采用生成式人工智能的成熟度模型 AWS

AWS 规范性指导

AWS 规范性指导: 采用生成式人工智能的成熟度模型 AWS

Table of Contents

简介 1
目标受众 1
瞄准业务目标 1
模型概述 2
成熟度等级 2
成熟度方面 5
采用的支柱 5
重点领域 6
主要活动 6
转型策略 6
第 1 级：Envision 7
重点和标准 7
主要活动 7
转型策略 10
第 2 级：实验 12
重点和标准 12
主要活动 12
转型策略 14
第 3 级：发射 16
重点和标准 16
主要活动 16
转型策略 18
第 4 级：比例 20
重点和标准 20
主要活动 20
继续旅程 23
后续步骤 24
资源 24
AWS 服务 文档 24
AWS 规范性指导 24
其他资源 25
贡献者 26
编写 26
正在审阅 26

技术写作 26
文档历史记录 27
术语表 28
..... 28
A 28
B 31
C 32
D 35
E 38
F 40
G 41
H 42
我 43
L 45
M 46
O 50
P 52
Q 54
R 55
S 57
T 60
U 61
V 62
W 62
Z 63
..... lxiv

在上面采用生成式人工智能的成熟度模型 AWS

亚马逊 Web Services（贡献者）

2025 年 6 月（文档历史记录）

生成式 AI 是 AI 模型的一个子集，这些模型已经过大量数据训练，可以生成新内容，包括文本、图像、音乐和视频。这些模型可以使用预训练的基础模型、自定义模型以及增强或专有数据集。生成式人工智能的影响遍及各行各业。它可以增强创造力，提高生产力，并实现新的商业模式。如果您的组织希望生成式人工智能来增强运营、推动创新和实现业务增长，那么结构化、分阶段的方法对于驾驭采用之旅至关重要。

根据首席信息官的一篇文章，88% 的人工智能飞行员未能投入生产。这会导致所谓的飞行员疲劳。文章说：“各公司只是厌倦了花更多的时间、金钱和精力来支持那些无法迅速或根本无法进入生产的试点项目。”这种疲劳会扼杀创新，阻碍对生成式人工智能的进一步实验。此外，根据一份McKinsey 报告，各组织正在努力应对人工智能实施中的重大数据质量和集成挑战。

本战略文件提供了一个结构化框架，以帮助组织实施生成式人工智能解决方案。该框架旨在帮助您应对技术采用的复杂性，并确保您不会忽视关键步骤或最佳实践。使用本指南中的建议来全面了解您的生成式 AI 成熟度。通过评估成熟度级别，您可以确定每个级别的重点领域，并启动 end-to-end生成式人工智能采用之旅。该框架探讨了四个成熟度级别，从最初的意识到全面的转型。它概述了每个级别的关键活动和基本实践。

目标受众

本文面向希望通过在组织中采用生成式人工智能来创造价值的高管、技术总监、商业领袖、数据科学家、生成人工智能和 AI/ML 专家、IT 专业人员和决策者。

瞄准业务目标

通过系统地推进生成式人工智能成熟度水平，组织可以实现以下关键业务成果：

- 通过经过验证的生成式 AI 用例进行战略业务流程创新
- 通过强大、可随时投入生产的 AI 解决方案实现卓越运营
- 通过标准化、可重复使用的 AI 组件提高企业级效率
- 通过战略转型和可扩展的人工智能能力获得竞争优势

生成式 AI 成熟度模型概述

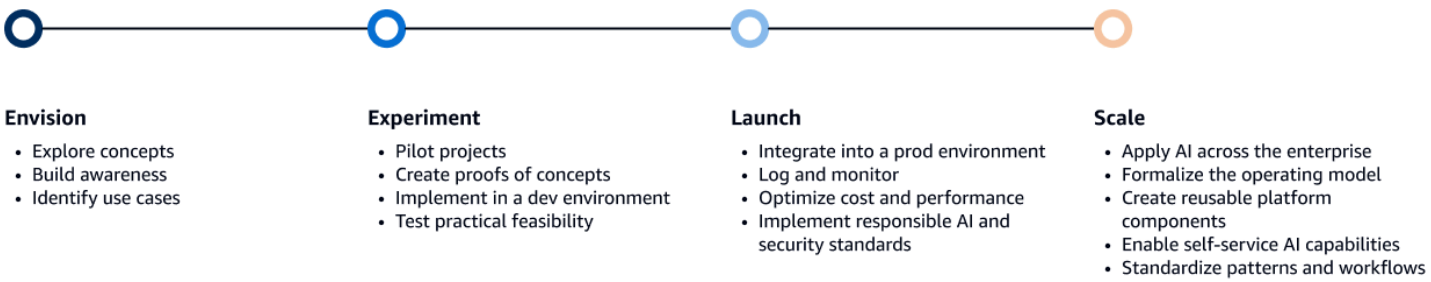
组织可以使用成熟度模型的框架来有效地整合生成式人工智能能力，避免常见的实施陷阱和实施差距。通过详细的成熟度评估，您可以清楚地了解您的组织在人工智能之旅中所处的位置，并查明需要关注的具体领域。进展跨越四个不同的层次，首先是基本的理解，最后是彻底的组织转型。每个关卡都包含有针对性的行动项目和战略指导方针，以推动成功。

本节包含以下主题：

- 生成式 AI 成熟度模型中的级别
- 生成式 AI 成熟度的各个方面

生成式 AI 成熟度模型中的级别

生成式 AI 成熟度模型分为四个主要层面。每个级别都代表了组织在使用生成式 AI 能力方面的进展。该模型可以帮助组织了解他们目前的立场，并引导他们迈向生成式人工智能之旅的下一步。下图显示了生成式 AI 成熟度模型的四个级别以及每个级别的关键活动。



以下是生成式 AI 成熟度模型中的四个级别：

- 第 1 级：Envision
- 第 2 级：实验
- 第 3 级：发射
- 第 4 级：比例

每个成熟度级别的标签反映了组织内部采用生成式人工智能的影响。当你确定组织在给定级别上的位置时，你可以深入了解下一个成熟度水平中的机会。较低的级别通常包含更具战术性的生成式人工智能用例，而较高的级别本质上往往更具战略性和变革性。

