Amazon Neptune 用の AWS Well-Architected フレームワークの適用

AWS 規範ガイダンス

AWS 規範ガイダンス: Amazon Neptune 用の AWS Well-Architected フレームワークの適用

Table of Contents

Amazon Neptune 用の AWS Well-Architected フレームワークの適用

Amazon Web Services (寄稿者)

2026 年 1 月 (ドキュメント履歴)

Amazon Neptune を使用して、Amazon Web Services (AWS) でグラフベースのソリューションを構築できます。このガイドは、Neptune のデプロイ計画にあたり AWS Well-Architected フレームワークの原則を適用するための規範ガイダンスを提供します。

AWS Well-Architected フレームワークは、さまざまなアプリケーションやワークロード向けに、安全で高性能、耐障害性、効率的なインフラストラクチャを構築するのに役立ちます。また、アーキテクチャを評価し、スケーラブルな設計を実装するための一貫したアプローチも提供します。

AWS Well-Architected フレームワークは、次の 6 つの柱を中心に構築されています。

- オペレーショナルエクセレンス
- セキュリティ
- 信頼性
- パフォーマンス効率
- コスト最適化
- 持続可能性

このガイドでは、 AWS Well-Architected フレームワークの設計の柱とベストプラクティス、および に Neptune をデプロイする際の考慮事項について説明します AWS。

対象者

このガイドは、 AWSでグラフを使用するソリューションを設計および実装するデータエンジニア、ソリューションアーキテクト、データアナリストを対象としています。

目的

このガイドは、お客様や組織が以下のことを行うのに役立ちます。

- ユースケースとクエリパターンに基づいて、サポートされているデプロイオプションとクエリ言語から選択します。
- 耐障害性とセキュリティの向上に役立つ AWS Well-Architected 設計パターンに従ってください。
- 最適なパフォーマンスとコスト削減のためのクエリを設計します。
- 本番環境で Neptune クラスターを管理する際に運用効率を高める方法について学びます。

運用上の優秀性の柱

AWS Well-Architected フレームワークの運用上の優秀性の柱は、システムの実行とモニタリング、プロセスと手順の継続的な改善に焦点を当てています。開発をサポートし、ワークロードを効率的に実行し、運用に関するインサイトを得て、ビジネス価値をもたらすサポートプロセスと手順を継続的に改善する能力が含まれます。人間の介入なしにほとんどの問題を検出して修復する自己修復ワークロードによって、運用上の複雑さを軽減できます。このセクションで説明されているベストプラクティスに従うことで、この目標を達成できます。Amazon Neptune メトリクス、API、各種メカニズムを使用して、ワークロードが予想される動作から逸脱した場合に適切に対応します。

運用上の優秀性の柱に関するこの説明では、以下の主要分野に焦点を当てています。

- Infrastructure as code (IaC)
- 変更管理
- レジリエンシー戦略
- インシデント管理
- コンプライアンスのための監査レポート
- ログ記録とモニタリング

IaC アプローチを使用してデプロイを自動化する

IaC を使用して Neptune でのデプロイを自動化するためのベストプラクティスは次のとおりです。

- 可能な限り、Infrastructure as Code (IaC) を適用して Neptune クラスターをデプロイします。一貫した環境設定を行うため、クラスターに必要なすべてのリソースを作成するには AWS CloudFormation テンプレート、AWS Cloud Development Kit (AWS CDK)、または HashiCorp Terraform を使用します。
- インスタンスのサイズ変更、リードレプリカの追加または削除、グローバルテーブルでの手動フェイルオーバーなど、Neptune の運用手順を可能な限り自動化します。
- 接続文字列をクライアントの外部に保存します。ブルー/グリーンデプロイ戦略、ディザスタリカバリ (DR)、新しいクラスターへのほぼゼロのダウンタイムでの移行を容易にするため、抽出、変換、ロード (ETL) プロセスを使用します。接続文字列は、AWS Secrets Manager、Amazon DynamoDB、または動的に変更できる任意の場所に保存できます。
- タグを使用してメタデータを Neptune リソースに追加し、タグに基づいて使用量を追跡します。詳細については、「Amazon Neptune リソースのタグ付け」を参照してください。

小規模かつ可逆的な変更を頻繁に行う

以下の推奨事項は、複雑さを最小限に抑え、ワークロードの中断の可能性を減らすための、小規模かつ可逆的な変更に焦点を当てています。

- IaC テンプレートとスクリプトを GitHub や GitLab といったソースコントロールサービスに保存します。

  > **Important**
  >
  > ソース管理に AWS 認証情報を保存しないでください。

- IaC デプロイでは、AWS CodePipeline や AWS CodeBuild などの継続的インテグレーションおよび継続的デリバリー (CI/CD) サービスを使用する必要があります。これらのサービスは、本番環境の Amazon Neptune クラスターに影響を与える前に、一時的な Neptune クラスターを含む非本番環境でコードをコンパイル、テスト、デプロイします。
- 本番環境にデプロイする前に、インフラストラクチャとアプリケーションのクエリをより低い環境でテストします。これにより、中断の可能性が最小限に抑えられ、ワークロードやスケールに合わせて適切に動作させることができます。

失敗を予測する

自己修復インフラストラクチャは、障害を予測し、介入なしで問題解決を試みることで、運用上の優秀性を実証します。以下の推奨事項は、Neptune でその成熟度を実現するのに役立ちます。

- Amazon CloudWatch メトリクスを使用して DB インスタンスの CPU とメモリの使用状況をモニタリングし、使用状況のパターンを理解するモニタリングプランを作成します。アプリケーションログにある主要なメトリクスと Neptune クライアントレスポンスを確認するための CloudWatch ダッシュボードとアラームを作成します。CPU 使用率の高低を示す指標の詳細については、Neptune ドキュメントの「Using CloudWatch to monitor DB instance performance in Neptune」を参照してください。

  クエリでout-of-memory例外が頻繁に発生する場合は、クエリが通過するノードの合計数を減らすか、RAM-to-CPU の比率が高い X2ファミリーのインスタンスの使用を検討してください。
- Neptune クラスターの状態をモニタリングするために通知を設定します。例えば、`BufferCacheHitRatio` は常に高い (99.9% を超える) 必要があります

が、MainRequestQueuePendingRequests は常に低い (理想的には 0 ですが、要件とレイテンシーの許容値によって異なります) 必要があります。

- Neptune 内で高可用性を実現するには、リードレプリカの使用を検討してください。フェイルオーバーイベント中にインスタンスが常に読み取りクエリを処理できるように、ライターインスタンスとは異なるアベイラビリティーゾーンに少なくとも 2 つのリードレプリカが必要です。
- 使用率メトリクスに基づいてリードレプリカを自動的にスケーリングします。詳細については、「Amazon Neptune DB クラスター内のレプリカの数の Auto-scaling」を参照してください。
- DB インスタンスのフェイルオーバーをテストすることで、そのプロセスでユースケースにかかる時間を把握します。
- アプリケーションが完全に AWS リージョン 停止した場合、DR プランの一部としてグローバルデータベースを使用することを検討してください。

すべての運用上の障害から学ぶ

自己修復インフラストラクチャは、まれな問題が発生したり、対応が意図したような効果を出さなかったりするたびに反復的に進化していく長期的な取り組みです。以下のプラクティスを採用することで、その目標に焦点を当てることができます。

- すべての障害から学ぶことで改善を推進します。
- チームと組織で教訓を共有します。組織内の複数のチームが Neptune を使用している場合は、共通のチャットルームまたはユーザーグループを作成して、学習とベストプラクティスを共有します。

ログ機能を使用して、不正または異常なアクティビティをモニタリングする

異常なパフォーマンスやアクティビティのパターンを観察するには、Amazon CloudWatch Logs にログを保存します。以下のベストプラクティスを考慮します。

- スロークエリロギングを有効にします。ログを定期的に確認し、特定のクエリが遅い理由を診断します。Gremlin、SPARQL、または openCypher 向けの Neptune の explain エンドポイントと profile エンドポイントを使用して、これらのクエリが遅い理由に関するインサイトを取得します。
- Neptune 監査ログを有効にし、ログに不正アクセスや異常がないか定期的に確認します。

- スロークエリログ記録または監査ログ記録を使用している場合は、CloudWatch Logs への発行を有効にします。これにより、インスタンスのディスク容量が不足するのを防ぐことができます。Neptune インスタンスのログストレージ容量は限られており、ログ容量を超えると古いログファイルは上書きされます。CloudWatch Logs は、ログの長期保持をサポートしています。CloudWatch Logs の強化されたモニタリング機能を使用すると、ログをクエリして問題を診断する能力が向上します。
- 監査ログの分析ツールを改善するため、監査ログデータを CloudWatch Logs のロググループに発行するように Neptune DB クラスターを設定できます。CloudWatch Logs を使用すると、ログデータのリアルタイム分析、CloudWatch を使用したアラームの作成およびメトリクスの表示、CloudWatch Logs を使用した、耐久性に優れたストレージへのログレコードの保存を行うことができます。詳細については、「Publishing Neptune logs to Amazon CloudWatch Logs」を参照してください。
- Neptune は、 AWS CloudTrailを使用したコントロールプレーンのアクションのログ記録をサポートしています。詳細については、「 を使用した Amazon Neptune API コールのログ記録 AWS CloudTrail」を参照してください。

セキュリティの柱

のクラウドセキュリティが最優先事項 AWS です。お客様は AWS 、セキュリティを最も重視する組織の要件を満たすように構築されたデータセンターとネットワークアーキテクチャを活用できます。

セキュリティは、 AWS お客様とお客様の間の責任共有です。責任共有モデルでは、これをクラウドのセキュリティおよびクラウド内のセキュリティとして説明しています。

- クラウドのセキュリティ – AWS は、 AWS のサービス で実行されるインフラストラクチャを保護する責任を担います AWS クラウド。 は、お客様が安全に使用できるサービス AWS も提供します。サードパーティーの監査者は、AWS コンプライアンスプログラムの一環として、 AWS セキュリティの有効性を定期的にテストおよび検証します。Amazon Neptune に適用されるコンプライアンスプログラムについては、コンプライアンスプログラムによるAWS 対象範囲内のサービスを参照してください。
- クラウドのセキュリティ – お客様の責任は、使用する によって決まり AWS のサービス ます。また、お客様は、お客様のデータの機密性、企業の要件、および適用可能な法律および規制などの他の要因についても責任を担います。データプライバシーの詳細については、「データプライバシーのよくある質問」を参照してください。欧州でのデータ保護の詳細については、「AWS Shared Responsibility Model and GDPR」のブログ記事を参照してください。

セキュリティの柱は、Neptune を使用する際に責任共有モデルを適用する方法を理解するのに役立ちます。以下のトピックでは、セキュリティおよびコンプライアンスの目的を達成するために Neptune を設定する方法を示します。また、Neptune リソースのモニタリングと保護 AWS のサービス に役立つ他の の使用方法についても説明します。

セキュリティの柱には、以下の主要な重点分野が含まれています。

- データセキュリティ
- ネットワークセキュリティ
- 認証と認可

データセキュリティの実装

データ漏洩やデータ侵害は顧客を危険にさらし、会社に大きな悪影響を与える可能性があります。以下のベストプラクティスは、顧客データを不注意による漏洩や悪意のある漏洩から保護するのに役立ちます。

- クラスター名、タグ、パラメータグループ、 AWS Identity and Access Management (IAM) ロール、およびその他のメタデータには、請求ログや診断ログに表示される可能性があるため、機密情報や機密情報を含めないでください。
- Neptune にデータとして保存されている外部サーバーへの URI またはリンクには、リクエストを検証するための認証情報を含めないでください。
- Neptune の暗号化されたインスタンスは、基になるストレージへの不正アクセスからデータを保護することで、データ保護のレイヤーを追加します。Neptune の暗号化を使用すると、クラウドにデプロイされたアプリケーションのデータ保護を強化できます。Neptune 暗号化は、保管中のデータのコンプライアンス要件を満たすために使用することもできます。

  新しい Neptune DB インスタンスの暗号化を有効にするには、Neptune コンソールの [暗号化の有効化] セクションで [はい] (デフォルトで選択) を選択するか、 CloudFormationで AWS::Neptune::DBCluster::StorageEncrypted プロパティを設定します。暗号化が有効になっている場合、Neptune はデフォルトで Amazon Relational Database Service (Amazon RDS) AWS マネージドキーを使用します。そうでない場合、カスタマーマネージドキーを作成できます。Neptune DB インスタンスの作成については、「Creating a new Neptune DB cluster」を参照してください。詳細については、「保管時の Neptune リソースの暗号化」を参照してください。自動スナップショットと手動スナップショットは、Neptune クラスターに選択したのと同じ暗号化を使用します。
- SPARQL 言語と openCypher 言語を使用する場合は、SQL インジェクションやその他の形式の攻撃を防ぐために、適切な入力検証手法およびパラメータ化の手法を実践してください。ユーザー指定の入力による文字列連結を使用するクエリは構築しないでください。パラメータ化されたクエリまたはプリペアドステートメントを使用して、入力パラメータをグラフデータベースに安全に渡します。詳細については、「openCypher のパラメーター化されたクエリの例」および「SPARQL インジェクション防御」を参照してください。
- Gremlin 言語の場合、潜在的なインジェクションの問題を回避するために、文字列ベースの Gremlin スクリプトを直接渡すのではなく Gremlin 言語バリアントを使用します。

ネットワークを保護する

Amazon Neptune DB クラスターは、 AWS上の仮想プライベートクラウド (VPC) でのみ作成できます。Neptune 1.4.6.0 まで、Neptune DB クラスターのエンドポイントには、その VPC 内でのみアクセスできました。Neptune 1.4.6.0 以降では、Neptune インスタンスをインターネット経由でパブリックにアクセスできるように設定できます。開発者が Neptune に簡単にアクセスできるようにするには、この機能を非本番環境でのみ使用することがベストプラクティスです (ただし、パブリックアクセシビリティを有効にするには で IAM 認証が常に必要です) 。パブリックアクセシビリティ

が有効になっている場合は、データベースポートのインバウンドセキュリティグループルールを既知の IP アドレストラフィックのみに設定することを検討してください。本番環境または機密データを含むクラスターでは、パブリックアクセシビリティを許可せず、Neptune DB クラスターがある VPC へのアクセスを制限することで、Neptune データを保護します。詳細については、「Amazon Neptune へのグラフへの接続」を参照してください。

転送中のデータを保護するために、Neptune は安全なプロトコルと暗号を使用して、HTTPS を介した SSL 接続を任意のインスタンスまたはクラスターエンドポイントに強制します。Neptune は Neptune DB インスタンスの SSL 証明書を提供します。Neptune SSL 証明書は、クラスターエンドポイント、リーダーエンドポイント、インスタンスエンドポイントのホスト名のみをサポートします。

ロードバランサーまたはプロキシサーバー (HAProxy など) を使用している場合は、SSL ターミネーションを使用して独自の SSL 証明書をプロキシサーバーに保存する必要があります。提供された SSL 証明書はプロキシサーバーのホスト名と一致しないため、SSL パススルーは機能しません。SSL を使用して Neptune エンドポイントに接続する方法の詳細については、Using the HTTP REST endpoint to connect to a Neptune DB instance」を参照してください。

認証と認可を実装する

Neptune DB クラスターと DB インスタンスで Neptune 管理 アクションを実行できるユーザーを制御するには、IAM データベース認証を有効にし、IAM 認証情報を使用します。IAM 認証情報を使用して AWS に接続するとき、IAM ロールには、Neptune 管理操作を実行するためのアクセス許可を付与する IAM ポリシーが必要です。最小特権の原則に従って、タスクを完了するために必要なアクセス許可のみを付与してください。詳細については、「Neptune へのアクセスを制御するためのさまざまな種類の IAM ポリシーの使用」および「一時的な認証情報を使用した IAM 認証」を参照してください。

Neptune クラスターに接続してデータをクエリできるユーザーを制御するには、IAM を使用して Neptune DB インスタンスまたは DB クラスターを認証します。Neptune DB クラスターで IAM 認証を有効にした場合、DB クラスターにアクセスするすべてのユーザーを最初に認証する必要があります。詳細については、「Enabling IAM database authentication in Neptune」を参照してください。

IAM データベース認証が有効になっている場合、各リクエストは AWS 署名バージョン 4 を使用して署名する必要があります。IAM 認証が有効になっているすべての Neptune エンドポイントに署名付きリクエストを送信する方法については、「Connecting and Signing with AWS Signature Version 4」を参照してください。awscurl などの多くのライブラリやツールは、 AWS 署名バージョン 4 を既にサポートしています。

他の とやり取りするために AWS のサービス、Amazon Neptune は IAM サービスにリンクされたロールを使用します。サービスにリンクされたロールは、Neptune に直接リンクされた一意のタイプの IAM ロールです サービスリンクロールは、Neptune によって事前定義されており、サービスがユーザーに代わって他の AWS のサービス を呼び出すために必要なすべてのアクセス許可が含まれています。詳細については、「Using Service-Linked Roles for Neptune」を参照してください。

