



使用 Amazon DynamoDB 对数据建模

AWS 规范性指导



AWS 规范性指导: 使用 Amazon DynamoDB 对数据建模

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

Amazon 的商标和商业外观不得用于任何非 Amazon 的商品或服务，也不得以任何可能引起客户混淆、贬低或诋毁 Amazon 的方式使用。所有非 Amazon 拥有的其他商标均为各自所有者的财产，这些所有者可能附属于 Amazon、与 Amazon 有关联或由 Amazon 赞助，也可能不是如此。

Table of Contents

简介	1
流程	2
RACI 矩阵	2
过程步骤	4
步骤 1：确定用例和逻辑数据模型	4
目标	4
流程	4
工具和资源	4
RACI	5
输出	5
步骤 2：创建初步的成本估算	5
目标	5
流程	5
工具和资源	6
RACI	6
输出	6
第 3 步。识别数据访问模式	6
目标	6
流程	6
工具和资源	7
RACI	7
输出	7
示例	7
步骤 4：确定技术要求	8
目标	8
流程	8
工具和资源	8
RACI	8
输出	8
步骤 5。创建 DynamoDB 数据模型	9
目标	9
流程	9
工具和资源	10
RACI	10

输出	10
示例	10
步骤 6. 创建数据查询	11
目标	11
流程	11
工具和资源	12
RACI	12
输出	12
示例	12
第 7 步。验证数据模型	12
目标	12
流程	13
工具和资源	13
RACI	13
输出	13
步骤 8 : 查看成本估算	13
目标	13
流程	14
工具和资源	14
RACI	14
输出	14
第 9 步。部署数据模型	14
目标	14
流程	15
工具和资源	15
RACI	15
输出	15
示例	15
模板	17
业务需求评测模板	17
技术需求评测模板	20
访问模式模板	24
模板	24
最佳实践	28
分层数据建模	29
第 1 步 : 确定使用案例和逻辑数据模型	29

第 2 步：创建初步的成本估算	31
第 3 步：识别数据访问模式	31
第 4 步：确定技术要求	32
第 5 步：创建 DynamoDB 数据模型	32
在表格中存储组件	33
该 GSI1 指数	34
该 GSI2 指数	34
第 6 步：创建数据查询	35
第 7 步：验证数据模型	38
第 8 步：查看成本估算信息	39
目标	39
流程	39
第 9 步：部署数据模型	40
其他资源	42
贡献者	44
文档历史记录	45
术语表	46
#	46
A	46
B	49
C	50
D	53
E	56
F	58
G	59
H	60
我	61
L	63
M	64
O	68
P	70
Q	72
R	73
S	75
T	78
U	79

V	80
W	80
Z	81
.....	lxxxii

使用 Amazon DynamoDB 对数据建模

流程、模板和最佳实践

Amazon Web Services ([贡献者](#))

2023 年 12 月 ([文档历史记录](#))

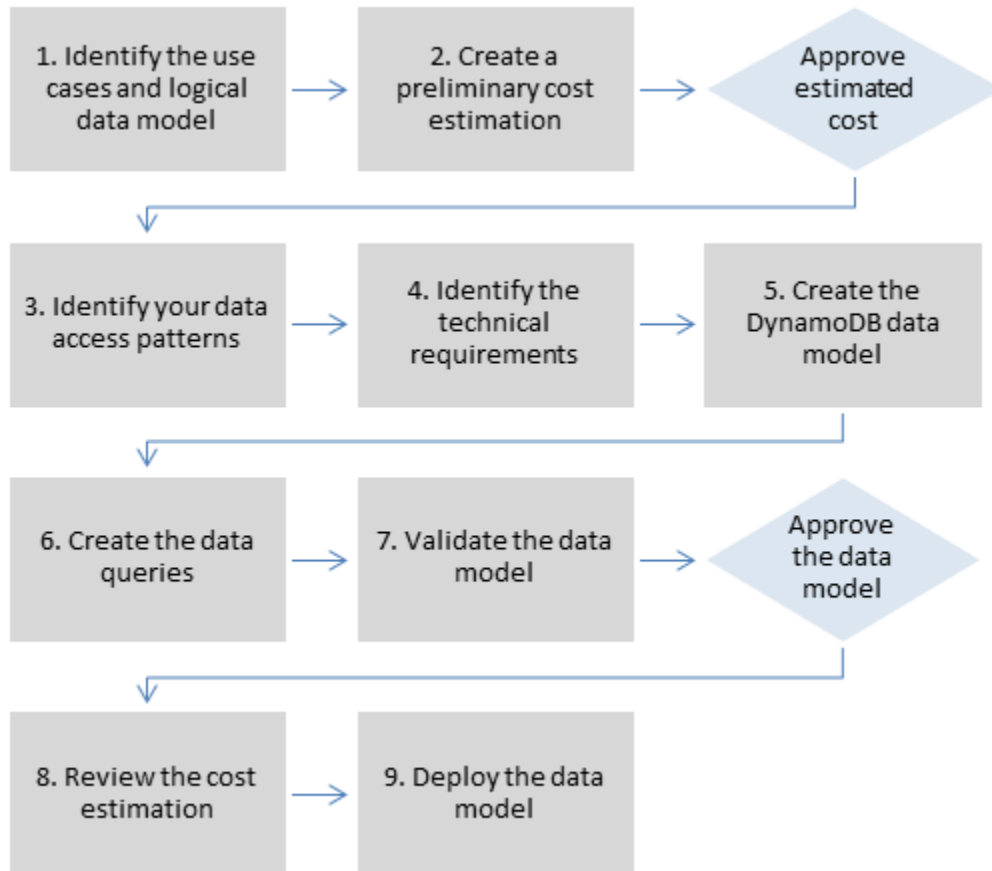
NoSQL 数据库为构建现代应用程序提供了灵活的架构。它们因其易于开发、功能强大和大规模性能而广受认可。Amazon DynamoDB 为 Amazon Web Services (AWS) 云中的 NoSQL 数据库提供快速、可预测的性能，能够实现无缝扩展。作为完全托管式数据库服务，DynamoDB 可助力减轻运营和扩展分布式数据库的管理负担。您不必担心硬件预置、设置和配置、复制、软件修补或集群扩展。

NoSQL 架构设计需要不同于传统关系数据库管理系统 (RDBMS) 设计的方法。RDBMS 数据模型侧重于数据的结构及其与其他数据的关系。NoSQL 数据建模侧重于访问模式或应用程序将如何使用数据，因此它以支持简单查询操作的方式存储数据。对于像 Microsoft SQL Server 或 IBM Db2 这样的 RDBMS，您可以创建一个标准化的数据模型，而无需过多考虑访问模式。之后再对其进行扩展以支持您的模式和查询。

本指南介绍了使用 DynamoDB 的数据建模流程，该流程提供了功能要求、性能和有效成本。该指南适用于计划使用 DynamoDB 作为其正在 AWS 运行的应用程序的操作数据库的数据库工程师。AWS 专业服务使用推荐的流程来帮助企业公司针对不同的用例和工作负载进行 DynamoDB 数据建模。

数据建模流程

我们建议在使用 Amazon DynamoDB 进行数据建模时采用以下流程。[本指南稍后](#)将详细讨论这些步骤。



RACI 矩阵

一些组织使用责任分配矩阵（也称为 RACI 矩阵）来描述一个特定项目或业务流程中涉及的各种角色。本指南提供了一个建议的 RACI 矩阵，可以帮助您的组织为 DynamoDB 数据建模过程确定合适的人员和正确的责任。对于该流程中的每个步骤，它列出了利益相关者及其参与情况：

- R：负责完成步骤
- A：负责批准和签署工作
- C：咨询以提供任务的意见
- I：了解了进展情况，但不直接参与任务

根据您的组织和项目团队的结构，以下 RACI 矩阵中的角色可以由同一个利益相关者担任。在某些情况下，利益相关方既要具体步骤负责，又要承担责任。例如，数据库工程师可以负责创建和批准数据模型，因为这是他们的域区域。

过程步骤	业务用户	业务分析师	解决方案架构师	数据库工程师	应用程序开发者	DevOps 工程师
1. 确定用例和逻辑数据模型	C	R/A	我	R		
2. 创建初步的成本估算	C	A	我	R		
3. 识别数据访问模式	C	A	我	R		
4. 确定技术要求	C	C	A	R		
5. 创建 DynamoDB 数据模型	我	我	我	R/A		
6. 创建数据查询	我	我	我	R/A	R	
7. 验证数据模型	A	R	我	C		
8. 查看成本估算	C	A	我	R		
9. 部署 DynamoDB 数据模型	我	我	C	C		R/A

数据建模流程步骤

本节详细介绍了 Amazon DynamoDB 推荐的数据建模过程的每个步骤。

主题

- [步骤 1：确定用例和逻辑数据模型](#)
- [步骤 2：创建初步的成本估算](#)
- [第 3 步。识别数据访问模式](#)
- [步骤 4：确定技术要求](#)
- [步骤 5。创建 DynamoDB 数据模型](#)
- [步骤 6. 创建数据查询](#)
- [第 7 步。验证数据模型](#)
- [步骤 8：查看成本估算](#)
- [第 9 步。部署数据模型](#)

步骤 1：确定用例和逻辑数据模型

目标

- 收集需要 NoSQL 数据库的业务需求和用例。
- 使用实体关系 (ER) 图定义逻辑数据模型。

流程

- 业务分析师会访谈业务用户，以确定用例和预期结果。
- 数据库工程师创建概念数据模型。
- 数据库工程师创建逻辑数据模型。
- 数据库工程师收集有关项目大小、数据量和预期读写吞吐量的信息。

工具和资源

- 业务需求评测 (请参阅[模板](#))

- 访问模式矩阵 (请参阅[模板](#))
- 您首选的图表创建工具

RACI

业务用户	业务分析师	解决方案架构师	数据库工程师	应用程序开发者	DevOps 工程师
C	R/A	我	R		

输出

- 记录用例和业务需求
- 逻辑数据模型 (ER 图)

步骤 2：创建初步的成本估算

目标

- 为 DynamoDB 制定初步的成本估算。

流程

- 数据库工程师使用 [DynamoDB 定价页面](#) 上提供的可用信息和示例创建初始成本分析。
 - 为按需容量创建成本估算 (请参阅[示例](#))。
 - 为预置容量创建成本估算 (请参阅[示例](#))。
 - 对于预置容量模型，请从计算器中获取估算成本，然后对预留容量应用折扣。
 - 比较两种容量模型的估计成本。
 - 创建所有环境 (开发、生产、质量保证) 的成本估算。
- 业务分析师审查并批准或拒绝初步成本估算。

工具和资源

- [AWS 定价计算器](#)

RACI

业务用户	业务分析师	解决方案架构师	数据库工程师	应用程序开发者	DevOps 工程师
C	A	我	R		

输出

- 初步的成本估算

第 3 步。识别数据访问模式

访问模式或查询模式定义了用户和系统如何访问数据以满足业务需求。

目标

- 记录数据访问模式。

流程

- 数据库工程师和业务分析师会见最终用户，以确定如何使用数据访问模式矩阵模板查询数据。
 - 对于新应用程序，他们查看有关活动和目标的用户案例。他们记录用例并分析用例所需的访问模式。
 - 对于现有应用程序，他们分析查询日志以了解人们目前使用该系统的方式，以及确认键访问模式。
- 数据库工程师确定了访问模式的以下属性：
 - 数据大小：了解一次存储和请求的数据量将有助于确定对数据进行分区的最有效方法（请参阅[博客帖子](#)）。
 - 数据形状：NoSQL 数据库处理查询时不会改变数据形状（RDBMS 系统会这样做），而是整理数据，使数据库中的数据形状与查询内容对应。这是加快速度并增强可扩展性的一个关键因素。

- **数据速度**：DynamoDB 通过增加可用于处理查询的物理分区数量，并在这些分区高效分布数据来进行扩展。提前了解查询负载峰值可能有助于确定如何对数据进行分区以获得最佳利用 I/O 容量。
- 企业用户对访问或查询模式进行优先排序。
 - 优先级查询通常是最常用或最相关的查询。识别需要较低响应延迟的查询也很重要。

工具和资源

- 访问模式矩阵 (请参阅 [模板](#))
- [选择正确的 DynamoDB 分区键](#) AWS (数据库博客)
- [适用于 DynamoDB 的 NoSQL 设计](#) (DynamoDB 文档)

RACI

业务用户	业务分析师	解决方案架构师	数据库工程师	应用程序开发者	DevOps 工程师
C	A	我	R		

输出

- 数据访问模式矩阵

示例

访问模式	优先级	读或写	描述	类型 (单个项目、多个项目或所有项目)	关键属性	筛选条件	结果排序
创建用户个人资料	高	写入	用户创建新的个人资料	单个项目	用户名	不适用	不适用

访问模式	优先级	读或写	描述	类型 (单个项目、多个项目或所有项目)	关键属性	筛选条件	结果排序
更新用户个人资料	中	写入	用户更新其个人资料	单个项目	用户名	用户名 = 当前用户	不适用

步骤 4：确定技术要求

目标

- 收集 DynamoDB 数据库的技术要求。

流程

- 业务分析师使用评估问卷采访业务用户和 DevOps 团队，以收集技术要求。

工具和资源

- 技术需求评测 (请参阅 [示例问卷](#))

