a seguir mostra como usar o [LiteDB](#) para armazenar dados em cache no armazenamento local. Você pode usar o LiteDB como alternativa ou complemento ao cache em memória. O código a seguir demonstra como usar a biblioteca LiteDB para armazenar dados em cache no armazenamento local. A classe `LocalStorageCache` contém as principais funções para gerenciar o cache.

```
using System;
using LiteDB;

public class LocalStorageCache
{
    private static string _liteDbPath = @"Filename=LocalCache.db";

    public static T GetData<T>(string key, Func<T> getData)
    {
        using (var db = new LiteDatabase(_liteDbPath))
        {
            var collection = db.GetCollection<T>("cache");
            var item = collection.FindOne(Query.EQ("_id", key));

            if (item != null)
            {
                return item;
            }
        }

        T data = getData();

        using (var db = new LiteDatabase(_liteDbPath))
        {
            var collection = db.GetCollection<T>("cache");
```

```
        collection.Upsert(new BsonValue(key), data);
    }

    return data;
}

public static void ClearCache(string key)
{
    using (var db = new LiteDatabase(_liteDbPath))
    {
        var collection = db.GetCollection("cache");
        collection.Delete(key);
    }
}

public class Program
{
    public static void Main()
    {
        string cacheKey = "sample_data";

        Func<string> getSampleData = () =>
        {
            // Replace this with your data retrieval logic
            return "Sample data";
        };

        string data = LocalStorageCache.GetData(cacheKey, getSampleData);
        Console.WriteLine("Data: " + data);
    }
}
```

Se você tem um cache ou arquivos estáticos que não mudam com frequência, é possível armazenar esses arquivos no armazenamento de objetos do Amazon Simple Storage Service (Amazon S3). A aplicação pode recuperar o arquivo de cache estático na inicialização para usar localmente. Para obter mais detalhes sobre como recuperar arquivos do Amazon S3 usando o .NET, consulte [Baixar objetos](#) na documentação do Amazon S3.

Armazenamento em cache com o DAX

Você pode usar uma camada de cache que pode ser compartilhada entre todas as instâncias da aplicação. O [DynamoDB Accelerator \(DAX\)](#) é um cache em memória totalmente gerenciado

e altamente disponível para o DynamoDB que pode proporcionar um aumento de dez vezes na performance. Você pode usar o DAX para reduzir custos diminuindo a necessidade de superprovisionar unidades de capacidade de leitura nas tabelas do DynamoDB. Isso é especialmente útil para workloads com muita leitura e que exigem leituras repetidas para chaves individuais.

O preço do DynamoDB é baseado no modelo sob demanda ou mediante capacidade provisionada, portanto, o número de leituras e gravações por mês contribui para o custo. Se você leu workloads pesadas, os clusters do DAX podem ajudar a cortar os custos ao reduzir o número de leituras nas tabelas do DynamoDB. Para obter instruções sobre como configurar o DAX, consulte [Aceleração em memória com o DynamoDB Accelerator \(DAX\)](#) na documentação do DynamoDB. Para obter informações sobre a integração de aplicativos.NET, assista [Integrando o Amazon DynamoDB DAX em seu aplicativo ASP.NET em YouTube](#)

Recursos adicionais do

- [Aceleração em memória com o DynamoDB Accelerator \(DAX\) - Amazon DynamoDB](#) (documentação do DynamoDB)
- [Integrando o Amazon DynamoDB DAX](#) em seu aplicativo ASP.NET () YouTube
- [Baixar objetos](#) (documentação do Amazon S3)

Considerar o .NET sem servidor

Visão geral do

A computação sem servidor tornou-se uma abordagem conhecida para criar e implantar aplicações. Isso se deve principalmente à escalabilidade e à agilidade que a abordagem sem servidor oferece ao criar uma arquitetura moderna. No entanto, é importante considerar o impacto nos custos da computação sem servidor em alguns cenários.

O Lambda é uma plataforma de computação sem servidor que permite que os desenvolvedores executem códigos sem a necessidade de servidores dedicados. O Lambda é uma opção particularmente atraente para desenvolvedores .NET que buscam reduzir os custos de infraestrutura. Com o Lambda, os desenvolvedores .NET podem desenvolver e implantar aplicações altamente escaláveis e com potencial para otimizar custos. Ao usar uma abordagem sem servidor, os desenvolvedores não provisionam mais servidores para lidar com solicitações de aplicações. Em vez disso, os desenvolvedores podem criar funções que são executadas sob demanda. Isso torna

uma abordagem sem servidor mais escalável, gerenciável e potencialmente mais econômica do que executar, gerenciar e escalar máquinas virtuais. Como resultado, você paga apenas pelos recursos usados pela aplicação, sem precisar se preocupar com os recursos subutilizados ou os custos de manutenção do servidor.

Os desenvolvedores podem usar versões modernas e multiplataforma do .NET para criar aplicações sem servidor que sejam rápidas, eficientes e econômicas. O .NET Core e as versões mais recentes são um framework gratuito e de código aberto que é mais adequado para execução em plataformas sem servidor em relação às versões anteriores do .NET Framework. Isso permite que os desenvolvedores reduzam o tempo de desenvolvimento e aumentem a performance da aplicação. O .NET Modern também é compatível com diversas linguagens de programação, incluindo C# e F#. Por esse motivo, é uma opção atraente para desenvolvedores que desejam criar arquiteturas modernas na nuvem.

Esta seção explica como você pode obter redução de custos usando o Lambda como uma opção sem servidor. Você pode otimizar ainda mais os custos fazendo o ajuste fino dos perfis de execução de suas funções do Lambda, dimensionando corretamente a alocação de memória das suas funções do Lambda, usando [AOT nativo](#) e migrando para funções baseadas em Graviton.

Impacto do custo

O quanto você pode reduzir custos depende de vários fatores, incluindo quantas execuções suas funções sem servidor executarão, além da quantidade de memória alocada e da duração de cada função. AWS Lambda oferece um nível gratuito, que inclui um milhão de solicitações gratuitas por mês e 400.000 GB de segundos de tempo de computação por mês. Você pode reduzir significativamente os custos mensais de workloads que estão dentro ou perto desses limites de nível gratuito.

Também pode haver custos adicionais ao usar um balanceador de carga com funções do Lambda como destino. Isso é calculado como a quantidade de dados processados pelo balanceador de carga para os [destinos do Lambda](#).

Recomendações de otimização de custos

Dimensionar corretamente suas funções do Lambda

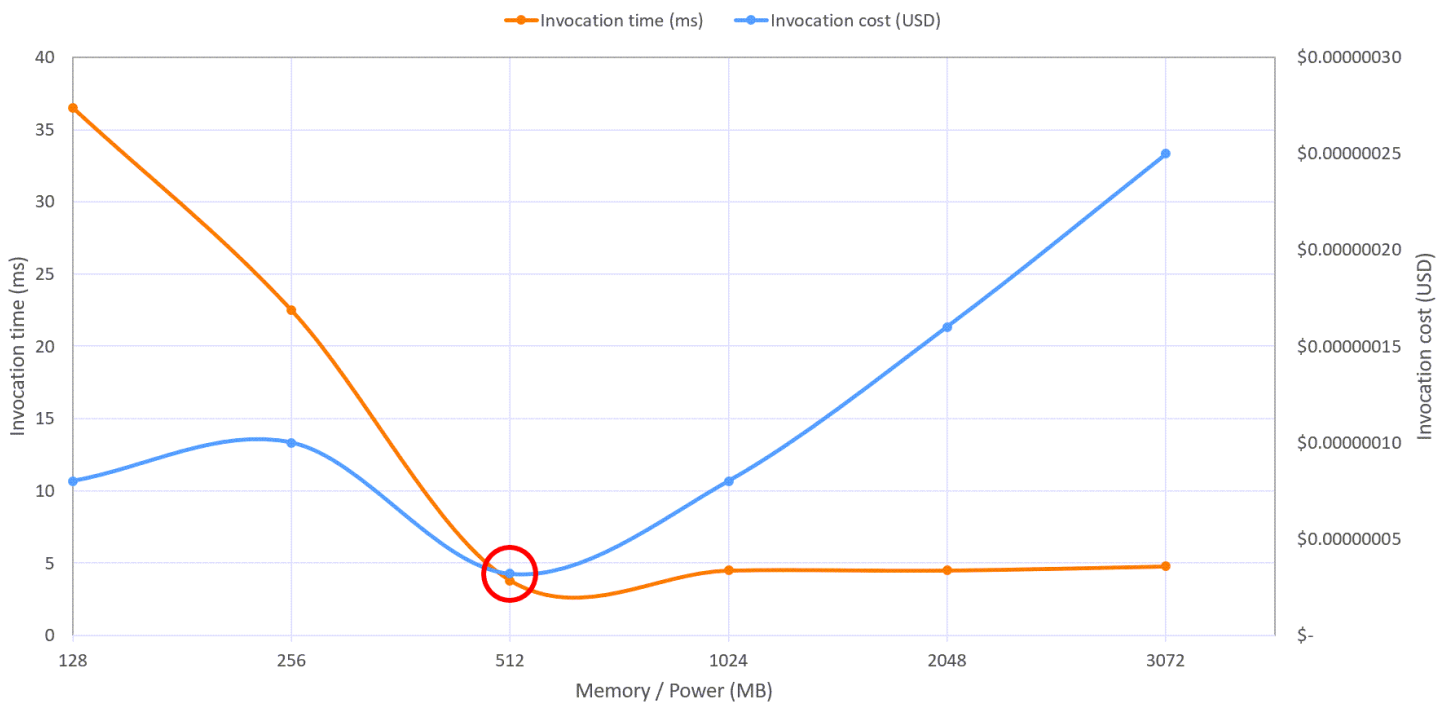
O dimensionamento correto é uma prática essencial para a otimização de custos em funções do Lambda baseadas em .NET. Esse processo envolve a identificação da configuração de memória ideal que equilibra performance e economia, sem exigir alterações no código.

Ao configurar a memória para uma função do Lambda, variando de 128 MB a 10.240 MB, você também ajusta a quantidade de vCPU disponível durante a invocação. Isso permite que aplicações vinculadas à memória ou à CPU acessem recursos adicionais durante a execução, levando a uma possível redução na duração da invocação e no custo geral.

No entanto, identificar a configuração ideal para suas funções do Lambda baseadas em .NET pode ser um processo manual e demorado, especialmente se as alterações forem frequentes. A [ferramenta Power Tuning do AWS Lambda](#) pode ajudar você a identificar a configuração apropriada analisando um conjunto de configurações de memória em relação a um exemplo de carga útil.

Por exemplo, aumentar a memória de uma função do Lambda baseada em .NET pode levar a um melhor tempo total de invocação e a um custo reduzido sem afetar a performance. A configuração de memória ideal para uma função pode variar. A ferramenta AWS Lambda Power Tuning pode ajudar a identificar a configuração mais econômica para cada função.

No gráfico de exemplo a seguir, o tempo total de invocação melhora à medida que a memória aumenta para essa função do Lambda. Isso leva a uma redução no custo da execução total sem afetar a performance original da função. Para essa função, a configuração de memória ideal para a função é de 512 MB, pois é aqui que a utilização de recursos é mais eficiente para o custo total de cada invocação. Isso varia de acordo com a função, e o uso da ferramenta em suas funções do Lambda pode identificar se elas se beneficiam do dimensionamento correto.



Recomendamos que você conclua esse exercício regularmente, como parte de qualquer teste de integração quando novas atualizações forem lançadas. Se houver atualizações com pouca frequência, faça esse exercício periodicamente para garantir que as funções estejam ajustadas e dimensionadas corretamente. Depois de identificar a configuração de memória apropriada para suas funções do Lambda, você pode adicionar o dimensionamento correto aos seus processos. A ferramenta AWS Lambda Power Tuning gera uma saída programática que pode ser usada por seus fluxos de trabalho de CI/CD durante o lançamento do novo código. Isso permite que você automatize a configuração de memória.

Você pode baixar a [ferramenta Power Tuning do AWS Lambda](#) gratuitamente. Para obter instruções sobre como usar a ferramenta, consulte [Como executar a máquina de estado](#) em GitHub.

O Lambda também é compatível com o AOT nativo, o que permite que aplicações .NET sejam pré-compiladas. Isso pode ajudar a reduzir custos ao diminuir os tempos de execução das funções .NET. Para obter mais informações sobre a criação de funções nativas de AOT, consulte [.NET functions with native AOT compilation](#) na documentação do Lambda.

Evitar tempo de espera ocioso

A duração da função do Lambda é uma dimensão usada para calcular o faturamento. Quando o código de função faz uma chamada de bloqueio, você é cobrado pelo tempo que ele espera para receber uma resposta. Esse tempo de espera pode aumentar quando as funções do Lambda são encadeadas ou quando uma função está atuando como orquestradora para outras funções. Se você tiver fluxos de trabalho, como operações em lote ou sistemas de entrega de pedidos, isso aumenta a sobrecarga de gerenciamento. Além disso, talvez não seja possível concluir toda a lógica do fluxo de trabalho e o tratamento de erros dentro do tempo limite máximo do Lambda de 15 minutos.

Em vez de lidar com essa lógica no código da função, recomendamos que você reestruture sua solução para usar o [AWS Step Functions](#) como orquestrador do fluxo de trabalho. Ao usar um fluxo de trabalho padrão, você é cobrado por cada transição de [estado](#) dentro do fluxo de trabalho, em vez da duração total do fluxo de trabalho. Além disso, você pode mover o suporte para novas tentativas, condições de espera, fluxos de trabalho de erro e [retornos de chamada](#) para a condição de estado a fim de permitir que suas funções do Lambda se concentrem na lógica de negócios. Para obter mais informações, consulte [Otimizando seus AWS Lambda custos — Parte 2](#) no blog sobre AWS computação.