许多组织会发现，多个成熟度级别的特征适用于其团队和用例。这是因为没有哪个级别本质上是优越或劣势的——适当的成熟度水平取决于组织的目标和准备程度。

> **Note**
>
> 这种生成式人工智能成熟度模型并不是要将组织或其生成人工智能能力归类为纯粹的初学者或变革性的。相反，应独立考虑采用生成式人工智能的各个方面。每个成熟度级别的特征代表该特定方面的连续性，但在其他方面不一定与同一级别相关。

下表概述了这四个级别。

类别	第 1 级：Envision	第 2 级：实验	第 3 级：发射	第 4 级：缩放
说明	Organizations 探索生成式人工智能概念，提高意识并识别潜在的用例。	Organizations通过结构化的试点项目和概念验证来验证生成式人工智能的潜力，同时建立核心技术能力和基础实施框架。	Organizations 系统地部署具有强大治理、监控和支持机制的生产就绪生成人工智能解决方案，在保持安全和合规标准的同时，提供一致的价值和卓越的运营。	Organizations 通过可重复使用的组件、标准化模式和自助服务平台在企业范围内建立生成式 AI 能力，从而在保持自动化治理和促进创新的同时加快采用速度。
聚焦	建立对生成式人工智能技术的认识和理解，探索潜在的应用，并确定人工智能可以为业务增加价值的领域	通过结构化的试点计划验证业务价值并培养核心能力	部署生产就绪型解决方案，通过强大的发布流程、全面的治理框架和绩效监控，提供可衡量的业务价值	创建可重复使用的组件和模式，加快生成式 AI 在整个企业中的采用
标准	• 对生成式 AI 概念有基本的了解	• 运行试点项目和概念验证	• 将一些生成式 AI 应用程序发布到生产环境中	• 在组织中的各个部门广泛采用生成式人工智能

类别	第 1 级：Envision	第 2 级：实验	第 3 级：发射	第 4 级：缩放
	• 没有正式的项目或资源分配 • 了解行业趋势和价值机会	• 组建小型团队探索生成式 AI 能力 • 建立基础和治理框架	• 为生成式 AI 应用实施风险、治理和负责任的人工智能政策 • 建立运营和支持小组	• 将许多生成式 AI 应用程序发布到生产环境中 • 优先考虑对生成式 AI 基础设施和工具的投资 • 正式确定运营模式和负责、负责、咨询、知情 (RACI) 矩阵
主要活动	• 参加 AI 意识培训、研讨会和会议 • 与 AI 主题专家和顾问接触 • 探索潜在的用例和业务优势 • 评估文化准备程度 • 评估生成式 AI 治理 • 积累知识	• 为试点项目定义和完善业务用例 • 开发概念证明 • 评估并选择合适的生成式 AI 模型和工具 • 衡量业务收益实现情况 • 培养内部能力和技术专业知识	• 初始化操作模型 • 创建解决方案架构治理 • 制定可随时投入生产的实施策略 • 建立监测和绩效跟踪机制 • 实施风险和治理管理 • 集成 IT 基础架构库 (ITIL) 框架 • 建立运营和支持结构	• 正式确定生成式 AI 运营模型和 RACI 矩阵 • 创建可重复使用的生成式 AI 功能和组件 • 标准化生成式 AI 用例模式 • 建立全组织范围的协作开发框架 • 将人工智能功能发展为内部开发平台 (IDP) 或软件即服务 (SaaS) • 共享知识并实现知识大众化

为了进一步解释和理解成熟度模型，重要的是要了解组织在生成式人工智能采用之旅中通常是如何进展的。这一进展不仅反映了组织如何使用生成式人工智能功能，还反映了促使他们推动采用的动力。在早期关卡中，许多用户可能根本没有正式化的人工智能流程。相反，他们将自己的工具视为来自各种内部来源的经过改进的功能集合。随着组织的成熟，这些能力的管理和标准化变得更加一致。最终，随着功能变得更加完善和易于发现，以及用户自然地选择使用人工智能功能，组织通常会偏离强制性或激励措施等外部动机。理想情况下，他们甚至开始将自己的精力投入到更广泛的人工智能创新和开发上。

生成式 AI 成熟度的各个方面

成功采用生成式人工智能需要对多个组织维度有全面的了解。本节探讨了组织在成熟过程中必须考虑和发展的四个关键方面：支持采用人工智能的基本支柱、指导战略优先事项的重点领域、推动实施的关键活动以及指导组织成熟度发展的转型策略。这些方面共同为评估和推进生成式人工智能能力提供了一个全面的框架。Organizations可以使用此框架来确定差距，确定投资的优先顺序，并制定可行的计划，以便在成熟度级别上取得进展。每个方面都是根据企业人工智能采用方面的丰富现场经验选择的。它们反映了区分成功实施和失败实施的关键要素。

本节包含以下主题：

- 采用的支柱
- 重点领域
- 主要活动
- 转型策略

采用的支柱

每个成熟度级别都根据以下采用支柱进行评估：

- 业务 — 战略调整和对业务目标的可衡量影响
- 人才 — 人才培养、技能培养和跨职能协作
- 治理 — 制定风险管理、合规和道德准则
- 平台 — 投资于可扩展的基础设施和平台，以实现生成式人工智能功能
- 安全 — 保护数据、隐私和生成式 AI 模型的部署
- 运营 — 管理生成式 AI 解决方案生命周期、优化部署、实施反馈机制和监控性能

这些支柱与AWS 云采用框架 (AWS CAF) 保持一致并进行了扩展，以满足生成式 AI 需求。本战略文档中的建议为每个支柱添加了特定于人工智能的元素，例如道德人工智能实施、模型生命周期管理和人工

智能基础设施要求。这种协调可以帮助组织使用现有的 AWS CAF 最佳实践，同时应对独特的 AI 采用挑战。

重点领域

每个成熟度级别的重点领域可帮助组织确定活动和投资的优先顺序。以下是四个重点领域：

- 创新和可行性 — 探索和验证创新的生成式人工智能用例以及所需数据集的可用性和质量
- 集成和效率 — 将生成式 AI 集成到现有业务流程中
- 可扩展性和优化 — 扩展生成式 AI 应用程序并持续提高性能
- 转型和领导力 — 使用生成式人工智能推动战略转变并获得竞争优势