RACI

业务用户	业务分析师	解决方案架构师	数据库工程师	应用程序开发者	DevOps 工程师
C	C	A	R		

输出

- 技术要求文档

步骤 5。创建 DynamoDB 数据模型

目标

- 创建 DynamoDB 数据模型。

流程

- 数据库工程师确定每个用例需要多少张表。我们建议在 DynamoDB 应用程序中保留尽可能少的表。
- 根据最常见的访问模式，确定可以是两种类型之一的主键：带有用于标识数据的分区键的主键，或带有分区键和排序键的主键。排序键是用于分组和组织数据的辅助索引，因此可以在分区内高效地对其进行查询。您可以利用排序键，在数据中定义层级关系，在任何层级查询（请参阅[博客帖子](#)）。
- 分区键设计
 - 定义分区键并评估其分布。
 - 确定是否需要[写入分片](#)均匀分配工作负载。
- 排序键设计
 - 确定排序键。
 - 确定是否需要复合排序键。
 - 确定版本控制的需求。
- 根据访问模式，确定二级索引以满足查询要求。
 - 确定是否需要[本地二级索引](#) (LSIs)。存在分区键与基表相同、但排序键不同的索引。
 - 对于带的表 LSIs，每个分区键值的大小限制为 10 GB。只要任何一个分区键值的总大小不超过 10 GB，则具有的表 LSIs 可以存储任意数量的项目。
 - 确定是否需要[全局二级索引](#) (GSIs)。存在分区键和排序键可以与基表上的分区键和排序键不同的索引（请参阅[博客帖子](#)）。
 - 定义索引投影。考虑减少投影属性的数量，尽可能减少写入索引的项目大小。在此步骤中，应确定是否要使用以下内容：
 - [稀疏索引](#)
 - [物化聚合查询](#)
 - [GSI 重载](#)
 - [GSI 分片](#)

- 数据库工程师决定数据是否包含大型项目。如果是，他们[通过使用压缩方法或在 Amazon Simple Storage Service \(Amazon S3\) 存储数据](#)，设计解决方案。
- 数据库工程师决定是否需要时间序列数据。如果是，他们将使用[时间序列设计模式](#)对数据进行建模。
- 数据库工程师确定 ER 模型是否包含 many-to-many 关系。如果是，他们将使用[相邻列表设计模式](#)对数据进行建模。

工具和资源

- [NoSQL Workbench for Amazon DynamoDB](#)：提供数据建模、数据可视化和查询开发及测试功能，可帮助您设计 DynamoDB 数据库
- [适用于 DynamoDB 的 NoSQL 设计](#) (DynamoDB 文档)
- [选择正确的 DynamoDB 分区键](#) AWS (数据库博客)
- [在 DynamoDB 中使用二级索引的最佳实践](#) (DynamoDB 文档)
- [如何设计亚马逊 DynamoDB 全球二级索引](#) AWS (数据库博客)

RACI

业务用户	业务分析师	解决方案架构师	数据库工程师	应用程序开发者	DevOps 工程师
我	我	我	R/A		

输出

- 满足您的访问模式和要求的 DynamoDB 表架构

示例

以下屏幕截图所示为 NoSQL Workbench。

RetailDatabase		GSI: BundleProducts-VendorProducts-CustomerOrdersByOrderId		GSI: VendorOrdersByStatusDate-ProductInventory		GSI: ProductCatalog-CustomerOrdersByProduct	
Primary Key		Attributes					
Partition Key: pk	Sort Key: sk						
P1	B1	GSI1-PK	GSI1-SK	name	desc		
		B1	P1	The Tiki Bundle	Everything you need for an island theme party.		
P4	B2	GSI1-PK	GSI1-SK	name	desc		
		B2	P4	Tiki Bar Set	Be the Mai Tai master with your very own Tiki Bar.		
P2	B1	name	desc	qty	GSI1-PK	GSI1-SK	location
		Tiki Torch	Bamboo tiki torch, 4 ft	6	B1	P2	W1-A9-S10-B52
	B2	name	desc	qty	GSI1-PK	GSI1-SK	location
		Tiki Torch	Bamboo tiki torch, 4 ft	2	B2	P2	W1-A9-S10-B52
P2	P2	name	desc	qty	location	reorderAt	GSI3-SK
		Tiki Torch	Bamboo tiki torch, 4 ft	656	W1-A9-S10-B52	100	/GardenOutdoor/OutdoorDecor/Lighting/LanternsT
	B1	name	desc	qty	GSI1-PK	GSI1-SK	location
		Tiki Statue - Pele	Tiki of the Hawaiian Fire Goddess Pele, 5 ft.	1	B1	P3	W1-A15-S6-B27

步骤 6. 创建数据查询

目标

- 创建主要查询以验证数据模型。

流程

- 数据库工程师在 AWS 区域或他们的计算机 (DynamoDB Local) 上手动创建 DynamoDB 表。
- 数据库工程师将示例数据添加到 DynamoDB 表中。
- [数据库工程师使用适用于 Amazon DynamoDB 的 NoSQL Workbench 或适用于 Java 或 Python AWS 的软件开发工具包来构建示例查询 \(参见博客文章 \) 。](#)

分面就像 DynamoDB 表的视图。

- 数据库工程师和云开发人员使用首选语言的 AWS Command Line Interface (AWS CLI) 或 AWS SDK 来构建示例查询。

工具和资源

- 一个活跃 AWS 账户，用于访问 DynamoDB 控制台
- [DynamoDB Local](#) (可选)，如果您想在不访问 DynamoDB 网络服务的情况下在计算机上构建数据库
- [NoSQL Workbench for Amazon DynamoDB](#) (下载和文档)
- AWS 使用你选择的语言 (Python JavaScript、PHP、.NET、Ruby、Java、Go、Node.js、C++ 和 SAP ABAP) [开发工具包](#)

RACI

业务用户	业务分析师	解决方案架构师	数据库工程师	应用程序开发者	DevOps 工程师
我	我	我	R/A	R	

输出

- 查询 DynamoDB 表的代码

示例

- [使用 AWS 适用于 Java 的 SDK 的 DynamoDB 示例](#)
- [Python 示例](#)
- [JavaScript 例子](#)

第 7 步。验证数据模型

目标

- 确保数据模型满足您的要求。

流程

- 数据库工程师用示例数据填充到 DynamoDB 表中。
- 数据库工程师运行代码来查询 DynamoDB 表。
- 数据库工程师收集查询结果。
- 数据库工程师收集查询性能指标。
- 企业用户验证查询结果是否满足业务需求。
- 业务分析师验证技术要求。

工具和资源

- 一个活跃 AWS 账户，用于访问 DynamoDB 控制台
- [DynamoDB Local](#) (可选)，如果您想在不访问 DynamoDB 网络服务的情况下在计算机上构建数据库
- AWS 使用您选择的语言进行 [SDK](#)

RACI

业务用户	业务分析师	解决方案架构师	数据库工程师	应用程序开发者	DevOps 工程师
A	R	我	C		

输出

- 批准的数据模型

步骤 8：查看成本估算

目标

- 定义容量模型并估计 DynamoDB 成本，以完善 [第 2 步](#) 中的成本估算。
- 获得业务分析师和利益相关者的最终财务审批。

流程

- 数据库工程师确定数据量的估计值。
- 数据库工程师确定数据传输要求。
- 数据库工程师定义所需的读取和写入容量单位。
- 业务分析师在[按需容量模式和预置容量模式](#)之间做出决定。
- 数据库工程师确定了 [DynamoDB 自动扩缩](#)的需求。
- 数据库工程师在 Simple Monthly Calculator 工具中输入参数。
- 数据库工程师向业务利益相关者提交最终的价格估算。
- 业务分析师和利益相关者批准或拒绝解决方案。

工具和资源

- [AWS 定价计算器](#)

RACI

业务用户	业务分析师	解决方案架构师	数据库工程师	应用程序开发者	DevOps 工程师
C	A	我	R		

输出

- 容量模式
- 订正成本估算

第 9 步。部署数据模型

目标

- 将一个或多个 DynamoDB 表部署到 AWS 区域。

流程

- DevOps 架构师为 DynamoDB 表 (或多个表) 创建 CloudFormation 模板或其他基础设施即代码 (IaC) 工具。CloudFormation 提供了一种自动配置和配置表及相关资源的方式。

工具和资源

- [CloudFormation](#)

RACI

业务用户	业务分析师	解决方案架构师	数据库工程师	应用程序开发者	DevOps 工程师
我	我	C	C		R/A

输出

- AWS CloudFormation 模板

示例

```
mySecondDDBTable:
  Type: AWS::DynamoDB::
  Table DependsOn: "myFirstDDBTable"
  Properties:
    AttributeDefinitions:
      - AttributeName: "ArtistId"
        AttributeType: "S"
      - AttributeName: "Concert"
        AttributeType: "S"
      - AttributeName: "TicketSales"
        AttributeType: "S"
    KeySchema:
      - AttributeName: "ArtistId"
        KeyType: "HASH"
```

```
- AttributeName: "Concert"  
  KeyType: "RANGE"  
ProvisionedThroughput:  
  ReadCapacityUnits:  
    Ref: "ReadCapacityUnits"  
  WriteCapacityUnits:  
    Ref: "WriteCapacityUnits"  
GlobalSecondaryIndexes:  
- IndexName: "myGSI"  
  KeySchema:  
  - AttributeName: "TicketSales"  
    KeyType: "HASH"  
  Projection:  
    ProjectionType: "KEYS_ONLY"  
  ProvisionedThroughput:  
  ReadCapacityUnits:  
    Ref: "ReadCapacityUnits"  
  WriteCapacityUnits:  
    Ref: "WriteCapacityUnits"  
Tags:  
- Key: mykey  
  Value: myvalue
```

模板

本节提供的模板基于 AWS 网站上[使用 Amazon DynamoDB 对游戏玩家数据进行建模](#)。

Note

本节中的表格使用 MM 作为百万的缩写，使用 K 作为千的缩写。

主题

- [业务需求评测模板](#)
- [技术需求评测模板](#)
- [访问模式模板](#)

业务需求评测模板

提供用例的描述：

描述

想象一下，您正在开发一款在线多人游戏。在您的游戏中，由 50 名玩家组成的小组加入一个会话来玩游戏，这通常需要大约 30 分钟才能玩完。在游戏过程中，您必须更新特定玩家的记录，以显示该玩家的游戏时间、他们的统计数据或他们是否赢得了比赛。用户希望看到他们玩过的旧游戏，要么查看游戏的获胜者，要么观看每款游戏动作的重播。

提供有关您的用户的信息：

用户	描述	预期数字
游戏玩家	在线游戏玩家。	100 万
开发团队	内部团队将使用游戏统计数据来改进游戏体验。	100

提供有关数据来源和数据摄取方式的信息：

源	描述	用户
在线游戏	游戏玩家将创建个人资料并开始新游戏。	游戏玩家
游戏应用程序	游戏应用程序将自动收集有关游戏的统计数据，例如开始和结束时间、玩家人数、每个玩家的位置以及游戏地图。	

提供有关如何使用数据的信息：

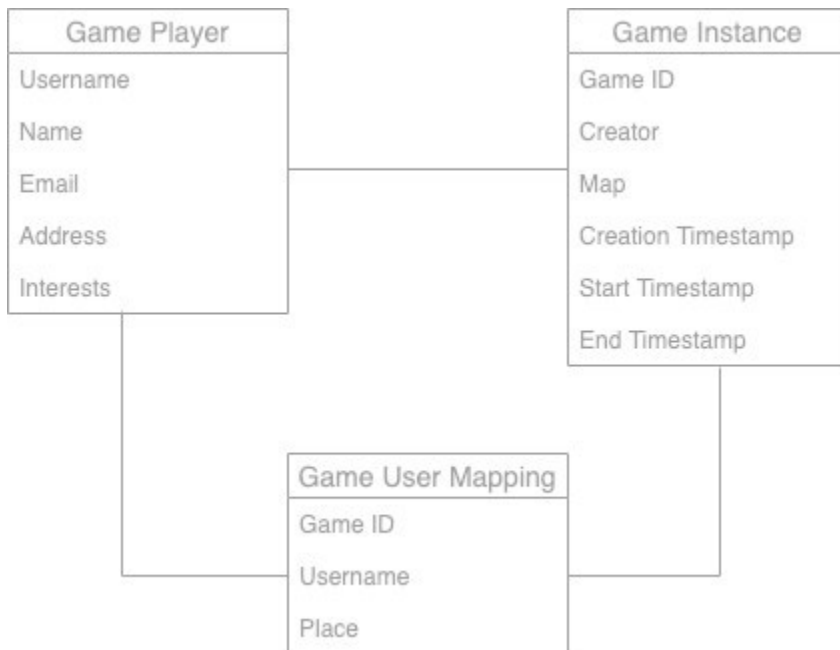
使用者	描述	用户
在线游戏	游戏玩家将查看个人资料并查看其游戏统计信息。	游戏玩家
数据分析	游戏开发团队将提取游戏统计数据进行分析并改善用户体验。数据将从数据存储库中导出并导入到 Amazon S3 中，以支持通过 Spark 应用程序进行分析。	开发团队

提供实体清单及其识别方式：

实体名称	描述	标识符
游戏玩家	存储每个用户（玩家）的身份、地址、人口统计信息、兴趣爱好等数据。	用户名：

游戏实例	提供有关玩过的每个游戏的信息，包括创建者、开始时间、结束时间以及使用的地图。	游戏 ID
游戏用户对应关系	表示用户与游戏之间的多对多关系。	游戏 ID 和用户名

为实体创建 ER 模型：



提供有关实体的高级统计数据：

实体名称	预计的记录数	记录大小	备注
游戏玩家	100 万	< 1 KB	游戏平台大约有 100 万用户。
游戏实例	600 万 (10 万/天 * 60 天)	< 1 KB	平均而言，每天有 10 万场游戏。我们需要存储过去 60 天的数据。
游戏用户对应关系	3 亿	< 1 KB	平均而言，每场比赛有 50 名玩家，我们需

(600 万场游戏 * 50
个玩家)

要存储他们的相关信
息。

技术需求评测模板

提供有关数据摄取类型的信息：

数据摄取类型	是/否	描述	频率。
应用程序访问	Y		
API 网关	Y		
数据流式传输	N		
批处理	N		
ETL	N		
导入数据	N		
时间序列	N		

提供有关数据使用类型的信息：

数据使用类型	是/否	描述	频率。
应用程序访问			
API 网关			
数据导出			
数据分析			
数据聚合			
Reporting			

搜索

数据流式传输

ETL

提供数据量估算值：

实体名称	预计的记录数	记录大小	数据量
游戏玩家	100 万	< 1 KB	1 GB (100 万 * 1 KB)
游戏实例	600 万 (10 万/天 * 60 天)	< 1 KB	~ 6 GB (600 万 * 1 KB)
游戏用户对对应关系	3 亿 (600 万场游戏 * 50 个玩家)	< 1 KB	~ 300 GB (3 亿 * 1 KB)

