SaaS 产品的网络连接选项已启 AWS 用

AWS 规范性指导

AWS 规范性指导: SaaS 产品的网络连接选项已启 AWS 用

Table of Contents

SaaS 产品的网络连接选项已启 AWS 用

Tomas Sykora 和 Luca Schumann，Amazon Web Services

2025 年 9 月（文档历史记录）

本指南探讨了将消费者应用程序连接到软件即服务 (SaaS) 提供商的常见场景。它讨论了如何连接到本地 AWS Cloud、在其他云服务提供商 (CSP) 云中或混合架构中的资源。这些场景包括以下内容：

- 通过 HTTPS 公开 Web 服务
- 公开基于 TCP 的服务
- AWS AppSync用于实现发布订阅 (Pub/Sub) 和 GraphQL APIs
- 使用 AWS 资源公开 WebSockets 实时应用程序
- 为交互式服务通信启用双向访问

通过与本指南中介绍的最佳实践保持一致，SaaS提供商可以提高客户的信任度，并支持对SaaS产品的可扩展、安全和弹性访问。

本指南还包括自我评估标准，可帮助您评估自己在满足SaaS产品的消费者网络要求方面的成功程度。除了连接模式外，您还可以找到 AWS 网络服务的全面比较、各种部署场景的高级架构图，以及有关如何根据您的特定业务环境选择正确方法的实用指南。该指南探讨了每种网络选项的安全注意事项，讨论了需要避免的常见陷阱，并提供了在技术要求和运营效率之间取得平衡的实施建议。此外，您还可以找到战略框架，使您的网络决策与您的业务模式、增长目标和监管合规需求保持一致。

目标受众

本指南适用于 SaaS 提供商。它可以帮助云架构师、产品经理和网络工程师，他们正在设计、实施和优化SaaS产品的网络连接 AWS Cloud。要理解本指南中的概念和建议，您应该熟悉 AWS 基础知识、SaaS 核心概念和高级网络原理。

目标

本指南讨论了网络架构选项和经过现场测试的最佳实践，可帮助消费者优化对 SaaS 产品的访问。实施本指南中的建议支持以下内容：

- 易于集成 — 提供从入职到生产的简单客户旅程，这样您就可以加快客户实现价值的时间，缩短他们的收入确认周期。

- 适应性 — 通过适应客户不断变化的需求，与他们现有的网络基础设施无缝集成。这增强了您产品的价值主张。
- 总拥有成本 — 标准化网络访问以降低变更成本和每个租户的成本。通过提高部署一致性，您还可以缩短执行根本原因分析或修复的时间。
- 依赖关系管理-了解不同网络访问选项的依赖关系、长期影响和权衡取舍。这有助于产品领导者做出明智的产品决策。
- 可@@ 组合性和可扩展性 — 将核心功能的开发与运营基础设施分开。这有助于开发团队更快地行动，专注于为客户创造价值。
- 提高信任 — 通过提供弹性、容错、安全和可扩展的 SaaS 产品访问权限，您可以降低监管风险，赢得对支持客户增长的能力的信任。

评估 SaaS 产品的网络访问决策

了解您的市场

你现在做出的网络决策决定了你的 SaaS 产品的价值主张能否交付给客户。尽管这些决策具有战略重要性，但提供对您的 SaaS 产品的访问权限通常被视为纯粹的技术话题。这种看法带来的风险包括收入确认周期延长、运营效率低下以及与业务战略不一致。例如，如果快速扩张是一项战略业务目标，那么决策过程的指导方针应该是你正在考虑的解决方案是否足够可扩展和灵活以支持扩张。即使您成功地发展了业务，运营管理也绝不能成为未来增长的障碍，成本结构错位可能会消耗您的所有利润。

例如，考虑以下市场考虑因素如何影响产品的技术方面，例如网络：

- 如果您的业务模式是基于订阅的，那么您的客户可能更喜欢具有可预测的经常性成本的解决方案，而不是大量的前期投资。
- 如果您的业务策略针对的是高价值的企业级客户，那么安全、治理和监管合规标准将决定是否考虑您的 SaaS 产品。
- 如果您的目标市场主要是初创公司，那么整合的便捷性、价值实现时间和适应性可能是重要因素。初创公司通常优先考虑速度和敏捷性。由于他们需要建立品牌并需要快速创造利润，因此他们可能更喜欢快速且易于集成、能够经济高效地扩展、减少对专家的依赖且不会占用宝贵周期的解决方案。
- 一些企业需要稳定、高吞吐量和低延迟的访问。这包括娱乐和媒体行业、制造业和金融交易处理。如果这些是您的目标客户，那么可靠性是他们最关心的问题。

在所有这些情况下，如果网络访问不是无缝的，客户可能会感受到原本健康的 SaaS 产品。如果网络成为障碍，那就不支持你的商业案例。如果您的客户无法可靠地访问您提供的服务，则您的SaaS产品的价值主张为零。

了解你的角色

你在支持业务目标方面的角色取决于你是谁、你的具体个人和团队目标是什么、你的客户是谁，以及什么对他们很重要。即使你不是通常与客户互动的团队的一员，你也需要关注他们的身份和需求。工程和开发团队还必须关心内部客户，尤其是那些定期与之互动的客户。通常，这些是运营和客户成功团队。

如果你是销售组织的一员，那么与产品和工程团队就网络进行沟通是至关重要的，尽管这是一个看似纯粹的技术话题。分享有关目标市场结构的见解。沟通现有和潜在客户和合作伙伴的痛点和需求。分享有关错失机会、每个细分市场的预测增长以及事件的数据和趣闻。提出挑战贵组织支持业务增长能力的问

题。这增加了机会的数量并提高了您业务的长期盈利能力。最终，这可以帮助您的组织为未来的扩张和发展提供资金。

如果您是工程组织的一员，则在尝试起草解决方案之前，请先了解组织的业务策略。与业务战略保持一致有助于您选择正确的指标来评估不同的网络接入选项。它还可以防止随着组织的发展而进行昂贵的大规模网络重新设计。业务协调可帮助您的团队保护和保留应对未来挑战所需的资源。您的团队人数、专业发展预算或获得尖端技术的机会将取决于您展示业务一致性的能力。理想情况下，您可以展示自己的决策如何为组织的业务成功做出了贡献。因此，我们建议您掌握决策过程，包括指标选择标准。定期查看您的指标，以确认它们与业务目标一致。这可以帮助你的团队获得应有的荣誉。定期审查还有助于验证您的团队不是基于假设或过时的历史原因做出决策。

以下各节中的指标列表与网络访问有关：

- 产品和商业指标
- 影响网络决策的工程指标

本指南自始至终使用这些指标中的一部分来帮助您确定适用于 SaaS 产品的最佳网络访问方法。选择与您的业务最重要和最相关的指标，然后根据这些指标评估方法。

影响网络决策的产品和商业指标

产品和商业团队使用成功标准来评估他们是否达到了业务目标。本节介绍您的组织所做的网络访问决策可能会对产品或商业指标产生正面或负面影响。

使用这些指标和自我评估问题来评估您的网络接入方法如何与您的业务定位和市场战略保持一致。此评估可帮助您确定您当前的网络决策是否支持贵公司的市场差异化、竞争优势和目标受众需求。