主要活动

组织可以使用生成式人工智能成熟度模型中的关键活动来导航其旅程，并成功定义和实施其生成式人工智能战略。这些活动从最初探索和理解生成式人工智能技术，到尝试原型，将人工智能解决方案集成到业务流程中，在整个组织中进行扩展，然后为持续改进和战略转型建立治理。关键活动分为以下类别之一：

- 探索和意识 — 发展生成式人工智能技术的基础知识，并确定采用的战略机会
- 实验和验证 — 促进和实施试点项目和原型，以评估技术可行性和商业价值
- 集成和实施 — 将生成式 AI 功能嵌入到现有业务流程中，并将解决方案部署到生产环境中
- 扩展和优化 — 在整个组织中集成生成式 AI 应用程序，持续提高其性能和效率
- 治理和领导力 — 建立框架和最佳实践，以管理生成式人工智能计划并将其用于战略转型

转型策略

每个级别的转型战略都侧重于指导组织进行渐进式改进。这包括制定生成式人工智能路线图和数据战略，与业务目标保持一致，投资人才和工具，以及实施治理框架。

生成式人工智能成熟度模型级别 1：Envision

这个基础级别是组织探索生成式人工智能概念、建立组织意识并确定与其业务目标一致的潜在用例的关键起点。通过建立这一基本基础，公司可以为其人工智能之旅制定清晰的愿景，同时解决业务、人员、治理、平台、安全和运营方面的关键考虑因素。

本节包括以下主题：

- 重点和标准
- 主要活动
- 更上一层楼的转型策略

重点和标准

该级别的目标是建立对生成式人工智能技术以及与该技术相关的新兴行业趋势的基本理解和认识。这包括评估潜在的应用，并确定生成式人工智能可以使业务受益的领域。该级别的重点是教育利益相关者了解生成式人工智能，并开始探索用例并进行风险和文化准备情况评估。

以下是达到此级别的标准：

- 该组织已经展示了生成式人工智能基础知识的基础知识。
- 该组织记录了人们对行业生成式人工智能应用和机会的认识。
- 该组织对人工智能的文化准备有了越来越多的认识。
- 该组织已对潜在的用例和优势进行了初步探索。
- 该组织已对治理和安全要求进行了初步考虑。

主要活动

下表显示了每个采用支柱的主要活动。

收养支柱	活动	
业务	• 了解生成式 AI 如何解决特定的业务问题。	

收养支柱	活动	
	• 将最初的生成式 AI 用例映射到业务目标，例如提高客户参与度或自动创建内容。 • 确定与选定用例相关的高价值数据源。	
人员	• 举办内部培训课程和知识共享研讨会。 • 确定组织内的人工智能拥护者，领导对生成式人工智能机会的探索。 • 评估贵组织的文化和变革管理准备情况，以备采用生成式 AI。 • 评估贵组织中当前的技术技能差距，并确定采用生成式人工智能所需的投资。 • 设计教育计划，帮助高级管理人员了解人工智能的战略潜力、技术能力、变革性的业务影响以及数据在生成式人工智能项目中的重要性。 • 参加行业论坛和会议，学习其他公司采用人工智能的经验。 • 组织内部黑客马拉松，鼓励实验和促进创新。	

收养支柱	活动
治理	• 探索采用生成式人工智能的伦理和监管注意事项，例如隐私和数据主权。 • 为组织中负责任地使用人工智能制定一套初步指导方针。
平台	• 探索采用生成式 AI 以符合组织标准的要求。 • 探索 AI/ML 模型和工具，例如用于访问基础模型的 Amazon Bedrock 和用于快速实验的 SageMaker Amazon AI。 • 评估和编目现有的内部和外部数据源。评估数据基础设施和质量，以确定生成式人工智能的可行性和潜在的实施要求。

收养支柱	活动	
安全性	• 了解与在组织中采用生成式 AI 相关的安全含义和任务，例如： • 数据隐私和保护风险，包括通过训练数据、提示和模型输出可能暴露敏感信息 • 访问控制和身份验证挑战，包括 AI 系统中用户验证和基于角色的权限的复杂性 • 对安全漏洞进行建模，包括对即时注入攻击的敏感性以及生成不安全或不当内容的可能性	
操作	• 了解与在组织中采用生成式 AI 相关的运营挑战，例如： • 规划 AI 解决方案的性能监控需求。 • 考虑治理和版本控制要求。 • 了解事件响应程序的要求。	

更上一层楼的转型策略

要升级到下一个成熟度级别，请考虑以下方面：

- 建立跨职能的生成式人工智能小组 — 组建具有明确角色和职责的跨职能生成人工智能小组。小组应包括能够领导实验 SMEs 工作的IT代表、业务代表、安全和治理利益相关者以及生成式人工智能。稍后，随着你扩大生成人工智能工作的规模，该小组将为更正式定义的卓越中心 (CoE) 奠定基础。

- 确定用例并确定其优先级 — 制定用例矩阵，帮助您根据可行性、业务影响以及与战略目标的一致性确定项目的优先顺序。要获得概念证明 (PoCs)，请创建热门用例的简短列表。
- 为试点项目分配资源 — 确保预算和人员用于小规模运营 PoCs。
- 培养生成式人工智能技能 — 提高员工对特定工具和技术的技能，例如亚马逊 Bedrock、A SageMaker I、Amazon Q Bus iness、A mazon Q Developer、提示工程、检索增强生成 (RAG) 以及代理人工智能和工作流程。
- 完成初步治理 — 建立初步治理，指导生成式 AI 的使用。它应涵盖合规、风险管理和道德方面的考虑。
- 文化准备 — 开始规划组织变革管理，以便在全公司范围内采用生成式人工智能。
- 确定成功指标-针对每个 PoC，定义成功标准以及业务和技术指标。