Note

数据留存期 60 天。60 天后，必须将数据存储到 Amazon S3 中进行分析，方法是使用 [DynamoDB 生存时间 \(TTL\) 自动将数据从 DynamoDB 移至 Amazon S3](#)。

回答以下有关时间模式的问题：

- 用户可以在什么时间范围内使用该应用程序（例如，全天候或工作日上午 9 点至下午 5 点）？
- 白天的使用量是否达到高峰？多少小时？应用程序使用百分比是多少？

指定写入吞吐量要求：

实体名称	写入数/天	小时/天	写入数/秒
------	-------	------	-------

游戏玩家	10000 次更新	18	< 1
游戏实例	300,000	18	< 5
游戏用户对对应关系	1800000000	18	~ 27.777

备注

游戏玩家写入操作：1% 的用户每天更新他们的个人资料，因此我们预计 100 万名用户会有 10,000 次更新。

游戏实例写入操作：100,000 个游戏/天。对于每款游戏，我们至少有 3 个写入操作（在创建、开始和结束时），因此总共有 300,000 个写入操作。

游戏用户映射写入操作：每款游戏每天有 100,000 个，每次有 50 个玩家。平均游戏时长为 30 分钟，玩家位置每 5 秒更新一次。我们估计每位玩家平均有 360 次更新，因此总数为 $100,000 * 50 * 360 = 18$ 亿次写入操作。

指定读取吞吐量要求：

实体名称	读取数/天	小时/天	读取数/秒
游戏玩家	200,000	18	~ 3
游戏实例	5000000	18	~ 77
游戏用户对对应关系	1800000000	18	~ 27.777

备注

游戏玩家读取操作：20% 的用户开始游戏，因此 $1000000 * 0.2 = 200000$ 。

游戏实例读取操作：100,000 个游戏/天。对于每款游戏，我们每位玩家至少有 1 次读取操作，每款游戏有 50 名玩家，因此读取操作总数为 500 万次。

游戏用户映射读取操作：每天 50 个玩家有 100,000 场游戏。平均游戏时长为 30 分钟，玩家位置每 5 秒更新一次。我们估计每位玩家平均需要 360 次更新，并且每次更新都需要读取操作，因此总数为 $100,000 * 50 * 360 = 18$ 亿次读取操作。

指定数据访问延迟要求：

操作	第 99 百分位数	最大延迟
读取	30 毫秒	100 毫秒
写入	10 毫秒	50 毫秒

指定数据可用性要求：

要求	是/否	指标	备注
高可用性	Y	99.9%	
RTO	Y	~1 小时	恢复时间目标
RPO	Y	~1 小时	恢复点目标
灾难恢复。	N		
区域内数据复制	N		
跨区域数据复制	N	3 秒延迟	哪个 AWS 区域？

指定安全要求：

要求	是/否	备注
敏感数据存储	N	受保护的健康信息 (PHI)、支付卡行业 (PCI) 信息、个人身份信息 (PII) ？
静态加密	Y	
传输中加密。	Y	
客户端加密	N	

任何专有或第三方供应商的加密库	N
用户访问日志记录	N
数据访问审计	N

访问模式模板

使用以下字段收集和记录有关用例访问模式的信息：

字段	描述
访问模式	提供访问模式的名称。
描述	提供对访问模式的更详细描述。
优先级	定义访问模式的优先级（高、中或低）。这定义了应用程序最相关的访问模式。
读或写	它是读取权限还是写入访问模式？
类型	该模式是访问单个项目、多个项目还是所有项目？
筛选条件	访问模式是否需要任何过滤器？
排序	结果需要任何排序吗？

模板

访问模式	描述	优先级 -	读或写	类型（单 个项目、 多个	关键属性	筛选条件	结果排序
------	----	-------	-----	--------------------	------	------	------

				多个项目 或所有项目 ()			
创建用户 个人资料	用户创建 新的个人 资料。	高。	写入	单个项目	用户名 :	不适用	(不适 用)
更新用户 个人资料	用户更新 其个人资 料。	中	写入	单个项目	用户名 :	用户名 = 当前用户	(不适 用)
获取用户 个人资料	用户查看 其个人资 料。	高。	读取	单个项目	用户名 :	用户名 = 当前用户	(不适 用)
创建游戏	用户创建 一个新游 戏。	高。	写入	单个项目	游戏 ID	不适用	(不适 用)
查找开放 的游戏	用户搜索 开放的游 戏。搜索 结果按起 始时间戳 降序排 列。	高。	读取	多项		游戏状态 = 开放	起始时间 戳降序
通过地图 查找开放 的游戏	用户使用 按开始时 间戳降序 排列的特 定地图来 搜索开放 的 游戏。	中	读取	多项		游戏状 态 = 开 放, 地图 = XYZ	起始时间 戳降序

查看游戏	用户查看游戏详情。	高。	读取	单个项目	游戏 ID	不适用	(不适用)
查看游戏中的用户	用户获取游戏中所有用户的列表。	中	读取	多项		游戏 ID = XYZ	(不适用)
将用户加入游戏	用户加入开放的游戏。	高。	写入	单个项目	游戏 ID 和用户名	游戏状态 = 开放	(不适用)
开始一场游戏	用户开始一场新游戏。	高。	写入	单个项目	游戏 ID	不适用	(不适用)
为用户更新游戏	更新用户在游戏中的位置。	中	写入	单个项目	游戏 ID 和用户名	不适用	(不适用)
更新游戏	游戏结束；更新统计数据。	中	写入	单个项目	游戏 ID	不适用	(不适用)
为用户查找玩过的所有游戏	按游戏开始时间戳顺序列出用户玩过的所有游戏。	低	读取	多项	用户名和游戏 ID	用户名 = 当前用户	开始时间戳。

导出数据 进行数据 分析	开发团队 运行批处 理作业， 将数据 导出到 Amazon S3。	低	读取	全部	不适用	不适用	(不适 用)
--------------------	---	---	----	----	-----	-----	-------------

最佳实践

考虑使用以下 DynamoDB 设计最佳实践：

- [分区键设计](#) - 使用高基数分区键均匀分配负载。
- [邻接列表设计模式](#) - 使用此设计模式管理一对多和多对多关系。
- [稀疏索引](#) - 对全局二级索引 (GSI) 使用稀疏索引。创建 GSI 时，您可以指定分区键和排序键 (可选)。只有基表中包含相应 GSI 分区键的项目才会显示在稀疏索引中。这有助于缩小 GSI。
- [索引过载](#) - 使用相同的 GSI 为各种类型的项目编制索引。
- [GSI 写入分片](#) - 明智地分片，将数据分布到各个分区，以实现更高效、更快速的查询。
- [大项目](#) - 仅在表中存储元数据，将 blob 保存在 Amazon S3 中，将引用保留在 DynamoDB 中。将大项目分成多个项目，并使用排序键有效地编制索引。

有关更多设计最佳实践，请参阅 [Amazon DynamoDB 文档](#)。

分层数据建模示例

以下各节以汽车公司为例，说明如何使用数据建模流程的步骤在 DynamoDB 中设计多级组件管理系统。

主题

- [第 1 步：确定使用案例和逻辑数据模型](#)
- [第 2 步：创建初步的成本估算](#)
- [第 3 步：识别数据访问模式](#)
- [第 4 步：确定技术要求](#)
- [第 5 步：创建 DynamoDB 数据模型](#)
- [第 6 步：创建数据查询](#)
- [第 7 步：验证数据模型](#)
- [第 8 步：查看成本估算信息](#)
- [第 9 步：部署数据模型](#)

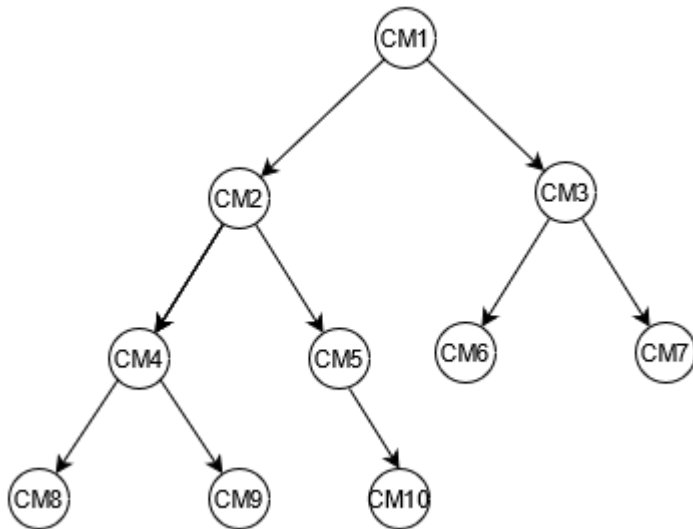
第 1 步：确定使用案例和逻辑数据模型

某汽车公司想要构建一个事务性组件管理系统，用于存储和搜索所有可用的汽车零件，并在不同的组件和零件之间建立关系。例如，一辆汽车包含多个电池，每个电池包含多个高级模块，每个模块包含多个单元，每个单元包含多个低级组件。

通常，对于构建层次关系模型，[Amazon Neptune](#) 之类的图形数据库是更好的选择。但在某些情况下，Amazon DynamoDB 因其灵活性、安全性、性能和可扩展性而成为分层数据建模的更好选择。

例如，您可以构建一个系统，其中 80-90% 的查询是事务性的，那么 DynamoDB 就非常适合。在本例中，另外 10-20% 的查询是关系型的，诸如 Neptune 之类的图形数据库更适合。这种情况下，在架构中增加一个额外的数据库，只完成 10-20% 的查询，可能导致成本增加。这还增加了维护多个系统并同步数据的运营负担。您可以在 DynamoDB 中对 10-20% 的关系查询进行建模。

绘制汽车组件的示例树有助于映射它们之间的关系。下图显示了一个包含四个级别的依赖关系图。CM1 是示例汽车本身的顶级组件。它有两个子组件，用于两个示例电池，CM2 和 CM3。每个电池都有两个子组件，即模块。CM2 有模块 CM4 和 CM5，CM3 有模块 CM6 和 CM7。每个模块有几个子组件，即单元。该 CM4 模块有两个单元，CM8 和 CM9。CM5 有一个细胞，CM10。CM6 而且还 CM7 没有任何关联的细胞。



本指南将使用此树及其组件标识符作为参考。顶部组件称为父级，子组件称为子级。例如，顶部组件 CM1 是 CM2 和 CM3 的父组件。CM2 是 CM4 和 CM5 的父项。这描绘了父子关系图。

从树状图中，您可以看到组件的完整依赖关系图。例如，依赖 CM8 于 CM4，依赖于 CM2，依赖于 CM1。树将完整的依赖关系图定义为路径。路径描述了两点：

- 依赖关系图
- 树中的位置

填写业务需求模板：

提供有关您的用户的信息：

User	描述
Employee	需要汽车及其组件信息的汽车公司内部员工

提供有关数据来源和数据摄取方式的信息：

源	描述	User
管理系统	此系统将存储与可用汽车零件及其与其他组件和零件的关系有关的所有数据。	Employee

提供有关如何使用数据的信息：

使用者	描述	User
管理系统	检索父组件 ID 的所有直接子组件。	Employee
管理系统	检索组件 ID 的所有子组件的递归列表。	Employee
管理系统	请参阅组件的原级。	Employee

第 2 步：创建初步的成本估算

估算应用程序所有环境的成本很重要，这样就可以检查解决方案在财务上是否可行。最佳做法是在继续开发和部署之前，先大致估算并取得业务分析师的批准。

- 数据库工程师使用 [DynamoDB 定价页面](#) 上提供的可用信息和示例创建初始成本分析。
 - 为按需容量创建成本估算（请参阅[示例](#)）。
 - 为预置容量创建成本估算（请参阅[示例](#)）。
 - 对于预置容量模型，请从计算器中获取估算成本，然后对预留容量应用折扣。
 - 比较两种容量模型的估计成本。
 - 创建所有环境（开发、生产、质量保证）的成本估算。
- 业务分析师审查并批准或拒绝初步成本估算。

您可以使用这些参考值，预估价格并提交审批。要创建预算，您可以使用 [DynamoDB 定价页面](#) 和 [AWS 定价计算器](#)

第 3 步：识别数据访问模式

此示例使用案例具有以下访问模式，用于管理不同汽车组件之间的关系。

访问模式	优先级	读或写	描述	Type	筛选条件	结果排序
直接子组件	高	读取	检索父组件 ID 的所	多个	Component ID	不适用

			有直接子组件。			
所有子组件	高	读取	检索组件 ID 的所有子组件的递归列表。	多个	Component ID	不适用
原级	高	读取	检索组件的原级。	多个	Component ID	不适用

第 4 步：确定技术要求

此示例没有具体的技术要求，这些超出了此示例的范围。在实际案例中，最佳实践是在开发和部署之前完成此步骤，并验证是否所有技术要求都已得到满足。可以使用[示例问卷](#)来完成业务案例中的此步骤。此外，我们建议验证 [DynamoDB 服务配额](#)，以确保您设计的解决方案没有硬性限制。

第 5 步：创建 DynamoDB 数据模型

为基表和全局二级索引定义分区键 (GSIs)：

- 遵循密钥设计最佳实践，在本示例中 ComponentId 用作基表的分区键。ComponentId 可以提供粒度，因其是唯一的。DynamoDB 使用分区键的哈希值来确定物理存储数据的分区。唯一的组件 ID 会生成不同的哈希值，这有助于在表内分布数据。可以使用 ComponentId 分区键查询基表。
- 要查找组件的直接子组件，请创建一个 GSI，其中 ParentId 是分区键，ComponentId 是排序键。您可以将 ParentId 用作分区键，以查询此 GSI。
- 要查找组件的所有递归子级，请创建一个 GSI，其中 GraphId 是分区键，Path 是排序键。您可以将 GraphId 用作分区键和排序键上的 BEGINS_WITH(Path, "\$path") 运算符来查询此 GSI。