信頼性の柱

信頼性の柱には、ワークロードが意図した機能を正しく一貫して実行する能力が含まれます。これには、ワークロードのライフサイクル全体を通じてワークロードを運用およびテストする能力が含まれます。

信頼性の高いワークロードの実現は、ソフトウェアとインフラストラクチャの両方について事前に設計を決定することから始まります。アーキテクチャの選択は、 AWS Well-Architected のすべての柱にわたってワークロードの動作に影響します。高い信頼性を保つには、特定のパターンに従う必要があります。

信頼性の柱は、次の主要分野に焦点を当てています。

- サービスクォータとデプロイパターンを含むワークロードアーキテクチャ
- 変更管理
- 障害管理

Neptune サービスクォータを理解する

Neptune クラスターボリュームは、クォータが 64 TiB である中国と GovCloud を除くすべてのサポート対象の AWS リージョン で、最大 128 tebibytes (TiB) のサイズまで増やすことができます。

128 TiB のクォータでは、グラフに約 2,000～4,000 億個のオブジェクトを十分に保存することができます。ラベル付きプロパティグラフ (LPG) では、オブジェクトはノード、エッジ、またはノードもしくはエッジ上のプロパティです。Resource Description Framework (RDF) グラフでは、オブジェクトは四角形です。

任意の Neptune サーバーレスクラスターでは、Neptune 容量ユニット (NCU) の最小数と最大数の両方を設定します。各 NCU は、2 ギビバイト (GiB) のメモリと、関連する vCPU およびネットワークで構成されます。最小および最大 NCU 値は、クラスター内のサーバーレスインスタンスに適用されます。設定できる最高の最大 NCU 値は 128.0 NCU であり、最低の最小値は 1.0 NCU です。Amazon CloudWatch メトリクス `ServerlessDatabaseCapacity` および `NCUUtilization` を監視し、一般的に実行される範囲をキャプチャしてその範囲内の望ましくない動作やコストを関連付けることで、アプリケーションに最適な NCU 範囲を最適化します。多くのワークロードでは、1.0 NCU の開始点が低すぎるため、非アクティブ状態が続くと動作の信頼性が低下します。ワークロードが十分に速くスケーリングしない場合は、スケーリング中の最初のサージに対して十分な処理を提供するように最小 NCU を増やします。

各 AWS アカウント には、作成できるデータベースリソースの数に対する各リージョンのクォータがあります。これらのリソースには、DB インスタンスと DB クラスターが含まれます。リソースの制限に達すると、リソースを作成するための追加の呼び出しは、例外が発生して失敗します。一部のクォータはソフトクォータで、リクエストにより引き上げることができます。Amazon Neptune と Amazon RDS、Amazon Aurora、Amazon DocumentDB (MongoDB 互換) の間で共有されるクォータのリストと、クォータの引き上げをリクエストするためのリンク (利用可能な場合) については、「Amazon RDS のクォータ」を参照してください。

Neptune デプロイパターンを理解する

Neptune DB クラスターには、1 つのプライマリ DB インスタンスと最大 15 の Neptune レプリカがあります。プライマリ DB インスタンスでは、読み書きオペレーションをサポートしており、クラスターボリュームに対してすべてのデータ変更を実行します。Neptune レプリカは、プライマリ DB インスタンスと同じストレージボリュームに接続し、読み取りオペレーションのみサポートしています。Neptune レプリカは、プライマリ DB インスタンスから読み取りワークロードをオフロードします。

高可用性を実現するには、リードレプリカを使用します。異なるアベイラビリティーゾーンで 1 つ以上のリードレプリカインスタンスを利用可能にすると、リードレプリカがプライマリインスタンスのフェイルオーバーターゲットとして機能するため、可用性が向上します。ライターインスタンスが失敗した場合、Neptune はリードレプリカインスタンスをプライマリインスタンスに昇格します。そうすると、プライマリインスタンスに対して行われた読み取り要求と書き込み要求が失敗して例外が発生すると、短時間の中断 (通常は 30 秒未満) が発生し、昇格したインスタンスは再起動されます。最高の信頼性を得るには、2 つのリードレプリカを別々のアベイラビリティーゾーンに使用することを検討してください。アベイラビリティーゾーン 1 のプライマリインスタンスがオフラインになった場合、アベイラビリティーゾーン 2 のインスタンスはプライマリに昇格されますが、その間はクエリを処理できません。したがって、移行中に読み取りクエリを処理するには、アベイラビリティーゾーン 3 のインスタンスが必要です。

Neptune サーバーレスを使用している場合、すべてのアベイラビリティーゾーンのリーダーインスタンスとライターインスタンスは、データベースの負荷に応じて、互いに独立してスケールアップおよびスケールダウンします。リーダーインスタンスの昇格階層を 0 または 1 に設定すると、ライターインスタンスの容量に合わせてスケールアップまたはスケールダウンできます。これにより、現在のワークロードをいつでも引き継ぐ準備が整います。

アプリケーションにグローバルなフットプリントがある場合、またはマルチリージョンフェイルオーバーが必要な場合は、Neptune グローバルデータベースの使用を検討してください。Amazon Neptune グローバルデータベースは複数のデータベースにまたがり AWS リージョン、低レイテン

シーのグローバル読み取りを可能にし、停止が 全体に影響を与えるまれなケースで高速リカバリを提供します AWS リージョン。Neptune グローバルデータベースは、1 つのリージョンにあるプライマリ DB クラスターと、異なるリージョンにある最大 5 つのセカンダリ DB クラスターで構成されます。

Neptune クラスターの管理とスケーリング

Neptune 自動スケーリングを使用すると、DB クラスター内の Neptune レプリカの数を自動的に調整して、CPU 利用率のしきい値に基づいて接続およびワークロードの要件を満たすことができます。自動スケーリングを使用すると、Neptune DB クラスターはワークロードの急激な増加を処理できます。ワークロードが減少すると、自動スケーリングは不要なレプリカを削除し、未使用の容量に対して料金が発生しないようにします。新しいインスタンスの起動には最大 15 分かかる場合があるため、自動スケーリングだけでは需要の急激な変化に対する十分なソリューションにはならないことに注意してください。

自動スケーリングは、既に 1 つのプライマリライターインスタンスと、少なくとも 1 つのリードレプリカインスタンスを持つ Neptune DB クラスターでのみ使用できます (「Amazon Neptune DB Clusters and Instances」を参照)。また、クラスター内のすべての read-replica インスタンスが使用可能な状態である必要があります。リードレプリカが使用可能以外の状態にある場合、クラスター内のすべてのリードレプリカが使用可能になるまで、Neptune 自動スケーリングは何もしません。

需要が急速に変化する場合は、サーバーレスインスタンスの使用を検討してください。サーバーレスインスタンスは短期間で垂直方向にスケールできる一方、自動スケーリングは長期間にわたり水平方向にスケールできます。この設定では、サーバーレスインスタンスが垂直方向にスケールし、自動スケーリングは新しいリードレプリカをインスタンス化して、単一のサーバーレスインスタンスの最大容量を超えるワークロードを処理するため、最適なスケーラビリティを提供します。Amazon Neptune サーバーレスの容量スケーリングの詳細については、「Neptune Serverless DB クラスタの容量スケーリング」を参照してください。

スケーリングのニーズが予測可能なタイミングで変化する場合は、最小インスタンス、最大インスタンス、しきい値に対する変更をスケジュールして、そうした変化するニーズをより適切に処理することができます。必要に応じて、少なくとも 15 分前にスケールアウトイベントをスケジュールして、インスタンスがオンラインになるようにしてください。

パラメータグループのパラメータを使用して、Amazon Neptune のデータベース設定を管理します。パラメータグループは、1 つ以上の DB インスタンスに適用されるエンジン設定値のコンテナとして機能します。パラメータグループのクラスターパラメータを変更するときは、静的パラメータと動的パラメータの違いと、それらを適用する方法とタイミングを理解してください。ステータスエンドポイントを使用して、現在適用されている設定を確認します。

バックアップとフェイルオーバーイベントを管理する

Neptune は、クラスターボリュームを自動的にバックアップし、バックアップ保持期間中、バックアップしたデータを保持します。Neptune のバックアップは連続的かつ増分的であるため、バックアップ保持期間内の任意の時点にすばやく復元できます。DB クラスターを作成または変更するときに、バックアップ保持期間を 1～35 日の間で指定できます。

バックアップ保持期間を超えたバックアップを保持するには、クラスターボリュームの中にデータのスナップショットを作成できます。スナップショットの保存には、Neptune の標準ストレージ料金がかかります。

DB クラスターの Amazon Neptune のスナップショットを作成すると、Neptune はクラスターのストレージボリュームのスナップショットを作成し、個々のインスタンスだけではなく、すべてのデータをバックアップします。新しい DB クラスターは、この DB クラスタースナップショットから復元することで後に作成できます。DB クラスターを復元する場合は、復元の元となる DB クラスタースナップショットの名前を指定し、復元によって作成される新しい DB クラスターの名前を指定します。

フェイルオーバーイベントに対するシステムの応答をテストします。Neptune API を使用してフェイルオーバーイベントを強制します。フェイルオーバーによる再起動が便利なのは、テスト用の DB インスタンスで障害をシミュレートするときや、フェイルオーバーの実行後にオペレーションを元のアベイラビリティーゾーンに復元するときです。詳細については、「マルチ AZ 配置の設定と管理」を参照してください。DB ライタークラスターを再起動すると、そのスタンバイレプリカにフェイルオーバーします。Neptune レプリカを再起動しても、フェイルオーバーは開始されません。

クライアントの信頼性を確保するよう設計します。フェイルオーバーイベント中に動作をテストします。エクスポネンシャルバックオフロジックを使用して、クライアントに再試行ロジックを実装します。このロジックを実装するコード例は、AWS Lambda Amazon Neptune の関数例のドキュメントにあります。

複数のデータベースエンジンに適用される一般的なバックアップ要件のセットがある場合は、AWS Backup の使用を検討してください。

パフォーマンス効率の柱

AWS Well-Architected フレームワークのパフォーマンス効率の柱は、データの取り込みまたはクエリ中にパフォーマンスを最適化する方法に焦点を当てています。パフォーマンスの最適化は、以下を段階的かつ継続的に実行するプロセスです。

- ビジネス要件を確認する
- ワークロードのパフォーマンスを測定する
- パフォーマンスの低いコンポーネントを特定する
- ビジネスニーズに合わせてコンポーネントを調整する

パフォーマンス効率の柱は、使用する適切なグラフデータモデルとクエリ言語を特定するのに役立つ、ユースケース固有のガイドラインを提供します。また、Amazon Neptune へデータを取り込み、Amazon Neptune からデータを使用する際に従うべきベストプラクティスも含まれています。

パフォーマンス効率の柱は、次の主要分野に焦点を当てています。

- グラフモデリング
- クエリの最適化
- クラスターの適切なサイズ設定
- 書き込みの最適化

グラフモデリングを理解する

Labeled Property Graph (LPG) モデルと Resource Description Framework (RDF) モデルの違いを理解します。ほとんどの場合は好みの問題ですが、あるモデルが他のモデルよりも適しているユースケースがいくつかあります。グラフ内の 2 つのノードを接続するパスに関する知識が必要な場合は、LPG を選択します。Neptune クラスターまたは他のグラフのトリプルストア間でデータをフェデレーションする場合は、RDF を選択します。

Software as a Service (SaaS) アプリケーション、またはマルチテナンシーを必要とするアプリケーションを構築する場合は、クラスターごとに 1 つのテナントを設定するのではなく、データモデルにテナントの論理的な分離を組み込むことを検討してください。このタイプの設計を実現するには、SPARQL の名前付きグラフとラベル付け戦略を使用できます。例えば、ラベルに顧客識別子を付加したり、テナント識別子を表すプロパティのキーと値のペアを追加したりできます。論理的な分

離を維持するために、クライアントレイヤーがこれらの値を挿入することを確認します。マルチテナンシーの推奨事項の詳細については、Amazon Neptune データベースを実行する ISVs」を参照してください。

クエリのパフォーマンスは、クエリの処理で評価する必要があるグラフオブジェクト (ノード、エッジ、プロパティ) の数によって異なります。そのため、グラフモデルはアプリケーションのパフォーマンスに大きな影響を与える可能性があります。可能な限り細かいラベルを使用し、パスの決定またはフィルタリングに必要なプロパティのみを保存します。より高いパフォーマンスを実現するには、要約ノードの作成や共通パスを接続するより直接的なエッジの作成など、グラフの一部を事前に計算することを検討してください。

同じラベルを持つエッジの数が異常に多いノード間の移動はなるべく避けます。このようなノードには多くの場合、何千ものエッジがあります (大部分のノードのエッジカウントは数十個です)。その結果、コンピューティングとデータの複雑さがはるかに高くなります。これらのノードは、一部のクエリパターンでは問題にならない場合がありますが、特に中間ステップとしてノード間を移動する場合は、このようなノードを回避するためにデータを異なる方法でモデリングすることをお勧めします。スロークエリログを使用して、これらのノード間を移動するクエリを識別できます。特にデバッグモードを使用すると、レイテンシーとデータアクセスメトリクスが平均的なクエリパターンよりもはるかに高くなることがわかります。

Neptune を使用して ID にランダムな GUID 値を割り当てる代わりに、確定的なノード ID をノードとエッジに使用します (ユースケースでサポートされている場合)。ID によるノードへのアクセスが最も効率的な方法です。

クエリを最適化する

LPG モデルでは、openCypher 言語と Gremlin 言語を同じように使用できます。パフォーマンスが最大の懸念事項である場合は、2 つの言語を互換的に使用することを検討してください。特定のクエリパターンでは、ある言語が他の言語よりもパフォーマンスに優れる可能性があるためです。

Neptune は、代替クエリエンジン (DFE) への変換を進めています。openCypher は DFE でのみ実行されますが、オプションで、クエリ注釈を使用して Gremlin クエリと SPARQL クエリの両方を DFE で実行するように設定できます。DFE を有効にした状態でクエリをテストし、DFE を使用しない場合のクエリパターンのパフォーマンスと比較することを検討してください。

Neptune は、グラフ全体を評価する分析クエリではなく、単一ノードまたは一連のノードから開始し、そこからファンアウトするトランザクションタイプのクエリに最適化されています。分析クエリワークロードには、Neptune Analytics を使用します。Neptune Analytics は、データ、分析、アル

ゴリズム処理に高速反復を必要とする調査、調査、またはデータサイエンスのワークロードに最適です。また、グラフデータに対してベクトル検索を実行し、Neptune データベースインスタンスから直接データをロードすることもできます。Neptune Analytics がニーズを満たさない場合は、 AWS SDK for Pandas または と AWS Glueまたは Amazon EMR を組み合わせた neptune-export の使用を検討することもできます。

モデルとクエリの非効率やボトルネックを特定するには、各クエリ言語の `profile` および `explain` API を使用して、クエリプランとクエリメトリクスの詳細な説明を取得します。詳細については、「Gremlin profile」、「openCypher explain」、および「SPARQL explain」を参照してください。

クエリパターンを理解します。グラフ内の個別のエッジの数が多くなると、Neptune のデフォルトのアクセス方式は効率が悪くなる場合があります。以下のクエリは、非常に非効率になる可能性があります。

- エッジラベルが指定されていない場合にエッジ間を後方に移動するクエリ。
- Gremlin の `.both()` など、内部的に同じパターンを使用する句、または任意の言語でノードをドロップする句 (ラベルを認識せずに受信エッジをドロップする必要があります)。
- プロパティラベルを指定せずにプロパティ値にアクセスするクエリ。これらのクエリは、非常に非効率になる可能性があります。これが使用パターンと一致する場合は、OSGP インデックス (目的語、主語、グラフ、述語) を有効にすることを検討してください。

スロークエリログを使用して、遅いクエリを識別します。クエリの遅延は、最適化されていないクエリプランや不必要に多いインデックスルックアップが原因で発生する可能性があり、I/O コストが増加する可能性があります。Gremlin、SPARQL、または openCypher 向けの Neptune の explain エンドポイントと profile エンドポイントは、これらのクエリが遅い理由を理解するのに役立ちます。句には次のものが含まれる場合があります。

- グラフ内の平均ノードと比較してエッジ数が異常に多いノード (数十に対して数千など) は、計算の複雑さが増し、レイテンシーが長くなり、リソースの消費量が増える可能性があります。これらのノードが正しくモデル化されているかどうか、または横断する必要があるエッジの数を減らすためにアクセスパターンを改善できるかどうかを決定します。
- 最適化されていないクエリには、特定のステップが最適化されていないという警告が含まれます。これらのクエリを書き換えて最適化されたステップを使用すると、パフォーマンスが向上する可能性があります。
- 冗長フィルターは、不要なインデックスルックアップを引き起こす可能性があります。同様に、冗長パターンは重複したインデックスルックアップを引き起こす可能性があり、これはクエリを