Migrar para funções baseadas em Graviton

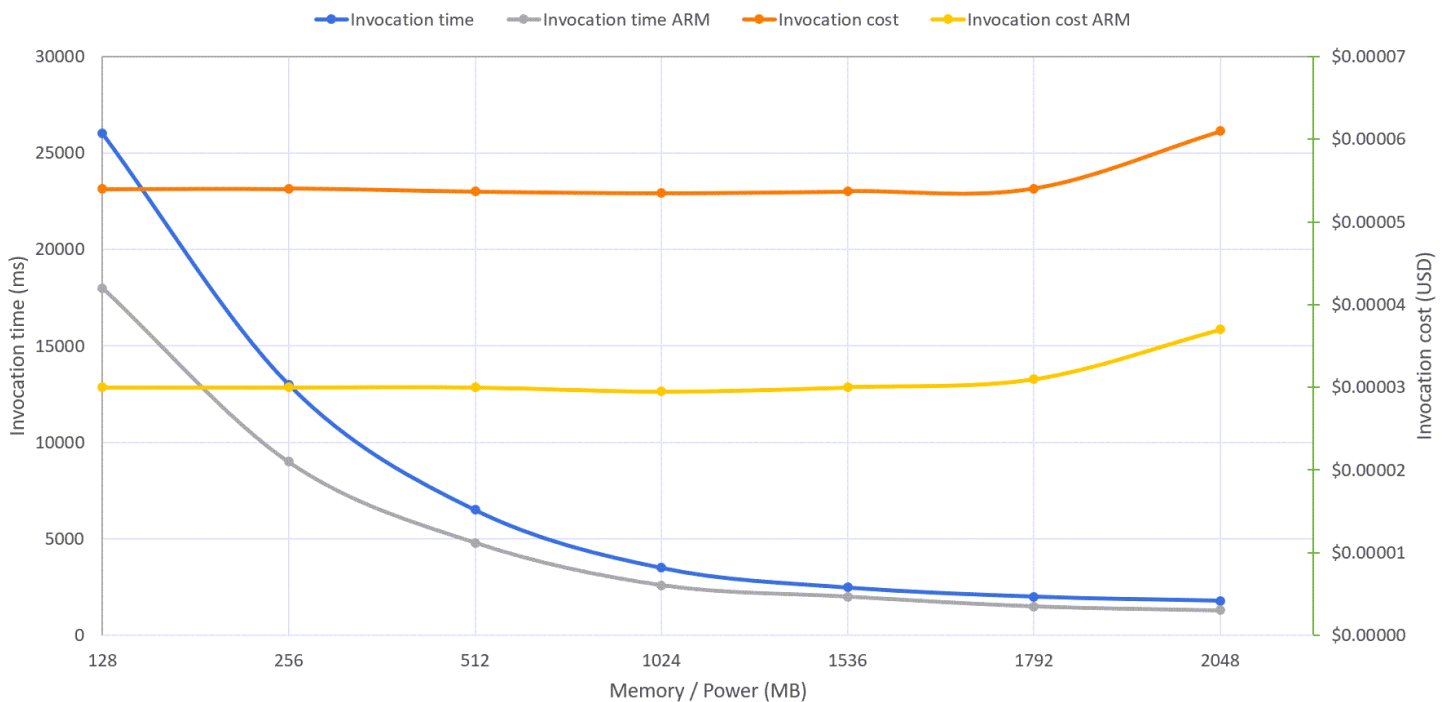
As funções do Lambda alimentadas pelos processadores Graviton2 de próxima geração agora estão disponíveis ao público em geral. As funções do Graviton2, usando uma arquitetura de processador baseada em ARM, foram projetadas para oferecer uma performance até 19% melhor a um custo 20% menor para uma variedade de workloads sem servidor. Com menor latência e melhor performance, as funções baseadas em processadores Graviton2 são ideais para alimentar aplicações sem servidor de missão crítica.

A migração para funções do Lambda baseadas em Graviton pode ser uma opção econômica para desenvolvedores .NET que desejam otimizar seus custos do Lambda. As funções baseadas em Graviton usam processadores baseados em ARM em vez dos processadores x86 tradicionais. Isso pode levar a uma economia significativa de custos sem sacrificar a performance.

Embora haja vários benefícios em migrar para funções baseadas em Graviton, também há vários desafios e considerações que recomendamos que você considere. Por exemplo, funções baseadas em Graviton exigem o uso do Amazon Linux 2, que pode não ser compatível com todas as aplicações .NET. Além disso, pode haver problemas de compatibilidade com bibliotecas ou dependências de terceiros que não são compatíveis com processadores baseados em ARM.

Se você estiver executando aplicações .NET Framework e quiser aproveitar as vantagens da tecnologia sem servidor com o Lambda, considere a possibilidade de portar as aplicações para o .NET Modern usando o [Assistente de Portabilidade para .NET](#). Isso pode ajudar você a acelerar a migração de aplicações .NET legadas para o .NET Modern, permitindo que a aplicação seja executada no Linux.

O gráfico a seguir compara os resultados das arquiteturas x86 e ARM/Graviton2 para uma função que calcula números primos.



A função está usando um único thread. A menor duração para as duas arquiteturas é relatada quando a memória é configurada com 1,8 GB. Acima disso, as funções do Lambda têm acesso a mais de uma vCPU, mas, nesse caso, a função não pode usar a capacidade adicional. Pelo mesmo motivo, os custos ficam estáveis com uma memória de até 1,8 GB. Com mais memória, os custos aumentam porque não há benefícios adicionais de performance para essa workload. O processador Graviton2 está claramente fornecendo melhor performance e custos mais baixos para essa função de computação intensiva.

Para configurar sua função para usar um processador baseado em ARM com Graviton, faça o seguinte:

1. Faça login no Console de gerenciamento da AWS e abra o console [Lambda](#).
2. Escolha a opção Criar função.
3. Em Function name (Nome da função), insira um nome.
4. Em Runtime, escolha .NET 6 (C#/ PowerShell).
5. Em Arquitetura, selecione amd64.
6. Faça as configurações adicionais necessárias e escolha Criar função.

Recursos adicionais do

- [Funções do Lambda como alvos \(documentação\)](#) AWS
- [Otimizando o AWS Lambda custo e o desempenho usando AWS Compute Optimizer](#) (AWS Compute Blog)
- [Otimizando seus AWS Lambda custos — Parte 1](#) (AWS Compute Blog)
- [Otimizando seus AWS Lambda custos — Parte 2](#) (AWS Compute Blog)
- [Criando aplicativos.NET sem servidor AWS Lambda usando o.NET 7](#) (AWS Compute Blog)

Considerar bancos de dados com propósito específico

Visão geral do

Um dos aspectos mais caros da execução de workloads baseadas na Microsoft é o licenciamento de bancos de dados comerciais, como o SQL Server. As empresas geralmente padronizam o SQL Server como a plataforma de banco de dados preferencial, e ele passa a fazer parte da cultura de desenvolvimento da organização. Os desenvolvedores geralmente escolhem um modelo relacional baseado no SQL Server, independentemente do caso de uso. Os motivos para isso incluem:

- A empresa já tem and/or licenças de instâncias do SQL Server disponíveis.
- As equipes se habituaram ao modelo de programação SQL por meio do uso de bibliotecas compartilhadas e lógica de negócios. ORMs
- A gerência não está ciente das alternativas.
- Os desenvolvedores não estão cientes das alternativas.

Bancos de dados com propósito específico podem acomodar os padrões de acesso aos dados do seu caso de uso. Esses bancos de dados são cada vez mais adotados pelas empresas à medida que adotam arquiteturas mais modernas (como microsserviços) e que o escopo de aplicações individuais é reduzido.

Um banco de dados com propósito específico não inviabiliza um modelo relacional nem exige um modelo NoSQL (não relacional). Na verdade, um banco de dados relacional é considerado como propósito específico quando selecionado em resposta às necessidades específicas de uma workload. O uso de banco de dados com propósito específico pode ajudar as equipes a reduzir os

custos de banco de dados associados às suas aplicações .NET, além de obter as vantagens padrão da nuvem, como escalabilidade, resiliência e redução esforço indiferenciado e pesado.

A tabela a seguir mostra os bancos de dados desenvolvidos especificamente oferecidos pelo. AWS

Banco de dados	Tipo	Características
Amazon Aurora PostgreSQL ou Amazon Aurora MySQL	Relacional	<p>Casos de uso em que os dados têm uma estrutura fixa</p> <p>Os bancos de dados relacionais mantêm naturalmente a consistência de dados por meio de transações ACID</p>
Amazon DynamoDB	Par de chave/valor	<p>Banco de dados NoSQL que armazena dados usando uma estrutura de dados de tabela hash</p> <p>Armazenamento e recuperação de dados não estruturados de alta performance</p> <p>Os casos de uso incluem perfis de usuário, estado da sessão e dados do carrinho de compras</p>
Amazon ElastiCache	Na memória	<p>Banco de dados NoSQL de alta performance que armazena dados não estruturados na memória com tempo de acesso inferior a um milissegundo</p> <p>Usado para dados efêmeros acessados com frequência, como sessões de usuário</p>

Banco de dados	Tipo	Características
		<p>e como uma camada de cache na frente de outros armazenamentos de dados mais lentos</p> <p>Inclui suporte para ElastiCache (Redis OSS) e ElastiCache (Memcached)</p>
Amazon MemoryDB	Em memória durável	Banco de dados com propósito específico compatível com Redis com armazenamento durável
Amazon Timestream	Séries temporais	<p>Banco de dados projetado para ingestão de dados de throughput em ordem temporal</p> <p>Os casos de uso incluem aplicações de Internet das Coisas (IoT) e armazenamento de dados de telemetria ou métricas.</p>
Amazon DocumentDB	Documento	<p>Banco de dados NoSQL que armazena dados sem uma estrutura prescrita ou relações forçadas com outros dados</p> <p>Geralmente usado para workloads intensivas em leitura, como catálogos de produtos</p>

Banco de dados	Tipo	Características
Amazon Neptune	Gráfico	<p>Banco de dados NoSQL que contém dados e uma representação das conexões entre itens de dados</p> <p>Os casos de uso incluem detecção de fraudes, mecanismos de recomendações e aplicações sociais</p>
Amazon Keyspaces	Coluna larga	<p>Banco de dados distribuído de alta performance baseado no Apache Cassandra</p> <p>Os casos de uso incluem aplicações de IoT, processamento de eventos e aplicações de jogos</p>

Um fator significativo para a adoção de bancos de dados com propósito específico pode ser atribuído à remoção da necessidade do licenciamento comercial. No entanto, a capacidade de ajuste de escala automático de bancos de dados como o [DynamoDB](#) (incluindo o [modo sob demanda](#)), [Aurora](#), [Amazon Neptune](#) e [Amazon Keyspaces](#) permite que você provisione capacidade para o caso médio, em vez de para o uso de pico. Bancos de dados com propósito específico, como o Timestream, usam tecnologia sem servidor e são escalados automaticamente para atender à demanda sem nenhum pré-provisionamento.

AWS oferece o [Babelfish for Aurora PostgreSQL](#) se você quiser usar um banco de dados relacional de código aberto compatível com um propósito específico, mas não puder ou não quiser fazer alterações significativas no código do seu aplicativo. Em alguns casos, o Babelfish permite que você use um código de acesso existente do SQL Server, quase sem alterações.

Ao escolher um banco de dados relacional com propósito específico para aplicações, é importante manter os mesmos recursos (ou funcionalmente equivalentes) de que você precisa para suas aplicações. Essa recomendação aborda bancos de dados com propósito específico como um

armazenamento de dados primário para aplicações. Aplicações específicas (como armazenamento em cache) são abordadas em outras recomendações.

Impacto do custo

A adoção de bancos de dados específicos para cargas de trabalho do.NET, embora seja improvável que afete diretamente a computação, pode influenciar consumption/cost diretamente o custo dos serviços de banco de dados consumidos pelos aplicativos.NET. Na verdade, a economia de custos pode ser uma meta secundária, quando comparada aos benefícios adicionais de agilidade, escalabilidade, resiliência e durabilidade dos dados.

Está fora do escopo deste guia explicar o processo completo de escolha de um banco de dados com propósito específico para aplicações e a redefinição da arquitetura de uma estratégia de dados para usá-los de forma eficaz. Para obter mais informações, consulte [Purpose-built databases](#) no diretório de tutoriais da AWS .

As tabelas a seguir mostram vários exemplos de como a substituição do SQL Server por um banco de dados com propósito específico pode alterar os custos da aplicação. Observe que estas são apenas estimativas aproximadas. São necessárias avaliações comparativas e otimização das workloads reais para calcular o custo exato de produção.

Estas são algumas estimativas comumente usadas para bancos de dados com propósito específico, que incluem computação sob demanda e SSD de 100 GB e bancos de dados de instância única em us-east-1. Os custos de licença incluem a licença do SQL Server e a garantia de software.

A tabela a seguir mostra os custos estimados para exemplos de bancos de dados comerciais.

mecanismo de banco de dados	Modelo de licenciamento	Tipo/espec. da instância	AWS custo de computação + armazenamento	Custo da licença	Custo mensal total
Edição SQL Server Standard no Amazon EC2	Licença incluída	r6i.2xlarge (8 CPU/64 GB de RAM)	\$1.345,36	\$0,00	\$1.345,36

mecanismo de banco de dados	Modelo de licenciamento	Tipo/espec. da instância	AWS custo de computação + armazenamento	Custo da licença	Custo mensal total
Edição SQL Server Enterprise no Amazon EC2	Licença incluída	r6i.2xlarge (8 CPU/64 GB de RAM)	\$2.834,56	\$0,00	\$2.834,56
Edição SQL Server Standard no Amazon EC2	BYOL	r6i.2xlarge (8 CPU/64 GB de RAM)	\$644,56	\$456,00	\$1.100,56
Edição SQL Server Enterprise no Amazon EC2	BYOL	r6i.2xlarge (8 CPU/64 GB de RAM)	\$644,56	\$1.750,00	\$2.394,56
Edição SQL Server Standard no Amazon RDS		db.r6i.2xlarge (8 CPU/64 GB de RAM)	\$2.318,30	\$0,00	\$2.318,30
Edição SQL Server Enterprise no Amazon RDS		db.r6i.2xlarge (8 CPU/64 GB de RAM)	\$3.750,56	\$0,00	\$3.750,56

A tabela a seguir mostra os custos estimados para exemplos com propósito específico.

mecanismo de banco de dados	Tipo/espec. da instância	AWS custo de computação + armazenamento	Custo da licença	Custo mensal total
Amazon Aurora PostgreSQL	r6g.2xlarge (8 CPU/64 GB de RAM)	\$855,87	\$0,00	\$855,87
DynamoDB	Base provisionada de 100 WCU/400 RCU	\$72,00		\$72,00
Amazon DocumentDB	db.r6i.2xlarge (8 CPU/64 GB de RAM)	\$778.60		\$778.60

Important

A tabela é baseada nos custos estimados de licenciamento do SQL Server com o Software Assurance, durante os primeiros três anos da compra. Para a edição SQL Server Standard: USD 4.100, pacote de 2 núcleos, 3 anos. Para a edição SQL Server Enterprise: USD 15.700, pacote de 2 núcleos, 3 anos.