通过采取这些行动，组织可以期望：

- 获得生成式 AI 技术的实践经验。
- 验证特定用例的可行性和潜在影响。
- 建立生成式 AI 方面的内部能力和专业知识。
- 确定与采用生成式人工智能相关的潜在挑战和风险。
- 提高生成式 AI 采用的准备程度，以便更上一层楼。

生成式 AI 成熟度模型第 2 级：实验

在上一级别建立的基础意识的基础上，实验级别标志着生成式人工智能技术从理论探索到实际实施的关键过渡。在这个层面上，组织超越了概念理解，转而参与实践PoC项目和试点计划。这些 PoC 和试点项目旨在验证业务价值并培养核心能力。该级别的特点是结构化实验，组织可以组建专门的团队，建立治理框架，并开始培养内部技术专业知识。通过精心控制的试点项目，组织可以检验他们对生成式人工智能潜力的假设，同时最大限度地降低风险并最大限度地利用学习机会。这为更广泛地实施和扩大成功举措奠定了基础。

本节包括以下主题：

- 重点和标准
- 主要活动
- 更上一层楼的转型策略

重点和标准

在这个层面上，组织从探索过渡到使用生成式人工智能技术的实践PoC实验和试点项目。重点是通过结构化的试点计划和培养核心能力来验证商业价值。这个级别强调实践学习、建立内部能力和技术专长，以及建立基础和治理框架。

以下是达到此级别的标准：

- 该组织正在积极开展试点项目，概念验证正在进行中。
- 专门的跨职能团队负责生成式人工智能计划。
- 制定了结构化的内部培训计划。
- 这些组织已经选择并验证了人工智能模型和工具。
- 该组织已经定义了其最初的治理和数据框架。

主要活动

下表显示了每个采用支柱的主要活动。

收养支柱	活动
业务	• 根据业务价值和可行性定义战略用例并确定其优先级。 • 为 PoCs此，建立衡量投资回报率 (ROI) 的成功指标和框架。 • 为每个 PoC 创建价值评估记分卡。 • 将范围限制在具有明确成功指标的可管理范围内。 PoCs • 对于每个 PoC，衡量投资回报率并评估其是否达到成功标准。
人员	• 在及时的工程、RAG 和模型微调中实施结构化培训计划。 • 创建生成式 AI 认证途径和职业发展框架。 • 聘请生成式 AI 和数据科学专家。 • 与外部专家（例如 AWS 生成式人工智能创新中心或 AWS专业服务）合作，共同构建 PoC、提供支持和转移知识。 • 建立 AI 认证途径和职业发展框架。
治理	• 制定初步框架，涵盖生成式 AI 的数据治理，例如用于矢量搜索的内容质量。 • 建立模型评估标准和质量控制。 • 为生成式 AI 项目制定风险评估协议。 • 为合乎道德和负责任地使用生成式人工智能制定指导方针。培训开发人员、数据科学家和生成式 AI 专家遵守这些指南。
平台	• 为 PoC 设置基础架构，例如 landing AWS zon e 和开发者所需的权限。 • 为生成式 AI 实验和 PoC 开发设置环境，例如 Amazon Bedrock 游乐场、亚马逊 A SageMaker I JupyterLab 空间或笔记本实例。

收养支柱	活动
	• 实施开发人员可以轻松使用的 RAG 方法或代理工作流程。对于 RAG 方法，可以考虑 Amazon Bedrock 知识库；对于代理工作流程，可以考虑 A ma zon Bedrock Agents。 • 设置管理提示、模型和提示评估的框架或管道。这些资源应该可以帮助开发人员快速评估 PoC 应用程序的结果和性能。 • 实施早期的数据集成工作，包括结构化和非结构化数据管道。为 RAG 实验设置矢量数据库。 • 根据成本、性能和用例适用性评估基础模型。你可以使用 Amazon Bedrock、Amazon SageMaker AI 和亚马逊 A SageMaker I JumpStart。
安全性	• 为训练生成式 AI 模型实施数据访问控制，并确保它们符合合规性要求。Amazon Q Bus iness 可以通过启用精细控制来简化 RAG 的实施，允许生成式 AI 工作负载仅检索用户有权访问的数据。 • 制定保护用于训练模型的数据集中的个人身份信息 (PII) 的策略。

操作	• 为以下内容创建文档和支持流程：
	  • PoC 的实施和学习
	  • 基本平台配置和安全控制
	  • 测试和评估程序
	  • 成功移交正在转向生产 PoCs 的成功流程

更上一层楼的转型策略

通过执行以下操作，组织可以过渡到下一个成熟度级别：

- 创建生产级基础设施以支持生成式 AI — 用于 AWS 服务 为生产部署实现 CI/CD 管道、标准化部署模式和适当的扩展机制。
- 实施治理 — 建立生产级治理框架，以管理持续的生成式 AI 使用和模型更新。
- 实现可观察性 — 实施专门针对生成式 AI 工作负载的可观察性、监控和日志记录实践。这包括模型性能指标、使用模式和响应质量评估。
- 专注于合规性 — 确保遵守数据隐私和安全的行业标准和法规。
- 组@@ 建专门的 AI 团队 — 组建一个团队，为生成式 AI 解决方案创建和维护标准化生产路径。
- 实施卓越运营-创建事件响应和上报流程。制定服务级别协议 (SLAs) 和性能指标。实施成本优化策略。

通过采取这些行动，组织可以：

- 验证生成式 AI 应用程序是否稳定、可靠，并持续为组织创造价值。
- 随着各部门需求和使用量的增加，Support 支持生成式 AI 解决方案的发展。
- 管理风险，保持监督，并使人工智能计划与监管标准保持一致，因为它们已成为业务运营不可或缺的一部分。
- 为生成式 AI 解决方案提供持续监控、改进和支持。这减少了对临时或临时项目团队的依赖。
- 让组织做好准备，从孤立的项目转向战略性和凝聚力的方法，让人工智能成为业务流程的核心推动力。该组织已准备好进一步扩大规模和更广泛地采用。