	分区键	排序键	映射属性
基表	ComponentId		ParentId, GraphId, Path
GSI1	ParentId	ComponentId	

GSi2	GraphId	Path	ComponentId
------	---------	------	-------------

在表格中存储组件

下一步是将每个组件存储在 DynamoDB 基表中。插入示例树中的所有组件后，您将得到如下基表。

ComponentId	ParentId	GraphId	路径
CM1		CM1#1	CM1
CM2	CM1	CM1#1	CM1 CM2
CM3	CM1	CM1#1	CM1 CM3
CM4	CM2	CM1#1	CM1 CM2 CM4
CM5	CM2	CM1#1	CM1 CM2 CM5
CM6	CM3	CM1#1	CM1 CM3 CM6
CM7	CM3	CM1#1	CM1 CM3 CM7
CM8	CM4	CM1#1	CM1 CM2 CM4 CM8

CM9	CM4	CM1#1	CM1 CM2 CM4 CM9
CM10	CM5	CM1#1	CM1 CM2 CM5 CM10

该 GSI1 指数

要检查组件的所有直接子组件，需要创建将 ParentId 用作分区键、将 ComponentId 用作排序键的索引。以下数据透视表表示该 GSI1 指数。您可以使用此索引通过父组件 ID 检索所有直接子组件。例如，您可以找出汽车中有多少可用电池 (CM1) 或模块 (CM4) 中有哪些电池可用。

ParentId	ComponentId
CM1	CM2
	CM3
	CM4
CM2	CM5
	CM6
CM3	CM7
	CM8
CM4	CM9
	CM10

该 GSI2 指数

以下数据透视表表示该 GSI2 指数。它配置为将 GraphId 用作分区键，将 Path 用作排序键。使用 GraphId 并对排序键 (Path) 执行 begins_with 操作，您可以在树中找到组件的完整谱系。

GraphId	路径	ComponentId
---------	----	-------------

CM1#1	CM1	CM1
	CM1 CM2	CM2
	CM1 CM3	CM3
	CM1 CM2 CM4	CM4
	CM1 CM2 CM5	CM5
	CM1 CM2 CM4 CM8	CM8
	CM1 CM2 CM4 CM9	CM9
	CM1 CM2 CM5 CM10	CM10
	CM1 CM3 CM6	CM6
	CM1 CM3 CM7	CM7

第 6 步：创建数据查询

定义访问模式并设计数据模型后，可以在 DynamoDB 数据库中查询层次数据。作为节省成本和确保性能的最佳实践，以下示例仅使用无 Scan 的查询操作。

- 查找组件的祖先。

要查找 CM8 组件的祖先（父母、祖父母、曾祖父母等），请使用查询基表。ComponentId = "CM8" 查询将返回以下记录。

要减小结果数据的大小，您可以使用投影表达式仅返回 Path 属性。

ComponentId	ParentId	GraphId	路径
CM8	CM4	CM1#1	CM1 CM2 CM4 CM8

路径

CM1|CM2|CM4|CM8

现在，使用管道 ("|") 分割路径，然后取前 N-1 个组件来获取原级。

查询结果：的祖先 CM8 是 CM1、CM2、CM4。

- 查找组件的直系子组件。

要获取该组件的所有直接子组件或下游单级组 CM2 件，请 GSI1 使用 ParentId = "CM2" 查询。查询将返回以下记录。

ParentId	ComponentId
CM2	CM4
	CM5

- 使用顶部组件查找所有下游子组件。

要获取顶级组件的所有子组件或下游组件 CM1，请 GSI2 使用 GraphId = "CM1#1" 和进行查询 begins_with("Path", "CM1|")，然后使用带的投影表达式 ComponentId。它将返回与该树相关的所有组件。

这个例子只有一棵树，CM1 作为顶部组件。实际上，同一个表中可能有数百万个顶部组件。

GraphId	ComponentId
	CM2
CM1#1	CM3
	CM4
	CM5
	CM8
	CM9
	CM10

CM6

CM7

- 使用中间组件查找所有下游子组件。

要递归获取组件的所有子组件或下游组件 CM2，您有两个选择。您可以逐级递归查询，也可以查询 GSI2 索引。

- 逐级递归查询 GSI1，直到达到子组件的最后一级。

1. GSI1 使用查询 `ParentId = "CM2"`。这将返回以下记录。

ParentId	ComponentId
CM2	CM4
	CM5

2. 再说一遍，GSI1 使用查询 `ParentId = "CM4"`。这将返回以下记录。

ParentId	ComponentId
CM4	CM8
	CM9

3. 再说一遍，GSI1 使用查询 `ParentId = "CM5"`。这将返回以下记录。

继续循环：查询每个 `ComponentId`，直至到达最后一级。当使用 `ParentId = "<ComponentId>"` 查询未返回任何结果时，上一个结果来自树的最后一级。

ParentId	ComponentId
CM5	CM10

4. 合并所有结果。

结果 = [CM4, CM5] + [,] + [CM8 CM9] + [] CM10

=[CM4, CM5, CM8, CM9, CM10]

- 查询 GSI2，它存储顶级组件（汽车或 CM1）的分层树。
 1. 首先，找到顶级组件或顶级祖先和 Path。CM2 为此，使用 ComponentId = "CM2" 查询基表，在层次树中找到该组件的路径。选择 GraphId 和 Path 属性。查询将返回以下记录。

GraphId	路径
CM1#1	CM1 CM2

2. GSI2 使用查询 GraphId = "CM1#1" AND BEGINS_WITH("Path", "CM1|CM2|")。查询将返回以下结果。

GraphId	路径	ComponentId
CM1#1	CM1 CM2 CM4	CM4
	CM1 CM2 CM5	CM5
	CM1 CM2 CM4 CM8	CM8
	CM1 CM2 CM4 CM9	CM9
	CM1 CM2 CM5 CM10	CM10

3. 选择要返回其所有子组件的 ComponentId 属性 CM2。

第 7 步：验证数据模型

在这一步，业务用户需验证查询结果并检查它们是否满足业务需求。可以使用下表根据用户的要求检查访问模式。

问题	基表/GSI	Query
作为用户，我想检索父组件 ID 的所有直接子组件。	GSI1	ParentId = "<ComponentId>" (查找组件的直接子组件。)

作为用户，我想检索组件 ID 的 GSI1 或者 GSI2 所有子组件的递归列表。

```
GSI1: ParentId =
"<ComponentId>"
```

或者

```
GSI2: GraphId =
"<TopLevelComponentId>#N" AND BEGINS_WITH("Path", "<PATH_OF_Component>")
```

(使用顶部组件查找所有下级子组件，使用中间组件查找所有下级子组件。)

作为用户，我想查看原级组件。 基表

```
ComponentId =
"<ComponentId>" ，然后选择 Path 属性。
```

(查找组件的原级。)

还可以使用任何编程语言实施脚本（测试），以直接查询 DynamoDB 并将结果与预期结果进行比较。

第 8 步：查看成本估算信息

再次查看并完善成本估算信息。此外，最好能让业务利益相关者审核确认，获得批准后再推进至下一步。

目标

- 定义容量模型并估算 DynamoDB 成本，以完善[第 2 步](#)中的成本估算信息。
- 获得业务分析师和利益相关者的最终财务审批。

流程

- 数据库工程师确定数据量的估计值。

- 数据库工程师确定数据传输要求。
- 数据库工程师定义所需的读取和写入容量单位。
- 业务分析师在[按需容量模式和预置容量模式](#)之间做出决定。
- 数据库工程师确定了 [DynamoDB 自动扩缩](#)的需求。
- 数据库工程师在 AWS 定价计算器中输入参数。
- 数据库工程师向业务利益相关者提交最终的价格估算。
- 业务分析师和利益相关者批准或拒绝解决方案。

第 9 步：部署数据模型

在这个具体示例中，模型是使用 [NoSQL Workbench](#) 部署的，NoSQL Workbench 是用于现代数据库开发和操作的应用程序。使用此工具，您可以选择创建数据模型、上传数据并将其直接部署到您的 AWS 账户。如果要实现此示例，则可以使用以下由 NoSQL Workbench 生成的 AWS CloudFormation 模板。

```
AWSTemplateFormatVersion: 2010-09-09
Resources:
  Components:
    Type: 'AWS::DynamoDB::Table'
    Properties:
      KeySchema:
        - AttributeName: ComponentId
          KeyType: HASH
      AttributeDefinitions:
        - AttributeName: ComponentId
          AttributeType: S
        - AttributeName: ParentId
          AttributeType: S
        - AttributeName: GraphId
          AttributeType: S
        - AttributeName: Path
          AttributeType: S
      GlobalSecondaryIndexes:
        - IndexName: GS1
          KeySchema:
            - AttributeName: ParentId
              KeyType: HASH
            - AttributeName: ComponentId
              KeyType: RANGE
```

```
Projection:
  ProjectionType: KEYS_ONLY
- IndexName: GSI2
  KeySchema:
    - AttributeName: GraphId
      KeyType: HASH
    - AttributeName: Path
      KeyType: RANGE
  Projection:
    ProjectionType: INCLUDE
    NonKeyAttributes:
      - ComponentId
BillingMode: PAY_PER_REQUEST
TableName: Components
```

其他资源

有关 DynamoDB 的更多信息

- [DynamoDB 定价](#)
- [DynamoDB 文档](#)
- [适用于 DynamoDB 的 NoSQL 设计](#)
- [写入分片](#)
- [本地二级索引 \(LSI\)](#)
- [全局二级索引 \(GSI\)](#)
- [重载 GSI](#)
- [GSI 分片](#)
- [使用 GSI 创建最终一致的副本](#)
- [稀疏索引](#)
- [物化聚合查询](#)
- [时间序列设计模式](#)
- [相邻列表设计模式](#)
- [按需和预置容量模式](#)
- [DynamoDB 自动扩缩](#)
- [DynamoDB 生存时间 \(TTL\)](#)
- [使用 DynamoDB 对游戏玩家数据进行建模 \(实验室\)](#)

AWS 服务

- [AWS CloudFormation](#)
- [Amazon S3 \(\)](#)

工具

- [AWS 定价计算器](#)
- [NoSQL Workbench for DynamoDB](#)
- [DynamoDB Local](#)

- [DynamoDB 和 AWS SDK](#)

最佳实践

- [使用 DynamoDB 进行设计和架构的最佳实践](#) (DynamoDB 文档)
- [使用二级索引的最佳实践](#) (DynamoDB 文档)
- [存储大型项目和属性的最佳实践](#) (DynamoDB 文档)
- [选择正确的 DynamoDB 分区键](#) (AWS 数据库博客)
- [如何设计 Amazon DynamoDBglobal 二级索引](#) (AWS 数据库博客)
- [适用于 Amazon DynamoDB 的 NoSQL Workbench 有哪些方面](#) (Medium 网站)

AWS 一般资源

- [AWS 规范指引网站](#)
- [AWS 文档](#)
- [AWS 一般参考](#)

贡献者

本指南的贡献者包括：

- Camilo Gonzalez，高级数据架构师，AWS
- Moinul Al-Mamun，高级大数据架构师，AWS
- Santiago Segura，专业服务顾问，AWS
- Satheish Kumar Chandraprakasam，云应用程序架构师，AWS

文档历史记录

下表介绍了本指南的一些重要更改。如果您希望收到有关未来更新的通知，可以订阅 [RSS 源](#)。

变更	说明	日期
添加了最佳实践部分和分层数据建模示例。	我们添加了 DynamoDB 最佳实践摘要以及设计和 step-by-step 验证分层模型的示例。	2023 年 12 月 5 日
初次发布	—	2020 年 10 月 26 日

AWS 规范性指导词汇表

以下是 AWS 规范性指导提供的策略、指南和模式中的常用术语。若要推荐词条，请使用术语表末尾的提供反馈链接。

数字

7 R

将应用程序迁移到云中的 7 种常见迁移策略。这些策略以 Gartner 于 2011 年确定的 5 R 为基础，包括以下内容：

- **重构/重新架构**：充分利用云原生功能来提高敏捷性、性能和可扩展性，以迁移应用程序并修改其架构。这通常涉及到移植操作系统和数据库。示例：将本地 Oracle 数据库迁移到 Amazon Aurora PostgreSQL 兼容版。
- **更换平台**：将应用程序迁移到云中，并进行一定程度的优化，以利用云功能。示例：将本地 Oracle 数据库迁移到 AWS Cloud 中的 Amazon Relational Database Service (Amazon RDS) for Oracle。
- **重新购买**：转换到其他产品，通常是从传统许可转向 SaaS 模式。示例：将客户关系管理 (CRM) 系统迁移到 Salesforce.com。
- **重新托管 (直接迁移)**：将应用程序迁移到云中，无需进行任何更改即可利用云功能。示例：将本地 Oracle 数据库迁移到 AWS Cloud 中 EC2 实例上的 Oracle。
- **重新放置 (虚拟机监控器级直接迁移)**：将基础设施迁移到云中，无需购买新硬件、重写应用程序或修改现有操作。您将服务器从本地平台迁移到同一平台的云服务中。示例：将 Microsoft Hyper-V 应用程序迁移到 AWS。
- **保留 (重访)**：将应用程序保留在源环境中。其中可能包括需要进行重大重构的应用程序，并且您希望将工作推迟到以后，以及您希望保留的遗留应用程序，因为迁移它们没有商业上的理由。
- **停用**：停用或删除源环境中不再需要的应用程序。

A

ABAC

请参阅[基于属性的访问控制](#)。

抽象服务

请参阅[托管服务](#)。

ACID

请参阅[原子性、一致性、隔离性、持久性](#)。

主动-主动迁移

一种数据库迁移方法，在这种方法中，源数据库和目标数据库保持同步（通过使用双向复制工具或双写操作），两个数据库都在迁移期间处理来自连接应用程序的事务。这种方法支持小批量、可控的迁移，而不需要一次性割接。它比[主动-被动迁移](#)更灵活，但工作量更大。

主动-被动迁移

一种数据库迁移方法，在这种方法中，源数据库和目标数据库保持同步，但在将数据复制到目标数据库时，只有源数据库处理来自连接应用程序的事务。目标数据库在迁移期间不接受任何事务。