Recomendamos considerar as implicações de custos antes de adotar bancos de dados com propósito específico. Por exemplo, o custo de atualizar aplicações para usar um banco de dados com propósito específico está relacionado à complexidade da aplicação e do banco de dados de origem. Certifique-se de considerar o custo total de propriedade ao planejar essa mudança de arquitetura. Isso inclui refatorar suas aplicações, requalificar a equipe em novas tecnologias e planejar cuidadosamente a performance e o consumo previstos para cada workload. Com base nisso, você pode determinar se o investimento vale a redução de custos. Na maioria dos casos, manter um end-of-support produto é um risco de segurança e conformidade, e o custo de remediá-lo vale o esforço e o investimento inicial.

Recomendações de otimização de custos

Para aplicações .NET que acessam o SQL Server, existem bibliotecas substitutas para bancos de dados relacionais com propósito específico. Você pode implementar essas bibliotecas em sua aplicação para substituir funcionalidades de aplicação semelhantes às do SQL Server.

A tabela a seguir destaca algumas bibliotecas que podem ser usadas em muitos cenários comuns.

Biblioteca	Banco de dados	Substituição para	Compatibilidade do Framework
Npgsql Entity Framework Core Provider	Amazon Aurora PostgreSQL	Entity Framework Core SQL Server Provider	Modern .NET
Npgsql Entity Framework 6 Provider	Amazon Aurora PostgreSQL	Entity Framework 6.0 SQL Server Provider	NET Framework
Npgsql (biblioteca PostgreSQL compatível com ADO.NET)	Amazon Aurora PostgreSQL	ADO.NET	Framework/ Modern .NET
MySQL Entity Framework Core Provider	Amazon Aurora MySQL	Entity Framework Core SQL Server Provider	Modern .NET
Pomelo.EntityFrameworkCore.MySql	Amazon Aurora MySQL	Entity Framework Core SQL Server Provider	Modern .NET

[Conectar-se ao Amazon Aurora PostgreSQL usando o Babelfish](#) não requer nenhuma codificação especial. No entanto, todo código deve ser exaustivamente testado antes de ser usado.

Outros bancos de dados com propósito específico têm bibliotecas para acessar bibliotecas compatíveis com o .NET que permitem acessar bancos de dados com propósito específico. Os exemplos incluem:

- [Usando bancos de dados NoSQL do Amazon DynamoDB \(documentação\)](#) AWS SDK para .NET

- [MongoDB C# Driver](#) (documentação do MongoDB)
- [.NET](#) (documentação do Timestream)
- [Using a Cassandra .NET Core client driver to access Amazon Keyspaces programmatically](#) (documentação do Amazon Keyspaces)
- [Using .NET to connect to a Neptune DB instance](#) (documentação do Neptune)

Se você migrar para bancos de dados criados especificamente, poderá usar essas ferramentas AWS para ajudar no processo de migração:

- O [AWS Schema Conversion Tool \(AWS SCT\)](#) pode ajudar a transformar os esquemas do SQL Server para o Amazon Aurora e no Amazon DynamoDB.
- O [AWS Database Migration Service \(AWS DMS\)](#) pode ajudar a migrar dados, uma vez ou continuamente, do SQL Server para o Aurora ou o DynamoDB.
- O [Babelfish Compass](#) pode ajudar a verificar a compatibilidade do seu banco de dados SQL Server para uso com o Babelfish para Aurora PostgreSQL.

Recursos adicionais do

- [Guidance for migrating SQL Server to Amazon Aurora PostgreSQL](#) (blog AWS Database)
- [Dia de Imersão na Modernização do Babelfish APP](#) (Workshop Studio)AWS
- [Dia de imersão do.NET](#) (AWS Workshop Studio)
- [Introdução ao Amazon Timestream com o.NET](#) () GitHub
- [Bancos de dados específicos para aplicativos.NET modernos em AWS](#) (apresentação)AWS

Próximas etapas

Depois de concluir a análise deste guia, recomendamos que você execute as próximas etapas para implementar a MACO:

1. Entre em contato com um especialista em MACO. Um especialista em MACO pode ajudar a responder às suas perguntas e a resolver suas preocupações. Se você já estiver trabalhando com uma equipe da conta da AWS, entre em contato com a equipe e solicite ajuda de um especialista em MACO. Se você não tiver uma equipe de conta, entre em contato com optimize-microsoft@amazon.com.
2. Aplique as recomendações. Aplique as recomendações, as práticas recomendadas e as estratégias que você aprendeu neste guia e ao falar com um especialista em MACO.
3. Acompanhe as mudanças de custos. Marque suas workloads e use serviços como o AWS Cost Explorer e o AWS Budgets para acompanhamento, monitoramento e controle detalhados de custos.

Histórico do documento

A tabela a seguir descreve alterações significativas feitas neste guia. Se desejar receber notificações sobre futuras atualizações, inscreva-se em um [feed RSS](#).

Alteração	Descrição	Data
Atualizações do SQL Server	Atualizamos a seção Otimizar CPU para cargas de trabalho do SQL Server para adicionar mais informações sobre o recurso Otimizar CPU para EC2 instâncias da Amazon.	22 de outubro de 2025
Atualizações do SQL Server	Atualizamos a seção Otimizar a CPU para workloads do SQL Server .	25 de outubro de 2024
Atualizações do SQL Server e de contêineres	Adicionamos as seções Otimizar o dimensionamento do SQL Server usando o Otimizador de Computação , as recomendações de Trusted Advisor revisão para cargas de trabalho do SQL Server e a seção Replataforma de aplicativos do Windows com o App2Container .	29 de junho de 2024
Otimização do licenciamento do SQL Server	Adicionamos a seção Otimizar o licenciamento do SQL Server usando o Compute Optimizer .	22 de maio de 2024
Publicação inicial	—	21 de dezembro de 2023

AWS Glossário de orientação prescritiva

A seguir estão os termos comumente usados em estratégias, guias e padrões fornecidos pela Orientação AWS Prescritiva. Para sugerir entradas, use o link Fornecer feedback no final do glossário.

Números

7 Rs

Sete estratégias comuns de migração para mover aplicações para a nuvem. Essas estratégias baseiam-se nos 5 Rs identificados pela Gartner em 2011 e consistem em:

- Refatorar/rearquitetar: mova uma aplicação e modifique sua arquitetura aproveitando ao máximo os recursos nativos de nuvem para melhorar a agilidade, a performance e a escalabilidade. Isso normalmente envolve a portabilidade do sistema operacional e do banco de dados. Exemplo: migrar seu banco de dados Oracle on-premises para o Amazon Aurora Edição Compatível com PostgreSQL.
- Redefinir a plataforma (mover e redefinir [mover e redefinir (lift-and-reshape)]): mova uma aplicação para a nuvem e introduza algum nível de otimização a fim de aproveitar os recursos da nuvem. Exemplo: migrar seu banco de dados Oracle on-premises para o Amazon Relational Database Service (Amazon RDS) para Oracle na Nuvem AWS.
- Recomprar (drop and shop): mude para um produto diferente, normalmente migrando de uma licença tradicional para um modelo SaaS. Exemplo: migrar seu sistema de gerenciamento de relacionamento com o cliente (CRM) para o Salesforce.com.
- Redefinir a hospedagem (mover sem alterações [lift-and-shift])mover uma aplicação para a nuvem sem fazer nenhuma alteração a fim de aproveitar os recursos da nuvem. Exemplo: migrar seu banco de dados Oracle on-premises para o Oracle em uma instância do EC2 na Nuvem AWS.
- Realocar (mover o hipervisor sem alterações [hypervisor-level lift-and-shift]): mover a infraestrutura para a nuvem sem comprar novo hardware, reescrever aplicações ou modificar suas operações existentes. Você migra servidores de uma plataforma on-premises para um serviço de nuvem para a mesma plataforma. Exemplo: migrar um Microsoft Hyper-V aplicativo para o. AWS
- Reter (revisitar): mantenha as aplicações em seu ambiente de origem. Isso pode incluir aplicações que exigem grande refatoração, e você deseja adiar esse trabalho para um

momento posterior, e aplicações antigas que você deseja manter porque não há justificativa comercial para migrá-las.

- Retirar: desative ou remova aplicações que não são mais necessárias em seu ambiente de origem.

A

ABAC

Consulte [controle de acesso baseado em atributo](#).

serviços abstraídos

Veja [serviços gerenciados](#).

ACID

Veja [atomicidade, consistência, isolamento, durabilidade](#).

migração ativa-ativa

Um método de migração de banco de dados no qual os bancos de dados de origem e de destino são mantidos em sincronia (por meio de uma ferramenta de replicação bidirecional ou operações de gravação dupla), e ambos os bancos de dados lidam com transações de aplicações conectadas durante a migração. Esse método oferece suporte à migração em lotes pequenos e controlados, em vez de exigir uma substituição única. É mais flexível, mas exige mais trabalho do que a [migração ativa-passiva](#).

migração ativa-passiva

Um método de migração de banco de dados em que os bancos de dados de origem e de destino são mantidos em sincronia, mas somente o banco de dados de origem manipula as transações das aplicações conectadas, enquanto os dados são replicados no banco de dados de destino. O banco de dados de destino não aceita nenhuma transação durante a migração.

AGGREGATE FUNCTION

Uma função SQL que opera em um grupo de linhas e calcula um único valor de retorno para o grupo. Exemplos de funções agregadas incluem SUM e MAX.

AI

Veja [inteligência artificial](#).

AIOps

Veja [operações de inteligência artificial](#).

anonimização

O processo de excluir permanentemente informações pessoais em um conjunto de dados. A anonimização pode ajudar a proteger a privacidade pessoal. Dados anônimos não são mais considerados dados pessoais.

antipadrões

Uma solução frequentemente usada para um problema recorrente em que a solução é contraproducente, ineficaz ou menos eficaz do que uma alternativa.

controle de aplicações

Uma abordagem de segurança que permite o uso somente de aplicações aprovadas para ajudar a proteger um sistema contra malware.

portfólio de aplicações

Uma coleção de informações detalhadas sobre cada aplicação usada por uma organização, incluindo o custo para criar e manter a aplicação e seu valor comercial. Essas informações são fundamentais para [o processo de descoberta e análise de portfólio](#) e ajudam a identificar e priorizar as aplicações a serem migradas, modernizadas e otimizadas.

inteligência artificial (IA)

O campo da ciência da computação que se dedica ao uso de tecnologias de computação para desempenhar funções cognitivas normalmente associadas aos humanos, como aprender, resolver problemas e reconhecer padrões. Para obter mais informações, consulte [O que é inteligência artificial?](#)

operações de inteligência artificial (AIOps)

O processo de usar técnicas de machine learning para resolver problemas operacionais, reduzir incidentes operacionais e intervenção humana e aumentar a qualidade do serviço. Para obter mais informações sobre como AIOps é usado na estratégia de AWS migração, consulte o [guia de integração de operações](#).

criptografia assimétrica

Um algoritmo de criptografia que usa um par de chaves, uma chave pública para criptografia e uma chave privada para descryptografia. É possível compartilhar a chave pública porque ela não é usada na descryptografia, mas o acesso à chave privada deve ser altamente restrito.

atomicidade, consistência, isolamento, durabilidade (ACID)

Um conjunto de propriedades de software que garantem a validade dos dados e a confiabilidade operacional de um banco de dados, mesmo no caso de erros, falhas de energia ou outros problemas.

controle de acesso por atributo (ABAC)

A prática de criar permissões minuciosas com base nos atributos do usuário, como departamento, cargo e nome da equipe. Para obter mais informações, consulte [ABAC AWS](#) na documentação AWS Identity and Access Management (IAM).

fonte de dados autorizada

Um local onde você armazena a versão principal dos dados, que é considerada a fonte de informações mais confiável. Você pode copiar dados da fonte de dados autorizada para outros locais com o objetivo de processar ou modificar os dados, como anonimizá-los, redigi-los ou pseudonimizá-los.

Zona de disponibilidade

Um local distinto dentro de um Região da AWS que está isolado de falhas em outras zonas de disponibilidade e fornece conectividade de rede barata e de baixa latência a outras zonas de disponibilidade na mesma região.

AWS Estrutura de adoção da nuvem (AWS CAF)

Uma estrutura de diretrizes e melhores práticas AWS para ajudar as organizações a desenvolver um plano eficiente e eficaz para migrar com sucesso para a nuvem. AWS O CAF organiza a orientação em seis áreas de foco chamadas perspectivas: negócios, pessoas, governança, plataforma, segurança e operações. As perspectivas de negócios, pessoas e governança têm como foco habilidades e processos de negócios; as perspectivas de plataforma, segurança e operações concentram-se em habilidades e processos técnicos. Por exemplo, a perspectiva das pessoas tem como alvo as partes interessadas que lidam com recursos humanos (RH), funções de pessoal e gerenciamento de pessoal. Nessa perspectiva, o AWS CAF fornece orientação para desenvolvimento, treinamento e comunicação de pessoas para ajudar a preparar a organização para a adoção bem-sucedida da nuvem. Para obter mais informações, consulte o [site da AWS CAF](#) e o [whitepaper da AWS CAF](#).