本节包含以下主题的指标和自我评估问题：

- 商业模式和市场定位
- 潜在市场总量、新客户获取率、增长和可扩展性
- 客户体验和留存率
- 效率和财务业绩
- 监管合规和风险管理
- 合作伙伴策略

商业模式和市场定位

这些指标与贵公司的市场地位有关，包括竞争差异化、市场覆盖面和品牌认知度。评估网络接入方法和业务模式之间的一致性至关重要。无论是基于订阅、基于使用量、免费增值、分级、市场、API 优先还是白标，都要进行评估。确保该模型支持组织的目标和客户的目标。

高分标准

网络接入方法与业务模式无缝对应。它简化了服务的采用和交付。它支持商业模式的长期财务可行性，并且成本结构与预期增长兼容。它最大限度地减少了客户或合作伙伴在采用该产品时遇到的任何摩擦。这增强了用户体验并鼓励更广泛地使用该服务。

低分指标

所选的网络接入方法与其应支持的业务模式不一致。成本结构和部署周转时间阻碍了目标市场的采用。持续的基础设施和运营成本抑制了任何潜在的利润。这阻碍了业务增长，并使其难以按预期规模运营。或者，由于监管原因，网络访问方法的特性可能会使客户无法考虑该服务。

自我评估问题

- 所选的网络接入方法对初始部署和持续交付的成本影响是什么？ 该方法的固定成本和可变成本是多少？
- 网络接入方法能否有效且高效地扩展以满足业务模式的增长需求？ 考虑个人租户规模和入驻租户数量。
- 网络接入方法是否施加了任何可能限制商业模式灵活性或适应性的技术或运营限制？
- 对于网络接入方法，部署提前期如何与业务模式所需的上市速度保持一致？

潜在市场总量、新客户获取率、增长和可扩展性

评估网络决策对组织向新市场扩张、有效获取客户和保持运营可扩展性的能力的影响至关重要。这些因素会影响转化率。它们还会影响网络接入方法是否支持向重要的细分市场扩张，还是仅限于为特定的客户类型提供服务。

高分标准

网络接入方法可以帮助组织进入目标市场的很大一部分，也可以将其与其他网络方法有效地结合使用以扩大市场覆盖范围。这种方法只需要最少的额外集成工作。该方法支持较短的部署交货时间、快速进入

市场和扩张。它允许进行大量的并行部署。集成对客户来说非常简单，这降低了采用门槛并增强了客户体验。该方法最大限度地减少了运营开销，保留了运营能力，并支持增长预测。

低分指标

网络接入方法仅支持目标市场的一小部分，或者主要适用于业务战略中未优先考虑的利基细分市场。它不能有效地补充其他已经支持的网络接入方法。部署的交货时间滞后于市场需求，这限制了市场扩张和新客户的获取。部署模式是顺序部署的，随着需求的增长，这会增加服务瓶颈的风险。复杂的整合流程会阻碍潜在客户，这会对获取率和转化率产生负面影响。大量的运营开销削弱了该组织的运营能力。这成为预期增长的障碍。

对于这些指标，评估引入新的网络接入方法是否可以帮助组织实现其战略业务目标。考虑一下新的网络接入方法是否会产生新的产品依赖关系或消耗运营资源而不带来预期的结果。

自我评估问题

- 当前方法中是否存在任何差距，使您无法进入目标市场的更大细分市场？
- 要覆盖目标市场的70-90%，您应该支持的最低限度的非重叠、标准化的网络接入方法列表是多少？
- 每种网络接入方法都能实现什么样的覆盖面？基础设施成本、运营周期和对专家的依赖等重要指标的相关增长有哪些？
- 网络基础设施的部署能力和服务限制如何与目标市场的增长预期保持一致？
- 网络集成是否会给新客户带来任何进入壁垒？ 如何解决这些问题以提高转化率？
- 管理网络的运营开销如何影响您的增长能力和可扩展性？
- 您可以实施哪些策略来缩短网络部署的交货时间，改善市场扩张和客户获取？
- 对专家资源的依赖是否会延迟部署或与客户生态系统的集成？

客户体验和留存率

本节中的指标可帮助您了解您的组织获取客户，最重要的是留住客户的能力。了解网络接入方法与客户满意度之间的关系可以帮助产品和工程团队根据数据做出决策。

高分标准

网络接入方法可靠且易于管理。它有助于获得较高的客户满意度 (CSAT) 和净推荐值 (NPS) 结果。这些分数表明了良好的品牌声誉和客户忠诚度。得益于与客户现有生态系统的无缝集成，采用摩擦很小，对专家的依赖也较低。您的组织始终如一地满足服务级别协议 (SLAs)，这增强了客户的信任和合同义务。由于客户享受稳定可靠的服务，因此您的客户保留率很高。

低分指标

集成困难和服务访问不一致通常会导致客户失望和负面反馈。这损害了品牌声誉。由于对专家的依赖或由于入职和整合时间过长，新客户无法从免费或试用计划转换为付费服务。经常未能满足要求会 SLAs 导致经济处罚和信誉丧失，从而有可能降低客户留存率。

自我评估问题

- 网络性能（例如速度、正常运行时间和延迟）如何直接影响 CSAT 和 NPS 结果？ 哪些具体的网络改进可以提高这些分数？
- 当前的网络延迟和正常运行时间指标如何影响初始用户体验和采用率？ 要优化这些指标，需要哪些具体的网络性能改进？
- 网络配置或安全设置中是否反复出现的问题会使新客户的集成复杂化？ 如何简化这些流程？
- 网络访问配置的便捷性对新用户的入门体验有何影响？ 是否有特定的网络接入点或交货时间可以优化以增强用户的初始印象？
- 为新客户自动配置网络服务面临哪些挑战。如何调整此过程以提高可扩展性和可靠性？
- 分析最近违反 SLA 的根本原因。它们是否与网络配置、容量规划或外部供应商问题有关？
- 网络问题导致您错过 SLA 承诺的频率有多高？ 最常见的与网络相关的故障有哪些？
- 过去，哪些网络性能改进对客户满意度产生了最显著的积极影响？

效率和财务业绩

该类别评估您业务的财务状况和盈利能力方面，例如成本效率、长期生存能力、盈利能力、投资回报率 (ROI) 和总拥有成本 (TCO)。通过标准化来简化网络运营，可以减少运营开销和维护成本。这为贵组织的增长目标提供了支持。

高分标准

网络接入方法的成本结构与业务模式非常吻合。它支持可持续增长，并通过节省大量成本来提高盈利能力。高效的网络访问使客户能够快速入门，从而缩短交付价值的时间并加快市场渗透率。这直接缩短了收入确认周期。