聚合函数

一种 SQL 函数，它对一组行进行操作并计算该组的单个返回值。聚合函数的示例包括 SUM 和 MAX。

AI

请参阅[人工智能](#)。

AIOps

请参阅[人工智能运营](#)。

匿名化

永久删除数据集中个人信息的过程。匿名化可以帮助保护个人隐私。匿名化数据不再被视为个人数据。

反模式

一种用于解决反复出现的问题的常用解决方案，而在这类问题中，此解决方案适得其反、无效或不如替代方案有效。

应用程序控制

一种安全方法，仅允许使用经批准的应用程序，以帮助保护系统免受恶意软件的侵害。

应用程序组合

有关组织使用的每个应用程序的详细信息的集合，包括构建和维护该应用程序的成本及其业务价值。这些信息是[产品组合发现和分析过程](#)的关键，有助于识别需要进行迁移、现代化和优化的应用程序并确定其优先级。

人工智能 (AI)

计算机科学领域致力于使用计算技术执行通常与人类相关的认知功能，例如学习、解决问题和识别模式。有关更多信息，请参阅[什么是人工智能？](#)

人工智能操作 (AIOps)

使用机器学习技术解决运营问题、减少运营事故和人为干预以及提高服务质量的过程。有关如何在 AIOps AWS 迁移策略中使用的更多信息，请参阅[操作集成指南](#)。

非对称加密

一种加密算法，使用一对密钥，一个公钥用于加密，一个私钥用于解密。您可以共享公钥，因为它不用于解密，但对私钥的访问应受到严格限制。

原子性、一致性、隔离性、持久性 (ACID)

一组软件属性，即使在出现错误、电源故障或其他问题的情况下，也能保证数据库的数据有效性和操作可靠性。

基于属性的访问权限控制 (ABAC)

根据用户属性（如部门、工作角色和团队名称）创建精细访问权限的做法。有关更多信息，请参阅 AWS Identity and Access Management (IAM) [文档](#) [AWS 中的 AB AC](#)。

权威数据来源

存储主要数据版本的位置，被认为是最可靠的信息源。您可以将数据从权威数据来源复制到其他位置，以便处理或修改数据，例如对数据进行匿名化、编辑或假名化。

可用区

中的一个不同位置 AWS 区域，不受其他可用区域故障的影响，并向同一区域中的其他可用区提供低成本、低延迟的网络连接。

AWS 云采用框架 (AWS CAF)

该框架包含指导方针和最佳实践 AWS，可帮助组织制定高效且有效的计划，以成功迁移到云端。AWS CAF 将指导分为六个重点领域，称为视角：业务、人员、治理、平台、安全和运营。业务、人员和治理角度侧重于业务技能和流程；平台、安全和运营角度侧重于技术技能和流程。例如，人

员角度针对的是负责人力资源 (HR)、人员配置职能和人员管理的利益相关者。从这个角度来看，AWS CAF 为人员发展、培训和沟通提供了指导，以帮助组织为成功采用云做好准备。有关更多信息，请参阅 [AWS CAF 网站](#) 和 [AWS CAF 白皮书](#)。

AWS 工作负载资格框架 (AWS WQF)

一种评估数据库迁移工作负载、推荐迁移策略和提供工作估算的工具。AWS WQF 包含在 AWS Schema Conversion Tool (AWS SCT) 中。它用来分析数据库架构和代码对象、应用程序代码、依赖关系和性能特征，并提供评测报告。

B

恶意机器人

一种旨在扰乱或伤害个人或组织的[机器人](#)。

BCP

请参阅[业务连续性计划](#)。

行为图

一段时间内资源行为和交互的统一交互式视图。您可以使用 Amazon Detective 的行为图来检查失败的登录尝试、可疑的 API 调用和类似的操作。有关更多信息，请参阅 Detective 文档中的[行为图中的数据](#)。

大端序系统

一个先存储最高有效字节的系统。另请参阅[字节顺序](#)。

二进制分类

一种预测二进制结果 (两个可能的类别之一) 的过程。例如，您的 ML 模型可能需要预测诸如“该电子邮件是否为垃圾邮件？”或“这个产品是书还是汽车？”之类的问题

bloom 筛选条件

一种概率性、内存高效的数据结构，用于测试元素是否为集合的成员。

蓝/绿部署

一种部署策略，您可以创建两个独立但完全相同的环境。在一个环境中运行当前应用程序版本 (蓝色)，在另一个环境中运行新应用程序版本 (绿色)。此策略可帮助您在影响最小的情况下快速回滚。

自动程序

一种通过互联网运行自动任务并模拟人类活动或交互的软件应用程序。有些机器人是有用或有益的，例如在互联网上索引信息的 Web 爬网程序。还有一些被称为恶意机器人的机器人，其目的是扰乱或伤害个人或组织。

僵尸网络

被[恶意软件](#)感染并受单方（称为僵尸网络控制者或僵尸网络操作者）控制的[僵尸网络](#)。僵尸网络是最著名的扩展机器人及其影响力的机制。

分支

代码存储库的一个包含区域。在存储库中创建的第一个分支是主分支。您可以从现有分支创建新分支，然后在新分支中开发功能或修复错误。为构建功能而创建的分支通常称为功能分支。当功能可以发布时，将功能分支合并回主分支。有关更多信息，请参阅[关于分支](#)（GitHub 文档）。

紧急（break-glass）访问

在特殊情况下，通过批准的流程，用户 AWS 账户 可以快速访问他们通常没有访问权限的内容。有关更多信息，请参阅 AWS Well-Architected Guidance 中的 [Implement break-glass procedures](#) 指示器。

棕地策略

您环境中的现有基础设施。在为系统架构采用棕地策略时，您需要围绕当前系统和基础设施的限制来设计架构。如果您正在扩展现有基础设施，则可以将棕地策略和[全新](#)策略混合。

缓冲区缓存

存储最常访问的数据的内存区域。

业务能力

企业如何创造价值（例如，销售、客户服务或营销）。微服务架构和开发决策可以由业务能力驱动。有关更多信息，请参阅[在 AWS 上运行容器化微服务](#)白皮书中的[围绕业务能力进行组织](#)部分。

业务连续性计划（BCP）

一项计划，旨在应对大规模迁移等破坏性事件对运营的潜在影响，并使企业能够快速恢复运营。

C

CAF

请参阅 [AWS 云采用框架](#)。

金丝雀部署

缓慢而渐进地向最终用户发布版本。当您确信无误后，即可部署新版本，并完全替换当前版本。

CCoE

请参阅[云卓越中心](#)。

CDC

请参阅[更改数据捕获](#)。

更改数据捕获 (CDC)

跟踪数据来源 (如数据库表) 的更改并记录有关更改的元数据的过程。您可以将 CDC 用于各种目的，例如审计或复制目标系统中的更改以保持同步。

混沌工程

故意引入故障或破坏性事件来测试系统的韧性。您可以使用 [AWS Fault Injection Service \(AWS FIS\)](#) 来执行实验，对您的 AWS 工作负载施加压力并评估其响应。

CI/CD

请参阅[持续集成和持续交付](#)。

分类

一种有助于生成预测的分类流程。分类问题的 ML 模型预测离散值。离散值始终彼此不同。例如，一个模型可能需要评估图像中是否有汽车。

客户端加密

在目标 AWS 服务 收到数据之前，对数据进行本地加密。

云卓越中心 (CCoE)

一个多学科团队，负责推动整个组织的云采用工作，包括开发云最佳实践、调动资源、制定迁移时间表、领导组织完成大规模转型。有关更多信息，请参阅 AWS Cloud 企业战略博客上的 [CCoE 帖子](#)。

云计算

通常用于远程数据存储和 IoT 设备管理的云技术。云计算通常连接到[边缘计算](#)技术。

云运营模型

在 IT 组织中，一种用于构建、完善和优化一个或多个云环境的运营模型。有关更多信息，请参阅[构建您的云运营模型](#)。

云采用阶段

组织迁移到 AWS Cloud 中时通常会经历四个阶段：

- 项目 - 出于概念验证和学习目的，开展一些与云相关的项目
- 基础 — 进行基础投资以扩大云采用率（例如，创建着陆区、定义 CCo E、建立运营模型）
- 迁移 - 迁移单个应用程序
- 重塑 - 优化产品和服务，在云中创新

Stephen Orban 在 AWS Cloud 企业战略博客的博客文章 [《云优先之旅和采用阶段》](#) 中定义了这些阶段。有关它们与 AWS 迁移策略的关系的信息，请参阅 [迁移准备指南](#)。

CMDB

请参阅 [配置管理数据库](#)。

代码存储库

通过版本控制过程存储和更新源代码和其他资产（如文档、示例和脚本）的位置。常见的云存储库包括 GitHub 或 Bitbucket Cloud。每个版本的代码都称为一个分支。在微服务结构中，每个存储库都专门用于一个功能。单个 CI/CD 管线可以使用多个存储库。

冷缓存

一种空的、填充不足或包含过时或不相关数据的缓冲区缓存。这会影响性能，因为数据库实例必须从主内存或磁盘读取，这比从缓冲区缓存读取要慢。

冷数据

很少访问的数据，且通常是历史数据。查询此类数据时，通常可以接受慢速查询。将这些数据转移到性能较低且成本更低的存储层或类别可以降低成本。

计算机视觉 (CV)

一种 [AI](#) 领域，它使用机器学习来分析和提取数字图像和视频等视觉格式中的信息。例如，Amazon SageMaker AI 为 CV 提供了图像处理算法。

配置偏移

对于工作负载而言，一种偏离预期状态的配置更改。这可能会导致工作负载变得不合规，且通常是渐进的，不是故意的。

配置管理数据库 (CMDB)

一种存储库，用于存储和管理有关数据库及其 IT 环境的信息，包括硬件和软件组件及其配置。您通常在迁移的产品组合发现和分析阶段使用来自 CMDB 的数据。

合规性包

一系列 AWS Config 规则和补救措施，您可以汇编这些规则和补救措施，以自定义您的合规性和安全性检查。您可以使用 YAML 模板将一致性包作为单个实体部署在 AWS 账户 和区域或整个组织中。有关更多信息，请参阅 AWS Config 文档中的[一致性包](#)。

持续集成和持续交付 (CI/CD)

自动执行软件发布过程的源代码、构建、测试、暂存和生产阶段的过程。CI/CD 通常被描述为管道。CI/CD 可以帮助您实现流程自动化、提高生产力、提高代码质量和更快地交付。有关更多信息，请参阅[持续交付的优势](#)。CD 也可以表示持续部署。有关更多信息，请参阅[持续交付与持续部署](#)。

CV

请参阅[计算机视觉](#)。

D

静态数据

网络中静止的数据，例如存储中的数据。

数据分类

根据网络中数据的关键性和敏感性对其进行识别和分类的过程。它是任何网络安全风险管理策略的关键组成部分，因为它可以帮助您确定对数据的适当保护和保留控制。数据分类是 Well-Architected AWS d Framework 中安全支柱的一个组成部分。有关详细信息，请参阅[数据分类](#)。

数据漂移

生产数据与用来训练机器学习模型的数据之间的有意义差异，或者输入数据随时间推移的有意义变化。数据漂移可能降低机器学习模型预测的整体质量、准确性和公平性。

传输中数据

在网络中主动移动的数据，例如在网络资源之间移动的数据。

数据网格

一种架构框架，可提供分布式、去中心化的数据所有权以及集中式管理和治理。

数据最少化

仅收集并处理绝对必要数据的原则。在中进行数据最小化 AWS Cloud 可以降低隐私风险、成本和分析碳足迹。

数据边界

AWS 环境中的一组预防性防护措施，可帮助确保只有可信身份才能访问来自预期网络的可信资源。有关更多信息，请参阅在[上构建数据边界](#)。AWS

数据预处理

将原始数据转换为 ML 模型易于解析的格式。预处理数据可能意味着删除某些列或行，并处理缺失、不一致或重复的值。

数据溯源

在数据的整个生命周期跟踪其来源和历史的过程，例如数据如何生成、传输和存储。

数据主体

正在收集和处理其数据的人。

数据仓库

一种支持商业智能（例如分析）的数据管理系统。数据仓库通常包含大量历史数据，通常用于查询和分析。

数据库定义语言（DDL）

在数据库中创建或修改表和对象结构的语句或命令。

数据库操作语言（DML）

在数据库中修改（插入、更新和删除）信息的语句或命令。

DDL

请参阅[数据库定义语言](#)。

深度融合

组合多个深度学习模型进行预测。您可以使用深度融合来获得更准确的预测或估算预测中的不确定性。

深度学习

一个 ML 子字段使用多层神经网络来识别输入数据和感兴趣的目标变量之间的映射。

defense-in-depth

一种信息安全方法，经过深思熟虑，在整个计算机网络中分层实施一系列安全机制和控制措施，以保护网络及其中数据的机密性、完整性和可用性。当你采用这种策略时 AWS，你会在 AWS

Organizations 结构的不同层面添加多个控件来帮助保护资源。例如，一种 defense-in-depth 方法可以结合多因素身份验证、网络分段和加密。

委派管理员

在中 AWS Organizations，兼容的服务可以注册 AWS 成员帐户来管理组织的帐户并管理该服务的权限。此帐户被称为该服务的委托管理员。有关更多信息和兼容服务列表，请参阅 AWS Organizations 文档中[使用 AWS Organizations 的服务](#)。

部署

使应用程序、新功能或代码修复在目标环境中可用的过程。部署涉及在代码库中实现更改，然后在应用程序的环境中构建和运行该代码库。

开发环境

请参阅[环境](#)。

侦测性控制

一种安全控制，在事件发生后进行检测、记录日志和发出提醒。这些控制是第二道防线，提醒您注意绕过现有预防性控制的安全事件。有关更多信息，请参阅在 AWS 上实施安全控制中的[侦测性控制](#)。

开发价值流映射 (DVSM)

用于识别对软件开发生命周期中的速度和质量产生不利影响的限制因素并确定其优先级的流程。DVSM 扩展了最初为精益生产实践设计的价值流映射流程。其重点关注在软件开发过程中创造和转移价值所需的步骤和团队。

数字孪生

真实世界系统的虚拟再现，如建筑物、工厂、工业设备或生产线。数字孪生支持预测性维护、远程监控和生产优化。

维度表

[星型架构](#)中的一种较小的表，其中包含事实表中定量数据的数据属性。维度表属性通常是文本字段或行为类似于文本的离散数字。这些属性通常用于查询约束、筛选和结果集标注。