AWS Estrutura de qualificação da carga de trabalho (AWS WQF)

Uma ferramenta que avalia as cargas de trabalho de migração do banco de dados, recomenda estratégias de migração e fornece estimativas de trabalho. AWS O WQF está incluído com AWS

Schema Conversion Tool (AWS SCT). Ela analisa esquemas de banco de dados e objetos de código, código de aplicações, dependências e características de performance, além de fornecer relatórios de avaliação.

B

bot malicioso

Um [bot](#) destinado a causar disrupção ou danos a indivíduos ou organizações.

BCP

Veja [planejamento de continuidade de negócios](#)

gráfico de comportamento

Uma visualização unificada e interativa do comportamento e das interações de recursos ao longo do tempo. É possível usar um gráfico de comportamento com o Amazon Detective para examinar tentativas de login malsucedidas, chamadas de API suspeitas e ações similares. Para obter mais informações, consulte [Dados em um gráfico de comportamento](#) na documentação do Detective.

sistema big-endian

Um sistema que armazena o byte mais significativo antes. Veja também [endianness](#).

classificação binária

Um processo que prevê um resultado binário (uma de duas classes possíveis). Por exemplo, seu modelo de ML pode precisar prever problemas como “Este e-mail é ou não é spam?” ou “Este produto é um livro ou um carro?”

filtro de bloom

Uma estrutura de dados probabilística e eficiente em termos de memória que é usada para testar se um elemento é membro de um conjunto.

blue/green deployment (implantação azul/verde)

Uma estratégia de implantação em que você cria dois ambientes separados, mas idênticos. Você executa a versão atual da aplicação em um ambiente (azul) e a nova versão da aplicação no outro ambiente (verde). Essa estratégia ajuda você a reverter rapidamente com o mínimo de impacto.

bot

Uma aplicação de software que executa tarefas automatizadas na internet e simula a atividade ou interação humana. Alguns bots são úteis ou benéficos, como crawlers da web que indexam informações na internet. Outros bots, conhecidos como bots maliciosos, têm como objetivo causar interrupção ou danos a indivíduos ou organizações.

botnet

Redes de [bots](#) infectadas por [malware](#) e sob o controle de uma única parte, conhecidas como bot herder ou operador de bots. Os botnets são o mecanismo mais conhecido para escalar bots e seu impacto.

ramo

Uma área contida de um repositório de código. A primeira ramificação criada em um repositório é a ramificação principal. Você pode criar uma nova ramificação a partir de uma ramificação existente e, em seguida, desenvolver recursos ou corrigir bugs na nova ramificação. Uma ramificação que você cria para gerar um recurso é comumente chamada de ramificação de recurso. Quando o recurso estiver pronto para lançamento, você mesclará a ramificação do recurso de volta com a ramificação principal. Para obter mais informações, consulte [Sobre filiais](#) (GitHub documentação).

Acesso de emergência

Em circunstâncias excepcionais e por meio de um processo aprovado, um meio rápido para um usuário obter acesso a um Conta da AWS que ele normalmente não tem permissão para acessar. Para obter mais informações, consulte o indicador [Implement break-glass procedures](#) nas orientações do AWS Well-Architected.

estratégia brownfield

A infraestrutura existente em seu ambiente. Ao adotar uma estratégia brownfield para uma arquitetura de sistema, você desenvolve a arquitetura de acordo com as restrições dos sistemas e da infraestrutura atuais. Se estiver expandindo a infraestrutura existente, poderá combinar as estratégias brownfield e [greenfield](#).

cache do buffer

A área da memória em que os dados acessados com mais frequência são armazenados.

capacidade de negócios

O que uma empresa faz para gerar valor (por exemplo, vendas, atendimento ao cliente ou marketing). As arquiteturas de microsserviços e as decisões de desenvolvimento podem

ser orientadas por recursos de negócios. Para obter mais informações, consulte a seção [Organizados de acordo com as capacidades de negócios](#) do whitepaper [Executar microsserviços containerizados na AWS](#).

planejamento de continuidade de negócios (BCP)

Um plano que aborda o impacto potencial de um evento disruptivo, como uma migração em grande escala, nas operações e permite que uma empresa retome as operações rapidamente.

C

CAF

Veja [AWS Cloud Adoption Framework](#).

implantação canário

O lançamento lento e incremental de uma versão para usuários finais. Quando estiver confiante, você implanta a nova versão e substitui a versão atual por completo.

CCoE

Veja [Centro de Excelência da Nuvem](#).

CDC

Veja [captura de dados de alteração](#).

captura de dados de alterações (CDC)

O processo de rastrear alterações em uma fonte de dados, como uma tabela de banco de dados, e registrar metadados sobre a alteração. É possível usar o CDC para várias finalidades, como auditar ou replicar alterações em um sistema de destino para manter a sincronização.

engenharia do caos

Introduzir intencionalmente falhas ou eventos disruptivos para testar a resiliência de um sistema. Você pode usar [AWS Fault Injection Service \(AWS FIS\)](#) para realizar experimentos que estressam suas AWS cargas de trabalho e avaliar sua resposta.

CI/CD

Veja [integração e entrega contínuas](#).

classificação

Um processo de categorização que ajuda a gerar previsões. Os modelos de ML para problemas de classificação predizem um valor discreto. Os valores discretos são sempre diferentes uns dos outros. Por exemplo, um modelo pode precisar avaliar se há ou não um carro em uma imagem.

criptografia no lado do cliente

Criptografia de dados localmente, antes que o alvo os AWS service (Serviço da AWS) receba.

Centro de excelência em nuvem (CCoE)

Uma equipe multidisciplinar que impulsiona os esforços de adoção da nuvem em toda a organização, incluindo o desenvolvimento de práticas recomendadas de nuvem, a mobilização de recursos, o estabelecimento de cronogramas de migração e a liderança da organização em transformações em grande escala. Para obter mais informações, consulte as [publicações CCo E](#) no blog de estratégia Nuvem AWS corporativa.

computação em nuvem

A tecnologia de nuvem normalmente usada para armazenamento de dados remoto e gerenciamento de dispositivos de IoT. A computação em nuvem é normalmente conectada à tecnologia de [computação de borda](#).

modelo operacional em nuvem

Em uma organização de TI, o modelo operacional usado para criar, amadurecer e otimizar um ou mais ambientes de nuvem. Para obter mais informações, consulte [Criar seu modelo operacional de nuvem](#).

estágios de adoção da nuvem

As quatro fases pelas quais as organizações normalmente passam ao migrar para a Nuvem AWS:

- Projeto: executar alguns projetos relacionados à nuvem para fins de prova de conceito e aprendizado
- Fundação — Fazer investimentos fundamentais para escalar sua adoção da nuvem (por exemplo, criar uma landing zone, definir um CCo E, estabelecer um modelo de operações)
- Migração: migrar aplicações individuais
- Reinvenção: otimizar produtos e serviços e inovar na nuvem

Esses estágios foram definidos por Stephen Orban na postagem do blog [The Journey Toward Cloud-First & the Stages of Adoption](#) no blog de estratégia Nuvem AWS empresarial. Para obter

informações sobre como eles se relacionam com a estratégia de AWS migração, consulte o [guia de preparação para migração](#).

CMDB

Veja [banco de dados de gerenciamento de configuração](#).

repositório de código

Um local onde o código-fonte e outros ativos, como documentação, amostras e scripts, são armazenados e atualizados por meio de processos de controle de versão. Os repositórios de nuvem comuns incluem o GitHub ou o Bitbucket Cloud. Cada versão do código é chamada de ramificação. Em uma estrutura de microsserviços, cada repositório é dedicado a uma única peça de funcionalidade. Um único pipeline de CI/CD pode usar vários repositórios.

cache frio

Um cache de buffer que está vazio, não está bem preenchido ou contém dados obsoletos ou irrelevantes. Isso afeta a performance porque a instância do banco de dados deve ler da memória principal ou do disco, um processo que é mais lento do que a leitura do cache do buffer.

dados frios

Dados que raramente são acessados e geralmente são históricos. Ao consultar esse tipo de dados, consultas lentas geralmente são aceitáveis. Mover esses dados para níveis ou classes de armazenamento de baixo desempenho e menos caros pode reduzir os custos.

visão computacional (CV)

Um campo de [IA](#) que usa machine learning para analisar e extrair informações de formatos visuais, como vídeos e imagens digitais. Por exemplo, a Amazon SageMaker AI fornece algoritmos de processamento de imagem para CV.

desvio de configuração

Em uma workload, uma alteração de configuração em relação ao estado esperado. Isso pode fazer com que a workload se torne incompatível e, normalmente, é gradual e não intencional.

banco de dados de gerenciamento de configuração (CMDB)

Um repositório que armazena e gerencia informações sobre um banco de dados e seu ambiente de TI, incluindo componentes de hardware e software e suas configurações. Normalmente, os dados de um CMDB são usados no estágio de descoberta e análise do portfólio da migração.

pacote de conformidade

Uma coleção de AWS Config regras e ações de remediação que você pode montar para personalizar suas verificações de conformidade e segurança. Você pode implantar um pacote de conformidade como uma entidade única em uma Conta da AWS região ou em uma organização usando um modelo YAML. Para obter mais informações, consulte [Pacotes de conformidade na documentação](#). AWS Config

integração contínua e entrega contínua (CI/CD)

O processo de automatizar os estágios de origem, criação, teste, preparação e produção do processo de lançamento do software. CI/CD é comumente descrito como um pipeline. CI/CD pode ajudá-lo a automatizar processos, melhorar a produtividade, melhorar a qualidade do código e entregar com mais rapidez. Para obter mais informações, consulte [Benefícios da entrega contínua](#). CD também pode significar implantação contínua. Para obter mais informações, consulte [Entrega contínua versus implantação contínua](#).

CV

Veja [visão computacional](#).

D

dados em repouso

Dados estacionários em sua rede, por exemplo, dados que estão em um armazenamento.

classificação de dados

Um processo para identificar e categorizar os dados em sua rede com base em criticalidade e confidencialidade. É um componente crítico de qualquer estratégia de gerenciamento de riscos de segurança cibernética, pois ajuda a determinar os controles adequados de proteção e retenção para os dados. A classificação de dados é um componente do pilar de segurança no AWS Well-Architected Framework. Para obter mais informações, consulte [Classificação de dados](#).

desvio de dados

Uma variação significativa entre os dados de produção e os dados usados para treinar um modelo de ML ou uma alteração significativa nos dados de entrada ao longo do tempo. O desvio de dados pode reduzir a qualidade geral, a precisão e a imparcialidade das previsões do modelo de ML.

dados em trânsito

Dados que estão se movendo ativamente pela sua rede, como entre os recursos da rede.

data mesh

Um framework de arquitetura que fornece propriedade de dados distribuída e descentralizada com gerenciamento e governança centralizados.

minimização de dados

O princípio de coletar e processar apenas os dados estritamente necessários. Praticar a minimização de dados no Nuvem AWS pode reduzir os riscos de privacidade, os custos e a pegada de carbono de sua análise.

perímetro de dados

Um conjunto de proteções preventivas em seu AWS ambiente que ajudam a garantir que somente identidades confiáveis acessem recursos confiáveis das redes esperadas. Para obter mais informações, consulte [Construindo um perímetro de dados em AWS](#)

pré-processamento de dados

A transformação de dados brutos em um formato que seja facilmente analisado por seu modelo de ML. O pré-processamento de dados pode significar a remoção de determinadas colunas ou linhas e o tratamento de valores ausentes, inconsistentes ou duplicados.

proveniência dos dados

O processo de rastrear a origem e o histórico dos dados ao longo de seu ciclo de vida, por exemplo, como os dados foram gerados, transmitidos e armazenados.

titular dos dados

Um indivíduo cujos dados estão sendo coletados e processados.

data warehouse

Um sistema de gerenciamento de dados compatível com business intelligence, como analytics. Os data warehouses geralmente contêm grandes quantidades de dados históricos e geralmente são usados para consultas e análises.

linguagem de definição de dados (DDL)

Instruções ou comandos para criar ou modificar a estrutura de tabelas e objetos em um banco de dados.

linguagem de manipulação de dados (DML)

Instruções ou comandos para modificar (inserir, atualizar e excluir) informações em um banco de dados.

DDL

Veja [linguagem de definição de banco de dados](#).

deep ensemble

A combinação de vários modelos de aprendizado profundo para gerar previsões. Os deep ensembles podem ser usados para produzir uma previsão mais precisa ou para estimar a incerteza nas previsões.