生成式 AI 成熟度模型级别 3：发布

在这个层面上，组织从 proof-of-concept举措过渡到有条不紊地将经过验证的精选生成式人工智能解决方案部署到生产环境中。这个级别代表着从实验转向专注于强大的治理协议、实时监控系统和专门的支持基础设施的关键转变。各公司专注于推出一些具有明显业务影响的生产级应用程序。该级别强调操作的严谨性——实施全面的启动框架，制定明确的治理指导方针，并保持严格的安全标准。发布可靠的生成式 AI 解决方案，提供可量化的结果，让组织为更广泛的采用做好准备。

本节包括以下主题：

- 重点和标准
- 主要活动
- 更上一层楼的转型策略

重点和标准

在这个层面上，组织系统地将生成式人工智能解决方案部署到生产环境中，并实施强大的治理、监控和支持机制。这些机制可提供一致的价值和卓越的运营，同时保持安全性和合规性标准。重点从实验性的生成式人工智能应用程序转移到部署生产就绪型解决方案，这些解决方案通过强大的发布流程、全面的治理框架和系统的性能监控来提供可衡量的商业价值。该级别侧重于部署一定数量的生产就绪生成式人工智能解决方案，这些解决方案可作为启动框架和治理机制的基础实现。

以下是达到此级别的标准：

- 生产就绪的生成式 AI 解决方案正在提供可衡量的业务成果。
- 该组织已经实施了基准安全、治理和负责任的人工智能框架。
- 建立了操作控制措施，包括自动监控和警报系统。
- 该组织已经定义了人工智能决策的 human-in-the-loop流程。
- 对于跨职能的人工智能团队，已经定义了初步角色和运营职责。

主要活动

下表显示了每个采用支柱的关键活动。

收养支柱	活动
业务	• 签署 RACI 矩阵的第一个版本，用于生成式 AI 运算。 • 确定平台架构、开发和支持所需的关键角色。 • 通过全面的仪表板衡量运营效率和业务价值。 • 跟踪和优化运营成本和资源利用率。
人员	• 创建生成式 AI 平台团队或小组，负责架构、开发和维护。 • 实施始终可用的分层支持结构和培训计划。
治理	• 获得企业架构审查委员会的正式架构认可。 • 建立负责任的人工智能政策框架并获得利益相关者的批准。 • 为人工智能实施审查设立跨职能监督委员会。 • 对于生成式 AI 解决方案，请保留有关治理批准、风险评估、标准化设计模式和技术规范的文档。
平台	• 为生成式 AI 解决方案实施自动化 CI/CD 管道。 • 部署基础设施即代码 (IaC) 来管理 AWS 资源。 • 记录生成式 AI 解决方案的设计模式和技术规范。 • 维护生成式 AI 平台组件的 CMDB 记录。
安全性	• 为生成式 AI 解决方案及其数据管道实施强大的安全控制。 • 实施负责任的人工智能的初步政策。 • 优化可扩展基础架构，以支持实时数据摄取、矢量搜索和微调。 • 定期进行安全评估和审计。

收养支柱	活动
	• 部署 Amazon Bedrock Guardrails，对生成式 AI 应用程序的安全和隐私控制进行标准化。
操作	• 建立 SLA 框架和绩效指标。 • 监控模型性能和护栏违规情况。设置警报。 • 创建具有自动警报系统的操作仪表板。 • 遵循 ITIL 流程进行变更管理和资产管理。 • 建立了包含操作手册、行动手册和故障排除指南的集中式知识库。 FAQs • 建立数据可观测性实践。跟踪数据沿袭、来源和质量指标，以便在扩展之前识别差距。 • 建立分层支持级别，有明确的上报路径。 • 定期进行绩效评估并分析客户反馈。

更上一层楼的转型策略

为了扩大生成式人工智能计划的规模，组织应：

- 正式确定生成式 AI 运营模式 — 正式确定整个组织的 RACI 矩阵。
- 增强生成式 AI 平台 — 对现有的生成式 AI 实现进行评估，以识别可重复使用的模式和组件。评估技术堆栈是否已准备好进行扩展。开始构想和设计模块化架构，该架构具有集中式提示管理、自动评估框架和标准化模式，可有效扩展生成式 AI 解决方案。
- 扩展用例 — 整合多个部门的 AI 功能并探索新的应用程序。
- 改善开发者体验 — 将现有平台转变为自助服务内部平台。该平台是一个全面的环境，可为整个企业的 AI 开发提供标准化工具、工作流程和治理。
- 共享知识 — 建立内部资源实践并创建组件市场，以便在团队之间共享可重复使用的 AI 资产。内部源代码实践是在组织内应用开源开发方法的策略。
- 设置运营扩展 — 通过自动事件响应和容量规划增强您的支持基础架构。这为基础架构做好了扩展准备，以便在企业范围内采用生成式人工智能。
- 投资高级分析 — 使用云端的高级分析工具，例如使用亚马逊 CloudWatch获取指标，使用 Amazon Quick Suite 进行可视化，使用数据分析进行持续改进。

- 审查数据治理模型 — 评估您的数据治理模型当前是否支持自助服务功能，同时保持标准化的策略和访问控制。过于严格或集中的方法可能会阻碍你将数据计划扩展到核心团队以外的能力，尤其是在不同的业务部门之间。

通过采取这些行动，组织可以：

- 在整个组织中扩展生成式 AI 计划，以产生广泛影响。
- 继续增强平台，同时寻找提高生产力和可重复使用性的机会。
- 改善开发者体验并减少认知负担。
- 培养数据驱动的文化。
- 通过将组织定位为生成式 AI 领导者，吸引顶尖人才。

生成式 AI 成熟度模型级别 4：规模

生成式人工智能成熟度模型的第 4 级，即规模级别，从卓越运营过渡到可扩展的创新。Organizations 开始超越单个生产部署，创建一个由可重复使用的组件、标准化模式和自动化工作流程组成的强大生态系统。该生态系统可帮助组织加快多个部门采用生成式人工智能，同时保持稳健的治理和成本优化。通过建立可扩展的架构和自助服务功能，这种成熟度使企业能够高效地部署大量生成式人工智能应用程序，从而最终推动整个组织的转型和可持续创新。