灾难

阻止工作负载或系统在其主要部署位置实现其业务目标的事件。这些事件可能是自然灾害、技术故障或人为操作的结果，例如无意的配置错误或恶意软件攻击。

灾难恢复 (DR)

您用来最大程度地减少由[灾难](#)造成的停机时间和数据丢失的策略和流程。有关更多信息，请参阅 Well-Architected Framework AWS work 中的“[工作负载灾难恢复：云端 AWS 恢复](#)”。

DML

请参阅[数据库操作语言](#)。

领域驱动设计

一种开发复杂软件系统的方法，通过将其组件连接到每个组件所服务的不断发展的领域或核心业务目标。Eric Evans 在其著作[领域驱动设计：软件核心复杂性应对之道](#) (Boston: Addison-Wesley Professional, 2003) 中介绍了这一概念。有关如何将领域驱动设计与 strangler fig 模式结合使用的信息，请参阅[使用容器和 Amazon API Gateway 逐步将原有的 Microsoft ASP.NET \(ASMX \) Web 服务现代化](#)。

DR

请参阅[灾难恢复](#)。

偏差检测

跟踪与基准配置的偏差。例如，您可以使用 AWS CloudFormation 来[检测系统资源中的偏差](#)，也可以使用 AWS Control Tower 来[检测着陆区中可能影响监管要求合规性的变化](#)。

DVSM

请参阅[开发价值流映射](#)。

E

EDA

请参阅[探索性数据分析](#)。

EDI

请参阅[电子数据交换](#)。

边缘计算

该技术可提高位于 IoT 网络边缘的智能设备的计算能力。与[云计算](#)比较时，边缘计算可以减少通信延迟并缩短响应时间。

电子数据交换 (EDI)

组织之间业务文件的自动交换。有关更多信息，请参阅[什么是电子数据交换](#)。

加密

一种将人类可读的纯文本数据转换为加密文字的计算流程。

加密密钥

由加密算法生成的随机位的加密字符串。密钥的长度可能有所不同，而且每个密钥都设计为不可预测且唯一。

字节顺序

字节在计算机内存中的存储顺序。大端序系统先存储最高有效字节。小端序系统先存储最低有效字节。

端点

请参阅[服务端点](#)。

端点服务

一种可以在虚拟私有云 (VPC) 中托管，与其他用户共享的服务。您可以使用其他 AWS 账户 或 AWS Identity and Access Management (IAM) 委托人创建终端节点服务，AWS PrivateLink 并向其授予权限。这些账户或主体可通过创建接口 VPC 端点来私密地连接到您的端点服务。有关更多信息，请参阅 Amazon Virtual Private Cloud (Amazon VPC) 文档中的[创建端点服务](#)。

企业资源规划 (ERP)

一种自动化和管理企业关键业务流程 (例如会计、[MES](#) 和项目管理) 的系统。

信封加密

用另一个加密密钥对加密密钥进行加密的过程。有关更多信息，请参阅 AWS Key Management Service (AWS KMS) 文档中的[信封加密](#)。

环境

正在运行的应用程序的实例。以下是云计算中常见的环境类型：

- 开发环境 — 正在运行的应用程序的实例，只有负责维护应用程序的核心团队才能使用。开发环境用于测试更改，然后再将其提升到上层环境。这类环境有时称为测试环境。
- 下层环境 — 应用程序的所有开发环境，比如用于初始构建和测试的环境。

- 生产环境 — 最终用户可以访问的正在运行的应用程序的实例。在 CI/CD 管道中，生产环境是最后一个部署环境。
- 上层环境 — 除核心开发团队以外的用户可以访问的所有环境。这可能包括生产环境、预生产环境和用户验收测试环境。

epic

在敏捷方法学中，有助于组织工作和确定优先级的功能类别。epics 提供了对需求和实施任务的总体描述。例如，AWS CAF 安全史诗包括身份和访问管理、侦探控制、基础设施安全、数据保护和事件响应。有关 AWS 迁移策略中 epics 的更多信息，请参阅[计划实施指南](#)。

ERP

请参阅[企业资源规划](#)。

探索性数据分析 (EDA)

分析数据集以了解其主要特征的过程。您收集或汇总数据，并进行初步调查，以发现模式、检测异常并检查假定情况。EDA 通过计算汇总统计数据 and 创建数据可视化得以执行。

F

事实表

[星型架构](#)中的中心表。它存储有关业务运营的定量数据。通常，事实表包含两种类型的列：包含度量的列和包含维度表外键的列。

快速失效机制

一种使用频繁且增量式的测试来缩短开发生命周期的理念。这是敏捷方法的关键部分。

故障隔离边界

在中 AWS Cloud，诸如可用区 AWS 区域、控制平面或数据平面之类的边界，它限制了故障的影响并有助于提高工作负载的弹性。有关更多信息，请参阅[AWS 故障隔离边界](#)。

功能分支

请参阅[分支](#)。

特征

您用来进行预测的输入数据。例如，在制造环境中，特征可能是定期从生产线捕获的图像。

特征重要性

特征对于模型预测的重要性。这通常表示为数值分数，可以通过各种技术进行计算，例如 Shapley 加法解释 (SHAP) 和积分梯度。有关更多信息，请参阅使用[机器学习模型的可解释性 AWS](#)。

功能转换

为 ML 流程优化数据，包括使用其他来源丰富数据、扩展值或从单个数据字段中提取多组信息。这使得 ML 模型能从数据中获益。例如，如果您将“2021-05-27 00:15:37”日期分解为“2021”、“五月”、“星期四”和“15”，则可以帮助学习与不同数据成分相关的算法学习精细模式。

少样本提示

在要求 [LLM](#) 执行类似任务之前，先向其提供少量示例，以演示任务和预期输出。此技术是上下文内学习的一种应用，其中模型可以从提示中嵌入的示例 (样本) 中学习。对于需要特定格式、推理或领域知识的任务，少样本提示可能非常有效。另请参阅[零样本提示](#)。

FGAC

请参阅[精细访问控制](#)。

精细访问控制 (FGAC)

使用多个条件允许或拒绝访问请求。

快闪迁移

一种数据库迁移方法，通过[更改数据捕获](#)使用连续数据复制，在极短的时间内迁移数据，而非使用分阶段方法。目标是将停机时间降至最低。

FM

请参阅[基础模型](#)。

基础模型 (FM)

一个大型深度学习神经网络，一直在广义和未标记数据的大量数据集上进行训练。FMs 能够执行各种各样的一般任务，例如理解语言、生成文本和图像以及用自然语言进行对话。有关更多信息，请参阅[什么是基础模型](#)。

G

生成式人工智能

[AI](#) 模型的一个子集，这些模型已经过大量数据训练，可以使用简单的文本提示来创建新的内容和构件，例如图像、视频、文本和音频。有关更多信息，请参阅[什么是生成式人工智能](#)。

地理阻止

请参阅[地理限制](#)。

地理限制 (地理阻止)

在 Amazon 中 CloudFront，一种阻止特定国家/地区的用户访问内容分发的选项。您可以使用允许列表或阻止列表来指定已批准和已禁止的国家/地区。有关更多信息，请参阅 CloudFront 文档[中的限制内容的地理分布](#)。

GitFlow 工作流程

一种方法，在这种方法中，下层和上层环境在源代码存储库中使用不同的分支。Gitflow 工作流程被认为是传统的工作流程，而[基于中继的工作流程](#)则是现代的、首选的方法。

黄金映像

系统或软件的快照，用作部署该系统或软件的新实例的模板。例如，在制造业中，黄金映像可用于在多个设备上预调配软件，并有助于提高设备制造操作的速度、可扩展性和生产效率。

全新策略

在新环境中缺少现有基础设施。在对系统架构采用全新策略时，您可以选择所有新技术，而不受对现有基础设施 (也称为[棕地](#)) 兼容性的限制。如果您正在扩展现有基础设施，则可以将棕地策略和全新策略混合。

防护机制

帮助管理各组织单位的资源、策略和合规性的高级规则 (OUs)。预防性防护机制会执行策略以确保符合合规性标准。它们是使用服务控制策略和 IAM 权限边界实现的。侦测性护栏会检测策略违规和合规性问题，并生成提醒以进行修复。它们通过使用 AWS Config、Amazon、AWS Security Hub CSPM GuardDuty AWS Trusted Advisor、Amazon Inspector 和自定义 AWS Lambda 支票来实现。

H

HA

请参阅[高可用性](#)。

异构数据库迁移

将源数据库迁移到使用不同数据库引擎的目标数据库 (例如，从 Oracle 迁移到 Amazon Aurora)。异构迁移通常是重新架构工作的一部分，而转换架构可能是一项复杂的任务。[AWS 提供了 AWS SCT](#) 来帮助实现架构转换。

高可用性 (HA)

在遇到挑战或灾难时，工作负载无需干预即可连续运行的能力。HA 系统旨在自动进行故障转移、持续提供良好性能，并以最小的性能影响处理不同负载和故障。

历史数据库现代化

一种用于实现运营技术 (OT) 系统现代化和升级以更好满足制造业需求的方法。历史数据库是一种用于收集和存储工厂中各种来源数据的数据库。

保留数据

从用于训练[机器学习](#)模型的数据集中保留的一部分标注的历史数据。通过将模型预测与保留数据进行比较，您可以使用保留数据来评估模型性能。

同构数据库迁移

将源数据库迁移到共享同一数据库引擎的目标数据库 (例如，从 Microsoft SQL Server 迁移到 Amazon RDS for SQL Server)。同构迁移通常是更换主机或更换平台工作的一部分。您可以使用本机数据库实用程序来迁移架构。

热数据

经常访问的数据，例如实时数据或近期的转化数据。这些数据通常需要高性能存储层或存储类别才能提供快速的查询响应。

修补程序

针对生产环境中关键问题的紧急修复。由于其紧迫性，修补程序通常是在典型的 DevOps 发布工作流程之外进行的。

hypercure 周期

割接之后，迁移团队立即管理和监控云中迁移的应用程序以解决任何问题的时间段。通常，这个周期持续 1-4 天。在 hypercure 周期结束时，迁移团队通常会将应用程序的责任移交给云运营团队。

我

laC

请参阅[基础设施即代码](#)。

基于身份的策略

附加到一个或多个 IAM 委托人的策略，用于定义他们在 AWS Cloud 环境中的权限。

空闲应用程序

90 天内平均 CPU 和内存使用率在 5% 到 20% 之间的应用程序。在迁移项目中，通常会停用这些应用程序或将其保留在本地。

IloT

请参阅[工业物联网](#)。

不可变基础设施

一种模型，可为生产工作负载部署新的基础设施，而不是更新、修补或修改现有基础设施。不可变基础设施本质上比[可变基础设施](#)更一致、更可靠、更可预测。有关更多信息，请参阅 AWS Well-Architected Framework 中的[使用不可变基础设施进行部署](#)最佳实践。

入站 (入口) VPC

在 AWS 多账户架构中，一种接受、检查和路由来自应用程序外部的网络连接的 VPC。[AWS 安全参考架构](#)建议设置您的网络帐户，包括入站、出站和检查，VPCs 以保护您的应用程序与更广泛的互联网之间的双向接口。

增量迁移

一种割接策略，在这种策略中，您可以将应用程序分成小部分进行迁移，而不是一次性完整割接。例如，您最初可能只将几个微服务或用户迁移到新系统。在确认一切正常后，您可以逐步迁移其他微服务或用户，直到停用遗留系统。这种策略降低了大规模迁移带来的风险。

工业 4.0

该术语由 [Klaus Schwab](#) 在 2016 年提出，指的是通过连接、实时数据、自动化、分析和 AI/ML 的进步来实现制造流程的现代化。

基础设施

应用程序环境中包含的所有资源和资产。

基础设施即代码 (IaC)

通过一组配置文件预调配和管理应用程序基础设施的过程。IaC 旨在帮助您集中管理基础设施、实现资源标准化和快速扩展，使新环境具有可重复性、可靠性和一致性。

工业物联网 (IloT)

在工业领域使用联网的传感器和设备，例如制造业、能源、汽车、医疗保健、生命科学和农业。有关更多信息，请参阅[制定工业物联网 \(IloT\) 数字化转型战略](#)。

检查 VPC

在 AWS 多账户架构中，一种集中式 VPC，用于管理对 VPCs（相同或不同 AWS 区域）、互联网和本地网络之间的网络流量的检查。[AWS 安全参考架构](#)建议设置您的网络帐户，包括入站、出站和检查，VPCs 以保护您的应用程序与更广泛的互联网之间的双向接口。

物联网 (IoT)

由带有嵌入式传感器或处理器的连接物理对象组成的网络，这些传感器或处理器通过互联网或本地通信网络与其他设备和系统进行通信。有关更多信息，请参阅[什么是 IoT ?](#)

可解释性

它是机器学习模型的一种特征，描述了人类可以理解模型的预测如何取决于其输入的程度。有关更多信息，请参阅使用[机器学习模型的可解释性 AWS](#)。

物联网

请参阅[物联网](#)。

IT 信息库 (ITIL)

提供 IT 服务并使这些服务符合业务要求的一套最佳实践。ITIL 是 ITSM 的基础。

IT 服务管理 (ITSM)

为组织设计、实施、管理和支持 IT 服务的相关活动。有关将云运营与 ITSM 工具集成的信息，请参阅[运营集成指南](#)。

ITIL

请参阅[IT 信息库](#)。

ITSM

请参阅[IT 服务管理](#)。

L

基于标签的访问控制 (LBAC)

强制访问控制 (MAC) 的一种实施方式，其中明确为用户和数据本身分配了安全标签值。用户安全标签和数据安全标签之间的交集决定了用户可以看到哪些行和列。

登录区

landing zone 是一个架构精良的多账户 AWS 环境，具有可扩展性和安全性。这是一个起点，您的组织可以从这里放心地在安全和基础设施环境中快速启动和部署工作负载和应用程序。有关登录区的更多信息，请参阅[设置安全且可扩展的多账户 AWS 环境](#)。

大语言模型 (LLM)