低分指标

客户正在转向您的竞争对手，以加快其应用程序和服务的交付。由于网络配置复杂多样，交货周期延长，您的组织增加了运营成本。成本结构和业务模式不一致，这可能会导致基于订阅的服务的前期成本过高。繁琐的入职流程会降低市场渗透率并推迟收入确认。

自我评估问题

- 新服务部署的当前交付周期是多少？它们对上市时间和收入确认有何影响？
- 标准化网络运营如何有效地降低开销和维护成本？
- 成功完成初始集成、日常运营、解决问题或实施变更是否需要专家资源？
- 就技术进步而言，当前网络投资的可持续性如何？ 您是否在投资符合预期市场发展的面向未来的技术？
- 您如何有效地分配和跟踪与网络流量和单个租户的使用情况相关的成本？

监管合规和风险管理

验证是否符合网络相关法规至关重要。这证实了您的经营是合法的，并且可以保持客户的信任。跨网络运营的标准化简化了合规流程，并促进了不同司法管辖区和地区之间的一致性。这些措施可帮助您扩展服务。

高分标准

网络运营始终如一地遵守法律标准，没有复杂性，这有助于市场扩张，减少采用摩擦并增强客户信任。证明遵守关键监管框架，例如《数字运营弹性法案》(DORA) 和美国国家标准与技术研究所 (NIST)，可帮助您赢得对监管合规性敏感的客户。持续了解您的合规状态可缩短完成审计所需的时间。

低分指标

网络合规方面的差距会导致高度的采用摩擦、服务启动延迟、法律挑战和潜在的罚款。这些挑战导致向新市场扩张的计划被推迟或取消。很难在不同的司法管辖区维持标准的合规实践，这会影响运营效率和市场声誉。

自我评估问题

- 您的网络运营在多大程度上符合适用的监管或行业准则？ 您最近必须进行的合规审计揭示了什么？
- 您如何遵守数字和网络安全领域的新法规？
- 您的文件和报告流程在满足不同监管机构的要求方面的有效性如何？
- 您制定了哪些风险管理策略，以便在潜在的合规风险导致法律挑战之前识别和解决这些风险？
- 您的网络管理团队需要什么级别的合规培训和意识才能支持您的网络接入方法？

合作伙伴策略

评估网络接入方法与由公认的合作伙伴、平台和市场组成的生态系统的协调程度。这一点至关重要，特别是如果您的增长战略依赖于通过合作伙伴进行扩展。

高分标准

网络访问方法已集成到您的合作伙伴生态系统中。它的成本结构与您的主要合作伙伴的业务模式非常吻合。合作伙伴拥有无缝集成您的 SaaS 产品所需的网络技能，并且他们可以提供持续的访问权限和功能。

低分指标

所选的网络接入方法需要专门的技能、资源或设备，这些技能、资源或设备稀缺或难以购买。它不同于平台和市场常用的标准网络访问协议。这导致成本结构不可预测，难以调和。网络接入方法与您的主要合作伙伴的业务模式不一致。

自我评估问题

- 网络接入方法对合作伙伴有哪些成本影响。这些成本如何与他们的商业模式保持一致？ 整合的哪一方承担了大部分成本，必须投入多少运营周期？
- 对于网络接入方法，是否存在任何可能影响合作伙伴关系或生态系统可扩展性的集成或维护障碍？
- 如何优化网络接入方法以增强整个生态系统的兼容性和易集成性？

影响网络决策的工程指标

与产品和商业团队一样，工程团队也使用成功标准来评估他们是否达到了业务目标。但是，这些指标有所不同，它们侧重于团队开发、运营和满足安全与合规要求的能力。本节介绍的工程指标可能会受到贵组织所做的网络访问决策的正面或负面影响。

使用这些指标和自我评估问题，根据您的业务需求和技术能力来评估您当前的网络接入方法。此评估可帮助您确定架构中的差距，并确定与您的战略目标一致的改进优先顺序。通过定期审查这些标准，您可以确保您的网络接入策略继续支持客户的需求和组织的增长计划。

本部分包含以下类别和主题的指标和自我评估问题：

- 开发指标
  - 部署频率、部署时间和冲刺速度

与 SaaS 产品的网络访问相关的开发指标

本节包含以下指标：

部署频率、部署时间和冲刺速度

为了优化开发周期的效率，了解网络堆栈配置对冲刺速度的影响至关重要。

高分标准

网络堆栈配置经过简化和自动化，并且需要最少的人工干预。它不会显著影响冲刺速度。任何团队成员都可以执行网络堆栈配置和重新部署。这减少了瓶颈和对专业资源的依赖。

低分指标

配置网络堆栈需要大量的故事点。这表明这是一个复杂而耗时的过程，会影响新功能的开发。频繁重新部署网络堆栈会产生大量的时间和成本开支。网络配置任务需要专业的工程专业知识，这会造成瓶颈并减慢开发周期。

自我评估问题

- 部署过程中涉及哪些手动步骤（如果有）。它们如何影响部署频率和时间？
- 部署失败时如何处理回滚。它们对部署频率和恢复时间有何影响？
- 在设置新环境时，配置网络堆栈需要多少故事点？
- 在开发过程中，频繁地重新部署网络堆栈会带来多少额外的成本和时间开销？
- 网络堆栈的配置取决于专业的工程专业知识，还是可以由任何团队成员管理的任务？

灵活性和功能交付

网络接入方法可以影响工程团队有效创新和部署新功能的能力。

高分标准

网络接入方法提供了快速、无缝部署功能所需的灵活性。它支持多种通信协议、单向和双向通信以及消息大小。它不会对开发过程或创新施加重大限制。

低分指标

由于缺乏支持的通信协议、消息大小不灵活或对特定技术和相关专家资源的依赖，网络访问方法限制了团队推出新功能的能力。这可能会导致开发周期变慢，并阻碍服务的发展。

自我评估问题

- 网络接入方法如何影响团队开发和部署新功能的灵活性？
- 网络访问方法中是否存在限制对某些通信协议或技术的支持的限制？
- 该方法如何促进或限制将新技术和创新整合到服务中？
- 网络接入方法如何影响开发时间表和产品路线图？

更改失败率

部署新服务或功能时，您选择的网络访问方法可能会影响更改失败率。更好的控制通常意味着更大的灵活性，但也会增加配置错误的可能性，例如在管理复杂的路由设置时。

高分标准

您可以对网络堆栈进行更改，同时将故障风险降至最低。存在足够的测试机制，存在高效的回滚机制，有效的监控可帮助您快速识别和解决问题。

低分指标

网络访问方法在变更过程中容易出现故障。测试选项有限，部署策略复杂，或者监控和故障排除能力不足。需要多方参与故障排除会议。这可能会导致停机时间增加并降低 SaaS 产品的可用性。