Aprendizado profundo

Um subcampo do ML que usa várias camadas de redes neurais artificiais para identificar o mapeamento entre os dados de entrada e as variáveis-alvo de interesse.

defense-in-depth

Uma abordagem de segurança da informação na qual uma série de mecanismos e controles de segurança são cuidadosamente distribuídos por toda a rede de computadores para proteger a confidencialidade, a integridade e a disponibilidade da rede e dos dados nela contidos. Ao adotar essa estratégia AWS, você adiciona vários controles em diferentes camadas da AWS Organizations estrutura para ajudar a proteger os recursos. Por exemplo, uma defense-in-depth abordagem pode combinar autenticação multifatorial, segmentação de rede e criptografia.

administrador delegado

Em AWS Organizations, um serviço compatível pode registrar uma conta de AWS membro para administrar as contas da organização e gerenciar as permissões desse serviço. Essa conta é chamada de administrador delegado para esse serviço. Para obter mais informações e uma lista de serviços compatíveis, consulte [Serviços que funcionam com o AWS Organizations](#) na documentação do AWS Organizations .

implantação

O processo de criar uma aplicação, novos recursos ou correções de código disponíveis no ambiente de destino. A implantação envolve a implementação de mudanças em uma base de código e, em seguida, a criação e execução dessa base de código nos ambientes da aplicação

ambiente de desenvolvimento

Veja [ambiente](#).

controle detectivo

Um controle de segurança projetado para detectar, registrar e alertar após a ocorrência de um evento. Esses controles são uma segunda linha de defesa, alertando você sobre eventos de segurança que contornaram os controles preventivos em vigor. Para obter mais informações, consulte [Controles detectivos](#) em Como implementar controles de segurança na AWS.

mapeamento do fluxo de valor de desenvolvimento (DVSM)

Um processo usado para identificar e priorizar restrições que afetam negativamente a velocidade e a qualidade em um ciclo de vida de desenvolvimento de software. O DVSM estende o processo de mapeamento do fluxo de valor originalmente projetado para práticas de manufatura enxuta. Ele se concentra nas etapas e equipes necessárias para criar e movimentar valor por meio do processo de desenvolvimento de software.

gêmeo digital

Uma representação virtual de um sistema real, como um prédio, fábrica, equipamento industrial ou linha de produção. Os gêmeos digitais oferecem suporte à manutenção preditiva, ao monitoramento remoto e à otimização da produção.

tabela de dimensões

Em um [esquema em estrela](#), uma tabela menor que contém atributos de dados sobre dados quantitativos em uma tabela de fatos. Os atributos da tabela de dimensões geralmente são campos de texto ou números discretos que se comportam como texto. Esses atributos normalmente são usados para restringir consultas, filtrar e rotular conjuntos de resultados.

desastre

Um evento que impede que uma workload ou sistema cumpra seus objetivos de negócios em seu local principal de implantação. Esses eventos podem ser desastres naturais, falhas técnicas ou o resultado de ações humanas, como configuração incorreta não intencional ou ataque de malware.

Recuperação de desastres (RD)

A estratégia e o processo que você usa para minimizar o tempo de inatividade e a perda de dados causados por um [desastre](#). Para obter mais informações, consulte [Recuperação de desastres de cargas de trabalho em AWS: Recuperação na nuvem no AWS Well-Architected Framework](#).

DML

Veja [linguagem de manipulação de banco de dados](#).

design orientado por domínio

Uma abordagem ao desenvolvimento de um sistema de software complexo conectando seus componentes aos domínios em evolução, ou principais metas de negócios, atendidos por cada componente. Esse conceito foi introduzido por Eric Evans em seu livro, *Design orientado por domínio: lidando com a complexidade no coração do software* (Boston: Addison-Wesley Professional, 2003). Para obter informações sobre como usar o design orientado por domínio com o padrão strangler fig, consulte [Modernizar incrementalmente os serviços web herdados do Microsoft ASP.NET \(ASMX\) usando contêineres e o Amazon API Gateway](#).

DR

Veja [recuperação de desastres](#).

Deteção da oscilação

Rastreamento de desvios de uma configuração de linha de base. Por exemplo, você pode usar AWS CloudFormation para [detectar desvios nos recursos do sistema](#) ou AWS Control Tower para [detectar mudanças em seu landing zone](#) que possam afetar a conformidade com os requisitos de governança.

DVSM

Veja [mapeamento do fluxo de valor de desenvolvimento](#).

E

EDA

Veja [análise exploratória de dados](#).

EDI

Veja [intercâmbio eletrônico de dados](#).

computação de borda

A tecnologia que aumenta o poder computacional de dispositivos inteligentes nas bordas de uma rede de IoT. Quando comparada com a [computação em nuvem](#), a computação de borda pode reduzir a latência da comunicação e melhorar o tempo de resposta.

intercâmbio eletrônico de dados (EDI)

A troca automatizada de documentos comerciais entre organizações. Para obter mais informações, consulte [O que é EDI \(Intercâmbio eletrônico de dados\)?](#).

criptografia

Um processo de computação que transforma dados de texto simples, legíveis por humanos, em texto cifrado.

chave de criptografia

Uma sequência criptográfica de bits aleatórios que é gerada por um algoritmo de criptografia. As chaves podem variar em tamanho, e cada chave foi projetada para ser imprevisível e exclusiva.

endianismo

A ordem na qual os bytes são armazenados na memória do computador. Os sistemas big-endian armazenam o byte mais significativo antes. Os sistemas little-endian armazenam o byte menos significativo antes.

endpoint

Veja [endpoint de serviço](#).

serviço de endpoint

Um serviço que pode ser hospedado em uma nuvem privada virtual (VPC) para ser compartilhado com outros usuários. Você pode criar um serviço de endpoint com AWS PrivateLink e conceder permissões a outros diretores Contas da AWS ou a AWS Identity and Access Management (IAM). Essas contas ou entidades principais podem se conectar ao serviço de endpoint de maneira privada criando endpoints da VPC de interface. Para obter mais informações, consulte [Criar um serviço de endpoint](#) na documentação do Amazon Virtual Private Cloud (Amazon VPC).

planejamento de recursos empresariais (ERP)

Um sistema que automatiza e gerencia os principais processos de negócios (como contabilidade, [MES](#) e gerenciamento de projetos) para uma empresa.

criptografia envelopada

O processo de criptografar uma chave de criptografia com outra chave de criptografia. Para obter mais informações, consulte [Criptografia de envelope](#) na documentação AWS Key Management Service (AWS KMS).

ambiente

Uma instância de uma aplicação em execução. Estes são tipos comuns de ambientes na computação em nuvem:

- ambiente de desenvolvimento: uma instância de uma aplicação em execução que está disponível somente para a equipe principal responsável pela manutenção da aplicação. Ambientes de desenvolvimento são usados para testar mudanças antes de promovê-las para ambientes superiores. Esse tipo de ambiente às vezes é chamado de ambiente de teste.
- ambientes inferiores: todos os ambientes de desenvolvimento para uma aplicação, como aqueles usados para compilações e testes iniciais.
- ambiente de produção: uma instância de uma aplicação em execução que os usuários finais podem acessar. Em um CI/CD pipeline, o ambiente de produção é o último ambiente de implantação.
- ambientes superiores: todos os ambientes que podem ser acessados por usuários que não sejam a equipe principal de desenvolvimento. Isso pode incluir um ambiente de produção, ambientes de pré-produção e ambientes para testes de aceitação do usuário.

epic

Em metodologias ágeis, categorias funcionais que ajudam a organizar e priorizar seu trabalho. Os epics fornecem uma descrição de alto nível dos requisitos e das tarefas de implementação. Por exemplo, os épicos de segurança AWS da CAF incluem gerenciamento de identidade e acesso, controles de detetive, segurança de infraestrutura, proteção de dados e resposta a incidentes. Para obter mais informações sobre epics na estratégia de migração da AWS, consulte o [guia de implementação do programa](#).

ERP

Veja [planejamento de recursos empresariais](#).

análise exploratória de dados (EDA)

O processo de analisar um conjunto de dados para entender suas principais características. Você coleta ou agrega dados e, em seguida, realiza investigações iniciais para encontrar padrões, detectar anomalias e verificar suposições. O EDA é realizado por meio do cálculo de estatísticas resumidas e da criação de visualizações de dados.

F

tabela de fatos

A tabela central em um [esquema em estrela](#). Ela armazena dados quantitativos sobre as operações comerciais. Normalmente, uma tabela de fatos contém dois tipos de colunas: as que contêm medidas e as que contêm uma chave externa para uma tabela de dimensões.

Antecipar-se à falha

Uma filosofia que usa testes frequentes e incrementais para reduzir o ciclo de vida do desenvolvimento. É uma parte essencial de uma abordagem ágil.

delimitação de isolamento contra falhas

No Nuvem AWS, um limite, como uma zona de disponibilidade, Região da AWS um plano de controle ou um plano de dados, que limita o efeito de uma falha e ajuda a melhorar a resiliência das cargas de trabalho. Para obter mais informações, consulte [AWS Fault Isolation Boundaries](#).

ramificação de recursos

Veja [ramificação](#).

recursos

Os dados de entrada usados para fazer uma previsão. Por exemplo, em um contexto de manufatura, os recursos podem ser imagens capturadas periodicamente na linha de fabricação.

importância do recurso

O quanto um recurso é importante para as previsões de um modelo. Isso geralmente é expresso como uma pontuação numérica que pode ser calculada por meio de várias técnicas, como Shapley Additive Explanations (SHAP) e gradientes integrados. Para obter mais informações, consulte [Interpretabilidade do modelo de aprendizado de máquina com AWS](#).

transformação de recursos

O processo de otimizar dados para o processo de ML, incluindo enriquecer dados com fontes adicionais, escalar valores ou extrair vários conjuntos de informações de um único campo de dados. Isso permite que o modelo de ML se beneficie dos dados. Por exemplo, se a data “2021-05-27 00:15:37” for dividida em “2021”, “maio”, “quinta” e “15”, isso poderá ajudar o algoritmo de aprendizado a aprender padrões diferenciados associados a diferentes componentes de dados.

prompt few shot

Fornecer a um [LLM](#) um pequeno número de exemplos que demonstram a tarefa e o resultado desejado antes de solicitar que ele execute uma tarefa semelhante. Essa técnica é uma aplicação do aprendizado em contexto, em que os modelos aprendem com exemplos (shots) incorporados aos prompts. Prompts few-shot podem ser eficazes para tarefas que exigem formatação, raciocínio ou conhecimento de domínio específicos. Veja também [prompts zero-shot](#).

FGAC

Veja [controle de acesso refinado](#).

Controle de acesso refinado (FGAC)

O uso de várias condições para permitir ou negar uma solicitação de acesso.

migração flash-cut

Um método de migração de banco de dados que usa replicação contínua de dados via [captura de dados de alteração](#) para migrar os dados no menor tempo possível, em vez de usar uma abordagem em fases. O objetivo é reduzir ao mínimo o tempo de inatividade.

FM

Veja [modelo de base](#).

modelo de base (FM)

Uma grande rede neural de aprendizado profundo que vem treinando em grandes conjuntos de dados generalizados e não rotulados. FMs são capazes de realizar uma ampla variedade de tarefas gerais, como entender a linguagem, gerar texto e imagens e conversar em linguagem natural. Para obter mais informações, consulte [O que são modelos de base?](#).

G

IA generativa

Um subconjunto de modelos de [IA](#) que foram treinados em grandes quantidades de dados e que podem usar um simples prompt de texto para criar novos artefatos e conteúdo, como imagens, vídeos, texto e áudio. Para obter mais informações, consulte [O que é IA generativa?](#).

bloqueio geográfico

Veja [restrições geográficas](#).

restrições geográficas (bloqueio geográfico)

Na Amazon CloudFront, uma opção para impedir que usuários em países específicos acessem distribuições de conteúdo. É possível usar uma lista de permissões ou uma lista de bloqueios para especificar países aprovados e banidos. Para obter mais informações, consulte [Restringir a distribuição geográfica do seu conteúdo](#) na CloudFront documentação.

Fluxo de trabalho do GitFlow

Uma abordagem na qual ambientes inferiores e superiores usam ramificações diferentes em um repositório de código-fonte. O fluxo de trabalho do Gitflow é considerado legado, e o [fluxo de trabalho trunk-based](#) é a abordagem moderna e preferencial.

golden image

Um snapshot de um sistema ou software usado como modelo para implantar novas instâncias desse sistema ou software. Por exemplo, na manufatura, uma golden image pode ser usada para provisionar software em vários dispositivos e ajudar a melhorar a velocidade, a escalabilidade e a produtividade nas operações de fabricação de dispositivos.

estratégia greenfield

A ausência de infraestrutura existente em um novo ambiente. Ao adotar uma estratégia greenfield para uma arquitetura de sistema, é possível selecionar todas as novas tecnologias sem a restrição da compatibilidade com a infraestrutura existente, também conhecida como [brownfield](#). Se estiver expandindo a infraestrutura existente, poderá combinar as estratégias brownfield e greenfield.

barreira de proteção

Uma regra de alto nível que ajuda a governar recursos, políticas e conformidade em todas as unidades organizacionais (OUs). Barreiras de proteção preventivas impõem políticas para garantir o alinhamento a padrões de conformidade. Elas são implementadas usando políticas de controle de serviço e limites de permissões do IAM. Barreiras de proteção detectivas detectam violações de políticas e problemas de conformidade e geram alertas para remediação. Eles são implementados usando AWS Config, AWS Security Hub CSPM, Amazon GuardDuty AWS Trusted Advisor, Amazon Inspector e verificações personalizadas AWS Lambda .

H

HA

Veja [alta disponibilidade](#).

migração heterogênea de bancos de dados

Migrar seu banco de dados de origem para um banco de dados de destino que usa um mecanismo de banco de dados diferente (por exemplo, Oracle para Amazon Aurora). A migração heterogênea geralmente faz parte de um esforço de redefinição da arquitetura, e converter

o esquema pode ser uma tarefa complexa. [O AWS fornece o AWS SCT](#) para ajudar nas conversões de esquemas.

alta disponibilidade (HA)

A capacidade de uma workload operar continuamente, sem intervenção, em caso de desafios ou desastres. Os sistemas AH são projetados para realizar o failover automático, oferecer consistentemente desempenho de alta qualidade e lidar com diferentes cargas e falhas com impacto mínimo no desempenho.

modernização de historiador

Uma abordagem usada para modernizar e atualizar os sistemas de tecnologia operacional (OT) para melhor atender às necessidades do setor de manufatura. Um historiador é um tipo de banco de dados usado para coletar e armazenar dados de várias fontes em uma fábrica.

dados de hold-out

Uma parte dos dados históricos rotulados que são retidos de um conjunto de dados usado para treinar um modelo de [machine learning](#). Você pode usar dados de hold-out para avaliar a performance do modelo comparando as previsões do modelo com os dados de retenção.

migração homogênea de bancos de dados

Migrar seu banco de dados de origem para um banco de dados de destino que compartilha o mesmo mecanismo de banco de dados (por exemplo, Microsoft SQL Server para Amazon RDS para SQL Server). A migração homogênea geralmente faz parte de um esforço de redefinição da hospedagem ou da plataforma. É possível usar utilitários de banco de dados nativos para migrar o esquema.

dados quentes

Dados acessados com frequência, como dados em tempo real ou dados translacionais recentes. Esses dados normalmente exigem uma camada ou classe de armazenamento de alto desempenho para fornecer respostas rápidas às consultas.

hotfix

Uma correção urgente para um problema crítico em um ambiente de produção. Devido à sua urgência, um hotfix geralmente é feito fora do fluxo de trabalho normal de DevOps lançamento.

período de hipercuidados

Imediatamente após a substituição, o período em que uma equipe de migração gerencia e monitora as aplicações migradas na nuvem para resolver quaisquer problemas. Normalmente,

a duração desse período é de 1 a 4 dias. No final do período de hipercuidados, a equipe de migração normalmente transfere a responsabilidade pelas aplicações para a equipe de operações de nuvem.

eu

laC

Veja [infraestrutura como código](#).