一种基于大量数据进行预训练的深度学习 [AI](#) 模型。LLM 可以执行多项任务，例如回答问题、总结文档、将文本翻译成其他语言以及完成句子。有关更多信息，请参阅[什么是 LLMs](#)。

大规模迁移

迁移 300 台或更多服务器。

LBAC

请参阅[基于标签的访问控制](#)。

最低权限

授予执行任务所需的最低权限的最佳安全实践。有关更多信息，请参阅 IAM 文档中的[应用最低权限许可](#)。

直接迁移

请参阅 [7 R](#)。

小端序系统

一个先存储最低有效字节的系统。另请参阅[字节顺序](#)。

LLM

请参阅[大型语言模型](#)。

下层环境

请参阅[环境](#)。

M

机器学习 (ML)

一种使用算法和技术进行模式识别和学习的人工智能。ML 对记录的数据 (例如物联网 (IoT) 数据) 进行分析和学习，以生成基于模式的统计模型。有关更多信息，请参阅[机器学习](#)。

主分支

请参阅[分支](#)。

恶意软件

旨在危害计算机安全或隐私的软件。恶意软件可能会破坏计算机系统、泄露敏感信息或获得未经授权的访问权限。恶意软件的示例包括病毒、蠕虫、勒索软件、木马、间谍软件和键盘记录器。

托管式服务

AWS 服务 它 AWS 运行基础设施层、操作系统和平台，您可以访问端点来存储和检索数据。Amazon Simple Storage Service (Amazon S3) 和 Amazon DynamoDB 就是托管服务的示例。这些服务也称为抽象服务。

制造执行系统 (MES)

一种软件系统，用于跟踪、监控、记录和控制将原材料转化为成品的生产过程。

MAP

请参阅[迁移加速计划](#)。

机制

一个完整的过程，您可以在其中创建工具，推动工具的采用，然后检查结果以进行调整。机制是一种在运作过程中自我强化和改善的循环。有关更多信息，请参阅在 Well-Architect AWS ed 框架中[构建机制](#)。

成员账户

AWS 账户 除属于组织中的管理账户之外的所有账户 AWS Organizations。一个账户一次只能是一个组织的成员。

MES

请参阅[制造执行系统](#)。

消息队列遥测传输 (MQTT)

[一种基于发布/订阅模式的轻量级 machine-to-machine \(M2M\) 通信协议，适用于资源受限的物联网设备。](#)

微服务

一种小型的独立服务，通过明确的定义进行通信 APIs ，通常由小型的独立团队拥有。例如，保险系统可能包括映射到业务能力 (如销售或营销) 或子域 (如购买、理赔或分析) 的微服务。微服务

的好处包括敏捷、灵活扩展、易于部署、可重复使用的代码和恢复能力。有关更多信息，请参阅[使用 AWS 无服务器服务集成微服务](#)。

微服务架构

一种使用独立组件构建应用程序的方法，这些组件将每个应用程序进程作为微服务运行。这些微服务使用轻量级通过定义明确的接口进行通信。APIs 该架构中的每个微服务都可以更新、部署和扩展，以满足对应用程序特定功能的需求。有关更多信息，请参阅[在上实现微服务](#)。AWS

迁移加速计划 (MAP)

AWS 该计划提供咨询支持、培训和服务，以帮助组织为迁移到云奠定坚实的运营基础，并帮助抵消迁移的初始成本。MAP 提供了一种以系统的方式执行遗留迁移的迁移方法，以及一套用于自动执行和加速常见迁移场景的工具。

大规模迁移

将大部分应用程序组合分波迁移到云中的过程，在每一波中以更快的速度迁移更多应用程序。本阶段使用从早期阶段获得的最佳实践和经验教训，实施由团队、工具和流程组成的迁移工厂，通过自动化和敏捷交付简化工作负载的迁移。这是[AWS 迁移策略](#)的第三阶段。

迁移工厂

跨职能团队，通过自动化、敏捷的方法简化工作负载迁移。迁移工厂团队通常包括运营、业务分析师和所有者、迁移工程师、开发人员和从事冲刺工作的 DevOps 专业人员。20% 到 50% 的企业应用程序组合由可通过工厂方法优化的重复模式组成。有关更多信息，请参阅本内容集中[有关迁移工厂的讨论](#)和[云迁移工厂指南](#)。

迁移元数据

有关完成迁移所需的应用程序和服务器器的信息。每种迁移模式都需要一套不同的迁移元数据。迁移元数据的示例包括目标子网、安全组和 AWS 账户。

迁移模式

一种可重复的迁移任务，详细列出了迁移策略、迁移目标以及所使用的迁移应用程序或服务。示例：使用 AWS 应用程序迁移服务重新托管向 Amazon EC2 的迁移。

迁移组合评测 (MPA)

一种在线工具，提供了用于验证迁移到 AWS Cloud 的业务案例的信息。MPA 提供了详细的组合评测（服务器规模调整、定价、TCO 比较、迁移成本分析）以及迁移计划（应用程序数据分析和数据收集、应用程序分组、迁移优先级排序和波次规划）。所有 AWS 顾问和 APN 合作伙伴顾问均可免费使用[MPA 工具](#)（需要登录）。

迁移准备情况评测 (MRA)

使用 AWS CAF 深入了解组织的云就绪状态、确定优势和劣势以及制定行动计划以缩小已发现差距的过程。有关更多信息，请参阅[迁移准备指南](#)。MRA 是 [AWS 迁移策略](#) 的第一阶段。

迁移策略

将工作负载迁移到 AWS Cloud 的方法。有关更多信息，请参见术语表中的 [7 R](#) 词条，以及[动员您的组织以加快大规模迁移](#)。

ML

请参阅[机器学习](#)。

现代化

将过时的（原有的或单体）应用程序及其基础设施转变为云中敏捷、弹性和高度可用的系统，以降低成本、提高效率和利用创新。有关更多信息，请参阅[在 AWS Cloud 中实现应用程序现代化的策略](#)。

现代化准备情况评估

一种评估方式，有助于确定组织应用程序的现代化准备情况；确定收益、风险和依赖关系；确定组织能够在多大程度上支持这些应用程序的未来状态。评估结果是目标架构的蓝图、详细说明现代化进程发展阶段和里程碑的路线图以及解决已发现差距的行动计划。有关更多信息，请参阅[在 AWS Cloud 中评估应用程序的现代化准备情况](#)。

单体应用程序 (单体式)

作为具有紧密耦合进程的单个服务运行的应用程序。单体应用程序有几个缺点。如果某个应用程序功能的需求激增，则必须扩展整个架构。随着代码库的增长，添加或改进单体应用程序的功能也会变得更加复杂。若要解决这些问题，可以使用微服务架构。有关更多信息，请参阅[将单体分解为微服务](#)。

MPA

请参阅[迁移组合评测](#)。

MQTT

请参阅[消息队列遥测传输](#)。

多分类器

一种帮助为多个类别生成预测（预测两个以上结果之一）的过程。例如，ML 模型可能会询问“这个产品是书、汽车还是手机？”或“此客户最感兴趣什么类别的产品？”

可变基础设施

一种用于更新和修改生产工作负载的现有基础设施的模型。为了提高一致性、可靠性和可预测性，Well-Architect AWS ed Framework 建议使用[不可变基础设施](#)作为最佳实践。

O

OAC

请参阅[来源访问控制](#)。

OAI

请参阅[来源访问身份](#)。

OCM

请参阅[组织变革管理](#)。

离线迁移

一种迁移方法，在这种方法中，源工作负载会在迁移过程中停止运行。这种方法会延长停机时间，通常用于小型非关键工作负载。

OI

请参阅[运营集成](#)。

OLA

请参阅[运营级别协议](#)。

在线迁移

一种迁移方法，在这种方法中，源工作负载无需离线即可复制到目标系统。在迁移过程中，连接工作负载的应用程序可以继续运行。这种方法的停机时间为零或最短，通常用于关键生产工作负载。

OPC-UA

请参阅[开放流程通信 – 统一架构](#)。

开放流程通信 – 统一架构 (OPC-UA)

一种用于工业自动化的 machine-to-machine (M2M) 通信协议。OPC-UA 提供了一个包含数据加密、身份验证和授权方案的互操作性标准。

运营级别协议 (OLA)

一项协议，阐明了 IT 职能部门承诺相互交付的内容，以支持服务水平协议 (SLA)。

运营准备情况审查 (ORR)

一份问题核对清单和关联的最佳实践，可帮助您了解、评估、预防或缩小事件和可能的故障的范围。有关更多信息，请参阅 [AWS Well-Architected Framework 中的运营准备情况审查 \(ORR \)](#)。

运营技术 (OT)

与物理环境配合使用以控制工业运营、设备和基础设施的硬件和软件系统。在制造业中，OT 和信息技术 (IT) 系统的集成是[工业 4.0](#) 转型的关键重点。

运营整合 (OI)

在云中实现运营现代化的过程，包括就绪计划、自动化和集成。有关更多信息，请参阅[运营整合指南](#)。

组织跟踪

由 AWS CloudTrail 此创建的跟踪记录组织 AWS 账户 中所有人的所有事件 AWS Organizations。该跟踪是在每个 AWS 账户 中创建的，属于组织的一部分，并跟踪每个账户的活动。有关更多信息，请参阅 CloudTrail 文档中的[为组织创建跟踪](#)。

组织变革管理 (OCM)

一个从人员、文化和领导力角度管理重大、颠覆性业务转型的框架。OCM 通过加快变革采用、解决过渡问题以及推动文化和组织变革，帮助组织为新系统和战略做好准备和过渡。在 AWS 迁移策略中，该框架被称为人员加速，因为云采用项目需要变更的速度。有关更多信息，请参阅[OCM 指南](#)。

来源访问控制 (OAC)

在中 CloudFront，一个增强的选项，用于限制访问以保护您的亚马逊简单存储服务 (Amazon S3) 内容。OAC 全部支持所有 S3 存储桶 AWS 区域、使用 AWS KMS (SSE-KMS) 进行服务器端加密，以及对 S3 存储桶的动态PUT和DELETE请求。

来源访问身份 (OAI)

在中 CloudFront，一个用于限制访问权限以保护您的 Amazon S3 内容的选项。当您使用 OAI 时，CloudFront 会创建一个 Amazon S3 可以对其进行身份验证的委托人。经过身份验证的委托人只能通过特定 CloudFront 分配访问 S3 存储桶中的内容。另请参阅[OAC](#)，其中提供了更精细和增强的访问控制。

ORR

请参阅[运营准备情况审查](#)。

OT

请参阅[运营技术](#)。

出站 (出口) VPC

在 AWS 多账户架构中，一种处理从应用程序内部启动的网络连接的 VPC。[AWS 安全参考架构](#)建议设置您的网络帐户，包括入站、出站和检查，VPCs 以保护您的应用程序与更广泛的互联网之间的双向接口。

P

权限边界

附加到 IAM 主体的 IAM 管理策略，用于设置用户或角色可以拥有的最大权限。有关更多信息，请参阅 IAM 文档中的[权限边界](#)。

个人身份信息 (PII)

直接查看其他相关数据或与之配对时可用于合理推断个人身份的信息。PII 的示例包括姓名、地址和联系信息。

PII

请参阅[个人身份信息](#)。

playbook

一套预定义的步骤，用于捕获与迁移相关的工作，例如在云中交付核心运营功能。playbook 可以采用脚本、自动化运行手册的形式，也可以是操作现代化环境所需的流程或步骤的摘要。

PLC

请参阅[可编程逻辑控制器](#)。

PLM

请参阅[产品生命周期管理](#)。

policy

一个对象，可以定义权限（请参阅[基于身份的策略](#)）、指定访问条件（请参阅[基于资源的策略](#)）或定义 AWS Organizations 的组织中所有账户的最大权限（请参阅[服务控制策略](#)）。

多语言持久性

根据数据访问模式和其他要求，独立选择微服务的数据存储技术。如果您的微服务采用相同的数据存储技术，它们可能会遇到实现难题或性能不佳。如果微服务使用最适合其需求的数据存储，则可以更轻松地实现微服务，并获得更好的性能和可扩展性。

组合评测

一个发现、分析和确定应用程序组合优先级以规划迁移的过程。有关更多信息，请参阅[评估迁移准备情况](#)。

谓词

返回 true 或 false 的查询条件，通常位于 WHERE 子句中。

谓词下推

一种数据库查询优化技术，可在传输之前筛选查询中的数据。这将减少从关系数据库检索和处理的数据量，并提高查询性能。

预防性控制

一种安全控制，旨在防止事件发生。这些控制是第一道防线，帮助防止未经授权的访问或对网络的意外更改。有关更多信息，请参阅在 AWS 上实施安全控制中的[预防性控制](#)。

主体

中 AWS 可以执行操作和访问资源的实体。此实体通常是 IAM 角色的根用户或用户。AWS 账户有关更多信息，请参阅 IAM 文档中[角色术语和概念](#)中的主体。

隐私设计

一种在整个开发过程中都考虑隐私的系统工程方法。

私有托管区

一个容器，其中包含有关您希望 Amazon Route 53 如何响应针对一个或多个 VPCs 域名及其子域名的 DNS 查询的信息。有关更多信息，请参阅 Route 53 文档中的[私有托管区的使用](#)。

主动控制

一种[安全控制](#)，旨在防止部署不合规资源。这些控制会在资源预置之前对其进行扫描。如果资源与控制不兼容，则不会预置它。有关更多信息，请参阅 AWS Control Tower 文档中的[控制参考指南](#)，并参见在上实施安全[控制中的主动控制](#) AWS。

产品生命周期管理 (PLM)

对产品在其整个生命周期内的数据和流程的管理，从设计、开发和发布，到增长和成熟，再到衰退和淘汰。

生产环境

请参阅[环境](#)。

可编程逻辑控制器 (PLC)

在制造业中，一种高度可靠、适应性强的计算机，用于监控机器并实现制造过程自动化。

提示串接

使用一个 [LLM](#) 提示的输出作为下一个提示的输入，以生成更好的响应。该技术用于将复杂的任务分解为子任务，或者迭代地完善或扩展初步响应。它有助于提高模型响应的准确性和相关性，并允许获得更精细的个性化结果。