自我评估问题

- 在更新网络堆栈时，有哪些措施可以降低变更失败的风险？
- 是否有全面的测试和验证流程？
- 系统能以多快的速度从失败的更改中恢复？ 是否有有效的回滚流程？
- 是否有主动监控和警报系统，以便在网络堆栈变更期间和之后快速检测和解决问题？
- 网络堆栈部署的历史更改失败率是多少。从过去的事件中吸取了哪些教训？
- 网络访问方法如何促进或限制变更的实施。这种方法能否最大限度地减少服务中断？
- 当您部署涉及网络访问方法的变更时，影响 SaaS 产品在生产环境中的可用性的风险是什么？

代码质量和工程团队绩效

网络访问方法可能会间接影响 SaaS 产品的代码质量。网络访问缺乏标准化可能会迫使工程团队支持多种集成方法，这可能会导致代码库膨胀。这反过来又会阻碍团队开发深度和控制代码质量的能力，而这正是维持高绩效工程团队所必需的。

高分标准

得益于代码模块化和跨支持的网络访问方法的可重用性，工程团队可以保持专注。网络访问方法与现有的部署管道和自动测试策略兼容。

低分指标

由于集成和维护过多的网络接入方法所带来的开销，工程团队的绩效会降低。有些方法会显著增加复杂性，产生技术债务，或者需要开发变通方法来解决能力缺失或不足的问题。

自我评估问题

- 网络接入方法如何管理网络可变性？
- 您是否需要开发其他代码来处理连接中断？
- 新的网络接入方法是与现有方法无缝集成，还是需要大量的定制开发？
- 采用新的网络接入方法需要在多大程度上进行变革？ 现有的代码库和自动测试能否得到有效利用？

- 使用所选的网络访问方法部署或重新部署服务有多容易或困难？ 可以经常这样做吗？ 是否存在对专家资源的依赖？
- 网络接入方法是促进还是使遵守编码标准和最佳实践变得复杂？
- 该方法 time-to-market对新功能或修复有何影响？

技术债务减免

在评估网络接入方法对技术债务的影响时，应考虑其可扩展性、可观察性和安全能力。

高分标准

随着客户群的扩大，该方法有效地简化了基础架构管理。它提供了强大的可观测性功能 out-of-the-box。这促进了高效的监控和维护。

低分指标

网络接入方法无法充分保护通信渠道，也缺乏足够的定性指标观测工具。随着客户群的增加，可能还需要对基础架构管理进行额外的开发，或者可能需要解决可靠性问题的变通方法。

自我评估问题

- 网络接入方法如何影响基础设施的长期可扩展性？ 它能否以最少的额外投资促进无缝增长？
- 随附的可观测性工具有多全面？ 它们是否允许主动监控和解决问题？
- 随着时间的推移，网络访问方法对代码库的维护和演进有何预期影响？
- 该方法能否很好地与现有和计划中的基础架构集成。是否需要进行重大更改或补充？

可扩展性、容量和性能

要确定网络接入方法对于 SaaS 产品的适用性，必须分析其在需求增加时如何保持最佳性能。

高分标准

网络接入方法无缝促进了扩展。它在请求处理期间保持低延迟，并且可以有效地处理流量峰值。无论流量水平如何增加，它都能提供稳定的性能，并且不会对增长施加运营限制。

低分指标

网络接入方法无法有效扩展，可能是由于固有的带宽限制或基础设施容量不足。资源配置和管理会增加复杂性或产生依赖关系。由于延迟、抖动和吞吐量变化增加，服务性能会降低，尤其是在拥挤的网络条件下。

自我评估问题

- 网络接入方法如何适应越来越多的租户及其数据量？
- 它本质上是否可以扩展以满足未来的需求？
- 采取了哪些措施来确保性能保持一致，即使在流量高峰期或快速扩展事件中也是如此？
- 该方法如何处理网络延迟和抖动？ 是否有优化数据吞吐量和最大限度减少延迟的机制？
- 网络接入方法能否适应不同的网络条件？ 它能否为每位客户提供单租户体验？
- 网络接入方法对底层基础设施有什么影响？ 它是否需要对现有系统进行重大升级或更改？

与 SaaS 产品网络接入相关的卓越运营指标

本节包含以下指标：

- 运营弹性和灾难恢复
- 服务和应用程序性能监控

运营弹性和灾难恢复

网络接入方法应有助于SaaS产品抵御各种类型的中断，并从任何灾难中快速恢复。

高分标准

既定和测试的灾难恢复计划始终表明，网络接入方法符合灾难恢复要求。网络访问方法支持高可用性配置，并支持自动、快速和可靠的故障转移机制。

低分指标

网络接入方法使得制定连贯的灾难恢复策略变得困难。您观察到中断后的恢复时间会延长。网络基础设施的频繁运行故障正在影响服务交付。

自我评估问题

- 上一次灾难恢复演习是什么时候，结果如何？
- 中断后恢复关键服务需要多长时间？ 需要重新部署网络基础设施的哪一部分？
- 为了简化您的灾难恢复计划，可以对网络基础设施进行哪些改进？
- 最关键的网络组件是否有冗余？
- 在严重中断之后，您是否已经自动完成了可能的网络基础设施重新部署？

- 网络接入方法如何支持容错和可靠性？是否有内置机制来处理网络中断和维护数据完整性？

服务和应用程序性能监控

网络访问方法可能会影响用于验证最佳操作和服务正常运行时间的性能监控工具。根据服务的不同，您可以访问低级指标（例如丢包率）或更高级别的指标（例如会话持续时间）。低级指标可以提供有关网络行为的详细技术见解，但解释起来可能很复杂。相比之下，更高级别的指标通常提供了一种更直接、更简单的方法来衡量整体用户体验。这是因为它们将底层网络状况的影响汇总为明确的服务质量指标。

高分标准

提供近乎实时的见解的全面监控工具随时可用。您拥有解决性能问题的自动警报和响应系统。您可以预测潜在的服务瓶颈或故障，以免它们影响用户。

低分指标

频繁的服务中断或性能问题在没有被观察或采取措施的情况下发生。缺乏对服务性能的可见性会导致对性能瓶颈的反应缓慢。需要多方团队来解决网络基础设施问题。

自我评估问题

- 目前有哪些监控工具和网络基础设施指标可用？它们在检测服务异常方面的效果如何？
- 您能以多快的速度识别和解决性能问题？
- 您是否有预测潜在性能问题的机制？
- 您可以做出哪些改进来增强可观测性能力？

与 SaaS 产品的网络访问相关的安全和治理指标

本节包含以下指标：

- 安全、合规和漏洞管理

安全、合规和漏洞管理

评估网络访问方法的安全方面至关重要，包括是否符合安全标准和漏洞管理。

高分标准

网络访问方法可帮助您的团队遵守安全框架，例如国际标准化组织 (ISO) 27001、系统和组织控制 2 (SOC 2) 或 NIST。它使定期进行安全审计变得容易。强大的加密和身份验证机制已经到位。网络是隔