Política baseada em identidade

Uma política anexada a um ou mais diretores do IAM que define suas permissões no Nuvem AWS ambiente.

aplicação ociosa

Uma aplicação que tem um uso médio de CPU e memória entre 5 e 20% em um período de 90 dias. Em um projeto de migração, é comum retirar essas aplicações ou retê-las on-premises.

IloT

Veja [Internet das Coisas Industrial](#).

infraestrutura imutável

Um modelo que implanta uma nova infraestrutura para workloads de produção em vez de atualizar, aplicar patches ou modificar a infraestrutura existente. Infraestruturas imutáveis são inerentemente mais consistentes, confiáveis e preditivas do que [infraestruturas mutáveis](#). Para obter mais informações, consulte a prática recomendada [Implantar usando infraestrutura imutável](#) no AWS Well-Architected Framework.

VPC de entrada (admissão)

Em uma arquitetura de AWS várias contas, uma VPC que aceita, inspeciona e roteia conexões de rede de fora de um aplicativo. A [Arquitetura de Referência de AWS Segurança](#) recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo e a Internet em geral.

migração incremental

Uma estratégia de substituição na qual você migra a aplicação em pequenas partes, em vez de realizar uma única substituição completa. Por exemplo, é possível mover inicialmente

apenas alguns microsserviços ou usuários para o novo sistema. Depois de verificar se tudo está funcionando corretamente, mova os microsserviços ou usuários adicionais de forma incremental até poder descomissionar seu sistema herdado. Essa estratégia reduz os riscos associados a migrações de grande porte.

Indústria 4.0

Um termo que foi introduzido por [Klaus Schwab](#) em 2016 para se referir à modernização dos processos de manufatura por meio de avanços em conectividade, dados em tempo real, automação, analytics e IA/ML.

infraestrutura

Todos os recursos e ativos contidos no ambiente de uma aplicação.

Infraestrutura como código (IaC)

O processo de provisionamento e gerenciamento da infraestrutura de uma aplicação por meio de um conjunto de arquivos de configuração. A IaC foi projetada para ajudar você a centralizar o gerenciamento da infraestrutura, padronizar recursos e escalar rapidamente para que novos ambientes sejam reproduzíveis, confiáveis e consistentes.

Internet industrial das coisas (IIoT)

O uso de sensores e dispositivos conectados à Internet nos setores industriais, como manufatura, energia, automotivo, saúde, ciências biológicas e agricultura. Para obter mais informações, consulte [Criando uma estratégia de transformação digital industrial da Internet das Coisas \(IIoT\)](#).

VPC de inspeção

Em uma arquitetura de AWS várias contas, uma VPC centralizada que gerencia as inspeções do tráfego de rede entre VPCs (na mesma ou em diferentes Regiões da AWS) a Internet e as redes locais. A [Arquitetura de Referência de AWS Segurança](#) recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo e a Internet em geral.

Internet das coisas (IoT)

A rede de objetos físicos conectados com sensores ou processadores incorporados que se comunicam com outros dispositivos e sistemas pela Internet ou por uma rede de comunicação local. Para obter mais informações, consulte [O que é IoT?](#)

interpretabilidade

Uma característica de um modelo de machine learning que descreve o grau em que um ser humano pode entender como as previsões do modelo dependem de suas entradas. Para obter mais informações, consulte [Interpretabilidade do modelo de aprendizado de máquina com AWS](#).

IoT

Veja [Internet das Coisas](#).

Biblioteca de informações de TI (ITIL)

Um conjunto de práticas recomendadas para fornecer serviços de TI e alinhar esses serviços a requisitos de negócios. A ITIL fornece a base para o ITSM.

Gerenciamento de serviços de TI (ITSM)

Atividades associadas a design, implementação, gerenciamento e suporte de serviços de TI para uma organização. Para obter informações sobre a integração de operações em nuvem com ferramentas de ITSM, consulte o [guia de integração de operações](#).

ITIL

Veja [biblioteca de informações de TI](#).

ITSM

Veja [gerenciamento de serviços de TI](#).

L

controle de acesso baseado em etiqueta (LBAC)

Uma implementação do controle de acesso obrigatório (MAC) em que os usuários e os dados em si recebem explicitamente um valor de etiqueta de segurança. A interseção entre a etiqueta de segurança do usuário e a etiqueta de segurança dos dados determina quais linhas e colunas podem ser vistas pelo usuário.

zona de pouso

Uma landing zone é um AWS ambiente bem arquitetado, com várias contas, escalável e seguro. Um ponto a partir do qual suas organizações podem iniciar e implantar rapidamente workloads e aplicações com confiança em seu ambiente de segurança e infraestrutura. Para obter mais

informações sobre zonas de pouso, consulte [Configurar um ambiente da AWS com várias contas seguro e escalável](#).

grande modelo de linguagem (LLM)

Um modelo de [IA](#) de aprendizado profundo pré-treinado em uma grande quantidade de dados. Um LLM pode realizar várias tarefas, como responder a perguntas, resumir documentos, traduzir texto para outros idiomas e completar frases. Para obter mais informações, consulte [O que são LLMs](#).

migração de grande porte

Uma migração de 300 servidores ou mais.

LBAC

Veja [controle de acesso baseado em rótulo](#).

privilégio mínimo

A prática recomendada de segurança de conceder as permissões mínimas necessárias para executar uma tarefa. Para obter mais informações, consulte [Aplicar permissões de privilégios mínimos](#) na documentação do IAM.

mover sem alterações (lift-and-shift)

Veja [7 Rs](#).

sistema little-endian

Um sistema que armazena o byte menos significativo antes. Veja também [endianness](#).

LLM

Veja [grande modelo de linguagem](#).

ambientes inferiores

Veja [ambiente](#).

M

machine learning (ML)

Um tipo de inteligência artificial que usa algoritmos e técnicas para reconhecimento e aprendizado de padrões. O ML analisa e aprende com dados gravados, por exemplo, dados da

Internet das Coisas (IoT), para gerar um modelo estatístico baseado em padrões. Para obter mais informações, consulte [Machine learning](#).

ramificação principal

Veja [ramificação](#).

Malware

Software projetado para comprometer a segurança ou a privacidade do computador. O malware pode interromper os sistemas do computador, vaziar informações sensíveis ou obter acesso não autorizado. Exemplos de malware incluem vírus, worms, ransomware, cavalos de Troia, spyware e keyloggers.

Serviços gerenciados

Serviços da AWS para o qual AWS opera a camada de infraestrutura, o sistema operacional e as plataformas, e você acessa os endpoints para armazenar e recuperar dados. O Amazon Simple Storage Service (Amazon S3) e o Amazon DynamoDB são exemplos de serviços gerenciados. Eles também são conhecidos como serviços abstraídos.

sistema de execução de manufatura (MES)

Um sistema de software para rastrear, monitorar, documentar e controlar processos de produção que convertem matérias-primas em produtos acabados no chão de fábrica.

MAP

Veja [Programa de Aceleração da Migração](#).

mecanismo

Um processo completo em que você cria uma ferramenta, impulsiona a adoção da ferramenta e, em seguida, inspeciona os resultados para fazer ajustes. Um mecanismo é um ciclo que se reforça e se aprimora à medida que opera. Para obter mais informações, consulte [Construindo mecanismos](#) no AWS Well-Architected Framework.

conta de membro

Todos, Contas da AWS exceto a conta de gerenciamento, que fazem parte de uma organização em AWS Organizations. Uma conta só pode ser membro de uma organização de cada vez.

MES

Veja [sistema de execução de manufatura](#).

Transporte de Telemetria de Enfileiramento de Mensagens (MQTT)

[Um protocolo de comunicação leve machine-to-machine \(M2M\), baseado no padrão de publicação/assinatura, para dispositivos de IoT com recursos limitados.](#)

microsserviço

Um serviço pequeno e independente que se comunica de forma bem definida APIs e normalmente é de propriedade de equipes pequenas e independentes. Por exemplo, um sistema de seguradora pode incluir microsserviços que mapeiam as capacidades comerciais, como vendas ou marketing, ou subdomínios, como compras, reclamações ou análises. Os benefícios dos microsserviços incluem agilidade, escalabilidade flexível, fácil implantação, código reutilizável e resiliência. Para obter mais informações, consulte [Integração de microsserviços usando serviços sem AWS servidor.](#)

arquitetura de microsserviços

Uma abordagem à criação de aplicações com componentes independentes que executam cada processo de aplicação como um microsserviço. Esses microsserviços se comunicam por meio de uma interface bem definida usando leveza. APIs Cada microsserviço nessa arquitetura pode ser atualizado, implantado e escalado para atender à demanda por funções específicas de uma aplicação. Para obter mais informações, consulte [Implementação de microsserviços em. AWS](#)

Programa de Aceleração da Migração (MAP)

Um AWS programa que fornece suporte de consultoria, treinamento e serviços para ajudar as organizações a criar uma base operacional sólida para migrar para a nuvem e ajudar a compensar o custo inicial das migrações. O MAP inclui uma metodologia de migração para executar migrações legadas de forma metódica e um conjunto de ferramentas para automatizar e acelerar cenários comuns de migração.

migração em escala

O processo de mover a maior parte do portfólio de aplicações para a nuvem em ondas, com mais aplicações sendo movidas em um ritmo mais rápido a cada onda. Essa fase usa as práticas recomendadas e lições aprendidas nas fases anteriores para implementar uma fábrica de migração de equipes, ferramentas e processos para agilizar a migração de workloads por meio de automação e entrega ágeis. Esta é a terceira fase da [estratégia de migração para a AWS.](#)

fábrica de migração

Equipes multifuncionais que simplificam a migração de workloads por meio de abordagens automatizadas e ágeis. As equipes da fábrica de migração geralmente incluem operações,

analistas e proprietários de negócios, engenheiros de migração, desenvolvedores e DevOps profissionais que trabalham em sprints. Entre 20 e 50% de um portfólio de aplicações corporativas consiste em padrões repetidos que podem ser otimizados por meio de uma abordagem de fábrica. Para obter mais informações, consulte [discussão sobre fábricas de migração](#) e o [guia do Cloud Migration Factory](#) neste conjunto de conteúdo.

metadados de migração

As informações sobre a aplicação e o servidor necessárias para concluir a migração. Cada padrão de migração exige um conjunto de metadados de migração diferente. Exemplos de metadados de migração incluem a sub-rede, o grupo de segurança e AWS a conta de destino.

padrão de migração

Uma tarefa de migração repetível que detalha a estratégia de migração, o destino da migração e a aplicação ou o serviço de migração usado. Exemplo: rehoste a migração para o Amazon EC2 AWS com o Application Migration Service.

Avaliação de Portfólio para Migração (MPA)

Uma ferramenta on-line que fornece informações para validar o caso de negócios para migrar para a Nuvem AWS. O MPA fornece avaliação detalhada do portfólio (dimensionamento correto do servidor, preços, comparações de TCO, análise de custos de migração), bem como planejamento de migração (análise e coleta de dados de aplicações, agrupamento de aplicações, priorização de migração e planejamento de ondas). A [ferramenta MPA](#) (requer login) está disponível gratuitamente para todos os AWS consultores e consultores parceiros da APN.

Avaliação de Preparação para Migração (MRA)

O processo de obter insights sobre o status de prontidão de uma organização para a nuvem, identificar pontos fortes e fracos e criar um plano de ação para fechar as lacunas identificadas, usando o CAF. AWS Para mais informações, consulte o [guia de preparação para migração](#). A MRA é a primeira fase da [estratégia de migração para a AWS](#).

estratégia de migração

A abordagem usada para migrar uma workload para a Nuvem AWS. Para obter mais informações, veja a entrada [7 Rs](#) neste glossário e consulte [Mobilize sua organização para acelerar migrações em grande escala](#).

ML

Veja [machine learning](#).

modernização

Transformar uma aplicação desatualizada (herdada ou monolítica) e sua infraestrutura em um sistema ágil, elástico e altamente disponível na nuvem para reduzir custos, ganhar eficiência e aproveitar as inovações. Para obter mais informações, consulte [Strategy for modernizing applications in the Nuvem AWS](#).

avaliação de preparação para modernização

Uma avaliação que ajuda a determinar a preparação para modernização das aplicações de uma organização. Ela identifica benefícios, riscos e dependências e determina o quão bem a organização pode acomodar o estado futuro dessas aplicações. O resultado da avaliação é um esquema da arquitetura de destino, um roteiro que detalha as fases de desenvolvimento e os marcos do processo de modernização e um plano de ação para abordar as lacunas identificadas. Para obter mais informações, consulte [Evaluating modernization readiness for applications in the Nuvem AWS](#).

aplicações monolíticas (monólitos)

Aplicações que são executadas como um único serviço com processos fortemente acoplados. As aplicações monolíticas apresentam várias desvantagens. Se um recurso da aplicação apresentar um aumento na demanda, toda a arquitetura deverá ser escalada. Adicionar ou melhorar os recursos de uma aplicação monolítica também se torna mais complexo quando a base de código cresce. Para resolver esses problemas, é possível criar uma arquitetura de microsserviços. Para obter mais informações, consulte [Decompor monólitos em microsserviços](#).

MPA

Veja [Avaliação do Portfólio para Migração](#).