改善することで最適化できます (プロファイル出力の「`Index Operations - Duplication ratio`」を参照)。

- Gremlin などの一部の言語には厳密に型指定された数値がなく、代わりに型の上位変換を使用します。例えば、値が 55 の場合、Neptune は、整数型、ロング型、浮動小数点型、およびその他の 55 に相当する数値型の値を検索します。これにより、追加のオペレーションが発生します。型が一致することが事前にわかっている場合は、クエリヒントを使用してこれを回避できます。
- グラフモデルはパフォーマンスに大きな影響を与える可能性があります。より細かいラベルを使用するか、マルチホップ線形パスへのショートカットを事前に計算して、評価する必要があるオブジェクトの数を減らすことを検討してください。

クエリの最適化だけではパフォーマンス要件を満たすことができない場合は、Neptune でさまざまなキャッシュ手法を使用してこれらの要件を満たすことを検討してください。

Neptune のパフォーマンスは、バージョンごとに継続的に向上しています。リリースノートを確認して、各リリースの改善点の詳細を確認してください。最適なパフォーマンスを実現するために、Neptune DB クラスターの定期的な更新を計画することを検討してください。新しいバージョンでは、新しいインスタンスもサポートされています。`r8g` インスタンスを利用できるようにするには、1.4.5.0 以降にアップグレードすることを検討してください。これがワークロードのパフォーマンスを向上させる方法の詳細については、「Amazon Amazon Neptune v1.4.5 を使用した AWS Graviton4 R8g インスタンスでの書き込みクエリの料金パフォーマンスが 4.7 倍向上しました」を参照してください。

クラスターを適切なサイズにする

同時実行要件とスループット要件に合わせてクラスターのサイズを設定します。クラスター内の各インスタンスで処理できる同時クエリの数は、そのインスタンスの仮想 CPU (vCPU) の数の 2 倍です。すべてのワーカースレッドが占有されている間に到着する追加のクエリは、サーバー側のキューに入れられます。これらのクエリは、ワーカースレッドが利用可能になったときに先入れ先出し (FIFO) ベースで処理されます。`MainRequestQueuePendingRequests` Amazon CloudWatch メトリクスには、各インスタンスの現在のキューの深さが表示されます。この値が頻繁に 0 を超える場合は、より多くの vCPU を持つインスタンスの選択を検討してください。キューの深さが 8,192 を超えると、Neptune は`ThrottlingException` エラーを返します。

各インスタンスの RAM の約 65% がバッファキャッシュ用に予約されています。バッファキャッシュは、データのワーキングデータセット (グラフ全体ではなく、クエリ対象のデータのみ) を保持します。ストレージではなくバッファキャッシュから取得されるデータの割合を確認するには、CloudWatch メトリクスの `BufferCacheHitRatio` をモニタリングします。このメトリクスが

99.9% を下回ることがよくある場合は、より多くのメモリを持つインスタンスを試して、レイテンシーと I/O コストが減少するかどうか判断することを検討してください。

リードレプリカは、ライターインスタンスと同じサイズである必要はありません。ただし、書き込みワークロードが多いと、レプリケーションに追いつくことができないため、小さいレプリカは遅くなり、再起動する可能性があります。したがって、レプリカのサイズはライターインスタンスと同じかそれ以上にすることをお勧めします。

リードレプリカに自動スケーリングを使用する場合は、新しいリードレプリカをオンラインにするのに最大 15 分かかる場合があることに注意してください。クライアントトラフィックが急に、ただし予測可能な範囲で増加する場合は、スケジュールに基づくスケーリングを使用して、その初期化時間を考慮してリードレプリカの最小数を高く設定することを検討してください。

サーバーレスインスタンスは、いくつかの異なるユースケースとワークロードをサポートしています。以下のシナリオでは、プロビジョニングインスタンスよりもサーバーレスを優先して検討します。

- ワークロードが 1 日を通して頻繁に変動する。
- 新しいアプリケーションを作成したが、ワークロードのサイズが不明。
- 開発とテストを実行している。

サーバーレスインスタンスは、同等のプロビジョニングされたインスタンスよりも、GB 単位の RAM ベースでコストがかかることに注意してください。各サーバーレスインスタンスは、2 GB の RAM と関連する vCPU およびネットワークで構成されます。想外の請求を回避するため、オプションごとのコストを分析します。一般的に、大量のワークロードが発生するのは毎日数時間のみで、残りの時間はほぼゼロの場合、またはワークロードが 1 日を通して大幅に変動する場合にのみ、サーバーレスでコスト削減を達成できます。

Amazon Neptune 料金計算ツールを使用して、1 queries-per-second (QPS) 要件などの要因に基づいてクラスターの正しい設定を評価します。

書き込みの最適化

書き込みを最適化するには、次の点を考慮してください。

- Neptune Bulk Loader は、最初にデータベースをロードしたり、既存のデータに追加したりする場合に最適な方法です。Neptune ローダーはトランザクションではなく、データを削除できないため、データの削除が要件である場合は使用しないでください。

- トランザクションの更新は、サポートされているクエリ言語を使用して行うことができます。書き込み I/O オペレーションを最適化するには、コミットごとに 50～100 個のオブジェクトを一括でデータを書き込みます。オブジェクトとは、LPG のノード、エッジ、またはノードもしくはエッジ上のプロパティ、または RDF のトリプルストアまたはクワッドです。
- すべてのトランザクション Neptune 書き込みオペレーションは、接続ごとにシングルスレッドになります。Neptune に大量のデータを送信するときは、それぞれがデータを書き込む複数の並列接続を設定することを検討してください。Neptune でプロビジョニングされたインスタンスを選択すると、インスタンスサイズは vCPU の数に関連付けられます。Neptune はインスタンス上の vCPU ごとに 2 つのデータベーススレッドを作成するため、最適な並列接続を決定するためにテストする際は vCPU の数の 2 倍から開始します。 サーバーレスインスタンスは、4 NCU ごとにおよそ 1 のレートで vCPU の数をスケーリングします。

  > **Note**
  >
  > これは一括ロード API には適用されず、直接接続にのみ適用されます。

- 任意の時間に単一の接続だけがデータを書き込んでいる場合でも、すべての書き込みプロセス中の ConcurrentModificationExceptions に備えて計画を立て、効率的に処理します。`ConcurrentModificationExceptions` が発生しても信頼性を損なわないようクライアントを設計します。
- すべてのデータを削除する場合は、同時削除クエリを発行するのではなく、高速リセット API を使用することを検討してください。後者は前者に比べてはるかに時間がかかり、かなりの I/O コストが発生します。
- ほとんどのデータを削除する場合は、neptune-export を使用して新しいクラスターにデータをロードすることで、保持するデータをエクスポートすることを検討してください。その後、元のクラスターを削除します。

コスト最適化の柱

AWS Well-Architected フレームワークのコスト最適化の柱は、不要なコストを回避することに重点を置いています。以下の推奨事項は、コスト最適化の設計原則とアーキテクチャのベストプラクティスを Amazon Neptune に実行するのに役立ちます。

コスト最適化の柱は、次の主要分野に焦点を当てています。

- 長期的な支出を把握し、資金配分を管理する
- 適切なタイプおよび数量のリソースを選択する
- 過剰な支出なしでスケールし、ビジネスニーズを満たす

必要な使用状況パターンとサービスの理解

データモデルに識別可能なグラフ構造があり、クエリで関係を調べて複数のホップを通過する必要がある場合、Neptune はワークロードに適しています。グラフデータベースは、次のパターンには適していません。

- 主にシングルホップクエリ (データをオブジェクトの属性として表現した方が適切かどうかを検討してください)
- プロパティとして保存される JSON または BLOB データ
- 多数のノードにわたる数値プロパティの合計の計算など、データセット全体で集計されるクエリ

特定のアクセスパターンに複数の目的別データベースを一緒に使用することで、すべてのニーズに対応できるかどうかを検討してください。例えば、次のようになります。

- 複雑なグラフナビゲーションをあまり必要とせず、単一ノードのプロパティを高度に同時取得する必要がある API では、Neptune、DynamoDB、または Amazon DocumentDB の 1 つ以上を使用することが最適な場合があります。
- リレーショナルデータベースは Neptune と共存して既存の機能を維持できますが、Neptune は、リレーショナルデータベースでうまく実行およびスケーリングされないマルチホップトラバーサルにのみ使用できます。

Neptune とやり取りするサービスや Neptune を補完するサービスに付随する以下のようなコストを把握してください。

- Neptune に一括ロードされるデータファイルの Amazon Simple Storage Service (Amazon S3) ストレージコスト
- クエリの挿入またはアップサート、クエリの読み取り、Neptune ストリーム処理に使用される Lambda 関数
- Amazon API Gateway または で (データベースに直接接続する代わりに) クライアントアプリケーションとやり取りするために Neptune 上に構築された API レイヤー AWS AppSync
- AWS Glue Neptune との間でデータを転送するために使用する ジョブ
- ストリーミングデータを受信して Neptune にほぼリアルタイムで取り込む Amazon Kinesis または Amazon Managed Streaming for Apache Kafka (Amazon MSK) インスタンス。
- AWS Database Migration Service Neptune へのリレーショナルデータの移行用
- Jupyter Notebook と Deep Graph Library 機械学習モデルの Amazon SageMaker ランタイムコスト

コストに注意してリソースを選択する

Neptune の料金は、時間単位のインスタンスコスト (またはサーバーレスに消費される Neptune Compute Units)、データ I/O、ストレージ使用量に基づいています。インスタンスは平均してコスト全体の 85% を占めるため、適切なサイズを選ぶことがコストに大きな影響を与える可能性があります。インスタンスのサイズを適正化する最善の方法は、さまざまなインスタンスでアプリケーションのパフォーマンスをテストし、次の要因を比較することです。

- `MainRequestQueuePendingRequests` CloudWatch メトリクスは一貫してゼロに近い低い数値のままですか?
- `BufferCacheHitRatio` CloudWatch メトリクスはほとんどの時間で 99.9% 以上を維持していていますか?
- インスタンスコストおよび関連するデータ I/O コストのコストとパフォーマンス曲線はどうですか? ストレージとのバッファキャッシュのスワップを頻繁に必要とするサイズの小さいインスタンスでは、データ読み取りコストが大幅に増加する可能性があります。`BufferCacheHitRatio` は、これらのシナリオでは頻繁に低下します。

インスタンスコストは、同じインスタンスファミリー内のサイズに応じて直線的にスケーリングされます。`db.r6i.2xlarge` インスタンスの時間単位のコストは `db.r6i.xlarge` インスタンスの 2 倍で、リソース割り当ても 2 倍です。`db.r6i.24xlarge` インスタンスは、`db.r6i.xlarge` インスタンスの時間単位のコストの 24 倍です。

サポートする必要がある同時クエリの数を見積もります。読み取り専用クエリを処理するために、0 ~ 15 個のリードレプリカを持つことができます。要件が時間帯、週、または月によって異なる場合は、複数の小さなインスタンスを使用してスケジュールに基づいてスケーリングできます。インスタンスの各 vCPU には、同時クエリを処理するための 2 つのスレッドが用意されています。それぞれ 4 つの vCPU を持つ 3 つの `db.r6i.xlarge` リードレプリカは、24 個の同時クエリを処理できます。

そうではなく、トラフィック量が 1 秒あたりのクエリ数 (QPS) で測定される場合は、実験でクエリの平均レイテンシーを決定する必要があります。Neptune クラスターがサポートできる 1 秒あたりのクエリ数は、`vCPU × 2 × (1 second/average query latency)` に等しくなります。例えば、4 つの vCPU があり、クエリレイテンシーが 100 ミリ秒 (0.1 秒) の場合、`QPS = 4 × 2 × (1s/0.1s) = 80 queries per second` となります。

プロビジョニングされたインスタンスは、継続的で安定した、予測可能なワークロードに対して、サーバーレスよりも安価です。サーバーレスは、1 日あたり数時間のみ非常に高い使用量を必要とするワークロード (`db.r6i.4xlarge` など) があり、残りの時間はほとんどトラフィックがない (1 Neptune Compute Unit など) 場合に、コスト最適化の機会が得られます。数時間スケールアップし、その後スケールダウンするサーバーレスインスタンスは、プロビジョニングされた `db.r6i.4xlarge` インスタンスを終日使用するよりも安価になります。

Neptune 1.4.5.0 以降にアップグレードし、 `r8g` インスタンスを使用して、 `r7g`や などの旧世代のインスタンスよりも低コストで読み取りと書き込みのスループットを向上させることを検討してくださいr6g。詳細については、Amazon Neptune v1.4.5 を使用した AWS Graviton4 R8g インスタンスでの書き込みクエリの料金パフォーマンスが 4.7 倍向上する」(AWS ブログ記事) を参照してください。

Neptune クラスターは、デフォルトで標準ストレージで作成されます (コンソールを使用して を作成すると、デフォルトで I/O 最適化ストレージが選択されます) 。I/O 最適化ストレージでは、ストレージとインスタンスのコストがわずかに高くなりますが、I/O コストはありません。これにより、予測可能な経常コストが発生しますが、I/O 使用率が一般的に低い場合は、標準ストレージを利用する方がコスト効率が良い可能性があります。最初に大量のデータをロードする場合は、I/O 最適化ストレージを選択し、初期データロードを実行してから、標準ストレージに切り替えることでコストを最適化できます。ストレージタイプは請求モデルにのみ影響し、Neptune DB クラスターまたはインスタンス設定に技術的な違いはありません。ストレージタイプは 30 日に 1 回変更できます。30 日後、詳細な Neptune コストを確認し、Neptune 料金ページを使用して、I/O 最適化ストレージを使用してコストが高くなったかどうかを計算します。そうであれば、引き続き標準ストレージを使用し、そうでない場合は I/O 最適化に切り替えます。

ワークロードに最適な Neptune インスタンス設定を選択する

2025 年 7 月 15 日より AWS アカウント 前に を作成した場合は、Neptune でのエントリレベルの実験に AWS 無料利用枠を使用できます。db.t3.medium および db.t4g.medium インスタンスの無料利用時間である 750 時間で、Neptune を低スケールで十分に理解できます。クラスターは無料トライアル期間終了後も残りますが、それ以降の使用量に対しては課金されます。

db.t3.medium および db.t4g.mediumインスタンスは、openCypher、Graph Explorer、またはさまざまな生成 AI 統合を使用していない低コストの開発環境に適しています。これらのインスタンスの RAM-to-vCPUの比率 (2:1) は、Rファミリーインスタンス (8:1) またはXファミリーインスタンス (16:1) よりも小さくなります。これにより、比率が削減され、openCypher パフォーマンス、GenAI 統合 (LLM にグラフスキーマを通知するため)、Graph Explorer を有効にする DFE エンジン統計が使用できなくなります。パフォーマンスプロファイルは、Tファミリーインスタンスを使用する場合、特に前述のワークロードでは大きく異なる場合があります。これらのインスタンスは、クエリがグラフの大部分をナビゲートOutOfMemoryExceptionsする場合の の発生を増やすこともできます。後者の条件が影響を受ける可能性があるかどうかを判断するには、BufferCacheHitRatio CloudWatch メトリクスを確認します。

本番環境を示唆しない一貫性のない結果が発生する可能性があるため、Tファミリーインスタンスでのパフォーマンステストや負荷テストを行わないことを強くお勧めします。

プロビジョニングされたインスタンスは、ワークロードがかなり安定していて予測可能な場合に、コストとパフォーマンスの最適な組み合わせを提供します。インスタンスサイズは、必要なリクエストの同時実行数とクエリの複雑さに基づいて選択します。同時実行数を増やすには、より多くの vCPU が必要です。クエリの複雑さが高いほど、より多くの RAM が必要になります。前者の影響を判断するには、MainRequestQueuePendingRequests CloudWatch メトリクスを使用します (0 より大きい場合は、同時リクエストの数が処理可能な数よりも多いことを表します)。後者の影響を判断するには、BufferCacheHitRatio CloudWatch メトリクスを使用します。比率が頻繁に 99.9% を下回る場合、評価対象のグラフの作業部分を含むのに十分な RAM がなく、キャッシュスワップの頻度が高くなることを示します。インスタンスの R ファミリーに十分な同時実行性があるものの、十分な RAM がない場合は、インスタンスのXファミリーを試すことを検討してください。

サーバーレスインスタンスの理想的なユースケースについては、Neptune ドキュメントで説明しています。プロビジョニングとサーバーレスのどちらが最適かが不明で、コストが主な懸念事項である場合は、サーバーレスでワークロードをテストして、使用される NCU の数を判断し、プロビジョニング (N hours × hourly provisioned cost) のコストをサーバーレス (sum of NCUs × hourly cost per NCU) と比較します。同等のサイズのプロビジョニングインスタンスがわからない場合、1 NCU は、約 2 GB の RAM および関連する vCPU とネットワークに相当します。プロ

ビジョニングされたインスタンスが r6iファミリーのものである場合、比率は 8 GB の RAM あたり 1 vCPU、または関連するネットワークとともに 4 NCUs。Amazon Neptune 料金計算ツールには、最適なコスト設定を決定するのに役立つ比較も用意されています。