本节包括以下主题：

- 重点和标准
- 主要活动

重点和标准

在这个层面上，组织从卓越运营过渡到可扩展的创新，专注于创建可重复使用的组件和模式，以加快生成式人工智能在整个企业中的采用。重点从个人生产部署转移到构建支持自助服务功能、标准化模式和自动化工作流程的功能，同时优化成本和维持大规模治理。与专注于特定生产工作负载的第 3 级不同，第 4 级支持通过标准化和可重复使用的组件快速部署大量生成式 AI 应用程序，从而提高企业范围的效率和生产力。

以下是达到此级别的标准：

- 多个部门已广泛使用生成式人工智能。
- 该组织已经建立了企业范围的生成式人工智能基础设施和工具生态系统。
- 定义并实现了操作模型和 RACI 矩阵。
- 可用的库包括标准化、可重复使用的 AI 组件、模式和应用程序。自助服务功能使整个组织都可以访问图书馆。
- 自动化治理机制在企业范围内运行。
- 该组织有持续创新实践和成果的证据。

主要活动

下表显示了每个采用支柱的主要活动。

收养支柱	活动
业务	• 使生成式 AI 项目与长期业务目标保持一致。专注于收入增长、成本降低和客户满意度。 • 通过可重复使用的组件和提供价值的标准化模式，推动企业范围内生成式 AI 的采用。 • 最终确定生成式 AI 运营模型和 RACI 矩阵，以实现规模化运营。 • 组建专门的平台架构、开发和维护小组。 • 创建标准化的治理和审批工作流程。 • 实施高级分析和监控以实现持续改进。 • 制定积极主动的方法，确定下一个创新且高价值的人工智能用例。考虑提高工作效率的内部用例和以产品为重点的外部用例。 • 评估复杂的决策自动化机会 • 评估个性化和产品增强的可能性
人员	• 对员工进行交叉培训，让他们使用生成式人工智能工具，培养持续学习和创新的文化。 • 在卓越中心内，制定指导计划，将知识从生成人工智能专家传递给其他团队成员。 • 使用内部资源或众源模型来帮助加速生成式 AI 可重复使用组件的开发。 • 通过卓越中心实施人工智能认证计划。
治理	• 建立涵盖数据使用、模型公平性和透明度的企业级 AI 治理和道德框架。 • 通过标准化框架和自动护栏扩展负责任的人工智能实践。 • 制定贡献准则和质量标准。
平台	• 开发可重复使用的 AI 组件，例如微服务架构和自动化管道，用于在人工监督下评估解决方案。

收养支柱	活动
	• 创建标准化解决方案模板，例如 RAG 实施和代理工作流程。 • 使用模型上下文协议 (MCP) 等行业标准，制定与第三方工具集成的标准化蓝图。 • 通过内部门户实现自助服务功能，例如 API 优先的集成架构和组件市场。
安全性	• 实施企业级安全控制和自动合规性验证。
操作	• 构建流程和指导方针，以支持内部资源或众包开发模型。 • 部署全面的可观测性框架。 • 创建可帮助您监控性能的仪表板。 • 实施自动化系统来收集反馈。

继续成熟之旅

对于在生成式 AI 成熟度模型中成功达到 4 级的组织，您可以继续提升到更高的复杂程度。要做到这一点，就需要一个超越技术实施的全面战略。这种进展需要战略举措，将生成人工智能深深嵌入组织的DNA中，将组织愿景、文化转型和卓越技术相结合。为了突破四个成熟度级别，组织必须加强内部能力，建立战略伙伴关系，并投资于尖端的研发。这种全面的晋升战略，加上对人才培养的高度重视，使企业能够从规模化运营转向变革性的人工智能领导力。这带来了更高的运营效率和可持续的竞争优势。

考虑采取以下措施以超越成熟度模型：

- 将生成人工智能嵌入组织的战略愿景 ——将生成人工智能定位为公司使命和愿景的核心组成部分。请务必利用其能力来推动战略计划并保持竞争优势。
- 培养持续创新的文化 — 鼓励员工探索生成式人工智能的新应用，并奖励符合业务目标的实验。
- 与行业合作伙伴和学术界合作 — 参与研究合作伙伴关系，并与外部专家合作，始终站在 AI 创新的前沿。
- 投资尖端的生成式人工智能研发 ——投入资源探索可以突破生成式人工智能界限的新方法，例如多模态人工智能和高级强化学习。
- 吸引和留住顶尖的生成人工智能人才 — 通过提供有吸引力的激励措施、专业发展机会和协作环境，专注于建立强大的人才渠道。

通过继续在整个组织中扩展生成式人工智能解决方案，企业可以获得以下好处：

- 跨业务部门的广泛影响 ——生成式人工智能解决方案嵌入到多个部门的日常运营中，从而提高生产力并提高效率。
- 增强决策 — 借助生成式 AI 的实时见解和预测能力，组织可以更快地做出数据驱动的决策。
- 战略竞争优势 — 通过使用生成式人工智能进行创新和优化，组织可以从竞争对手中脱颖而出，开辟新的收入来源。
- 成熟的生成式人工智能 platform/blueprints 和优化的资源管理 — 通过自动化流程和改善生成解决方案的管理，您可以降低运营成本并提高可扩展性。

后续步骤

生成式人工智能成熟度模型为组织提供了一种结构化的方法，以指导其生成式人工智能的采用之旅 AWS。了解不同的成熟度级别和活动有助于组织评估其准备情况，并采取明智的措施来实现生成式人工智能的全部潜力。该框架可帮助组织制定与其独特业务目标相一致的量身定制的策略，从而使生成式人工智能成为增长和创新的关键驱动力。

重要的是要认识到，生成式人工智能的采用不是一个 one-size-fits-all过程。每个组织的旅程都是独一无二的，它会受到行业、业务目标和现有技术能力等因素的影响。但是，这份战略文件可作为宝贵的指南。它为组织提供了一个框架，可以评估其准备情况，找出差距，并实施必要的措施，以成功利用生成式人工智能的变革潜力。

随着组织踏上生成式人工智能采用之旅，他们应该保持敏捷性和适应能力。不断重新评估您的成熟度水平，并相应地调整您的策略。人工智能领域的快速创新步伐要求我们致力于持续学习、技能发展和采用最佳实践。