MQTT

Veja [Transporte de Telemetria de Enfileiramento de Mensagens](#).

classificação multiclasse

Um processo que ajuda a gerar previsões para várias classes (prevendo um ou mais de dois resultados). Por exemplo, um modelo de ML pode perguntar “Este produto é um livro, um carro ou um telefone?” ou “Qual categoria de produtos é mais interessante para este cliente?”

infraestrutura mutável

Um modelo que atualiza e modifica a infraestrutura existente para workloads de produção. Para melhorar a consistência, confiabilidade e previsibilidade, o AWS Well-Architected Framework recomenda o uso de infraestrutura [imutável](#) como uma prática recomendada.

O

OAC

Veja [controle de acesso de origem](#).

OAI

Veja [identidade de acesso de origem](#).

OCM

Veja [gerenciamento de alterações organizacionais](#).

migração offline

Um método de migração no qual a workload de origem é desativada durante o processo de migração. Esse método envolve tempo de inatividade prolongado e geralmente é usado para workloads pequenas e não críticas.

OI

Veja [integração de operações](#).

Ola

Veja [acordo de nível operacional](#).

migração online

Um método de migração no qual a workload de origem é copiada para o sistema de destino sem ser colocada offline. As aplicações conectadas à workload podem continuar funcionando durante a migração. Esse método envolve um tempo de inatividade nulo ou mínimo e normalmente é usado para workloads essenciais para a produção.

OPC-UA

Veja [Open Process Communications - Unified Architecture](#).

Open Process Communications - Unified Architecture (OPC-UA)

Um protocolo de comunicação machine-to-machine (M2M) para automação industrial. O OPC-UA fornece um padrão de interoperabilidade com esquemas de criptografia, autenticação e autorização de dados.

acordo de nível operacional (OLA)

Um acordo que esclarece o que os grupos funcionais de TI prometem oferecer uns aos outros para apoiar um acordo de serviço (SLA).

análise de prontidão operacional (ORR)

Uma lista de verificação de perguntas e práticas recomendadas associadas que ajudam você a entender, avaliar, prevenir ou reduzir o escopo de incidentes e possíveis falhas. Para obter mais informações, consulte [Operational Readiness Reviews \(ORR\)](#) no AWS Well-Architected Framework.

tecnologia operacional (TO)

Sistemas de hardware e software que trabalham com o ambiente físico para controlar operações, equipamentos e infraestrutura industriais. Na manufatura, a integração dos sistemas de tecnologia da informação (TI) e tecnologia operacional (TO) é o foco principal das transformações da [Indústria 4.0](#).

integração de operações (OI)

O processo de modernização das operações na nuvem, que envolve planejamento de preparação, automação e integração. Para obter mais informações, consulte o [guia de integração de operações](#).

trilha organizacional

Uma trilha criada por ela AWS CloudTrail registra todos os eventos de todas as Contas da AWS em uma organização em AWS Organizations. Essa trilha é criada em cada Conta da AWS que faz parte da organização e monitora a atividade em cada conta. Para obter mais informações, consulte [Criação de uma trilha para uma organização](#) na CloudTrail documentação.

gerenciamento de alterações organizacionais (OCM)

Uma estrutura para gerenciar grandes transformações de negócios disruptivas de uma perspectiva de pessoas, cultura e liderança. O OCM ajuda as organizações a se prepararem e fazerem a transição para novos sistemas e estratégias, acelerando a adoção de alterações, abordando questões de transição e promovendo mudanças culturais e organizacionais. Na estratégia de AWS migração, essa estrutura é chamada de aceleração de pessoas, devido à velocidade de mudança exigida nos projetos de adoção da nuvem. Para obter mais informações, consulte o [guia do OCM](#).

controle de acesso de origem (OAC)

Em CloudFront, uma opção aprimorada para restringir o acesso para proteger seu conteúdo do Amazon Simple Storage Service (Amazon S3). O OAC oferece suporte a todos os buckets S3 Regiões da AWS, criptografia do lado do servidor com AWS KMS (SSE-KMS) e solicitações dinâmicas ao bucket S3. PUT DELETE

Identidade do acesso de origem (OAI)

Em CloudFront, uma opção para restringir o acesso para proteger seu conteúdo do Amazon S3. Quando você usa o OAI, CloudFront cria um principal com o qual o Amazon S3 pode se autenticar. Os diretores autenticados podem acessar o conteúdo em um bucket do S3 somente por meio de uma distribuição específica. CloudFront Veja também [OAC](#), que fornece um controle de acesso mais granular e aprimorado.

ORR

Veja [análise de prontidão operacional](#).

OT

Veja [tecnologia operacional](#).

VPC de saída (egresso)

Em uma arquitetura de AWS várias contas, uma VPC que gerencia conexões de rede que são iniciadas de dentro de um aplicativo. A [Arquitetura de Referência de AWS Segurança](#) recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo e a Internet em geral.

P

limite de permissões

Uma política de gerenciamento do IAM anexada a entidades principais do IAM para definir as permissões máximas que o usuário ou perfil podem ter. Para obter mais informações, consulte [Limites de permissões](#) na documentação do IAM.

Informações de identificação pessoal (PII)

Informações que, quando visualizadas diretamente ou combinadas com outros dados relacionados, podem ser usadas para inferir razoavelmente a identidade de um indivíduo. Exemplos de PII incluem nomes, endereços e informações de contato.

PII

Veja [informações de identificação pessoal](#).

manual

Um conjunto de etapas predefinidas que capturam o trabalho associado às migrações, como a entrega das principais funções operacionais na nuvem. Um manual pode assumir a forma de scripts, runbooks automatizados ou um resumo dos processos ou etapas necessários para operar seu ambiente modernizado.

PLC

Veja [controlador lógico programável](#).

PLM

Veja [gerenciamento do ciclo de vida do produto](#).

política

Um objeto que pode definir permissões (veja [política baseada em identidade](#)), especificar condições de acesso (veja [política baseada em recurso](#)) ou definir as permissões máximas para todas as contas em uma organização no AWS Organizations (veja [política de controle de serviços](#)).

persistência poliglota

Escolher de forma independente a tecnologia de armazenamento de dados de um microsserviço com base em padrões de acesso a dados e outros requisitos. Se seus microsserviços tiverem a mesma tecnologia de armazenamento de dados, eles poderão enfrentar desafios de implementação ou apresentar baixa performance. Os microsserviços serão implementados com mais facilidade e alcançarão performance e escalabilidade melhores se usarem o armazenamento de dados mais bem adaptado às suas necessidades.

avaliação do portfólio

Um processo de descobrir, analisar e priorizar o portfólio de aplicações para planejar a migração. Para obter mais informações, consulte [Avaliar a preparação para a migração](#).

predicado

Uma condição de consulta que retorna `true` ou `false`, normalmente localizada em uma cláusula `WHERE`.

pushdown de predicados

Uma técnica de otimização de consultas de banco de dados que filtra os dados na consulta antes da transferência. Isso reduz a quantidade de dados que devem ser recuperados e processados do banco de dados relacional e melhora a performance das consultas.

controle preventivo

Um controle de segurança projetado para evitar que um evento ocorra. Esses controles são a primeira linha de defesa para ajudar a evitar acesso não autorizado ou alterações indesejadas em sua rede. Para obter mais informações, consulte [Controles preventivos](#) em Como implementar controles de segurança na AWS.

principal (entidade principal)

Uma entidade AWS que pode realizar ações e acessar recursos. Essa entidade geralmente é um usuário raiz para um Conta da AWS, uma função do IAM ou um usuário. Para obter mais informações, consulte Entidade principal em [Termos e conceitos de perfis](#) na documentação do IAM.

Privacidade por design

Uma abordagem em engenharia de sistemas que leva em consideração a privacidade em todo o processo de desenvolvimento.

zonas hospedadas privadas

Um contêiner que contém informações sobre como você deseja que o Amazon Route 53 responda às consultas de DNS para um domínio e seus subdomínios em um ou mais VPCs. Para obter mais informações, consulte [Como trabalhar com zonas hospedadas privadas](#) na documentação do Route 53.

controle proativo

Um [controle de segurança](#) desenvolvido para evitar a implantação de recursos não conformes. Esses controles verificam os recursos antes de serem provisionados. Se o recurso não estiver em conformidade com o controle, ele não será provisionado. Para obter mais informações, consulte o [guia de referência de controles](#) na AWS Control Tower documentação e consulte [Controles proativos](#) em Implementação de controles de segurança em AWS.

gerenciamento do ciclo de vida do produto (PLM)

O gerenciamento de dados e processos de um produto em todo o seu ciclo de vida, desde a concepção, o desenvolvimento e o lançamento, passando pelo crescimento e maturidade, até o declínio e a remoção.

ambiente de produção

Veja [ambiente](#).

controlador lógico programável (PLC)

Na manufatura, um computador altamente confiável e adaptável que monitora as máquinas e automatiza os processos de fabricação.

encadeamento de prompts

Uso da saída de um prompt do [LLM](#) como entrada para o próximo prompt para gerar respostas melhores. Essa técnica é usada para dividir uma tarefa complexa em subtarefas, ou para refinar ou expandir iterativamente uma resposta preliminar. Isso ajuda a melhorar a precisão e a relevância das respostas de um modelo e permite resultados mais granulares e personalizados.

pseudonimização

O processo de substituir identificadores pessoais em um conjunto de dados por valores de espaço reservado. A pseudonimização pode ajudar a proteger a privacidade pessoal. Os dados pseudonimizados ainda são considerados dados pessoais.

publish/subscribe (pub/sub)

Um padrão que permite comunicações assíncronas entre microsserviços para melhorar a escalabilidade e a capacidade de resposta. Por exemplo, em um [MES](#) baseado em microsserviços, um microsserviço pode publicar mensagens de eventos em um canal em que outros microsserviços possam assinar. O sistema pode adicionar novos microsserviços sem alterar o serviço de publicação.

Q

plano de consulta

Uma série de etapas, como instruções, usadas para acessar os dados em um sistema de banco de dados relacional SQL.

regressão de planos de consultas

Quando um otimizador de serviço de banco de dados escolhe um plano menos adequado do que escolhia antes de uma determinada alteração no ambiente de banco de dados ocorrer. Isso pode ser causado por alterações em estatísticas, restrições, configurações do ambiente, associações de parâmetros de consulta e atualizações do mecanismo de banco de dados.

R

Matriz RACI

Veja [responsável, aprovador, consultado, informado \(RACI\)](#).

RAG

Veja [geração aumentada via recuperação](#).

ransomware

Um software mal-intencionado desenvolvido para bloquear o acesso a um sistema ou dados de computador até que um pagamento seja feito.

Matriz RASCI

Veja [responsável, aprovador, consultado, informado \(RACI\)](#).

RCAC

Veja [controle de acesso por linha e coluna](#).

réplica de leitura

Uma cópia de um banco de dados usada somente para leitura. É possível encaminhar consultas para a réplica de leitura e reduzir a carga no banco de dados principal.

Redefinir arquitetura

Veja [7 Rs](#).

objetivo de ponto de recuperação (RPO).

O máximo período de tempo aceitável desde o último ponto de recuperação de dados. Isso determina o que é considerado uma perda aceitável de dados entre o último ponto de recuperação e a interrupção do serviço.

objetivo de tempo de recuperação (RTO)

O máximo atraso aceitável entre a interrupção e a restauração do serviço.

refatorar

Veja [7 Rs](#).

Região

Uma coleção de AWS recursos em uma área geográfica. Cada um Região da AWS é isolado e independente dos outros para fornecer tolerância a falhas, estabilidade e resiliência. Para obter informações, consulte [Specify which Regiões da AWS your account can use](#).

regressão

Uma técnica de ML que prevê um valor numérico. Por exemplo, para resolver o problema de “Por qual preço esta casa será vendida?” um modelo de ML pode usar um modelo de regressão linear para prever o preço de venda de uma casa com base em fatos conhecidos sobre a casa (por exemplo, a metragem quadrada).

redefinir a hospedagem

Veja [7 Rs](#).

versão

Em um processo de implantação, o ato de promover mudanças em um ambiente de produção.

realocar

Veja [7 Rs](#).

redefinir a plataforma

Veja [7 Rs](#).

recomprar

Veja [7 Rs](#).

resiliência

A capacidade de uma aplicação de resistir ou se recuperar de interrupções. [Alta disponibilidade](#) e [recuperação de desastres](#) são considerações comuns ao planejar a resiliência na Nuvem AWS. Para obter mais informações, consulte [Nuvem AWS Resilience](#).

política baseada em recurso

Uma política associada a um recurso, como um bucket do Amazon S3, um endpoint ou uma chave de criptografia. Esse tipo de política especifica quais entidades principais têm acesso permitido, ações válidas e quaisquer outras condições que devem ser atendidas.

matriz responsável, accountable, consultada, informada (RACI)

Uma matriz que define as funções e responsabilidades de todas as partes envolvidas nas atividades de migração e nas operações de nuvem. O nome da matriz é derivado dos tipos de responsabilidade definidos na matriz: responsável (R), responsabilizável (A), consultado (C) e informado (I). O tipo de suporte (S) é opcional. Se você incluir suporte, a matriz será chamada de matriz RASCI e, se excluir, será chamada de matriz RACI.

controle responsivo

Um controle de segurança desenvolvido para conduzir a remediação de eventos adversos ou desvios em relação à linha de base de segurança. Para obter mais informações, consulte [Controles responsivos](#) em Como implementar controles de segurança na AWS.

reter

Veja [7 Rs](#).

Retirada

Veja [7 Rs](#).

Geração Aumentada de Recuperação (RAG)

Uma tecnologia de [IA generativa](#) em que um [LLM](#) faz referência a uma fonte de dados autorizada que está fora de suas fontes de dados de treinamento antes de gerar uma resposta. Por exemplo, um modelo RAG pode realizar uma pesquisa semântica na base de conhecimento ou nos dados personalizados de uma organização. Para obter mais informações, consulte [O que é RAG \(geração aumentada via recuperação\)?](#).

alternância

O processo de atualizar periodicamente um [segredo](#) para dificultar o acesso de um invasor às credenciais.

controle de acesso por linha e coluna (RCAC)

O uso de expressões SQL básicas e flexíveis que tenham regras de acesso definidas. O RCAC consiste em permissões de linha e máscaras de coluna.