プライマリインスタンスとレプリカインスタンスにサーバーレスを使用する場合、昇格階層 0 と 1 のリードレプリカは、フェイルオーバーイベントが発生した場合に適切にスケーリングされるように、ライターインスタンスに合わせて NCU をスケーリングすることに注意してください。これらのインスタンスの NCU の制限は、ライターまたはリーダーのどちらのインスタンスが最も多くのトラフィックを受信するかに基づいて設定します。

クラスターが 1 日 24 時間、週 7 日必要でない環境では、使用していないときに Neptune インスタンスをオフにし、使用する前に再度起動するスクリプトを記述することを検討してください。必要なメンテナンスアップデートが適用されるように、Neptune インスタンスは 7 日ごとに自動的に再起動されます。インスタンスを長期間オフのままにする場合は、週次スクリプトを使用して再度シャットダウンします。

適切なサイズのデータストレージと転送

より効率的なクエリ (触れる必要のあるグラフ内のノード、エッジ、プロパティが少ないクエリなど) では、必要な I/O 転送が少なくなり、バッファキャッシュが少なくて済むため、より小さなインスタンスを使用できる可能性があります。クエリ言語のプロファイルエンドポイントまたは説明エンドポイントを使用してクエリを最適化し、クエリのパフォーマンスに合わせてグラフモデルを最適化することを検討してください。

Neptune は大きな文字列にディクショナリエンコードを使用し、そのディクショナリは効率ではなくパフォーマンスに最適化されています。プロパティに対して大きな BLOBs、JSON、または頻繁に変更される文字列がある場合は、それらを Neptune の外部の Amazon S3、Amazon DynamoDB、または Amazon DocumentDB に保存し、Neptune ノードには参照のみを保存することを検討してください。

場合によっては、インスタンスサイズを大きくするとコストを下げることができます。`BufferCacheHitRatio` が低いために I/O コストが非常に高い場合、バッファキャッシュが大きくすると、そのコストが大幅に削減される可能性があります。これは、データがストレージから頻繁にスワップされ、I/O 転送レートが発生する代わりに、すべてのデータがキャッシュに収まるためです。

Neptune はコピーオンライトでクローンを作成します。クローンを作成してグラフを複数のシャードに分割する場合、クローンされたクラスター上の不要なデータを削除しない方が効率的です。これ

には新しいデータページの作成が含まれ、ストレージコストが増加するためです。クローン作成イベント前から変更されていないデータは、2 つのクラスター間で共有される単独のデータページに存在し、その単独のコピーに対してのみ課金されます。

OSGP インデックスを有効にしたり、R5d インスタンスを使用したりしないでください。ただし、ワークロードに大きな違いが生じることをテストで確認している場合は除きます。どちらもめったに発生しないシナリオ向けに設計されており、メリットがほとんどないにもかかわらずコストが増加する可能性があります。

持続可能性の柱

持続可能性の柱は、クラウドワークロードの実行による環境への影響を最小限に抑えることに焦点を当てています。主なトピックは、持続可能性に関する責任共有モデル、影響の把握、必要なリソースを最小限に抑えてダウンストリームへの影響を軽減するための使用率の最大化です。

持続可能性の柱には、以下の主要な重点分野が含まれています。

- ユーザーによる影響
- 持続可能性の目標
- 使用率の最大化
- より効率的な新しいハードウェアおよびソフトウェア製品の予測および採用
- マネージドサービスの使用
- ダウンストリームの影響軽減

このガイドでは、ユーザーによる影響に焦点を当てます。その他の持続可能性設計原則の詳細については、AWS 「Well-Architected フレームワーク」を参照してください。

ユーザーによる選択と要件は、環境に影響を及ぼします。ユーザーが炭素集約度の低い AWS リージョン を選択し、ユーザーの要件が単に稼働時間と耐久性を最大化するのでなく、実際のワークロードのニーズを反映している場合、ワークロードの持続可能性が向上します。ワークロードの設計および継続的な運用に、次のセクションで説明するベストプラクティスと考慮事項を採用すると、環境に良い影響を及ぼすことができます。

AWS リージョン 選択

一部の AWS リージョン は Amazon 再生可能エネルギープロジェクトの近くにあるか、グリッドの炭素強度が他のものよりも低い公開されている場所にあります。リージョンを選ぶ際は、ワークロードにとって有効な持続可能性への影響を考慮し、リストを Neptune が利用可能なリージョンと相互参照します。

ユーザー行動パターンに基づく使用量

ユーザーのトラフィックと動作に合わせて使用量を適切なサイズにすることで、 AWS が環境に与える影響を最小限に抑えることができます。ソリューションを設計するときは、次のベストプラクティスを考慮してください。

- CPUUtilization、MainRequestQueuePendingRequests、TotalRequestsPerSec などの Amazon CloudWatch メトリクスをモニタリングして、需要が最も高いときと最も低いときを判断し、そうしたタイミングにおいてクラスターリソースのサイズが適切に設定されていることを確認します。
- 使用していない時間帯は非本番環境を停止するよう自動化します。詳細については、「Automate the stopping and starting of Amazon Neptune environment resources using resource tags」を参照してください。
- トラフィックパターンの変動が頻繁かつ予測不可能な場合は、ピークトラフィック用にプロビジョニングされたインスタンスを使用する代わりに、需要に応じてスケールアップ/ダウンする Neptune サーバーレスインスタンスを使用することを検討してください。
- サービスレベル契約を、ビジネス継続性の目標だけでなく持続可能性の目標に合わせることを検討してください。マルチリージョンディザスタリカバリ、高可用性、長期バックアップ保持などの簡単な要件については、特に非本番環境やミッションクリティカルでないワークロードの場合、これらの目標を達成するために必要なリソースの量を減らすことができます。

ソフトウェア開発とアーキテクチャパターンを最適化する

無駄を防ぐには、モデルとクエリを最適化し、コンピューティングリソースを共有して、Neptune インスタンスとクラスターで使用可能なすべてのリソースを活用します。具体的なベストプラクティスは次のとおりです。

- 開発者に、それぞれが独自のインスタンスを作成するのではなく、Neptune インスタンスと Jupyter Notebook アプリケーションのインスタンスを共有してもらいます。マルチテナンシーパーティショニング戦略を使用して、各デベロッパーに単一の Neptune クラスターに独自の論理パーティションを提供し、単一の Jupyter インスタンスで各デベロッパーに個別のノートブックフォルダを作成します。
- リソースの使用率を最大化し、アイドル時間を最小限に抑えるパターンを実装します。例えば、データをロードするための並列スレッドや、レコードをより大きなトランザクションにまとめてバッチ処理するなどです。
- クエリとグラフモデルを最適化して、結果の計算に必要なリソースを最小限に抑えます。
- Gremlin クエリの結果については、結果キャッシュ機能を使用して、ページ分割されたクエリまたは頻繁に繰り返されるクエリの再計算に費やされるリソースを最小限に抑えます。
- Neptune 環境を最新の状態に保ちます。Neptune の最新バージョンは、Graviton などの最新の Amazon EC2 インスタンスをより効率的にサポートします。また、クエリ最適化の改善とバグ修正により、クエリの計算に必要なリソースの量を削減できます。

リソース

リファレンス

- AWS Well-Architected
- AWS Well-Architected フレームワークのドキュメント
- Neptune の最新更新
- ベストプラクティス: Neptune を最大限に活用する
- Amazon Neptune 料金計算ツール

ブログ記事

- Automated testing of Amazon Neptune data access with Apache TinkerPop Gremlin
- Automate the stopping and starting of Amazon Neptune environment resources using resource tags
- Fine Grained Access Control for Amazon Neptune data plane actions
- Amazon Neptune v1.4.5 を使用した AWS Graviton4 R8g インスタンスでの書き込みクエリの料金パフォーマンスが 4.7 倍向上
- Orca Security が Amazon Neptune データベースのパフォーマンスを最適化する方法
- Amazon Neptune パブリックエンドポイントを使用してグラフアプリケーションを迅速に構築する
- 新しい Amazon Neptune エンジンバージョンはopenCypher クエリパフォーマンスのために最大 9 倍高速および 10 倍高いスループットを提供します。

無料 AWS スキルビルダーコース

- Amazon Neptune の開始方法
- Building Applications on Amazon Neptune
- Data Modeling for Amazon Neptune

寄稿者

このガイドの寄稿者は次のとおりです。

- Brian O'Keefe、プリンシパル Neptune ソリューションアーキテクト、 AWS
- Abhishek Mishra、シニア Neptune ソリューションアーキテクト、 AWS
- Ganesh Sawhney、戦略的パートナーサクセスソリューションアーキテクトチームリーダー AWS
- マイケル・ヘイビー、シニア Neptune ソリューションアーキテクト、 AWS
- Kevin Phillips、Neptune ソリューションアーキテクト、 AWS
- Melissa Kwok、Neptune ソリューションアーキテクト、 AWS
- プリンシパルソリューションアーキテクト、Sakti Mishra AWS
- Javed Ali、シニアソリューションアーキテクト、 AWS

ドキュメント履歴

以下の表は、本ガイドの重要な変更点について説明したものです。今後の更新に関する通知を受け取る場合は、RSS フィードをサブスクライブできます。

変更	説明	日付
Neptune リリースの更新	Amazon Neptune 1.4.6.0 以降に関する情報を含めるようにドキュメントを更新しました。	2026 年 1 月 2 日
初版発行	—	2023 年 9 月 27 日

AWS 規範ガイダンスの用語集

以下は、 AWS 規範ガイダンスによって提供される戦略、ガイド、パターンで一般的に使用される用語です。エントリを提案するには、用語集の最後のフィードバックの提供リンクを使用します。

数字

7 Rs

アプリケーションをクラウドに移行するための 7 つの一般的な移行戦略。これらの戦略は、ガートナーが 2011 年に特定した 5 Rs に基づいて構築され、以下で構成されています。

- リファクタリング/アーキテクチャの再設計 — クラウドネイティブ特徴を最大限に活用して、俊敏性、パフォーマンス、スケーラビリティを向上させ、アプリケーションを移動させ、アーキテクチャを変更します。これには、通常、オペレーティングシステムとデータベースの移植が含まれます。例: オンプレミスの Oracle データベースを Amazon Aurora PostgreSQL 互換エディションに移行する。
- リプラットフォーム (リフトアンドリシェイプ) — アプリケーションをクラウドに移行し、クラウド機能を活用するための最適化レベルを導入します。例: お客様のオンプレミスの Oracle データベースを AWS クラウドの Oracle 用の Amazon Relational Database Service (Amazon RDS) に移行する。
- 再購入 (ドロップアンドショップ) — 通常、従来のライセンスから SaaS モデルに移行して、別の製品に切り替えます。例: 顧客関係管理 (CRM) システムを Salesforce.com に移行する。
- リホスト (リフトアンドシフト) — クラウド機能を活用するための変更を加えずに、アプリケーションをクラウドに移行します。例: お客様のオンプレミスの Oracle データベースを AWS クラウドの EC2 インスタンス上の Oracle に移行する。
- 再配置 (ハイパーバイザーレベルのリフトアンドシフト) – 新しいハードウェアを購入したり、アプリケーションを書き換えたり、既存の運用を変更したりすることなく、インフラストラクチャをクラウドに移行できます。オンプレミスプラットフォームから同じプラットフォームのクラウドサービスにサーバーを移行します。例： Microsoft Hyper-Vアプリケーションを に移行します AWS。
- 保持 (再アクセス) — アプリケーションをお客様のソース環境で保持します。これには、主要なリファクタリングを必要とするアプリケーションや、お客様がその作業を後日まで延期したいアプリケーション、およびそれらを移行するためのビジネス上の正当性がないため、お客様が保持するレガシーアプリケーションなどがあります。
- 廃止 — お客様のソース環境で不要になったアプリケーションを停止または削除します。

A

ABAC

「属性ベースのアクセス制御」をご覧ください。

抽象化されたサービス

「マネージドユーザー」をご覧ください。

ACID

「原子性、一貫性、分離性、耐久性 (ACID)」をご覧ください。

アクティブ/アクティブ移行

(双方向レプリケーションツールまたは二重書き込み操作を使用して) ソースデータベースとターゲットデータベースを同期させ、移行中に両方のデータベースが接続アプリケーションからのトランザクションを処理するデータベース移行方法。この方法では、1 回限りのカットオーバーの必要がなく、管理された小規模なバッチで移行できます。アクティブ/パッシブ移行よりも柔軟な方法ですが、さらに多くの作業が必要となります。

アクティブ/パッシブ移行

ソースデータベースとターゲットデータベースを同期させながら、データがターゲットデータベースにレプリケートされている間、接続しているアプリケーションからのトランザクションをソースデータベースのみで処理するデータベース移行方法。移行中、ターゲットデータベースはトランザクションを受け付けません。

集計関数

複数行に処理を行い、グループ全体を対象に単一の戻り値を計算する SQL 関数。集計関数の例としては、SUM や MAX などがあります。

AI

「人工知能」をご覧ください。

AIOps

「AI オペレーション」をご覧ください。

匿名化

データセット内の個人情報を完全に削除するプロセス。匿名化は個人のプライバシー保護に役立ちます。匿名化されたデータは、もはや個人データとは見なされません。

アンチパターン

繰り返し起こる問題に対して頻繁に用いられる解決策で、その解決策が逆効果であったり、効果がなかったり、代替案よりも効果が低かったりするもの。

アプリケーション制御

マルウェアからシステムを保護するために、承認されたアプリケーションのみを使用できるようにするセキュリティアプローチ。

アプリケーションポートフォリオ

アプリケーションの構築と維持にかかるコスト、およびそのビジネス価値を含む、組織が使用する各アプリケーションに関する詳細情報の集まり。この情報は、ポートフォリオの検出と分析プロセス の重要な要素であり、移行、モダナイズ、最適化するアプリケーションを特定し、優先順位を付けるのに役立ちます。

人工知能 (AI)

コンピューティングテクノロジーを使用し、学習、問題の解決、パターンの認識など、通常は人間に関連づけられる認知機能の実行に特化したコンピュータサイエンスの分野。詳細については、「人工知能 (AI) とは何ですか?」をご覧ください。

AI オペレーション (AIOps)

機械学習技術を使用して運用上の問題を解決し、運用上のインシデントと人の介入を減らし、サービス品質を向上させるプロセス。 AWS 移行戦略での AIOps の使用方法については、オペレーション統合ガイド を参照してください。

非対称暗号化

暗号化用のパブリックキーと復号用のプライベートキーから成る 1 組のキーを使用した、暗号化のアルゴリズム。パブリックキーは復号には使用されないため共有しても問題ありませんが、プライベートキーの利用は厳しく制限する必要があります。

原子性、一貫性、分離性、耐久性 (ACID)

エラー、停電、その他の問題が発生した場合でも、データベースのデータ有効性と運用上の信頼性を保証する一連のソフトウェアプロパティ。

属性ベースのアクセス制御 (ABAC)

部署、役職、チーム名など、ユーザーの属性に基づいてアクセス許可をきめ細かく設定する方法。詳細については、 AWS Identity and Access Management (IAM) ドキュメントの「 の ABAC AWS」を参照してください。

信頼できるデータソース

最も信頼性のある情報源とされるデータのプライマリーバージョンを保存する場所。匿名化、編集、仮名化など、データを処理または変更する目的で、信頼できるデータソースから他の場所にデータをコピーすることができます。

アベイラビリティーゾーン (AZ)

他のアベイラビリティーゾーンの障害から AWS リージョン 隔離され、同じリージョン内の他のアベイラビリティーゾーンへの低コストで低レイテンシーのネットワーク接続を提供する 内の別の場所。

AWS クラウド導入フレームワーク (AWS CAF)

組織がクラウドへの移行を成功させるための効率的で効果的な計画を立て AWS るための、 のガイドラインとベストプラクティスのフレームワークです。 AWS CAF は、ビジネス、人材、ガバナンス、プラットフォーム、セキュリティ、運用という 6 つの重点分野にガイダンスを整理しています。ビジネス、人材、ガバナンスの観点では、ビジネススキルとプロセスに重点を置き、プラットフォーム、セキュリティ、オペレーションの視点は技術的なスキルとプロセスに焦点を当てています。例えば、人材の観点では、人事 (HR)、人材派遣機能、および人材管理を扱うステークホルダーを対象としています。この観点から、 AWS CAF は人材開発、トレーニング、コミュニケーションに関するガイダンスを提供し、組織がクラウド導入を成功させるための準備を支援します。詳細については、AWS CAF ウェブサイトと AWS CAF のホワイトペーパー を参照してください。

AWS ワークロード認定フレームワーク (AWS WQF)

データベース移行ワークロードを評価し、移行戦略を推奨し、作業見積もりを提供するツール。AWS WQF は AWS Schema Conversion Tool (AWS SCT) に含まれています。データベーススキーマとコードオブジェクト、アプリケーションコード、依存関係、およびパフォーマンス特性を分析し、評価レポートを提供します。