通过遵循此指导并使用 AWS 人工智能/机器学习服务，组织可以在日益由人工智能驱动的世界中发掘新的机会，提高效率并获得持续的竞争优势。

资源

以下资源可以帮助您了解有关采用生成式 AI 的更多信息。

AWS 服务 文档

- Amazon Bedrock
- Amazon 基岩护栏
- Amazon Q Business
- Amazon Q 开发者版
- 亚马逊 SageMaker AI

AWS 规范性指导

- 利用生成式 AI 加快软件开发生命周 AWS 期
- 生成式 AI 工作量评估
- 检索增强生成选项和架构 AWS

- 利用生成式 AI 转变应用程序开发和维护运营模式 AWS

其他资源

- 人工智能的现状：组织如何进行重组以获取价值（McKinsey 报告）
- 88% 的人工智能飞行员未能进入生产阶段，但这还不是全部 IT（首席信息官文章）

贡献者

编写

- 冯浩飞，高级交付顾问， AWS
- 刘斌，高级交付顾问， AWS
- 克里斯·多灵顿，首席交付顾问， AWS
- Melanie Li，高级解决方案架构师， AWS
- Romain Vivier，高级解决方案架构师经理， AWS
- 山姆·爱德华兹，解决方案架构师， AWS
- 陈新，高级交付顾问， AWS

正在审阅

- Melchi Salins，高级解决方案架构师， AWS
- Junaid Baba，高级交付顾问， AWS

技术写作

- Lilly AbouHarb，高级技术撰稿人， 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
初次发布	—	2025 年 6 月 4 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：将您的本地 Oracle 数据库迁移到兼容 Amazon Aurora PostgreSQL 的版本。
- 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：在中将您的本地 Oracle 数据库迁移到适用于 Oracle 的亚马逊关系数据库服务 (Amazon RDS) AWS 云。
- 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：将您的客户关系管理 (CRM) 系统迁移到 Salesforce.com。
- 更换主机（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：在中的 EC2 实例上将您的本地 Oracle 数据库迁移到 Oracle AWS 云。
- 重新定位（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。您可以将服务器从本地平台迁移到同一平台的云服务。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

参见托管服务。

ACID

参见原子性、一致性、隔离性、耐久性。

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。与主动-被动迁移相比，它更灵活，但需要更多的工作。

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

一个 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

参见人工智能。

AIOps

参见人工智能操作。

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 (I AM) 文档 AWS中的 AB AC。

权威数据源

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

坏机器人

旨在破坏个人或组织或对其造成伤害的机器人。

BCP

参见业务连续性计划。

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。另请参见字节顺序。

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的应用程序版本（蓝色），在另一个环境中运行新的应用程序版本（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互动的软件应用程序。有些机器人是有用或有益的，例如在互联网上索引信息的网络爬虫。其他一些被称为恶意机器人的机器人旨在破坏个人或组织或对其造成伤害。

僵尸网络

被恶意软件感染并受单方（称为机器人牧民或机器人操作员）控制的机器人网络。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）。

破碎的玻璃通道

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 Well -Architected 指南中的 “实施破碎玻璃程序” 指示 AWS 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

参见AWS 云采用框架。

金丝雀部署

向最终用户缓慢而渐进地发布版本。当您确信时，可以部署新版本并全部替换当前版本。

CCoE

参见云卓越中心。

CDC

请参阅变更数据捕获。

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弹性。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

查看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 云 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。云计算通常与边缘计算技术相关。

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

组织迁移到以下阶段时通常会经历四个阶段 AWS 云：

- 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 云 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

参见配置管理数据库。

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 管道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)

人工智能领域，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。

配置偏差

对于工作负载，配置会从预期状态发生变化。这可能会导致工作负载变得不合规，而且通常是渐进的，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

参见计算机视觉。

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-Architecte 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 云 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个人。

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

参见数据库定义语言。

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 子字段使用多层人工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。

委托管理员

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。此账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。

后

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

参见环境。

侦测性控制

一种安全控制，在事件发生后进行检测、记录日志和发出警报。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

在星型架构中，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)

您用来最大限度地减少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 Fr ame AWS work 中的 “工作负载灾难恢复：云端 AWS 恢复”。

DML

参见数据库操作语言。

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

参见灾难恢复。

漂移检测

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

参见开发价值流映射。

E

EDA

参见探索性数据分析。

EDI

参见电子数据交换。

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。与云计算相比，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)

组织之间自动交换业务文档。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密文的计算过程。

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

参见服务端点。

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

参见企业资源规划。

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和创建数据可视化得以执行。

F

事实表

星形架构中的中心表。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。

失败得很快

一种使用频繁和增量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 云，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

参见分支。

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。

少量提示

在要求法学硕士执行类似任务之前，向其提供少量示例，以演示该任务和所需的输出。这种技术是情境学习的应用，模型可以从提示中嵌入的示例（镜头）中学习。对于需要特定格式、推理或领域知识的任务，Few-shot 提示可能非常有效。另请参见零镜头提示。

FGAC

请参阅精细的访问控制。

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，它使用连续的数据复制，通过更改数据捕获在尽可能短的时间内迁移数据，而不是使用分阶段的方法。目标是将停机时间降至最低。

FM

参见基础模型。

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

人工智能模型的子集，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工件，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 AI。

地理封锁

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，而基于主干的工作流程是现代的首选方法。

金色影像

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配置软件，并有助于提高设备制造运营的速度、可扩展性和生产力。

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。侦测性防护机制会检测策略违规和合规性问题，并生成警报以进行修复。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

参见高可用性。

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。

抵制数据

从用于训练机器学习模型的数据集中扣留的一部分带有标签的历史数据。通过将模型预测与抵制数据进行比较，您可以使用抵制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。

hypercare 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 hypercare 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

IaC

参见基础设施即代码。

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 云 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。

IIoT

参见工业物联网。

不可变的基础架构

一种为生产工作负载部署新基础架构，而不是更新、修补或修改现有基础架构的模型。不可变基础设施本质上比可变基础架构更一致、更可靠、更可预测。有关更多信息，请参阅 Well-Architected Framework 中的使用不可变基础架构 AWS 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

该术语由克劳斯·施瓦布（Klaus Schwab）于2016年推出，指的是通过连接、实时数据、自动化、分析和人工智能/机器学习的进步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置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 (IIoT)