假名化

用占位符值替换数据集中个人标识符的过程。假名化可以帮助保护个人隐私。假名化数据仍被视为个人数据。

publish/subscribe (pub/sub)

一种支持微服务间异步通信的模式，可提高可扩展性和响应能力。例如，在基于微服务的 [MES](#) 中，微服务可以将事件消息发布到其他微服务可以订阅的频道。系统可以在不更改发布服务的情况下添加新的微服务。

Q

查询计划

一系列用于访问 SQL 关系数据库系统中的数据的步骤，类似于指令。

查询计划回归

当数据库服务优化程序选择的最佳计划不如数据库环境发生特定变化之前时。这可能是由统计数据、约束、环境设置、查询参数绑定更改和数据库引擎更新造成的。

R

RACI 矩阵

请参阅[责任、问责、咨询和知情 \(RACI \)](#)。

RAG

请参阅[检索增强生成](#)。

勒索软件

一种恶意软件，旨在阻止对计算机系统或数据的访问，直到付款为止。

RASCI 矩阵

请参阅[责任、问责、咨询和知情 \(RACI \)](#)。

RCAC

请参阅[行列访问控制](#)。

只读副本

用于只读目的的数据库副本。您可以将查询路由到只读副本，以减轻主数据库的负载。

重新架构

请参阅 [7 R](#)。

恢复点目标 (RPO)

自上一个数据恢复点以来可接受的最长时间。这决定了从上一个恢复点到服务中断之间可接受的数据丢失情况。

恢复时间目标 (RTO)

服务中断和服务恢复之间可接受的最大延迟。

重构

请参阅 [7 R](#)。

Region

地理区域内的 AWS 资源集合。每一个 AWS 区域 都相互隔离，彼此独立，以提供容错、稳定性和弹性。有关更多信息，请参阅[指定您的账户可以使用的 AWS 区域](#)。

回归

一种预测数值的 ML 技术。例如，要解决“这套房子的售价是多少？”的问题 ML 模型可以使用线性回归模型，根据房屋的已知事实（如建筑面积）来预测房屋的销售价格。

重新托管

请参阅 [7 R](#)。

版本

在部署过程中，推动生产环境变更的行为。

重新放置

请参阅 [7 R](#)。

更换平台

请参阅 [7 R](#)。

重新购买

请参阅 [7 R](#)。

韧性

应用程序抵御中断或从中断中恢复的能力。在 AWS Cloud 中规划韧性时，[高可用性](#)和[灾难恢复](#)是常见的考虑因素。有关更多信息，请参阅 [AWS Cloud 韧性](#)。

基于资源的策略

一种附加到资源的策略，例如 AmazonS3 存储桶、端点或加密密钥。此类策略指定了允许哪些主体访问、支持的操作以及必须满足的任何其他条件。

责任、问责、咨询和知情 (RACI) 矩阵

定义参与迁移活动和云运营的所有各方的角色和责任的矩阵。矩阵名称源自矩阵中定义的责任类型：负责 (R)、问责 (A)、咨询 (C) 和知情 (I)。支持 (S) 类型是可选的。如果包括支持，则该矩阵称为 RASCI 矩阵，如果将其排除在外，则称为 RACI 矩阵。

响应性控制

一种安全控制，旨在推动对不良事件或偏离安全基线的情况进行修复。有关更多信息，请参阅在 AWS 上实施安全控制中的[响应性控制](#)。

保留

请参阅 [7 R](#)。

停用

请参阅 [7 R](#)。

检索增强生成 (RAG)

一种[生成式人工智能](#)技术，其中 [LLM](#) 在生成响应之前引用其训练数据来源之外的权威数据来源。例如，RAG 模型可以对组织的知识库或自定义数据执行语义搜索。有关更多信息，请参阅[什么是 RAG](#)。

轮换

定期更新[密钥](#)以使攻击者更难访问凭证的过程。

行列访问控制 (RCAC)

使用已定义访问规则的基本、灵活的 SQL 表达式。RCAC 由行权限和列掩码组成。

RPO

请参阅[恢复点目标](#)。

RTO

请参阅[恢复时间目标](#)。

运行手册

执行特定任务所需的一套手动或自动程序。它们通常是为了简化重复性操作或高错误率的程序而设计的。

S

SAML 2.0

许多身份提供商 (IdPs) 使用的开放标准。此功能支持联合单点登录 (SSO)，因此用户无需在 IAM 中为组织中的所有人创建用户即可登录 AWS 管理控制台 或调用 AWS API 操作。有关基于 SAML 2.0 的联合身份验证的更多信息，请参阅 IAM 文档中的[关于基于 SAML 2.0 的联合身份验证](#)。

SCADA

请参阅[监督控制和数据采集](#)。

SCP

请参阅[服务控制策略](#)。

机密密钥

在中 AWS Secrets Manager，您以加密形式存储的机密或受限信息，例如密码或用户凭证。它由密钥值及其元数据组成。密钥值可以是二进制、单个字符串或多个字符串。有关更多信息，请参阅 Secrets Manager 文档中的[什么是 Amazon Secrets Manager 密钥？](#)。

安全设计

一种在整个开发过程中都考虑安全的系统工程方法。

安全控制

一种技术或管理防护机制，可防止、检测或降低威胁行为体利用安全漏洞的能力。安全控制有以下四种类型：[预防性](#)、[检测性](#)、[响应性](#)和[主动性](#)。

安全固化

缩小攻击面，使其更能抵御攻击的过程。这可能包括删除不再需要的资源、实施授予最低权限的最佳安全实践或停用配置文件中不必要的功能等操作。

安全信息和事件管理 (SIEM) 系统

结合了安全信息管理 (SIM) 和安全事件管理 (SEM) 系统的工具和服务。SIEM 系统会收集、监控和分析来自服务器、网络、设备和其他来源的数据，以检测威胁和安全漏洞，并生成警报。

安全响应自动化

一种预定义的程序化操作，旨在自动响应或修复安全事件。这些自动化可作为[侦探或响应式](#)安全控制措施，帮助您实施 AWS 安全最佳实践。自动响应操作的示例包括修改 VPC 安全组、修补 Amazon EC2 实例或轮换凭证。

服务器端加密

由接收数据的人在目的地对数据 AWS 服务 进行加密。

服务控制策略 (SCP)

一种策略，用于集中控制组织中所有账户的权限 AWS Organizations。SCPs 定义防护措施或限制管理员可以委托给用户或角色的操作。您可以使用 SCPs 允许列表或拒绝列表来指定允许或禁止哪些服务或操作。有关更多信息，请参阅 AWS Organizations 文档中的[服务控制策略](#)。

服务端点

的入口点的 URL AWS 服务。您可以使用端点，通过编程方式连接到目标服务。有关更多信息，请参阅 AWS 一般参考 中的[AWS 服务 端点](#)。

服务水平协议 (SLA)

一份协议，阐明了 IT 团队承诺向客户交付的内容，比如服务正常运行时间和性能。

服务水平指示器 (SLI)

对服务性能方面的衡量，例如错误率、可用性或吞吐量。

服务水平目标 (SLO)

代表服务运行状况的目标指标，由[服务水平指示器](#)衡量。

责任共担模式

描述您在云安全与合规方面共同承担 AWS 的责任的模型。AWS 负责云的安全，而您则负责云中的安全。有关更多信息，请参阅[责任共担模式](#)。

SIEM

请参阅[安全信息和事件管理系统](#)。

单点故障 (SPOF)

应用程序的单个关键组件出现故障，可能会中断系统。

SLA

请参阅[服务水平协议](#)。

SLI

请参阅[服务水平指示器](#)。

SLO

请参阅[服务水平目标](#)。

split-and-seed 模型

一种扩展和加速现代化项目的模式。随着新功能和产品发布的定义，核心团队会拆分以创建新的产品团队。这有助于扩展组织的能力和服务，提高开发人员的工作效率，支持快速创新。有关更多信息，请参阅[在 AWS Cloud 中实现应用程序现代化的分阶段方法](#)。

SPOF

请参阅[单点故障](#)。

星型架构

一种数据库组织结构，它使用一个大型事实表来存储事务数据或测量数据，并使用一个或多个较小的维度表来存储数据属性。此结构专为在[数据仓库](#)中使用或用于商业智能目的而设计。

strangler fig 模式

一种通过逐步重写和替换系统功能直至可以停用原有的系统来实现单体系统现代化的方法。这种模式用无花果藤作为类比，这种藤蔓成长为一棵树，最终战胜并取代了宿主。该模式是由 [Martin Fowler](#) 提出的，作为重写单体系统时管理风险的一种方法。有关如何应用此模式的示例，请参阅[使用容器和 Amazon API Gateway 逐步将原有的 Microsoft ASP.NET \(ASMX \) Web 服务现代化](#)。

子网

您的 VPC 内的一个 IP 地址范围。子网必须位于单个可用区中。

监督控制和数据采集 (SCADA)

在制造业中，一种使用硬件和软件来监控实物资产和生产操作的系统。

对称加密

一种加密算法，它使用相同的密钥来加密和解密数据。

综合测试

以模拟用户交互的方式测试系统，以检测潜在问题或监控性能。您可以使用 [Amazon S CloudWatch ynthetic](#) 来创建这些测试。

系统提示

一种为 [LLM](#) 提供上下文、说明或准则以指导其行为的技术。系统提示有助于设置上下文并制定与用户交互的规则。

T

标签

键值对，用作组织资源的元数据。AWS 标签有助于您管理、识别、组织、搜索和筛选 资源。有关更多信息，请参阅[标记您的 AWS 资源](#)。

目标变量

您在监督式 ML 中尝试预测的值。这也被称为结果变量。例如，在制造环境中，目标变量可能是产品缺陷。

任务列表

一种通过运行手册用于跟踪进度的工具。任务列表包含运行手册的概述和要完成的常规任务列表。对于每项常规任务，它包括预计所需时间、所有者和进度。

测试环境

请参阅[环境](#)。

训练

为您的 ML 模型提供学习数据。训练数据必须包含正确答案。学习算法在训练数据中查找将输入数据属性映射到目标（您希望预测的答案）的模式。然后输出捕获这些模式的 ML 模型。然后，您可以使用 ML 模型对不知道目标的新数据进行预测。

中转网关

一个网络传输中心，可用于将您的网络 VPCs 和本地网络互连。有关更多信息，请参阅 AWS Transit Gateway 文档中的[什么是公交网关](#)。

基于中继的工作流程

一种方法，开发人员在功能分支中本地构建和测试功能，然后将这些更改合并到主分支中。然后，按顺序将主分支构建到开发、预生产和生产环境。

可信访问权限

向您指定的服务授予权限，该服务可以代表您在其账户中执行任务。AWS Organizations 当需要服务相关的角色时，受信任的服务会在每个账户中创建一个角色，为您执行管理任务。有关更多信息，请参阅 AWS Organizations 文档中的[AWS Organizations 与其他 AWS 服务一起使用](#)。

优化

更改训练过程的各个方面，以提高 ML 模型的准确性。例如，您可以通过生成标签集、添加标签，并在不同的设置下多次重复这些步骤来优化模型，从而训练 ML 模型。

双披萨团队

一个小 DevOps 团队，你可以用两个披萨来喂食。双披萨团队的规模可确保在软件开发过程中充分协作。

U

不确定性

这一概念指的是不精确、不完整或未知的信息，这些信息可能会破坏预测式 ML 模型的可靠性。不确定性有两种类型：认知不确定性是由有限的、不完整的数据造成的，而偶然不确定性是由数据中固有的噪声和随机性导致的。有关更多信息，请参阅[量化深度学习系统中的不确定性指南](#)。

无差别任务

也称为繁重工作，即创建和运行应用程序所必需的工作，但不能为最终用户提供直接价值或竞争优势。无差别任务的示例包括采购、维护和容量规划。

上层环境

请参阅[环境](#)。

V

vacuum 操作

一种数据库维护操作，包括在增量更新后进行清理，以回收存储空间并提高性能。

版本控制

跟踪更改的过程和工具，例如存储库中源代码的更改。

VPC 对等连接

两者之间的连接 VPCs，允许您使用私有 IP 地址路由流量。有关更多信息，请参阅 Amazon VPC 文档中的[什么是 VPC 对等连接](#)。

漏洞

损害系统安全的软件缺陷或硬件缺陷。

W

热缓存

一种包含经常访问的当前相关数据的缓冲区缓存。数据库实例可以从缓冲区缓存读取，这比从主内存或磁盘读取要快。

暖数据

不常访问的数据。查询此类数据时，通常可以接受中速查询。

窗口函数

一种对与当前记录有某种关联的一组行执行计算的 SQL 函数。窗口函数对于处理任务很有用，例如计算移动平均值或根据当前行的相对位置访问行的值。

工作负载

一系列资源和代码，它们可以提供商业价值，如面向客户的应用程序或后端过程。

工作流

迁移项目中负责一组特定任务的职能小组。每个工作流都是独立的，但支持项目中的其他工作流。例如，组合工作流负责确定应用程序的优先级、波次规划和收集迁移元数据。组合工作流将这些资产交付给迁移工作流，然后迁移服务器和应用程序。

WORM

请参阅[一次写入多次读取](#)。

WQF

请参阅[AWS 工作负载资格鉴定框架](#)。

一次写入多次读取 (WORM)

一种存储模型，可一次写入数据并防止数据被删除或修改。授权用户可以根据需要多次读取数据，但无法对其进行更改。此数据存储基础设施被认为[不可变](#)。

Z

零日漏洞利用

一种利用[零日漏洞](#)的攻击，通常为恶意软件。

零日漏洞

生产系统中不可避免的缺陷或漏洞。威胁主体可能利用这种类型的漏洞攻击系统。开发人员经常因攻击而意识到该漏洞。

零样本提示

为[LLM](#)提供执行任务的说明，但没有可以帮助指导的示例（样本）。LLM 必须使用预先训练的知识来处理任务。零样本提示的有效性取决于任务的复杂性和提示的质量。另请参阅[少样本提示](#)。

僵尸应用程序

平均 CPU 和内存使用率低于 5% 的应用程序。在迁移项目中，通常会停用这些应用程序。

本文属于机器翻译版本。若本译文内容与英语原文存在差异，则一律以英文原文为准。