B

不正なボット

個人や組織に混乱や損害を与えることを目的としたボット。

BCP

「ビジネス継続性計画 (BCP)」をご覧ください。

動作グラフ

リソースの動作とインタラクションを経時的に示した、一元的なインタラクティブビュー。Amazon Detective の動作グラフを使用すると、失敗したログオンの試行、不審な API 呼び出し、その他同様のアクションを調べることができます。詳細については、Detective ドキュメントの「動作グラフのデータ」を参照してください。

ビッグエンディアンシステム

最上位バイトを最初に格納するシステム。「エンディアン性」もご覧ください。

二項分類

バイナリ結果 (2 つの可能なクラスのうちの 1 つ) を予測するプロセス。例えば、お客様の機械学習モデルで「この E メールはスパムですか、それともスパムではありませんか」などの問題を予測する必要があるかもしれません。または「この製品は書籍ですか、車ですか」などの問題を予測する必要があるかもしれません。

ブルームフィルター

要素がセットのメンバーであるかどうかをテストするために使用される、確率的でメモリ効率の高いデータ構造。

ブルー/グリーンデプロイ

それぞれが独立しているが、同一の環境を 2 つ作成するデプロイ戦略。現在のアプリケーションバージョンを 1 つの環境 (ブルー) で実行し、新しいアプリケーションバージョンを別の環境 (グリーン) で実行します。この戦略は、最小限の影響で迅速にロールバックするのに役立ちます。

ボット

インターネット経由で自動タスクを実行し、人間のアクティビティややり取りをシミュレートするソフトウェアアプリケーション。インターネット上の情報のインデックスを作成するウェブクローラーなど、一部のボットは有用または有益です。悪質なボットと呼ばれる他のボットの中には、個人や組織を混乱させたり、損害を与えたりすることを意図したものもあります。

ボットネット

マルウェアに感染しており、ボットハーダーまたはボットオペレーターと呼ばれる単一の当事者によって制御されているボットのネットワーク。ボットネットは、ボットとその影響力を拡大する仕組みとして、非常によく知られています。

ブランチ

コードリポジトリに含まれる領域。リポジトリに最初に作成するブランチは、メインブランチといいます。既存のブランチから新しいブランチを作成し、その新しいブランチで機能を開発した

り、バグを修正したりできます。機能を構築するために作成するブランチは、通常、機能ブランチと呼ばれます。機能をリリースする準備ができたら、機能ブランチをメインブランチに統合します。詳細については、「ブランチの概要」(GitHub ドキュメント) を参照してください。

ブレークグラスアクセス

例外的な状況では、承認されたプロセスを通じて、ユーザーが AWS アカウント 通常アクセス許可を持たない にすばやくアクセスできるようにします。詳細については、 AWS Well-Architected ガイダンスの「ブレークグラス手順の実装」インジケータを参照してください。

ブラウンフィールド戦略

環境の既存インフラストラクチャ。システムアーキテクチャにブラウンフィールド戦略を導入する場合、現在のシステムとインフラストラクチャの制約に基づいてアーキテクチャを設計します。既存のインフラストラクチャを拡張している場合は、ブラウンフィールド戦略とグリーンフィールド戦略を融合させることもできます。

バッファキャッシュ

アクセス頻度が最も高いデータが保存されるメモリ領域。

ビジネス能力

価値を生み出すためにビジネスが行うこと (営業、カスタマーサービス、マーケティングなど)。マイクロサービスのアーキテクチャと開発の決定は、ビジネス能力によって推進できます。詳細については、AWSでのコンテナ化されたマイクロサービスの実行ホワイトペーパーの「ビジネス機能を中心に組織化」セクションを参照してください。

ビジネス継続性計画 (BCP)

大規模移行など、中断を伴うイベントが運用に与える潜在的な影響に対処し、ビジネスを迅速に再開できるようにする計画。

C

CAF

「AWS クラウド導入フレームワーク」を参照してください

カナリアデプロイ

エンドユーザーへのバージョンリリースを、時間をかけて段階的に行うこと。確信が持てたら新規バージョンをデプロイして、現在のバージョン全体を置き換えます。

CCoE

「Cloud Center of Excellence」を参照してください。

CDC

「変更データキャプチャ」を参照してください。

変更データキャプチャ (CDC)

データソース (データベーステーブルなど) の変更を追跡し、その変更に関するメタデータを記録するプロセス。CDC は、ターゲットシステムでの変更を監査またはレプリケートして同期を維持するなど、さまざまな目的に使用できます。

カオスエンジニアリング

障害や破壊的なイベントを意図的に導入して、システムの耐障害性をテストすること。AWS Fault Injection Service (AWS FIS) を使用して、 AWS ワークロードにストレスを与え、その応答を評価する実験を実行できます。

CI/CD

「継続的インテグレーションと継続的デリバリー」を参照してください。

分類

予測を生成するのに役立つ分類プロセス。分類問題の機械学習モデルは、離散値を予測します。離散値は、常に互いに区別されます。例えば、モデルがイメージ内に車があるかどうかを評価する必要がある場合があります。

クライアント側の暗号化

ターゲットがデータ AWS のサービス を受信する前のローカルでのデータの暗号化。

Cloud Center of Excellence (CCoE)

クラウドのベストプラクティスの作成、リソースの移動、移行のタイムラインの確立、大規模変革を通じて組織をリードするなど、組織全体のクラウド導入の取り組みを推進する学際的なチーム。詳細については、 AWS クラウド エンタープライズ戦略ブログの CCoE 投稿を参照してください。

クラウドコンピューティング

リモートデータストレージと IoT デバイス管理に通常使用されるクラウドテクノロジー。クラウドコンピューティングは、一般的に、エッジコンピューティングに接続されています。

クラウド運用モデル

IT 組織において、1 つ以上のクラウド環境を構築、成熟、最適化するために使用される運用モデル。詳細については、「クラウド運用モデルの構築」を参照してください。

導入のクラウドステージ

組織が、 AWS クラウドへの移行時に通常実行する 4 つの段階。

- プロジェクト — 概念実証と学習を目的として、クラウド関連のプロジェクトをいくつか実行する
- 基礎固め — お客様のクラウドの導入を拡大するための基礎的な投資 (ランディングゾーンの作成、CCoE の定義、運用モデルの確立など)
- 移行 — 個々のアプリケーションの移行
- 再発明 — 製品とサービスの最適化、クラウドでのイノベーション

これらのステージは、 AWS クラウド エンタープライズ戦略ブログのブログ記事「クラウドファーストへのジャーニー」と「導入のステージ」で Stephen Orban によって定義されました。AWS 移行戦略との関連性については、「移行準備ガイド」を参照してください。

CMDB

「構成管理データベース (CMDB)」を参照してください。

コードリポジトリ

ソースコードやその他の資産 (ドキュメント、サンプル、スクリプトなど) が保存され、バージョン管理プロセスを通じて更新される場所。一般的なクラウドリポジトリには、GitHub や Bitbucket Cloud があります。コードの各バージョンはブランチと呼ばれます。マイクロサービスの構造では、各リポジトリは 1 つの機能専用です。1 つの CI/CD パイプラインで複数のリポジトリを使用できます。

コールドキャッシュ

空である、または、かなり空きがある、もしくは、古いデータや無関係なデータが含まれているバッファキャッシュ。データベースインスタンスはメインメモリまたはディスクから読み取る必要があり、バッファキャッシュから読み取るよりも時間がかかるため、パフォーマンスに影響します。

コールドデータ

めったにアクセスされず、通常は過去のデータです。この種類のデータをクエリする場合、通常は低速なクエリでも問題ありません。このデータを低パフォーマンスで安価なストレージ階層またはクラスに移動すると、コストを削減することができます。

コンピュータビジョン (CV)

機械学習を使用してデジタルイメージやビデオといった、ビジュアル形式の情報を分析および抽出する AI の分野。例えば、Amazon SageMaker AI では、CV 用の画像処理アルゴリズムを利用できます。

設定ドリフト

ワークロードにおいて、設定が想定した状態から変化すること。これによって、ワークロードが非準拠になる可能性がありますが、この状態は、徐々に生じ、意図的なものではありません。

構成管理データベース (CMDB)

データベースとその IT 環境 (ハードウェアとソフトウェアの両方のコンポーネントとその設定を含む) に関する情報を保存、管理するリポジトリ。通常、CMDB のデータは、移行のポートフォリオの検出と分析の段階で使用します。

コンフォーマンスパック

コンプライアンスチェックとセキュリティチェックをカスタマイズするためにアセンブルできる AWS Config ルールと修復アクションのコレクション。YAML テンプレートを使用して、コンフォーマンスパックを AWS アカウント および リージョンの単一のエンティティとしてデプロイすることも、組織全体にデプロイすることもできます。詳細については、 AWS Config ドキュメントの「コンフォーマンスパック」を参照してください。

継続的インテグレーションと継続的デリバリー (CI/CD)

ソフトウェアリリースプロセスのソース、ビルド、テスト、ステージング、本番の各ステージを自動化するプロセス。CI/CD は一般的にパイプラインと呼ばれます。プロセスの自動化、生産性の向上、コード品質の向上、配信の加速化を可能にします。詳細については、「継続的デリバリーの利点」を参照してください。CD は継続的デプロイ (Continuous Deployment) の略語でもあります。詳細については「継続的デリバリーと継続的なデプロイ」を参照してください。

CV

「コンピュータビジョン」を参照してください。

D

保管中のデータ

ストレージ内にあるデータなど、常に自社のネットワーク内にあるデータ。

データ分類

ネットワーク内のデータを重要度と機密性に基づいて識別、分類するプロセス。データに適した保護および保持のコントロールを判断する際に役立つため、あらゆるサイバーセキュリティのリスク管理戦略において重要な要素です。データ分類は、 AWS Well-Architected フレームワークのセキュリティの柱のコンポーネントです。詳細については、「データ分類」を参照してください。

データドリフト

実稼働データと ML モデルのトレーニングに使用されたデータとの間に有意な差異が生じたり、入力データが時間の経過と共に有意に変化したりすることです。データドリフトは、ML モデル予測の全体的な品質、精度、公平性を低下させる可能性があります。

転送中のデータ

ネットワーク内 (ネットワークリソース間など) を活発に移動するデータ。

データメッシュ

非一元的で分散型のデータ所有権を持つとともに、一元的な管理およびガバナンスを行えるアーキテクチャフレームワーク。

データ最小化

厳密に必要なデータのみを収集し、処理するという原則。でデータ最小化を実践 AWS クラウドすることで、プライバシーリスク、コスト、分析のカーボンフットプリントを削減できます。

データ境界

AWS 環境内の一連の予防ガードレール。信頼できる ID のみが、期待されるネットワークから信頼できるリソースにアクセスできるようにします。詳細については、「 でのデータ境界の構築 AWS」を参照してください。

データの前処理

raw データをお客様の機械学習モデルで簡単に解析できる形式に変換すること。データの前処理とは、特定の列または行を削除して、欠落している、矛盾している、または重複する値に対処することを意味します。

データ出所

データの生成、送信、保存の方法など、データのライフサイクル全体を通じてデータの出所と履歴を追跡するプロセス。

データ件名

データを収集、処理している個人。

データウェアハウス

分析などのビジネスインテリジェンスをサポートするデータ管理システム。データウェアハウスには、一般的に、大量の履歴データが含まれており、多くの場合、それらはクエリや分析に使用されます。

データベース定義言語 (DDL)

データベース内のテーブルやオブジェクトの構造を作成または変更するためのステートメントまたはコマンド。

データベース操作言語 (DML)

データベース内の情報を変更 (挿入、更新、削除) するためのステートメントまたはコマンド。

DDL

「データベース定義言語」を参照してください。

ディープアンサンブル

予測のために複数の深層学習モデルを組み合わせます。ディープアンサンブルを使用して、より正確な予測を取得したり、予測の不確実性を推定したりできます。

深層学習

人工ニューラルネットワークの複数層を使用して、入力データと対象のターゲット変数の間のマッピングを識別する機械学習サブフィールド。

多層防御

一連のセキュリティメカニズムとコントロールをコンピュータネットワーク全体に層状に重ねて、ネットワークとその内部にあるデータの機密性、整合性、可用性を保護する情報セキュリティの手法。この戦略を採用するときは AWS、リソースの保護に役立つように、 AWS Organizations 構造の異なるレイヤーに複数のコントロールを追加します。たとえば、多層防御アプローチでは、多要素認証、ネットワークセグメンテーション、暗号化を組み合わせることができます。

委任管理者

では AWS Organizations、互換性のあるサービスが AWS メンバーアカウントを登録して組織のアカウントを管理し、そのサービスのアクセス許可を管理できます。このアカウントを、そのサービスの委任管理者と呼びます。詳細、および互換性のあるサービスの一覧は、 AWS

Organizations ドキュメントの「AWS Organizationsで使用できるサービス」を参照してください。

トラブルシューティング

アプリケーション、新機能、コードの修正をターゲットの環境で利用できるようにするプロセス。デプロイでは、コードベースに変更を施した後、アプリケーションの環境でそのコードベースを構築して実行します。

開発環境

「環境」を参照してください。

検出管理

イベントが発生したときに、検出、ログ記録、警告を行うように設計されたセキュリティコントロール。これらのコントロールは副次的な防衛手段であり、実行中の予防的コントロールをすり抜けたセキュリティイベントをユーザーに警告します。詳細については、「 AWSでのセキュリティコントロールの実装」の「検出的コントロール」を参照してください。

開発バリューストリームマッピング (DVSM)

ソフトウェア開発ライフサイクルのスピードと品質に悪影響を及ぼす制約を特定し、優先順位を付けるために使用されるプロセス。DVSM は、もともとリーンマニュファクチャリング・プラクティスのために設計されたバリューストリームマッピング・プロセスを拡張したものです。ソフトウェア開発プロセスを通じて価値を創造し、動かすために必要なステップとチームに焦点を当てています。

デジタルツイン

建物、工場、産業機器、生産ラインなど、現実世界のシステムを仮想的に表現したものです。デジタルツインは、予知保全、リモートモニタリング、生産最適化をサポートします。

ディメンションテーブル

スタースキーマにおいて、ファクトテーブルの定量データに関するデータ属性が含まれる小さいテーブル。ディメンションテーブルの属性は、通常、テキストフィールド、またはテキストのように扱える個別の数値で示されます。これらの属性は、一般的に、クエリの制約、フィルタリング、結果セットのラベル付けに使用されます。

ディザスタ

ワークロードまたはシステムが、導入されている主要な場所でのビジネス目標の達成を妨げるイベント。これらのイベントは、自然災害、技術的障害、または意図しない設定ミスやマルウェア攻撃などの人間の行動の結果である場合があります。

ディザスタリカバリ (DR)

ディザスタによるダウンタイムとデータ損失を最小限に抑えるための戦略とプロセス。詳細については、 AWS Well-Architected フレームワークの「 でのワークロードのディザスタリカバリ AWS: クラウドでのリカバリ」を参照してください。

DML

「データベース操作言語」を参照してください。

ドメイン駆動型設計

各コンポーネントが提供している変化を続けるドメイン、またはコアビジネス目標にコンポーネントを接続して、複雑なソフトウェアシステムを開発するアプローチ。この概念は、エリック・エヴァンスの著書、Domain-Driven Design: Tackling Complexity in the Heart of Software (ドメイン駆動設計:ソフトウェアの中心における複雑さへの取り組み) で紹介されています (ボストン: Addison-Wesley Professional、2003)。strangler fig パターンでドメイン駆動型設計を使用する方法の詳細については、「コンテナと Amazon API Gateway を使用して、従来の Microsoft ASP.NET (ASMX) ウェブサービスを段階的にモダナイズ」を参照してください。

DR

「ディザスタリカバリ」を参照してください。

ドリフト検出

ベースライン設定からの偏差を追跡します。たとえば、 AWS CloudFormation を使用してシステムリソースのドリフトを検出したり、 を使用して AWS Control Tower 、ガバナンス要件のコンプライアンスに影響を与える可能性のあるランディングゾーンの変更を検出したりできます。

DVSM

「開発バリューストリームマッピング」を参照してください。

E

EDA

「探索的データ分析」を参照してください。

EDI

「電子データ交換」を参照してください。

エッジコンピューティング

IoT ネットワークのエッジにあるスマートデバイスの計算能力を高めるテクノロジー。クラウドコンピューティングと比較すると、エッジコンピューティングは通信レイテンシーを短縮し、応答時間を改善できます。

電子データ交換 (EDI)

組織間で行う、ビジネスドキュメントの自動交換。詳細については、「電子データ交換とは」を参照してください。

暗号化

人間が読み取り可能なプレーンテキストデータを暗号文に変換するコンピューティング処理。

暗号化キー

暗号化アルゴリズムが生成した、ランダム化されたビットからなる暗号文字列。キーの長さは決まっておらず、各キーは予測できないように、一意になるように設計されています。

エンディアン

コンピュータメモリにバイトが格納される順序。ビッグエンディアンシステムでは、最上位バイトが最初に格納されます。リトルエンディアンシステムでは、最下位バイトが最初に格納されます。