离的，只有必要的资源才会暴露给客户的基础架构。您可以近乎实时地发现网络异常，而不会产生过多的开销。

低分指标

网络访问方法容易反复出现安全漏洞或漏洞，并且不符合关键安全标准。您经常观察到安全事件的检测和响应延迟。

自我评估问题

- 最近是否存在与所选网络接入方法相关的安全漏洞，我们从中学到了什么？
- 您的网络接入方法如何符合全球安全标准？
- 检测和应对安全威胁需要多长时间？ 网络访问如何帮助或限制这种能力？
- 对网络接入方法进行安全评估的频率如何？ 您能否使用常用工具来评估网络接入方法的安全性，还是需要专门的软件？
- 网络接入方法固有的安全级别是什么？它如何与行业最佳实践和监管要求保持一致？

SaaS 产品的 AWS 网络服务概述

本节讨论本指南中提及的 AWS 网络服务。它还比较了它们的功能并描述了每项服务的安全注意事项。

本节包含以下主题：

AWS 网络服务

以下是本 AWS 服务 指南中一致讨论的内容。

AWS PrivateLink

AWS PrivateLink是一项云原生服务，如果您的客户已经在 SaaS 中运营，则可以访问您的 SaaS 产品。 AWS Cloud您的客户通过接口 VPC 终端节点连接到 SaaS 产品。这是在客户的一个或多个子网中配置的端点网络接口。 AWS 账户在本指南的场景中，流量通过接口 VPC 终端节点，到达您账户中的 Network Load Bal ancer。Network Load Balancer 将流量转发到您已注册为终端节点服务的 SaaS 应用程序。通过资源 VPC 终端节点， AWS PrivateLink 还可以帮助您访问其他资源，例如数据库。

Amazon VPC Lattice

Amazon VPC Lattice 是一项应用程序联网服务，可帮助 SaaS 提供商安全、高效地向跨多个 VPCs 和 AWS 账户运营的客户提供服务。客户通过VPC Lattice访问您的SaaS产品，该产品可提供一致的网络连接、强大的访问控制和高级流量管理。在这些场景中，流量通过 VPC Lattice 流向您注册的应用程序服务。无论您使用哪种计算服务，它都能提供可扩展且安全的通信。

VPC 对等连接

VPC 对等互连是两个虚拟私有云 (VPCs) 之间的网络连接，它使用私有 IPv4 地址或 IPv6地址在它们之间路由流量。VPC 对等连接通常在可信实体之间使用，例如同一组织内的实体。您的客户向您的客户创建了对等互连请求。 VPCs当您接受它时，流量可以在两个 VPCs 方向之间流动。这种连接方法因其独特性而引人注目，因为它涉及两者之间的直接通信， VPCs 无需管理任何中介服务或基础架构。

AWS Transit Gateway

AWS Transit Gateway是一个集中式网络传输中心，可以连接 VPCs、虚拟专用网 (VPN) 连接、AWS Direct Connect 网关、VPC 中的第三方虚拟设备以及其他传输网关。传输网关的每个连接可以有不同的路由表。这为路由提供了最大的灵活性，还可以帮助您隔离网络。它通常用于将多个连接在 VPCs 一起或用于集中检查。

AWS Site-to-Site VPN

AWS Site-to-Site VPN可以使用 Internet 协议安全 (IPsec) 技术在本地网络、远程办公室、工厂、其他云提供商和 AWS 全球网络之间建立连接。连接是从 VPC 中的虚拟私有网关或传输网关与物理或基于软件的客户网关建立的，后者可以位于本地或其他 CSP 的云中。 AWS Cloud AWS Cloud可以通过互联网或物理 AWS Direct Connect 连接进行连接。也可以使用加速 Site-to-Site VPN 连接 AWS Global Accelerator。加速连接可将流量路由到 AWS 边缘位置，从而减少延迟并提高性能。

AWS Direct Connect

AWS Direct Connect在本地数据中心和之间建立高速私有连接 AWS Cloud。通过绕过公共互联网，Direct Connect 提供更可靠、更安全、更稳定的低延迟连接。 AWS Cloud客户连接到其中一个Direct Connect 地点，然后选择托管或专用连接 AWS。尽管对于SaaS产品来说，这是一种不常见的架构选择，但它可能非常适合那些只有很少但大型企业消费者的SaaS提供商。

比较服务能力

下表概述了本指南中讨论的支持的功能。 AWS 服务 以下是此表中包含的功能的描述：

- 重叠 CIDR 范围-可以连接两个或多个 CIDR 范围相同或重叠的网络
- 双向通信 — 可以支持双向通信渠道，这样 SaaS 使用者就可以向 SaaS 提供商公开内部资源，例如数据库
- IPv6— 可以支持单 IPv6堆栈或双堆栈
- 巨型帧 — 可以支持帧大小不超过 8,500 字节的巨型帧
- 混合云 — 可以支持与本地网络的连接
- 多云 — 可以支持不同云服务提供商的网络之间的连接

服务或方法	重叠的 CIDR 范围	双向通信	IPv6	巨型画面	混合云	多云

VPC 对等连接	没有	是	是	是的 [5]	没有	没有
AWS PrivateLink	是	是[1]	是	是	No [6]	No [6]
Amazon VPC Lattice	是	是[1]	是	是	No [6]	No [6]
AWS Transit Gateway	没有	是	是	是	是的 [3]	是的 [3]
AWS Site-to-Site VPN	没有	是	是	没有	是	是
AWS Direct Connect	没有	是	是	是[2]	是	是
公共互联网接入 [4]	不适用	没有	是	是	是	是

1. 在 Amazon VPC 中使用 VPC 资源莱迪思
2. 仅适用于私有和传输虚拟接口
3. 使用 Site-to-Site VPN 或 AWS Direct Connect 附件
4. 作为使应用程序可公开访问的 AWS 资源的统称，例如 Application Load Balancer
5. 仅适用于同一个内部的对等连接 AWS 区域
6. 可通过环境之间预先存在的第 3 层连接实现

安全功能和注意事项

下表概述了本指南 AWS 服务 中讨论的安全功能。

- 身份验证方式 — 如何确保只有您的客户才能连接到您的服务。通常仍需要对传入的请求进行另一级别的身份验证，尤其是在共享租户环境中。
- 传输中的加密-描述默认情况下是否提供传输中的加密。本机加密描述了为数据中心内 VPCs VPCs、跨数据中心或跨数据中心的所有流量 AWS 提供加密功能。补充加密描述的是您可以控制的加密，并且可以由相应的服务停止这些加密。

服务或方法	身份验证方式	传输中加密
VPC 对等连接	您可以向客户的 AWS 账户 和 VPC 发起对等互连请求或接受他们发起的请求。请参阅接受或拒绝 VPC 对等连接。	仅限本机加密
AWS PrivateLink	您可以选择允许 AWS 账户 哪些服务创建终端节点。这些账户被称为允许的委托人。请参阅接受或拒绝连接请求。	仅限本机加密
Amazon VPC Lattice	您与客户 AWS 账户共享VPC 莱迪思服务或服务网络。请参阅共享您的 VPC 莱迪思实体。	本机加密和补充 TLS 加密
AWS Transit Gateway	您的客户向其创建对等连接请求 AWS 账户，或者您发起请求。请参阅 Amazon VPC 传输网关中的公交网关对等连接附件。	本机加密和带有 VPN 附件的补充 IPsec 加密
AWS Site-to-Site VPN	您可以在客户的设备上使用 IPsec 预共享密钥或私有证书。请参阅AWS Site-to-Site VPN 隧道身份验证选项。	补充 IPsec 加密
AWS Direct Connect	您的客户从他们那里创建了一个虚拟接口请求 AWS 账户。请参阅Direct Connect 虚拟接口和托管虚拟接口。	在选定站点可以进行第 2 层补充加密。参见Direct Connect 地点。