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 (IIoT) 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。

IoT

参见物联网。

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。

大型语言模型 (LLM)

一种基于大量数据进行预训练的深度学习 AI 模型。法学硕士可以执行多项任务，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

见 7 R。

小端序系统

一个先存储最低有效字节的系统。另请参见字节顺序。

LLM

参见大型语言模型。

下层环境

参见环境。

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

参见分支。

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。恶意软件的示例包括病毒、蠕虫、勒索软件、特洛伊木马、间谍软件和键盘记录器。

托管服务

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。亚马逊简单存储服务 (Amazon S3) Service 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、记录和控制在车间将原材料转化为成品的生产过程。

MAP

参见迁移加速计划。

机制

一个完整的过程，在此过程中，您可以创建工具，推动工具的采用，然后检查结果以进行调整。机制是一种在运行过程中自我增强和改进的循环。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

参见制造执行系统。

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 DevOps 人员和冲刺专业人员。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： EC2 使用 AWS 应用程序迁移服务重新托管向 Amazon 的迁移。

迁移组合评测（MPA）

一种在线工具，可提供信息，用于验证迁移到的业务案例。 AWS 云 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

用于将工作负载迁移到的方法 AWS 云。有关更多信息，请参阅此词汇表中的 7 R 条目和动员组织以加快大规模迁移。

ML

参见机器学习。

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，请参阅中的应用程序现代化策略。 AWS 云

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，请参阅中的评估应用程序的现代化准备情况 AWS 云。

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

参见迁移组合评估。

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？”

可变基础架构

一种用于更新和修改现有生产工作负载基础架构的模型。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

请参阅源站访问控制。

OAI

参见源访问身份。

OCM

参见组织变更管理。

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

参见运营集成。

OLA

参见运营层协议。

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

参见开放流程通信-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)

一份问题清单和相关的最佳实践，可帮助您理解、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 Well-Architecte AWS d Frame work 中的运营准备情况评估 (ORR)。

操作技术 (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）

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

参见运营准备情况审查。

OT

参见运营技术。

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

查看个人身份信息。

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

参见可编程逻辑控制器。

PLM

参见产品生命周期管理。

policy

一个对象，可以在中定义权限（参见基于身份的策略）、指定访问条件（参见基于资源的策略）或定义组织中所有账户的最大权限 AWS Organizations （参见服务控制策略）。

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。有关更多信息，请参阅在微服务中实现数据持久性。

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`或的查询条件`false`，通常位于子`WHERE`句中。

谓词下推

一种数据库查询优化技术，可在传输前筛选查询中的数据。这减少了必须从关系数据库检索和处理的数据量，并提高了查询性能。

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。

通过设计保护隐私

一种在整个开发过程中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

一种安全控制措施，旨在防止部署不合规的资源。这些控件会在资源配置之前对其进行扫描。如果资源与控件不兼容，则不会对其进行配置。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)

在产品的整个生命周期中，从设计、开发和上市，到成长和成熟，再到衰落和移除，对产品进行数据和流程的管理。

生产环境

参见环境。

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。

提示链接

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，以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

一系列步骤，例如指令，用于访问 SQL 关系数据库系统中的数据。

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

参见 “负责任、负责、咨询、知情”（RACI）。

RAG

请参见检索增强生成。

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

参见 “负责任、负责、咨询、知情”（RACI）。

RCAC

请参阅行和列访问控制。

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。

重新架构师

见 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

见 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 AWS 区域 您的账户可以使用的账户。

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

见 7 R。

版本

在部署过程中，推动生产环境变更的行为。

搬迁

见 7 R。

更换平台

见 7 R。

回购

见 7 R。

故障恢复能力

应用程序抵御中断或从中断中恢复的能力。在中规划弹性时，高可用性和灾难恢复是常见的考虑因素。 AWS 云有关更多信息，请参阅AWS 云 弹性。

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

见 7 R。

退休

见 7 R。

检索增强生成（RAG）

一种生成式人工智能技术，其中法学硕士在生成响应之前引用其训练数据源之外的权威数据源。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据的过程。

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

参见恢复点目标。

RTO

参见恢复时间目标。

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

参见监督控制和数据采集。

SCP

参见服务控制政策。

secret

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 Secret s Manager 密钥中有什么？ 在 Secrets Manager 文档中。

安全性源于设计

一种在整个开发过程中考虑安全性的系统工程方法。

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。安全控制主要有四种类型：预防性、侦测、响应式和主动式。

安全加固

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

一种预定义和编程的操作，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 实例或轮换证书。

服务器端加密

在目的地对数据进行加密，由接收方 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。

服务级别指示器 (SLI)

对服务性能方面的衡量，例如其错误率、可用性或吞吐量。

服务级别目标 (SLO)

代表服务运行状况的目标指标，由服务级别指标衡量。

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

参见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

参见服务级别协议。

SLI

参见服务级别指标。

SLO

参见服务级别目标。

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，请参阅中的分阶段实现应用程序现代化的方法。 AWS 云

恶作剧

参见单点故障。

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交易数据或测量数据，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。

监控和数据采集 (SCADA)

在制造业中，一种使用硬件和软件来监控有形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。你可以使用 Amazon S CloudWatch ynthetics 来创建这些测试。

系统提示符

一种向法学硕士提供上下文、说明或指导方针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，充当用于组织资源的元数据。 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

参见环境。

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

参见环境。

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

一个 SQL 函数，用于对一组以某种方式与当前记录相关的行进行计算。窗口函数对于处理任务很有用，例如计算移动平均线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。

蠕虫

参见一次写入，多读。

WQF

请参阅AWS 工作负载资格框架。

一次写入，多次读取 (WORM)

一种存储模型，它可以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，但他们无法对其进行更改。这种数据存储基础架构被认为是不可变的。

Z

零日漏洞利用

一种利用未修补漏洞的攻击，通常是恶意软件。

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。

零镜头提示

向法学硕士提供执行任务的说明，但没有示例（镜头）可以帮助指导任务。法学硕士必须使用其预先训练的知识来处理任务。零镜头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 few-shot 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。