エンドポイント

「サービスエンドポイント」を参照してください。

エンドポイントサービス

仮想プライベートクラウド (VPC) 内でホストして、他のユーザーと共有できるサービス。を使用してエンドポイントサービスを作成し AWS PrivateLink 、他の AWS アカウント または AWS Identity and Access Management (IAM) プリンシパルにアクセス許可を付与できます。これらのアカウントまたはプリンシパルは、インターフェイス VPC エンドポイントを作成することで、エンドポイントサービスにプライベートに接続できます。詳細については、Amazon Virtual Private Cloud (Amazon VPC) ドキュメントの「エンドポイントサービスを作成する」を参照してください。

エンタープライズリソースプランニング (ERP)

エンタープライズの主要なビジネスプロセス (会計、MES、プロジェクト管理など) を自動化および管理するシステム。

エンベロープ暗号化

暗号化キーを、別の暗号化キーを使用して暗号化するプロセス。詳細については、 AWS Key Management Service (AWS KMS) ドキュメントの「エンベロープ暗号化」を参照してください。

環境

実行中のアプリケーションのインスタンス。クラウドコンピューティングにおける一般的な環境の種類は以下のとおりです。

- 開発環境 — アプリケーションのメンテナンスを担当するコアチームのみが使用できる、実行中のアプリケーションのインスタンス。開発環境は、上位の環境に昇格させる変更をテストするときに使用します。このタイプの環境は、テスト環境と呼ばれることもあります。
- 下位環境 — 初期ビルドやテストに使用される環境など、アプリケーションのすべての開発環境。
- 本番環境 — エンドユーザーがアクセスできる、実行中のアプリケーションのインスタンス。CI/CD パイプラインでは、本番環境が最後のデプロイ環境になります。
- 上位環境 — コア開発チーム以外のユーザーがアクセスできるすべての環境。これには、本番環境、本番前環境、ユーザー承認テスト環境などが含まれます。

エピック

アジャイル方法論で、お客様の作業の整理と優先順位付けに役立つ機能カテゴリ。エピックでは、要件と実装タスクの概要についてハイレベルな説明を提供します。例えば、 AWS CAF セキュリティエピックには、ID とアクセスの管理、検出コントロール、インフラストラクチャセキュリティ、データ保護、インシデント対応が含まれます。 AWS 移行戦略のエピックの詳細については、プログラム実装ガイドを参照してください。

ERP

「エンタープライズリソース計画」を参照してください。

探索的データ分析 (EDA)

データセットを分析してその主な特性を理解するプロセス。お客様は、データを収集または集計してから、パターンの検出、異常の検出、および前提条件のチェックのための初期調査を実行します。EDA は、統計の概要を計算し、データの可視化を作成することによって実行されます。

F

ファクトテーブル

スタースキーマの中央にあるテーブル。ビジネスオペレーションに関する定量的データが保存されます。一般的に、ファクトテーブルは、2 種類の列で構成されます。1 つは測定値が含まれる列、もう 1 つはディメンションテーブルへの外部キーが含まれる列です。

フェイルファスト

開発ライフサイクルを短縮するために、頻繁かつ段階的にテストを行う哲学であり、アジャイルアプローチでは、この考え方がきわめて重要です。

障害分離境界

では AWS クラウド、障害の影響を制限し、ワークロードの耐障害性を高めるのに役立つアベイラビリティーゾーン AWS リージョン、コントロールプレーン、データプレーンなどの境界。詳細については、「AWS 障害分離境界」を参照してください。

機能ブランチ

「ブランチ」を参照してください。

特徴量

お客様が予測に使用する入力データ。例えば、製造コンテキストでは、特徴量は製造ラインから定期的にキャプチャされるイメージの可能性もあります。

特徴量重要度

モデルの予測に対する特徴量の重要性。これは通常、Shapley Additive Deskonations (SHAP) や積分勾配など、さまざまな手法で計算できる数値スコアで表されます。詳細については、「 を使用した機械学習モデルの解釈可能性 AWS」を参照してください。

機能変換

追加のソースによるデータのエンリッチ化、値のスケーリング、単一のデータフィールドからの複数の情報セットの抽出など、機械学習プロセスのデータを最適化すること。これにより、機械学習モデルはデータの恩恵を受けることができます。例えば、「2021-05-27 00:15:37」の日付を「2021 年」、「5 月」、「木」、「15」に分解すると、学習アルゴリズムがさまざまなデータコンポーネントに関連する微妙に異なるパターンを学習するのに役立ちます。

数ショットプロンプト

LLM に、タスクと望ましい出力を示す例を少数提示した後に、類似のタスクを実行させること。この手法は、プロンプトに記述された例 (ショット) からモデルが学習する「インコンテキスト学

習」の一種です。数ショットプロンプトは、特定のフォーマット、推論、専門知識が必要なタスクに効果的です。「ゼロショットプロンプト」も参照してください。

FGAC

「きめ細かなアクセス制御」を参照してください。

きめ細かなアクセス制御 (FGAC)

複数の条件を使用してアクセス要求を許可または拒否すること。

フラッシュカット移行

変更データのキャプチャによる継続的なデータ複製を利用して、段階的なアプローチではなく、可能な限り短時間でデータを移行するデータベース移行方法。目的はダウンタイムを最小限に抑えることです。

FM

「基盤モデル」を参照してください。

基盤モデル (FM)

大規模な深層学習ニューラルネットワークであり、一般化およびラベル付けされていないデータからなる大規模データセットでトレーニングされています。FM により、言語理解、テキストおよび画像生成、自然言語での会話といった、一般的な各種タスクを実行できます。詳細については、「基盤モデルとは何ですか?」を参照してください。

G

生成 AI

AI モデルのサブセット。大量のデータでトレーニングされており、シンプルなテキストプロンプトを使用して、画像、動画、テキスト、オーディオなどの新しいコンテンツやアーティファクトを作成できます。詳細については、「生成 AI とは何ですか?」を参照してください。

ジオブロッキング

「地理的制限」を参照してください。

地理的制限 (ジオブロッキング)

特定の国のユーザーがコンテンツ配信にアクセスできないようにするための、Amazon CloudFront のオプション。アクセスを許可する国と禁止する国は、許可リストまたは禁止リスト

を使って指定します。詳細については、CloudFront ドキュメントの「コンテンツの地理的ディストリビューションの制限」を参照してください。

Gitflow ワークフロー

下位環境と上位環境が、ソースコードリポジトリでそれぞれ異なるブランチを使用する方法。Gitflow ワークフローは古いと見なされている方法であり、トランクベースのワークフローは推奨されている新しい方法です。

ゴールデンイメージ

システムまたはソフトウェアのスナップショットであり、システムまたはソフトウェアの新規インスタンスをデプロイするテンプレートとして使用されます。製造の例で言えば、ゴールデンイメージを使用すると、複数のデバイスにソフトウェアをプロビジョニングして、デバイス製造オペレーションの速度、スケーラビリティ、生産性を向上させることができます。

グリーンフィールド戦略

新しい環境に既存のインフラストラクチャが存在しないこと。システムアーキテクチャにグリーンフィールド戦略を導入する場合、既存のインフラストラクチャ (別名ブラウンフィールド) との互換性の制約を受けることなく、あらゆる新しいテクノロジーを選択できます。既存のインフラストラクチャを拡張している場合は、ブラウンフィールド戦略とグリーンフィールド戦略を融合させることもできます。

ガードレール

組織単位 (OU) 全般のリソース、ポリシー、コンプライアンスを管理するのに役立つ概略的なルール。予防ガードレールは、コンプライアンス基準に一致するようにポリシーを実施します。これらは、サービスコントロールポリシーと IAM アクセス許可の境界を使用して実装されます。検出ガードレールは、ポリシー違反やコンプライアンス上の問題を検出し、修復のためのアラートを発信します。これらは AWS Config、、 AWS Security Hub CSPM、Amazon GuardDuty、、 AWS Trusted Advisor Amazon Inspector、およびカスタム AWS Lambda チェックを使用して実装されます。

H

HA

「高可用性」を参照してください。

異種混在データベースの移行

別のデータベースエンジンを使用するターゲットデータベースへお客様の出典データベースの移行 (例えば、Oracle から Amazon Aurora)。異種間移行は通常、アーキテクチャの再設計作業の一部であり、スキーマの変換は複雑なタスクになる可能性があります。AWS は、スキーマの変換に役立つ AWS SCTを提供します。

高可用性 (HA)

課題や災害が発生した場合に、介入なしにワークロードを継続的に運用できること。HA システムは、自動的にフェイルオーバーし、一貫して高品質のパフォーマンスを提供し、パフォーマンスへの影響を最小限に抑えながらさまざまな負荷や障害を処理するように設計されています。

ヒストリアンのモダナイゼーション

製造業のニーズによりよく応えるために、オペレーションテクノロジー (OT) システムをモダナイズし、アップグレードするためのアプローチ。ヒストリアンは、工場内のさまざまなソースからデータを収集して保存するために使用されるデータベースの一種です。

ホールドアウトデータ

機械学習モデルのトレーニング用データセットから保留される、ラベル付き履歴データの一部。ホールドアウトデータを使用すると、モデル予測をホールドアウトデータと比較して、モデルのパフォーマンスを評価できます。

同種データベースの移行

お客様の出典データベースを、同じデータベースエンジンを共有するターゲットデータベース (Microsoft SQL Server から Amazon RDS for SQL Server など) に移行する。同種間移行は、通常、リホストまたはリプラットフォーム化の作業の一部です。ネイティブデータベースユーティリティを使用して、スキーマを移行できます。

ホットデータ

リアルタイムデータや最近の翻訳データなど、頻繁にアクセスされるデータ。通常、このデータには高速なクエリ応答を提供する高性能なストレージ階層またはクラスが必要です。

ホットフィックス

本番環境の重大な問題を修正するために緊急で配布されるプログラム。緊急性が高いため、通常の DevOps のリリースワークフローからは外れた形で実施されます。

ハイパーケア期間

カットオーバー直後、移行したアプリケーションを移行チームがクラウドで管理、監視して問題に対処する期間。通常、この期間は 1～4 日です。ハイパーケア期間が終了すると、アプリケーションに対する責任は一般的に移行チームからクラウドオペレーションチームに移ります。

I

IaC

「Infrastructure as Code」を参照してください。

ID ベースのポリシー

AWS クラウド 環境内のアクセス許可を定義する 1 つ以上の IAM プリンシパルにアタッチされたポリシー。

アイドル状態のアプリケーション

90 日間の平均的な CPU およびメモリ使用率が 5～20% のアプリケーション。移行プロジェクトでは、これらのアプリケーションを廃止するか、オンプレミスに保持するのが一般的です。

IIoT

「インダストリアル IoT」を参照してください。

イミュータブルインフラストラクチャ

既存インフラストラクチャの更新、パッチ適用、変更などを行わずに、本番環境ワークロードに使用する新規インフラストラクチャをデプロイするモデル。本質的に、イミュータブルインフラストラクチャは、ミュータブルインフラストラクチャよりも一貫性、信頼性、予測性に優れています。詳細については、 AWS Well-Architected フレームワークにある「イミュータブルインフラストラクチャを使用してデプロイする」のベストプラクティスを参照してください。

インバウンド (受信) VPC

AWS マルチアカウントアーキテクチャでは、アプリケーションの外部からネットワーク接続を受け入れ、検査し、ルーティングする VPC。AWS Security Reference Architecture では、アプリケーションとより広範なインターネット間の双方向のインターフェイスを保護するために、インバウンド、アウトバウンド、インスペクションの各 VPC を使用してネットワークアカウントを設定することを推奨しています。

増分移行

アプリケーションを 1 回ですべてカットオーバーするのではなく、小さい要素に分けて移行するカットオーバー戦略。例えば、最初は少数のマイクロサービスまたはユーザーのみを新しいシステムに移行する場合があります。すべてが正常に機能することを確認できたら、残りのマイクロサービスやユーザーを段階的に移行し、レガシーシステムを廃止できるようにします。この戦略により、大規模な移行に伴うリスクが軽減されます。

インダストリー 4.0

2016 年に Klaus Schwab 氏が提唱した用語で、接続、リアルタイムデータ、オートメーション、分析、AI/ML の進歩による、ビジネスプロセスのモダナイズを意味します。

インフラストラクチャ

アプリケーションの環境に含まれるすべてのリソースとアセット。

Infrastructure as Code (IaC)

アプリケーションのインフラストラクチャを一連の設定ファイルを使用してプロビジョニングし、管理するプロセス。IaC は、新しい環境を再現可能で信頼性が高く、一貫性のあるものにするため、インフラストラクチャを一元的に管理し、リソースを標準化し、スケールを迅速に行えるように設計されています。

インダストリアル IoT (IIoT)

製造、エネルギー、自動車、ヘルスケア、ライフサイエンス、農業などの産業部門におけるインターネットに接続されたセンサーやデバイスの使用。詳細については、「インダストリアル IoT (IIoT) デジタルトランスフォーメーション戦略の構築」を参照してください。

インスペクション VPC

AWS マルチアカウントアーキテクチャでは、VPC (同一または異なる 内 AWS リージョン)、インターネット、オンプレミスネットワーク間のネットワークトラフィックの検査を管理する一元化された VPCs。AWS Security Reference Architecture では、アプリケーションとより広範なインターネット間の双方向のインターフェイスを保護するために、インバウンド、アウトバウンド、インスペクションの各 VPC を使用してネットワークアカウントを設定することを推奨しています。

IoT

インターネットまたはローカル通信ネットワークを介して他のデバイスやシステムと通信する、センサーまたはプロセッサが組み込まれた接続済み物理オブジェクトのネットワーク。詳細については、「IoT とは」を参照してください。

解釈可能性

機械学習モデルの特性で、モデルの予測がその入力にどのように依存するかを人間が理解できる度合いを表します。詳細については、「 を使用した機械学習モデルの解釈可能性 AWS」を参照してください。

IoT

「IoT」を参照してください。

IT 情報ライブラリ (ITIL)

IT サービスを提供し、これらのサービスをビジネス要件に合わせるための一連のベストプラクティス。ITIL は ITSM の基盤を提供します。

IT サービス管理 (ITSM)

組織の IT サービスの設計、実装、管理、およびサポートに関連する活動。クラウドオペレーションと ITSM ツールの統合については、オペレーション統合ガイドを参照してください。

ITIL

「IT 情報ライブラリ」を参照してください。

ITSM

「IT サービス管理」を参照してください。

L

ラベルベースアクセス制御 (LBAC)

強制アクセス制御 (MAC) の実装で、ユーザーとデータ自体にそれぞれセキュリティラベル値が明示的に割り当てられます。ユーザーセキュリティラベルとデータセキュリティラベルが交差する部分によって、ユーザーに表示される行と列が決まります。

ランディングゾーン

ランディングゾーンは、スケーラブルで安全な、適切に設計されたマルチアカウント AWS 環境です。これは、組織がセキュリティおよびインフラストラクチャ環境に自信を持ってワークロードとアプリケーションを迅速に起動してデプロイできる出発点です。ランディングゾーンの詳細については、「安全でスケーラブルなマルチアカウント AWS 環境のセットアップ」を参照してください。

大規模言語モデル (LLM)

大量のデータで事前トレーニングされた深層学習 AI モデル。LLM では、質問への回答、ドキュメントの要約、他言語へのテキスト翻訳、文を完成させるなど、さまざまなタスクを実行できます。詳細については、「大規模言語モデル (LLM) とは何ですか?」を参照してください。

大規模な移行

300 台以上のサーバの移行。

LBAC

「ラベルベースアクセス制御」を参照してください。

最小特権

タスクの実行には必要最低限の権限を付与するという、セキュリティのベストプラクティス。詳細については、IAM ドキュメントの「最小特権アクセス許可を適用する」を参照してください。

リフトアンドシフト

「7 Rs」を参照してください。

リトルエンディアンシステム

最下位バイトを最初に格納するシステム。「エンディアン性」もご覧ください。

LLM

「大規模言語モデル」を参照してください。

下位環境

「環境」を参照してください。

M

機械学習 (ML)

パターン認識と学習にアルゴリズムと手法を使用する人工知能の一種。ML は、モノのインターネット (IoT) データなどの記録されたデータを分析して学習し、パターンに基づく統計モデルを生成します。詳細については、「機械学習」を参照してください。

メインブランチ

「ブランチ」を参照してください。

マルウェア

コンピュータのセキュリティやプライバシーを侵害するように設計されたソフトウェア。マルウェアは、コンピュータシステムの中断、機密情報の漏洩、不正アクセスを招く可能性があります。マルウェアの例には、ウイルス、ワーム、ランサムウェア、トロイの木馬、スパイウェア、キーロガーなどがあります。

マネージドサービス

AWS のサービス はインフラストラクチャレイヤー、オペレーティングシステム、プラットフォーム AWS を運用し、エンドポイントにアクセスしてデータを保存および取得します。マネージドサービスの例として、Amazon Simple Storage Service (Amazon S3) と Amazon DynamoDB が挙げられます。このサービスは、抽象化されたサービスとも呼ばれます。