RPO

Veja [objetivo de ponto de recuperação](#).

RTO

Veja [objetivo de tempo de recuperação](#).

runbook

Um conjunto de procedimentos manuais ou automatizados necessários para realizar uma tarefa específica. Eles são normalmente criados para agilizar operações ou procedimentos repetitivos com altas taxas de erro.

S

SAML 2.0

Um padrão aberto que muitos provedores de identidade (IdPs) usam. Esse recurso permite o login único federado (SSO), para que os usuários possam fazer login no Console de gerenciamento da AWS ou chamar as operações da AWS API sem que você precise criar um usuário no IAM para todos em sua organização. Para obter mais informações sobre a federação baseada em SAML 2.0, consulte [Sobre a federação baseada em SAML 2.0](#) na documentação do IAM.

SCADA

Veja [controle de supervisão e aquisição de dados](#).

SCP

Veja [política de controle de serviço](#).

secret

Em AWS Secrets Manager, informações confidenciais ou restritas, como uma senha ou credenciais de usuário, que você armazena de forma criptografada. Consiste no valor secreto e em seus metadados. O valor secreto pode ser binário, uma única string ou várias strings. Para obter mais informações, consulte [What's in a Secrets Manager secret?](#) na documentação do Secrets Manager.

segurança desde a concepção

Uma abordagem em engenharia de sistemas que leva em consideração a segurança em todo o processo de desenvolvimento.

controle de segurança

Uma barreira de proteção técnica ou administrativa que impede, detecta ou reduz a capacidade de uma ameaça explorar uma vulnerabilidade de segurança. Existem quatro tipos primários de controles de segurança: [preventivos](#), [detectivos](#), [responsivos](#) e [proativos](#).

hardening da segurança

O processo de reduzir a superfície de ataque para torná-la mais resistente a ataques. Isso pode incluir ações como remover recursos que não são mais necessários, implementar a prática recomendada de segurança de conceder privilégios mínimos ou desativar recursos desnecessários em arquivos de configuração.

sistema de gerenciamento de eventos e informações de segurança (SIEM)

Ferramentas e serviços que combinam sistemas de gerenciamento de informações de segurança (SIM) e gerenciamento de eventos de segurança (SEM). Um sistema SIEM coleta, monitora e analisa dados de servidores, redes, dispositivos e outras fontes para detectar ameaças e violações de segurança e gerar alertas.

automação de resposta de segurança

Uma ação predefinida e programada projetada para responder ou remediar automaticamente um evento de segurança. Essas automações servem como controles de segurança [responsivos](#) ou [detectivos](#) que ajudam você a implementar as melhores práticas AWS de segurança. Exemplos de ações de resposta automatizada incluem a modificação de um grupo de segurança da VPC, a aplicação de patches em uma instância do Amazon EC2 ou a alternância de credenciais.

Criptografia do lado do servidor

Criptografia dos dados em seu destino, por AWS service (Serviço da AWS) quem os recebe.

política de controle de serviços (SCP)

Uma política que fornece controle centralizado sobre as permissões de todas as contas em uma organização em AWS Organizations. SCPs defina barreiras ou estabeleça limites nas ações que um administrador pode delegar a usuários ou funções. Você pode usar SCPs como listas de permissão ou listas de negação para especificar quais serviços ou ações são permitidos ou proibidos. Para obter mais informações, consulte [Políticas de controle de serviço](#) na AWS Organizations documentação.

service endpoint (endpoint de serviço)

O URL do ponto de entrada para um AWS service (Serviço da AWS). Você pode usar o endpoint para se conectar programaticamente ao serviço de destino. Para obter mais informações, consulte [Endpoints do AWS service \(Serviço da AWS\)](#) na Referência geral da AWS.

acordo de serviço (SLA)

Um acordo que esclarece o que uma equipe de TI promete fornecer aos clientes, como tempo de atividade e performance do serviço.

indicador de nível de serviço (SLI)

Uma avaliação de um aspecto de performance de um serviço, como taxa de erro, disponibilidade ou throughput.

objetivo de nível de serviço (SLO)

Uma métrica alvo que representa a integridade de um serviço, conforme avaliado por um [indicador de nível de serviço](#).

modelo de responsabilidade compartilhada

Um modelo que descreve a responsabilidade com a qual você compartilha AWS pela segurança e conformidade na nuvem. AWS é responsável pela segurança da nuvem, enquanto você é responsável pela segurança na nuvem. Para obter mais informações, consulte o [Modelo de responsabilidade compartilhada](#).

SIEM

Veja [sistema de gerenciamento de eventos e informações de segurança](#).

ponto único de falha (SPOF)

Uma falha em um único componente crítico de uma aplicação que pode interromper o sistema.

SLA

Veja [acordo de serviço](#).

SLI

Veja [indicador de nível de serviço](#).

SLO

Veja [objetivo de nível de serviço](#).

split-and-seed modelo

Um padrão para escalar e acelerar projetos de modernização. À medida que novos recursos e lançamentos de produtos são definidos, a equipe principal se divide para criar novas equipes de produtos. Isso ajuda a escalar os recursos e os serviços da sua organização, melhora a produtividade do desenvolvedor e possibilita inovações rápidas. Para obter mais informações, consulte [Phased approach to modernizing applications in the Nuvem AWS](#).

SPOF

Veja [ponto único de falha](#).

esquema em estrela

Uma estrutura organizacional de banco de dados que usa uma grande tabela de fatos para armazenar dados transacionais ou medidos e usa uma ou mais tabelas dimensionais menores para armazenar atributos de dados. Essa estrutura foi projetada para ser usada em um [data warehouse](#) ou para fins de inteligência comercial.

padrão strangler fig

Uma abordagem à modernização de sistemas monolíticos que consiste em reescrever e substituir incrementalmente a funcionalidade do sistema até que o sistema herdado possa ser desativado. Esse padrão usa a analogia de uma videira que cresce e se torna uma árvore estabelecida e, eventualmente, supera e substitui sua hospedeira. O padrão foi [apresentado por Martin Fowler](#) como forma de gerenciar riscos ao reescrever sistemas monolíticos. Para ver um exemplo de como aplicar esse padrão, consulte [Modernizar incrementalmente os serviços Web herdados do Microsoft ASP.NET \(ASMX\) usando contêineres e o Amazon API Gateway](#).

sub-rede

Um intervalo de endereços IP na VPC. Cada sub-rede fica alocada em uma única zona de disponibilidade.

controle supervisão e aquisição de dados (SCADA)

Na manufatura, um sistema que usa hardware e software para monitorar ativos físicos e operações de produção.

symmetric encryption (criptografia simétrica)

Um algoritmo de criptografia que usa a mesma chave para criptografar e descriptografar dados.

testes sintéticos

Testar um sistema de forma que simule as interações do usuário para detectar possíveis problemas ou monitorar a performance. Você pode usar o [Amazon CloudWatch Synthetics](#) para criar esses testes.

prompt do sistema

Uma técnica para fornecer contexto, instruções ou orientações a um [LLM](#) a fim de direcionar seu comportamento. Os prompts do sistema ajudam a definir o contexto e a estabelecer regras para interações com os usuários.

T

tags

Pares de valores-chave que atuam como metadados para organizar seus recursos. AWS As tags podem ajudar você a gerenciar, identificar, organizar, pesquisar e filtrar recursos da . Para obter mais informações, consulte [Marcar seus recursos do AWS](#).

variável-alvo

O valor que você está tentando prever no ML supervisionado. Ela também é conhecida como variável de resultado. Por exemplo, em uma configuração de fabricação, a variável-alvo pode ser um defeito do produto.

lista de tarefas

Uma ferramenta usada para monitorar o progresso por meio de um runbook. Uma lista de tarefas contém uma visão geral do runbook e uma lista de tarefas gerais a serem concluídas. Para cada tarefa geral, ela inclui o tempo estimado necessário, o proprietário e o progresso.

ambiente de teste

Veja [ambiente](#).

treinamento

O processo de fornecer dados para que seu modelo de ML aprenda. Os dados de treinamento devem conter a resposta correta. O algoritmo de aprendizado descobre padrões nos dados de treinamento que mapeiam os atributos dos dados de entrada no destino (a resposta que você deseja prever). Ele gera um modelo de ML que captura esses padrões. Você pode usar o modelo de ML para obter previsões de novos dados cujo destino você não conhece.

gateway de trânsito

Um hub de trânsito de rede que você pode usar para interconectar sua rede com VPCs a rede local. Para obter mais informações, consulte [O que é um gateway de trânsito](#) na AWS Transit Gateway documentação.

fluxo de trabalho baseado em troncos

Uma abordagem na qual os desenvolvedores criam e testam recursos localmente em uma ramificação de recursos e, em seguida, mesclam essas alterações na ramificação principal. A ramificação principal é então criada para os ambientes de desenvolvimento, pré-produção e produção, sequencialmente.

Acesso confiável

Conceder permissões a um serviço que você especifica para realizar tarefas em sua organização AWS Organizations e em suas contas em seu nome. O serviço confiável cria um perfil vinculado ao serviço em cada conta, quando esse perfil é necessário, para realizar tarefas de gerenciamento para você. Para obter mais informações, consulte [Usando AWS Organizations com outros AWS serviços](#) na AWS Organizations documentação.

tuning (ajustar)

Alterar aspectos do processo de treinamento para melhorar a precisão do modelo de ML. Por exemplo, você pode treinar o modelo de ML gerando um conjunto de rótulos, adicionando rótulos e repetindo essas etapas várias vezes em configurações diferentes para otimizar o modelo.

equipe de duas pizzas

Uma pequena DevOps equipe que você pode alimentar com duas pizzas. Uma equipe de duas pizzas garante a melhor oportunidade possível de colaboração no desenvolvimento de software.

U

incerteza

Um conceito que se refere a informações imprecisas, incompletas ou desconhecidas que podem minar a confiabilidade dos modelos preditivos de ML. Há dois tipos de incertezas: a incerteza epistêmica é causada por dados limitados e incompletos, enquanto a incerteza aleatória é causada pelo ruído e pela aleatoriedade inerentes aos dados. Para obter mais informações, consulte o guia [Como quantificar a incerteza em sistemas de aprendizado profundo](#).

tarefas indiferenciadas

Também conhecido como trabalho pesado, trabalho necessário para criar e operar um aplicativo, mas que não fornece valor direto ao usuário final nem oferece vantagem competitiva. Exemplos de tarefas indiferenciadas incluem aquisição, manutenção e planejamento de capacidade.

ambientes superiores

Veja [ambiente](#).

V

aspiração

Uma operação de manutenção de banco de dados que envolve limpeza após atualizações incrementais para recuperar armazenamento e melhorar a performance.

controle de versões

Processos e ferramentas que rastreiam mudanças, como alterações no código-fonte em um repositório.

emparelhamento da VPC

Uma conexão entre duas VPCs que permite rotear o tráfego usando endereços IP privados. Para ter mais informações, consulte [O que é emparelhamento de VPC?](#) na documentação da Amazon VPC.

Vulnerabilidade

Uma falha de software ou hardware que compromete a segurança do sistema.

W

cache quente

Um cache de buffer que contém dados atuais e relevantes que são acessados com frequência. A instância do banco de dados pode ler do cache do buffer, o que é mais rápido do que ler da memória principal ou do disco.

dados mornos

Dados acessados raramente. Ao consultar esse tipo de dados, consultas moderadamente lentas geralmente são aceitáveis.

função de janela

Uma função SQL que executa um cálculo em um grupo de linhas que se relacionam de alguma forma com o registro atual. As funções de janela são úteis para processar tarefas, como calcular uma média móvel ou acessar o valor das linhas com base na posição relativa da linha atual.

workload

Uma coleção de códigos e recursos que geram valor empresarial, como uma aplicação voltada para o cliente ou um processo de backend.

workstreams

Grupos funcionais em um projeto de migração que são responsáveis por um conjunto específico de tarefas. Cada workstream é independente, mas oferece suporte aos outros workstreams do projeto. Por exemplo, o workstream de portfólio é responsável por priorizar aplicações, planejar ondas e coletar metadados de migração. O workstream de portfólio entrega esses ativos ao workstream de migração, que então migra os servidores e as aplicações.

WORM

Veja [gravação única e várias leituras](#).

WQF

Veja [AWS Workload Qualification Framework](#).

gravação única e várias leituras (WORM)

Um modelo de armazenamento que grava dados uma única vez e evita que os dados sejam excluídos ou modificados. Os usuários autorizados podem ler os dados quantas vezes forem necessárias, mas não podem alterá-los. Essa infraestrutura de armazenamento de dados é considerada [imutável](#).

Z

exploração de dia zero

Um ataque, normalmente malware, que tira proveito de uma [vulnerabilidade zero-day](#).

vulnerabilidade de dia zero

Uma falha ou vulnerabilidade não mitigada em um sistema de produção. Os agentes de ameaças podem usar esse tipo de vulnerabilidade para atacar o sistema. Os desenvolvedores frequentemente ficam cientes da vulnerabilidade como resultado do ataque.

prompt zero shot

Fornecer a um [LLM](#) instruções para realizar uma tarefa, mas sem exemplos (shots) que possam ajudar a orientá-lo. O LLM deve usar seu conhecimento pré-treinado para lidar com a tarefa. A

eficácia dos prompts zero-shot depende da complexidade da tarefa e da qualidade do prompt.

Veja também [prompts few-shot](#).

aplicação zumbi

Uma aplicação que tem um uso médio de CPU e memória inferior a 5%. Em um projeto de migração, é comum retirar essas aplicações.

As traduções são geradas por tradução automática. Em caso de conflito entre o conteúdo da tradução e da versão original em inglês, a versão em inglês prevalecerá.