公共互联网接入 [1]	需要自定义身份验证。	可以进行补充 TLS 加密

1. 作为使应用程序可公开访问的 AWS 资源的统称，例如 Application Load Balancer

评估 SaaS 产品的网络访问选项

对您的组织至关重要的指标将取决于您的客户是谁、您的业务策略和您的组织目标。本指南提供了可用于选择网络访问方法的指标，但您应该优先考虑那些满足您的用例独特要求的指标。

本节包含以下主题：

评估指标

有些指标在不同组织和用例之间是一致的，这些指标是我们可以帮助您进行评分的指标。以下是这些指标：

- 易于集成 — 您能以多快和轻松的方式吸引新客户？
- 总拥有成本 (TCO)-成本结构是什么？ 除了固定和可变基础架构成本外，还有一些与运营开销、对专家的依赖、实施变更的成本和合规性相关的额外成本考虑因素。想要了解更多信息，请参阅总拥有成本部分。
- 可扩展性 — 您的网络接入方法是否能够扩展以支持公司的发展？ 扩大客户群有重要的架构和组织考虑因素。考虑一下如何进行扩展，以容纳的客户数量是当今支持的 5-100 倍。
- 适应性 — 你能轻松实现变更吗？ 更改可能包括新应用程序、新功能、不同的平台或不同的网络。
- 网络隔离 — 您向客户暴露了多少网络基础架构？ 您提供的访问权限恰到好处，还是暴露了整个网络？ 如果您尽早隔离网络资源，以后可以更轻松地提供安全、隐私和合规保证。
- 可观察性 — 您检测服务故障或降级的能力如何？ 识别问题有多容易和快速？ 你能以多快的速度帮助客户了解他们的故障点并帮助他们解决故障（开销有多大）？
- 修复时间 — 从检测到服务故障或性能下降到恢复运营之间的交货时间是多少？ 影响这种能力的因素有哪些？

其他指标是您的组织或产品所独有的，因为它们与您的业务运营、战略或目标有关。只有您可以对这些指标进行评分。以下是这些指标：

- 业务模式协调 ——您的业务模式是什么，个人访问方法与之对应程度如何？

- 总可寻址市场 (TAM) — 您当前和未来的市场是什么，网络接入方法覆盖的程度如何？
- 投资回报率 (ROI) — 您预计盈利能力和利润率会有哪些改善？ 预期的经济收益是否足以满足您对适应性强、灵活的服务访问的需求？
- 监管合规 — 适用什么样的监管要求，适用于哪个市场？
- 服务级别协议 (SLAs) — 客户是否需要您的 SaaS 产品具有高可用性？ 根据合同，你有义务遵守什么样的承诺？

总拥有成本

本节探讨总拥有成本 (TCO)，这是用于比较网络接入方法的评估指标之一。TCO 是一个综合指标，包括固定和可变基础架构成本、运营开销、专家依赖性、变更成本和合规成本。

每种网络访问方法的 TCO 评级可能因您的用例而异。例如，拥有简单网络服务和五个租户的SaaS提供商的变更成本与拥有复杂、相互关联的产品组合和成百上千个租户的SaaS提供商的变更成本不同。此外，并非所有组件的重量都相同。例如，聘请网络专家通常比支持个性化部署服务的基础设施成本更高。使用下表中的值作为初始方向，并作为进一步讨论的参考点。

访问方法	固定基础设施成本	可变的基础设施成本	运营开销	专家依赖	变更成本	合规成本
VPC 对等连接	无	无	高	低	高	中
AWS PrivateLink	低	低	低	无	低	低
Amazon VPC Lattice	中	中	低	低	低	低
AWS Transit Gateway	中	中	低	低	低	中
AWS Site-to-Site VPN	中	高	高	中	中	低

AWS Direct Connect	高	中	中	高	高	低
公共互联网访问	低	高	中	低	低	高

VPC 对等互连费用

VPC 对等连接没有直接的基础设施成本。当流量停留在同一可用区内时，不收取数据传输费用。但是，运营开销可能很大，因为每增加一个对等连接，管理和复杂性就会呈指数级增长。对网络的一些基本了解足以建立对等连接，但是如果有多个对等连接，则很难在网络上进行更改。合规成本略高，因为双方都将整个 VPC 暴露给对方，而不是单独的服务。

AWS PrivateLink 成本

AWS PrivateLink 通常是一种经济实惠的解决方案，操作开销很小。这是因为 SaaS 提供商只能管理 Network Load Balancer，而使用者只能管理 VPC 终端节点。您可以透明地在双方进行更改，从而减少昂贵且资源密集型的跨组织协作。合规成本往往很低，因为SaaS提供商只公开他们想要的服务，而不是整个网络。

亚马逊 VPC Lattice 成本

Amazon VPC Lattice 提供平衡的成本结构，固定和可变基础设施成本适中。作为一个完全托管的服务网络，它通过自动执行多个 VPCs服务发现、流量管理和访问控制，显著降低了运营开销。与手动网络配置相比，这简化了初始部署和持续管理。您可以通过基于策略的控制来实施更改，而无需进行复杂的路由更新，从而减少了对网络专家的依赖。合规成本往往低于传统联网方法，因为VPC Lattice通过内置的监控和日志功能提供了精细的访问控制和全面的可见性。这可以更轻松地证明合规性。

AWS Transit Gateway 成本

AWS Transit Gateway 每小时和数据处理费用都高于 AWS PrivateLink，但操作开销相似。要正确设置所有路由表，您必须 AWS 对 AWS Transit Gateway 服务和路由有更深入的了解。基础设施变更可能需要路由或 DNS 更新。合规成本与 VPC 对等连接类似，因为双方都可能将子网或整个子网暴露 VPCs 给对方。 AWS Transit Gateway 还需要谨慎处理路由表，因为它们由多个使用者共享，并且您不得允许它们之间有任何流量。

AWS Site-to-Site VPN 成本

由于 Site-to-Site VPN 本质上是将流量发送到互联网，因此由于数据传输费用，相比之下，可变成本最高。尽管它是一项托管虚拟专用网络 (VPN) 服务，但它会带来巨大的运营开销，尤其是在客户网关上。配置和操作需要高级的网络知识，而更改通常需要双方采取行动。合规成本通常很低，因为安全团队通常无需额外审查即可预先批准 IPsec隧道。