製造実行システム (MES)

生産プロセスを追跡、モニタリング、文書化、制御するソフトウェアシステムであり、工場では、これによって、原材料から製品を完成させます。

MAP

「Migration Acceleration Program」を参照してください。

メカニズム

ツールを作成してその導入を推進し、導入結果を調べて調整を行うための包括的なプロセス。メカニズムとは、運用中にそれ自体を強化し改善するサイクルを意味します。詳細については、AWS 「 Well-Architected フレームワーク」の「メカニズムの構築」を参照してください。

メンバーアカウント

組織の一部である管理アカウント AWS アカウント 以外のすべて AWS Organizations。 アカウントが 組織のメンバーになることができるのは、一度に 1 つのみです。

MES

「製造実行システム」を参照してください。

Message Queuing Telemetry Transport (MQTT)

発行/サブスクライブのパターンに基づく、軽量のマシンツーマシン (M2M) 通信プロトコルであり、リソースに限りのある IoT デバイスに使用されます。

マイクロサービス

明確に定義された API を介して通信し、通常は小規模な自己完結型のチームが所有する、小規模で独立したサービスです。例えば、保険システムには、販売やマーケティングなどのビジネス

機能、または購買、請求、分析などのサブドメインにマッピングするマイクロサービスが含まれる場合があります。マイクロサービスの利点には、俊敏性、柔軟なスケーリング、容易なデプロイ、再利用可能なコード、回復力などがあります。詳細については、AWS「サーバーレスサービスを使用したマイクロサービスの統合」を参照してください。

マイクロサービスアーキテクチャ

各アプリケーションプロセスをマイクロサービスとして実行する独立したコンポーネントを使用してアプリケーションを構築するアプローチ。これらのマイクロサービスは、軽量 API を使用して、明確に定義されたインターフェイスを介して通信します。このアーキテクチャの各マイクロサービスは、アプリケーションの特定の機能に対する需要を満たすように更新、デプロイ、およびスケーリングできます。詳細については、「 でのマイクロサービスの実装 AWS」を参照してください。

Migration Acceleration Program (MAP)

組織がクラウドに移行するための強力な運用基盤を構築し、移行の初期コストを相殺するのに役立つコンサルティングサポート、トレーニング、サービスを提供する AWS プログラム。MAP には、組織的な方法でレガシー移行を実行するための移行方法論と、一般的な移行シナリオを自動化および高速化する一連のツールが含まれています。

大規模な移行

アプリケーションポートフォリオの大部分を次々にクラウドに移行し、各ウェーブでより多くのアプリケーションを高速に移動させるプロセス。この段階では、以前の段階から学んだベストプラクティスと教訓を使用して、移行ファクトリー チーム、ツール、プロセスのうち、オートメーションとアジャイルデリバリーによってワークロードの移行を合理化します。これは、AWS 移行戦略 の第 3 段階です。

移行ファクトリー

自動化された俊敏性のあるアプローチにより、ワークロードの移行を合理化する部門横断的なチーム。移行ファクトリーチームには、通常、運用、ビジネスアナリストおよび所有者、移行エンジニア、デベロッパー、およびスプリントで作業する DevOps プロフェッショナルが含まれます。エンタープライズアプリケーションポートフォリオの 20～50% は、ファクトリーのアプローチによって最適化できる反復パターンで構成されています。詳細については、このコンテンツセットの移行ファクトリーに関する解説とCloud Migration Factory ガイドを参照してください。

移行メタデータ

移行を完了するために必要なアプリケーションおよびサーバーに関する情報。移行パターンごとに、異なる一連の移行メタデータが必要です。移行メタデータの例としては、ターゲットサブネット、セキュリティグループ、 AWS アカウントなどがあります。

移行パターン

移行戦略、移行先、および使用する移行アプリケーションまたはサービスを詳述する、反復可能な移行タスク。例: AWS Application Migration Service を使用して Amazon EC2 への移行をリホストします。

Migration Portfolio Assessment (MPA)

オンラインツール。これによって、 AWS クラウドに移行するビジネスケースの検証に必要な情報を得られます。MPA は、詳細なポートフォリオ評価 (サーバーの適切なサイジング、価格設定、TCO 比較、移行コスト分析) および移行プラン (アプリケーションデータの分析とデータ収集、アプリケーションのグループ化、移行の優先順位付け、およびウェーブプランニング) を提供します。MPA ツール (ログインが必要) は、すべての AWS コンサルタントと APN パートナーコンサルタントが無料で利用できます。

移行準備状況評価 (MRA)

AWS CAF を使用して、組織のクラウド準備状況に関するインサイトを取得し、長所と短所を特定し、特定されたギャップを埋めるためのアクションプランを構築するプロセス。詳細については、移行準備状況ガイドを参照してください。MRA は、AWS 移行戦略の第一段階です。

移行戦略

ワークロードを AWS クラウドに移行するために使用するアプローチ。詳細については、この用語集の 7 Rs エントリと、「組織を動員して大規模な移行を加速する」を参照してください。

ML

「機械学習」を参照してください。

モダナイゼーション

古い (レガシーまたはモノリシック) アプリケーションとそのインフラストラクチャをクラウド内の俊敏で弾力性のある高可用性システムに変換して、コストを削減し、効率を高め、イノベーションを活用します。詳細については、「AWS クラウドでのアプリケーションのモダナイズ戦略」を参照してください。

モダナイゼーション準備状況評価

組織のアプリケーションのモダナイゼーションの準備状況を判断し、利点、リスク、依存関係を特定し、組織がこれらのアプリケーションの将来の状態をどの程度適切にサポートできるかを決定するのに役立つ評価。評価の結果として、ターゲットアーキテクチャのブループリント、モダナイゼーションプロセスの開発段階とマイルストーンを詳述したロードマップ、特定されたギャップに対処するためのアクションプランが得られます。詳細については、「AWS クラウドでのアプリケーションのモダナイゼーションの準備状況を評価する」を参照してください。

モノリシックアプリケーション (モノリス)

緊密に結合されたプロセスを持つ単一のサービスとして実行されるアプリケーション。モノリシックアプリケーションにはいくつかの欠点があります。1 つのアプリケーション機能エクスペリエンスの需要が急増する場合は、アーキテクチャ全体をスケーリングする必要があります。モノリシックアプリケーションの特徴を追加または改善することは、コードベースが大きくなると複雑になります。これらの問題に対処するには、マイクロサービスアーキテクチャを使用できます。詳細については、「モノリスをマイクロサービスに分解する」を参照してください。

MPA

「Migration Portfolio Assessment」を参照してください。

MQTT

「Message Queuing Telemetry Transport」を参照してください。

多クラス分類

複数のクラスの予測を生成するプロセス (2 つ以上の結果の 1 つを予測します)。例えば、機械学習モデルが、「この製品は書籍、自動車、電話のいずれですか?」 または、「このお客様にとって最も関心のある商品のカテゴリはどれですか?」と聞くかもしれません。

ミュータブルなインフラストラクチャ

本番ワークロードに使用する既存のインフラストラクチャを更新および変更するためのモデル。Well-Architected AWS フレームワークでは、一貫性、信頼性、予測可能性を向上させるために、イミュータブルインフラストラクチャの使用をベストプラクティスとして推奨しています。

O

OAC

「オリジンアクセス制御」を参照してください。

OAI

「オリジンアクセスアイデンティティ」を参照してください。

OCM

「組織変更管理」を参照してください。

オフライン移行

移行プロセス中にソースワークロードを停止させる移行方法。この方法はダウンタイムが長くなるため、通常は重要ではない小規模なワークロードに使用されます。

OI

「オペレーション統合」を参照してください。

Ola

「オペレーショナルレベルアグリーメント」を参照してください。

オンライン移行

ソースワークロードをオフラインにせずにターゲットシステムにコピーする移行方法。ワークロードに接続されているアプリケーションは、移行中も動作し続けることができます。この方法はダウンタイムがゼロから最小限で済むため、通常は重要な本番稼働環境のワークロードに使用されます。

OPC-UA

「Open Process Communications - Unified Architecture」を参照してください。

Open Process Communications - Unified Architecture (OPC-UA)

産業オートメーション用のマシンツーマシン (M2M) 通信プロトコル。OPC-UA により、相互運用の際に、データ暗号化、認証、認可の各スキームを標準化できます。

オペレーショナルレベルアグリーメント (OLA)

サービスレベルアグリーメント (SLA) をサポートするために、どの機能的 IT グループが互いに提供することを約束するかを明確にする契約。

運用準備状況レビュー (ORR)

質問と関連するベストプラクティスのチェックリスト。インシデントや起こり得る障害を理解、評価、防止したり、その範囲を縮小したりする際に役立ちます。詳細については、 AWS Well-Architected フレームワークの「Operational Readiness Reviews (ORR)」を参照してください。

運用テクノロジー (OT)

産業オペレーション、機器、インフラストラクチャを制御するために物理環境と連携させるハードウェアおよびソフトウェアシステム。製造分野では、Industry 4.0 への変革を進める上で、OT と情報技術 (IT) システムの統合に焦点が当てられています。

オペレーション統合 (OI)

クラウドでオペレーションをモダナイズするプロセスには、準備計画、オートメーション、統合が含まれます。詳細については、オペレーション統合ガイドを参照してください。

組織の証跡

組織 AWS アカウント 内のすべての のすべてのイベント AWS CloudTrail を ログに記録する によって作成された証跡 AWS Organizations。証跡は、組織に含まれている各 AWS アカウント に作成され、各アカウントのアクティビティを追跡します。詳細については、CloudTrail ドキュメントの「組織の証跡の作成」を参照してください。

組織変更管理 (OCM)

人材、文化、リーダーシップの観点から、主要な破壊的なビジネス変革を管理するためのフレームワーク。OCM は、変化の導入を加速し、移行問題に対処し、文化や組織の変化を推進することで、組織が新しいシステムと戦略の準備と移行するのを支援します。 AWS 移行戦略では、クラウド導入プロジェクトに必要な変化のスピードにより、このフレームワークは人材アクセラレーションと呼ばれます。詳細については、OCM ガイドを参照してください。

オリジンアクセス制御 (OAC)

Amazon Simple Storage Service (Amazon S3) コンテンツを保護するための、CloudFront のアクセス制限の強化オプション。OAC は AWS リージョン、すべての S3 バケット、 AWS KMS (SSE-KMS) によるサーバー側の暗号化、S3 バケットへの動的 PUT および DELETEリクエストをサポートします。

オリジンアクセスアイデンティティ (OAI)

CloudFront の、Amazon S3 コンテンツを保護するためのアクセス制限オプション。OAI を使用すると、CloudFront が、Amazon S3 に認証可能なプリンシパルを作成します。認証されたプリンシパルは、S3 バケット内のコンテンツに、特定の CloudFront ディストリビューションを介してのみアクセスできます。OAC も併せて参照してください。OAC では、より詳細な、強化されたアクセス制御が可能です。

ORR

「運用準備状況レビュー」を参照してください。

OT

「運用テクノロジー」を参照してください。

アウトバウンド (送信) VPC

AWS マルチアカウントアーキテクチャでは、アプリケーション内から開始されたネットワーク接続を処理する VPC。AWS Security Reference Architecture では、アプリケーションとより広範なインターネット間の双方向のインターフェイスを保護するために、インバウンド、アウトバウンド、インスペクションの各 VPC を使用してネットワークアカウントを設定することを推奨しています。

P

アクセス許可の境界

ユーザーまたはロールが使用できるアクセス許可の上限を設定する、IAM プリンシパルにアタッチされる IAM 管理ポリシー。詳細については、IAM ドキュメントのアクセス許可の境界を参照してください。

個人を特定できる情報 (PII)

直接閲覧した場合、または他の関連データと組み合わせた場合に、個人の身元を合理的に推測するために使用できる情報。PII の例には、氏名、住所、連絡先情報などがあります。

PII

「個人を特定できる情報」を参照してください。

プレイブック

クラウドでのコアオペレーション機能の提供など、移行に関連する作業を取り込む、事前定義された一連のステップ。プレイブックは、スクリプト、自動ランブック、またはお客様のモダナイズされた環境を運用するために必要なプロセスや手順の要約などの形式をとることができます。

PLC

「プログラマブルロジックコントローラー」を参照してください。

PLM

「製品ライフサイクル管理」を参照してください。

ポリシー

次の操作を可能にするオブジェクト: アクセス許可を定義する (ID ベースのポリシーを参照)。アクセス条件を指定する (リソースベースのポリシーを参照)。 AWS Organizations の組織における全アカウントにアクセス許可の上限を定義する (サービスコントロールポリシーを参照)。

多言語の永続性

データアクセスパターンやその他の要件に基づいて、マイクロサービスのデータストレージテクノロジーを個別に選択します。マイクロサービスが同じデータストレージテクノロジーを使用している場合、実装上の問題が発生したり、パフォーマンスが低下する可能性があります。マイクロサービスは、要件に最も適合したデータストアを使用すると、より簡単に実装でき、パフォーマンスとスケーラビリティが向上します。

ポートフォリオ評価

移行を計画するために、アプリケーションポートフォリオの検出、分析、優先順位付けを行うプロセス。詳細については、「移行の準備状況の評価」を参照してください。

述語

`true` または `false` を返すためのクエリ条件。一般的に、`WHERE` 句に記述されます。

述語プッシュダウン

データベースクエリを最適化する手法。これによって、転送前にクエリ内のデータをフィルタリングします。この手法を取ると、リレーショナルデータベースから取得し処理する必要のあるデータの量が減少するため、クエリのパフォーマンスが向上します。

予防的コントロール

イベントの発生を防ぐように設計されたセキュリティコントロール。このコントロールは、ネットワークへの不正アクセスや好ましくない変更を防ぐ最前線の防御です。詳細については、「AWSでのセキュリティコントロールの実装」の「予防的コントロール」を参照してください。

プリンシパル

アクションを実行し AWS 、リソースにアクセスできる のエンティティ。このエンティティは通常、、IAM AWS アカウントロール、または ユーザーのルートユーザーです。詳細については、IAM ドキュメントの「ロールに関する用語と概念」にあるプリンシパルを参照してください。

プライバシーバイデザイン

開発プロセス全体を通してプライバシーが考慮されているシステムエンジニアリングのアプローチ。

プライベートホストゾーン

1 つ以上の VPC 内のドメインとそのサブドメインへの DNS クエリに対し、Amazon Route 53 がどのように応答するかに関する情報を保持するコンテナ。詳細については、Route 53 ドキュメントの「プライベートホストゾーンの使用」を参照してください。

プロアクティブコントロール

非準拠リソースのデプロイ防止を目的としたセキュリティコントロール。このコントロールにより、プロビジョニング前にリソースをスキャンします。コントロールに準拠していないリソースは、プロビジョニングされません。詳細については、 AWS Control Tower ドキュメントの「コントロールリファレンスガイド」および「 セキュリティコントロールの実装」の「プロアクティブコントロール」を参照してください。 AWS

製品ライフサイクル管理 (PLM)

製品の設計、開発、発売から、成長、成熟、衰退、廃棄に至る、製品のライフサイクル全体を通してデータとプロセスを管理すること。

本番環境

「環境」を参照してください。

プログラマブルロジックコントローラー (PLC)

製造分野で使用される、信頼性と適応性に優れたコンピュータであり、これによって、マシンをモニタリングするとともに、製造プロセスを自動化します。

プロンプトチェイニング

1 つの LLM プロンプトによる出力を次のプロンプトの入力に使用して、より良いレスポンスを生成します。この手法を使用すると、複雑なタスクをサブタスクに分割したり、事前レスポンスを繰り返し改良または拡張したりできます。これによって、モデルのレスポンスの精度と関連性が向上し、粒度の高いパーソナライズされた結果を得られます。

仮名化

データセット内の個人識別子をプレースホルダー値に置き換えるプロセス。仮名化は個人のプライバシー保護に役立ちます。仮名化されたデータは、依然として個人データとみなされます。

発行/サブスクライブ (pub/sub)

マイクロサービス間の非同期通信を可能にするパターン。これにより、スケーラビリティと応答性を向上させます。例えば、マイクロサービスベースの MES の場合、マイクロサービスは、他のマイクロサービスがサブスクライブ可能なチャネルにイベントメッセージを発行できます。このシステムでは、発行サービスの変更なしに、新規マイクロサービスを追加できます。

Q

クエリプラン

手順などの一連のステップであり、SQL リレーショナルデータベースシステムのデータにアクセスするために使用されます。

クエリプランのリグレッション

データベースサービスのオプティマイザーが、データベース環境に特定の変更が加えられる前に選択されたプランよりも最適性の低いプランを選択すること。これは、統計、制限事項、環境設定、クエリパラメータのバインディングの変更、およびデータベースエンジンの更新などが原因である可能性があります。

R

RACI マトリックス

「実行責任者、説明責任者、協業先、報告先 (RACI)」を参照してください。

RAG

「検索拡張生成」を参照してください。

ランサムウェア

決済が完了するまでコンピュータシステムまたはデータへのアクセスをブロックするように設計された、悪意のあるソフトウェア。

RASCI マトリックス

「実行責任者、説明責任者、協業先、報告先 (RACI)」を参照してください。

RCAC

「行と列のアクセス制御」を参照してください。

リードレプリカ

読み取り専用に使用されるデータベースのコピー。クエリをリードレプリカにルーティングして、プライマリデータベースへの負荷を軽減できます。

リアーキテクト

「7 Rs」を参照してください。

目標復旧時点 (RPO)

最後のデータリカバリポイントからの最大許容時間です。これにより、最後の回復時点からサービスが中断されるまでの間に許容できるデータ損失の程度が決まります。

目標復旧時間 (RTO)

サービスの中断から復旧までの最大許容遅延時間。

リファクタリング

「7 Rs」を参照してください。

リージョン

地理的エリア内の AWS リソースのコレクション。各 AWS リージョン は、耐障害性、安定性、耐障害性を提供するために、他の から分離され、独立しています。詳細については、「アカウントが使用できる AWS リージョン を指定する」を参照してください。

リグレッション

数値を予測する機械学習手法。例えば、「この家はどれくらいの値段で売れるでしょうか?」という問題を解決するために、機械学習モデルは、線形回帰モデルを使用して、この家に関する既知の事実 (平方フィートなど) に基づいて家の販売価格を予測できます。

リホスト

「7 Rs」を参照してください。

リリース

デプロイプロセスで、変更を本番環境に昇格させること。

再配置

「7 Rs」を参照してください。

リプラットフォーム

「7 Rs」を参照してください。

再購入

「7 Rs」を参照してください。

回復性

中断に抵抗または中断から回復するアプリケーションの機能。 AWS クラウドでの回復力を計画する際には、一般的に、高可用性とディザスタリカバリが考慮されます。詳細については、「AWS クラウド の耐障害性」を参照してください。

リソースベースのポリシー

Amazon S3 バケット、エンドポイント、暗号化キーなどのリソースにアタッチされたポリシー。このタイプのポリシーは、アクセスが許可されているプリンシパル、サポートされているアクション、その他の満たすべき条件を指定します。

実行責任者、説明責任者、協業先、報告先 (RACI) に基づくマトリックス

移行活動とクラウド運用に関わるすべての関係者の役割と責任を定義したマトリックス。マトリックスの名前は、マトリックスで定義されている責任の種類、すなわち責任 (R) 、説明責任 (A) 、協議 (C) 、情報提供 (I) に由来します。サポート (S) タイプはオプションです。サポートが含まれる場合は RASCI マトリックスと呼ばれ、含まれない場合は RACI マトリックスと呼ばれます。

レスポンシブコントロール

有害事象やセキュリティベースラインからの逸脱について、修復を促すように設計されたセキュリティコントロール。詳細については、「 AWSでのセキュリティコントロールの実装」の「レスポンシブコントロール」を参照してください。

保持

「7 Rs」を参照してください。

廃止

「7 Rs」を参照してください。

検索拡張生成 (RAG)

生成 AI の技術。これにより、LLM では、レスポンスの生成前に、トレーニングデータソースの外部にある信頼できるデータソースが参照されます。例えば、RAG モデルによって、組織のナレッジベースまたはカスタムデータのセマンティック検索を実行できる場合があります。細については、「RAG (検索拡張生成) とは何ですか?」を参照してください。

ローテーション

定期的にシークレット情報を更新して、攻撃者が認証情報にアクセスするのをより困難にするプロセス。

行と列のアクセス制御 (RCAC)

アクセスルールが定義された、基本的で柔軟な SQL 表現の使用。RCAC は行権限と列マスクで構成されています。

RPO

「目標復旧時点」を参照してください。

RTO

「目標復旧時間」を参照してください。

ランブック

特定のタスクを実行するために必要な手動または自動化された一連の手順。これらは通常、エラー率の高い反復操作や手順を合理化するために構築されています。

S

SAML 2.0

多くの ID プロバイダー (IdP) が使用しているオープンスタンダード。この機能を使用すると、フェデレーティッドシングルサインオン (SSO) が有効になるため、ユーザーは組織内のすべてのユーザーを IAM で作成しなくても、 AWS マネジメントコンソール にログインしたり AWS 、API オペレーションを呼び出すことができます。SAML 2.0 ベースのフェデレーションの詳細については、IAM ドキュメントの「SAML 2.0 ベースのフェデレーションについて」を参照してください。

SCADA

「監視制御とデータ取得」を参照してください。

SCP

「サービスコントロールポリシー」を参照してください。

シークレット

暗号化された形式で保存する AWS Secrets Managerパスワードやユーザー認証情報などの機密情報または制限付き情報。シークレット値とそのメタデータで構成されます。シークレット値には、バイナリ、1 つの文字列、複数の文字列を指定できます。詳細については、Secrets Manager ドキュメントの「Secrets Manager シークレットの概要」を参照してください。

セキュリティバイデザイン

開発プロセス全体を通してセキュリティが考慮されているシステムエンジニアリングのアプローチ。

セキュリティコントロール

脅威アクターによるセキュリティ脆弱性の悪用を防止、検出、軽減するための、技術上または管理上のガードレール。セキュリティコントロールには、主に 4 つの種類があります。4 つとは、予防、検出、レスポンシブ、プロアクティブです。

セキュリティ強化

アタックサーフェスを狭めて攻撃への耐性を高めるプロセス。このプロセスには、不要になったリソースの削除、最小特権を付与するセキュリティのベストプラクティスの実装、設定ファイル内の不要な機能の無効化、といったアクションが含まれています。

Security Information and Event Management (SIEM) システム

セキュリティ情報管理 (SIM) とセキュリティイベント管理 (SEM) のシステムを組み合わせたツールとサービス。SIEM システムは、サーバー、ネットワーク、デバイス、その他ソースからデータを収集、モニタリング、分析して、脅威やセキュリティ違反を検出し、アラートを発信します。

セキュリティレスポンスの自動化

セキュリティイベントへの自動レスポンスまたは自動修復を目的として、事前定義およびプログラムされたアクション。これらの自動化は、セキュリティのベストプラクティスを実装するのに役立つ検出的または応答的な AWS セキュリティコントロールとして機能します。自動レスポンスアクションの例には、VPC セキュリティグループの変更、Amazon EC2 インスタンスへのパッチ適用、認証情報の更新などがあります。

サーバー側の暗号化

送信先で、それ AWS のサービス を受け取る によるデータの暗号化。

サービスコントロールポリシー (SCP)

AWS Organizationsの組織内の、すべてのアカウントのアクセス許可を一元的に管理するポリシー。SCP は、管理者がユーザーまたはロールに委任するアクションに、ガードレールを定義したり、アクションの制限を設定したりします。SCP は、許可リストまたは拒否リストとして、許可または禁止するサービスやアクションを指定する際に使用できます。詳細については、 AWS Organizations ドキュメントの「サービスコントロールポリシー」を参照してください。

サービスエンドポイント

のエントリポイントの URL AWS のサービス。ターゲットサービスにプログラムで接続するには、エンドポイントを使用します。詳細については、「AWS 全般のリファレンス」の「AWS のサービス エンドポイント」を参照してください。

サービスレベルアグリーメント (SLA)

サービスのアップタイムやパフォーマンスなど、IT チームがお客様に提供すると約束したものを明示した合意書。

サービスレベルインジケータ (SLI)

エラー率、可用性、スループットといった、サービスパフォーマンス面の指標。

サービスレベル目標 (SLO)

サービスレベルインジケータによって測定され、サービスの状態を表すターゲットメトリクス。

責任共有モデル

クラウドのセキュリティとコンプライアンス AWS について と共有する責任を説明するモデル。AWS はクラウドのセキュリティを担当しますが、 はクラウドのセキュリティを担当します。詳細については、「責任共有モデル」を参照してください。

SIEM

「Security Information and Event Management システム」を参照してください。

単一障害点 (SPOF)

特定のアプリケーションを構成する単一の重要なコンポーネントで発生し、システム稼働に支障をきたす可能性のある障害。

SLA

「サービスレベルアグリーメント」を参照してください。

SLI

「サービスレベルインジケータ」を参照してください。

SLO

「サービスレベルの目標」を参照してください。

スプリットアンドシードモデル

モダナイゼーションプロジェクトのスケーリングと加速のためのパターン。新機能と製品リリースが定義されると、コアチームは解放されて新しい製品チームを作成します。これにより、お客様の組織の能力とサービスの拡張、デベロッパーの生産性の向上、迅速なイノベーションのサポートに役立ちます。詳細については、「AWS クラウドでのアプリケーションをモダナイズするための段階的アプローチ」を参照してください。

SPOF

「単一障害点」を参照してください。

スタースキーマ

データベースの編成構造を意味し、1 つの大きいファクトテーブルにトランザクションデータまたは測定データが保存され、1 つ以上の小さいディメンションテーブルにデータ属性が保存されます。この構造は、データウェアハウスやビジネスインテリジェンスを用途とするように設計されています。

strangler fig パターン

レガシーシステムが廃止されるまで、システム機能を段階的に書き換えて置き換えることにより、モノリシックシステムをモダナイズするアプローチ。このパターンは、宿主の樹木から根を成長させ、最終的にその宿主を包み込み、宿主に取って代わるイチジクのつるを例えています。そのパターンは、モノリシックシステムを書き換えるときのリスクを管理する方法として Martin Fowler により提唱されました。このパターンの適用方法の例については、「コンテナと Amazon API Gateway を使用して、従来の Microsoft ASP.NET (ASMX) ウェブサービスを段階的にモダナイズ」を参照してください。

サブネット

VPC 内の IP アドレスの範囲。サブネットは、1 つのアベイラビリティーゾーンに存在する必要があります。

監視制御とデータ取得 (SCADA)

製造分野において、ハードウェアとソフトウェアを使用して物理アセットと本番運用をモニタリングするシステム。

対称暗号化

データの暗号化と復号に同じキーを使用する暗号化のアルゴリズム。

合成テスト

ユーザーとのやり取りをシミュレートして、起こり得る問題を検出したり、パフォーマンスをモニタリングしたりすることで、システムをテストします。Amazon CloudWatch Synthetics を使用すると、こうしたテストを作成できます。

システムプロンプト

コンテキスト、指示、ガイドラインなどを提示して、LLM に動作を指示する手法。システムプロンプトは、コンテキストを設定して、ユーザーとやり取りするルールを確立するのに有用です。

T

タグ

AWS リソースを整理するためのメタデータとして機能するキーと値のペア。タグは、 リソースの管理、識別、整理、検索、フィルタリングに役立ちます。詳細については、「AWS リソースのタグ付け」を参照してください。

ターゲット変数

監督された機械学習でお客様が予測しようとしている値。これは、結果変数のことも指します。例えば、製造設定では、ターゲット変数が製品の欠陥である可能性があります。

タスクリスト

ランブックの進行状況を追跡するために使用されるツール。タスクリストには、ランブックの概要と完了する必要のある一般的なタスクのリストが含まれています。各一般的なタスクには、推定所要時間、所有者、進捗状況が含まれています。

テスト環境

「環境」を参照してください。

トレーニング

お客様の機械学習モデルに学習するデータを提供すること。トレーニングデータには正しい答えが含まれている必要があります。学習アルゴリズムは入力データ属性をターゲット (お客様が予測したい答え) にマッピングするトレーニングデータのパターンを検出します。これらのパターンをキャプチャする機械学習モデルを出力します。そして、お客様が機械学習モデルを使用して、ターゲットがわからない新しいデータでターゲットを予測できます。

トランジットゲートウェイ

VPC とオンプレミスネットワークを相互接続するために使用できる、ネットワークの中継ハブ。詳細については、 AWS Transit Gateway ドキュメントの「トランジットゲートウェイとは」を参照してください。

トランクベースのワークフロー

デベロッパーが機能ブランチで機能をローカルにビルドしてテストし、その変更をメインブランチにマージするアプローチ。メインブランチはその後、開発環境、本番前環境、本番環境に合わせて順次構築されます。

信頼されたアクセス

ユーザーに代わって AWS Organizations およびそのアカウントで組織内でタスクを実行するために指定したサービスにアクセス許可を付与します。信頼されたサービスは、サービスにリンクされたロールを必要なときに各アカウントに作成し、ユーザーに代わって管理タスクを実行します。詳細については、 ドキュメントの「Using AWS Organizations with other AWS services AWS Organizations 」を参照してください。

チューニング

機械学習モデルの精度を向上させるために、お客様のトレーニングプロセスの側面を変更する。例えば、お客様が機械学習モデルをトレーニングするには、ラベル付けセットを生成し、ラベルを追加します。これらのステップを、異なる設定で複数回繰り返して、モデルを最適化します。

ツーピザチーム

2 枚のピザを分け合えることができるくらい小さな DevOps チーム。ツーピザチームの規模では、ソフトウェア開発におけるコラボレーションに最適な機会が確保されます。

U

不確実性

予測機械学習モデルの信頼性を損なう可能性がある、不正確、不完全、または未知の情報を指す概念。不確実性には、次の 2 つのタイプがあります。認識論的不確実性は、限られた、不完全なデータによって引き起こされ、弁論的不確実性 は、データに固有のノイズとランダム性によって引き起こされます。詳細については、深層学習システムにおける不確実性の定量化ガイドを参照してください。

未分化なタスク

ヘビーリフティングとも呼ばれ、アプリケーションの作成と運用には必要だが、エンドユーザーに直接的な価値をもたらさなかったり、競争上の優位性をもたらしたりしない作業です。未分化なタスクの例としては、調達、メンテナンス、キャパシティプランニングなどがあります。

上位環境

「環境」を参照してください。

V

バキューミング

ストレージを再利用してパフォーマンスを向上させるために、増分更新後にクリーンアップを行うデータベースのメンテナンス操作。

バージョンコントロール

リポジトリ内のソースコードへの変更など、変更を追跡するプロセスとツール。

VPC ピアリング

プライベート IP アドレスを使用してトラフィックをルーティングできる、2 つの VPC 間の接続。詳細については、Amazon VPC ドキュメントの「VPC ピア機能とは」を参照してください。

脆弱性

システムのセキュリティを脅かすソフトウェアまたはハードウェアの欠陥。

W

ウォームキャッシュ

頻繁にアクセスされる最新の関連データを含むバッファキャッシュ。データベースインスタンスはバッファキャッシュから、メインメモリまたはディスクからよりも短い時間で読み取りを行うことができます。

ウォームデータ

アクセス頻度の低いデータ。この種類のデータをクエリする場合、通常は適度に遅いクエリでも問題ありません。

ウィンドウ関数

現在のレコードに何らかの形で関連している行のグループに計算を実行する SQL 関数。ウィンドウ関数は、移動平均を計算したり、現在の行の相対位置に基づいて他の行の値にアクセスするといったタスクの処理に役立ちます。

ワークロード

ビジネス価値をもたらすリソースとコード (顧客向けアプリケーションやバックエンドプロセスなど) の総称。

ワークストリーム

特定のタスクセットを担当する移行プロジェクト内の機能グループ。各ワークストリームは独立していますが、プロジェクト内の他のワークストリームをサポートしています。たとえば、ポートフォリオワークストリームは、アプリケーションの優先順位付け、ウェーブ計画、および移行メタデータの収集を担当します。ポートフォリオワークストリームは、これらの設備を移行ワークストリームで実現し、サーバーとアプリケーションを移行します。

WORM

「Write-Once-Read-Many」を参照してください。

WQF

「AWS ワークロード資格フレームワーク」を参照してください

Write-Once-Read-Many (WORM)

データを 1 回のみ書き込むことで、データの削除や変更を防ぐストレージモデル。承認済みユーザーは、必要な回数だけデータを読み取ることができますが、変更することはできません。このデータストレージインフラストラクチャは、イミュータブルと見なされます。

Z

ゼロデイエクスプロイト

ゼロデイ脆弱性を悪用した攻撃 (一般的にマルウェアによる)。

ゼロデイ脆弱性

実稼働システムにおける未解決の欠陥または脆弱性。脅威アクターは、このような脆弱性を利用してシステムを攻撃する可能性があります。開発者は、よく攻撃の結果で脆弱性に気付きます。

ゼロショットプロンプト

LLM にタスク実行の手順は提示するが、実行のガイドとして役立つ例 (ショット) は提示しない方法。LLM は、事前トレーニング済みの知識を使用してタスクを処理する必要があります。ゼロショットプロンプトの有効性は、タスクの複雑さとプロンプトの品質によって異なります。「数ショットプロンプト」も参照してください。

ゾンビアプリケーション

平均 CPU およびメモリ使用率が 5% 未満のアプリケーション。移行プロジェクトでは、これらのアプリケーションを廃止するのが一般的です。

翻訳は機械翻訳により提供されています。提供された翻訳内容と英語版の間で齟齬、不一致または矛盾がある場合、英語版が優先します。