AWS Direct Connect 成本

AWS Direct Connect 固定基础设施成本最高，因为它是直接连接到. 的私有物理连接 AWS Cloud。设置和操作边界网关协议 (BGP) 会话（如果需要）、操作 VPN 连接以及执行流量工程都需要专业知识。该服务减少了安全团队的工作量，因为它将私有连接与媒体访问控制安全 (MACsec) 和 IPsec 加密功能相结合。

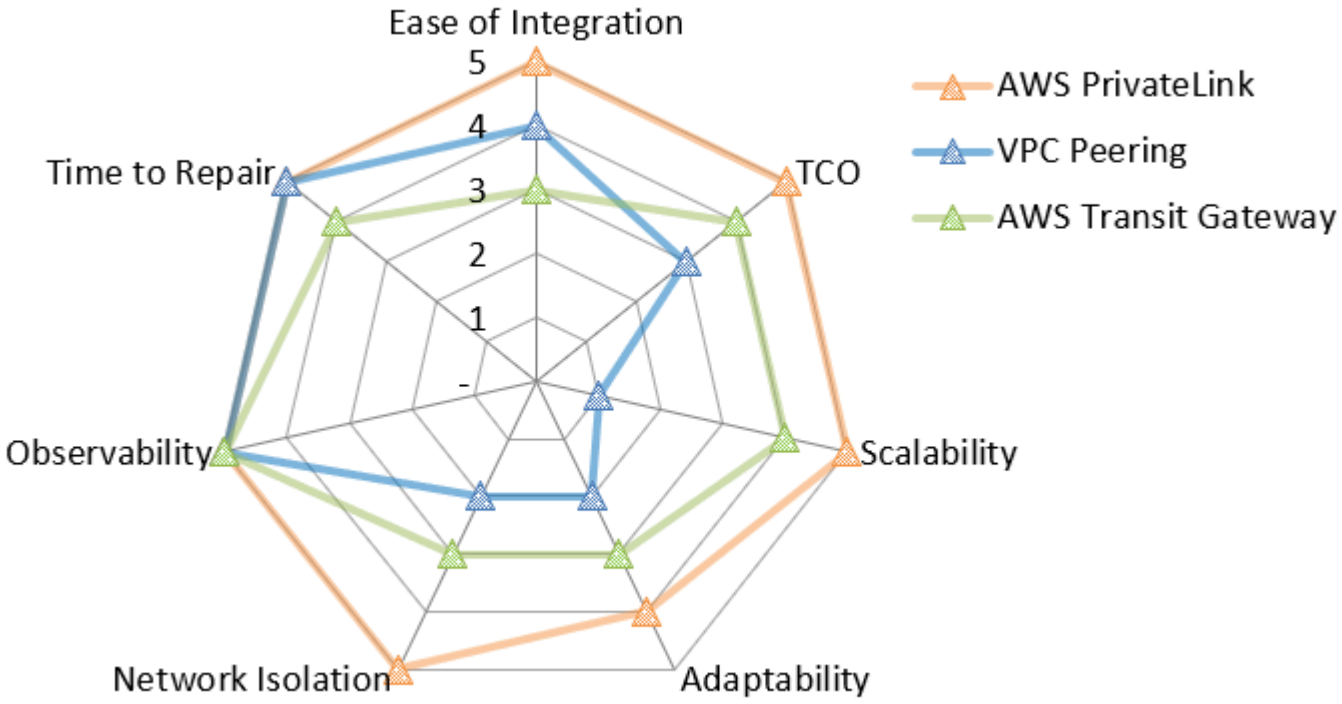
公共互联网接入费用

公共互联网访问是指可用于公开访问应用程序的 AWS 资源，例如 Application Load Balancer。对于这种方法，与提供服务访问权限相关的费用是可变的，包括将数据传输到互联网的费用。运营开销和合规成本可能很高，因为您将服务暴露在互联网上，并且需要额外的安全和身份验证机制。但是，不涉及复杂的路由，任何一方都不必知道彼此基础设施的详细信息。

网络价值地图

为了帮助您了解大局并做出明智的决策，本指南包括每个场景的网络价值图。由于不同场景的评分不同，因此同一个服务在两种情况下的评分可能会有所不同。价值图是雷达图，其中假设的满分是所有类别的五分。

例如，下图显示了雷达图示例。它仅包括我们可以帮助评估的指标。我们建议您创建自己的价值地图，其中包含只有您才能评估的其他指标。


Ease of Integration
5
4
3
2
1
-
TCO
Scalability
Adaptability
Network Isolation
Observability
Time to Repair
AWS PrivateLink
VPC Peering
AWS Transit Gateway

中 SaaS 产品的网络访问场景 AWS Cloud

本节介绍了 SaaS 产品的不同网络访问选项 AWS Cloud。它从消费者的角度讨论了这些方法，他们可能在内部 AWS Cloud、本地数据中心或其他云服务提供商那里有连接需求（CSPs）。此外，您可能需要支持来自多种类型的消费者环境的访问。

了解这些不同环境中的网络连接要求对于制定全面的接入策略至关重要。您的架构决策必须考虑不同的安全模型、性能预期和技术限制，同时保持运营效率。正确的方法可以提供安全、可靠的连接，该连接可随着您的业务增长而扩展，并最大限度地降低实施复杂性和持续的管理开销。

在评估网络接入选项时，请考虑每种方法如何影响您的总拥有成本，不仅包括基础设施成本，还包括运营开销和合规性要求。有些方法在可扩展性方面表现出色，但可能会带来复杂性，而另一些方法则以牺牲网络隔离为代价，优先考虑易于集成。消费者的技术能力和资源在确定最合适的解决方案方面也起着重要作用。

对于消费者而言 AWS Cloud，诸如 AWS PrivateLink 此类的服务在安全性和可扩展性方面具有显著的优势。本地消费者可以从中受益于稳 AWS Direct Connect 定的性能，也可以从 Site-to-Site VPN中受益以实现经济高效的连接。多云场景通常需要仔细考虑互操作性挑战，您可以使用传输 VPC 架构来标准化访问模式。在所有情况下，您的设计都应预测未来的消费者和流量增长，以便随着SaaS产品的发展，您的网络架构保持弹性和适应性。

本节包含以下场景：

- SaaS 消费者在以下平台上进行操作 AWS
- 在场所运营的服务消费者
- 在其他云服务提供商上运营的 SaaS 消费者
- 支持混合环境

SaaS 消费者在以下平台上进行操作 AWS

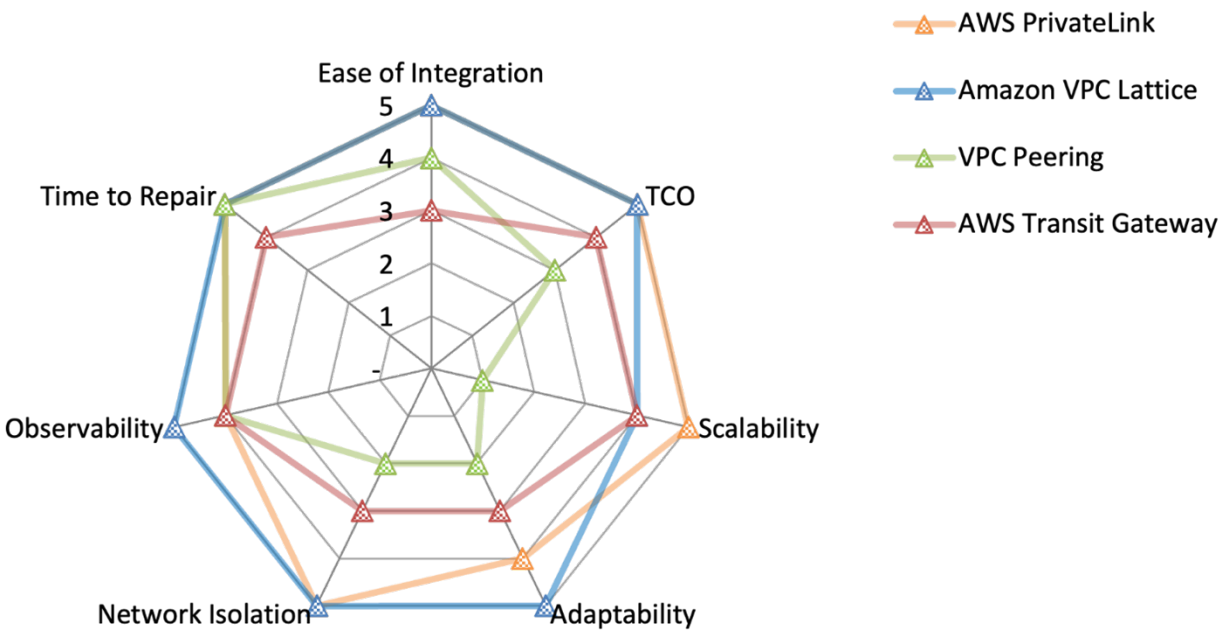
本节讨论如果您和您的消费者都在中操作时的连接选项 AWS Cloud。这种情况提供了最大的灵活性，因为许多方案都是 AWS 服务 本地集成的，而且双方都可以访问整个 AWS 服务 产品组合。

本节讨论以下网络访问方法：

- 与集成 AWS PrivateLink
- 共享 Amazon VPC 莱迪思服务
- 创建 VPC 对等连接

- VPCs 与... 连接 AWS Transit Gateway

以下网络价值图汇总了每个评估指标中每个选项的得分情况。有关评估指标的更多信息，请参阅本指南中的评估指标。在地图中，五表示最高分数，例如最低的 TCO、最佳的网络隔离或最低的修复时间。有关如何阅读此雷达图的更多信息，请参阅本指南网络价值地图中的。



雷达图显示以下值。

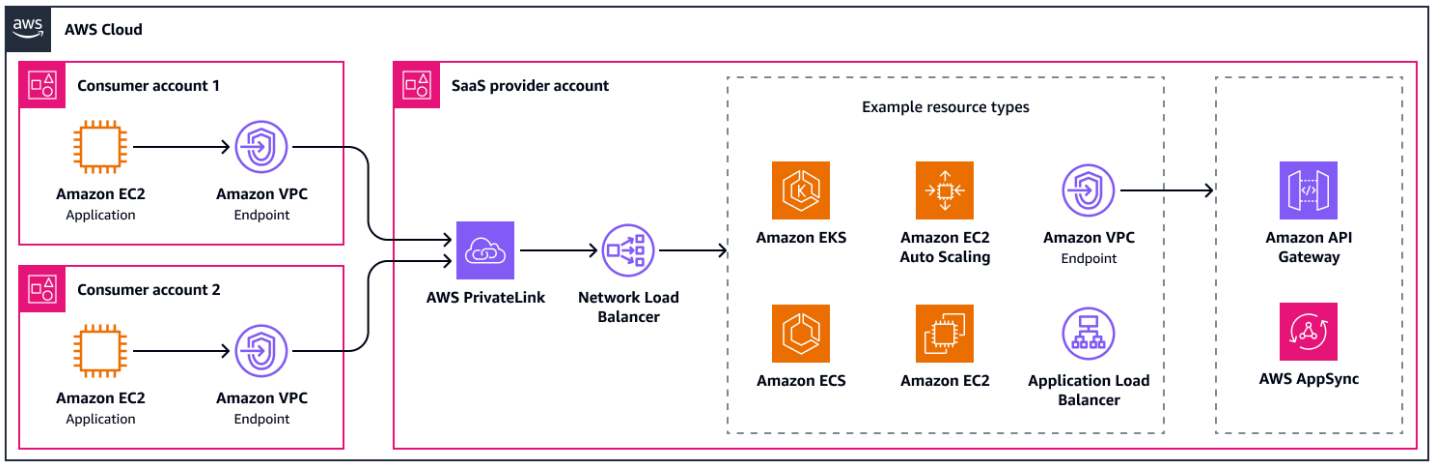
评估指标	AWS PrivateLink	Amazon VPC Lattice	VPC 对等连接	AWS Transit Gateway
易于集成	5	5	4	3
TCO	5	5	3	4
可扩展性	5	4	1	4
适应性	4	5	2	3
网络隔离	5	5	2	3
可观察性	4	5	4	4

评估指标	AWS PrivateLink	Amazon VPC Lattice	VPC 对等连接	AWS Transit Gateway
是时候修理了	5	5	5	4

与集成 AWS PrivateLink

AWS PrivateLink是集成 SaaS 产品的最云原生方式。SaaS 提供商可以将其应用程序托管在 Network Load Balancer 后面。网络负载均衡器直接与应用程序负载均衡器、亚马逊弹性容器服务 (Amazon ECS)、亚马逊 Elastic Kubernetes Service (Amazon EKS) 和 Auto Scalin g 组集成。也可以将流量从 Network Load Balancer 路由到 SaaS 提供商账户中的接口 VPC 终端节点。这可以帮助您使用 API 访问应用程序，例如通过 Amazon API Gateway 或AWS AppSync。如果您的应用程序需要访问客户环境中未进行负载平衡的资源（例如数据库），则可以使用资源 VPC 终端节点。

AWS PrivateLink 每个可用区支持高达 100 Gbps 的带宽。下图显示了基本配置，其中包含一些可能的集成。它通过将两个消费者帐户连接到SaaS提供商帐户 AWS PrivateLink。消费者账户中有服务端点，SaaS 提供商账户中有一个 Network Load Balancer。



这种方法的优点如下：

- 易于集成：无需更改路由表
- 易于集成：您可以通过以下方式提供端点服务 AWS Marketplace
- 易于集成：VPC 终端节点支持友好的 DNS 名称
- 可扩展性：它可以扩展到成千上万的 SaaS 用户
- 适应性：Support 支持重叠的 CIDR 范围
- 适应性：Support for IPv6

- 适应性：跨区域支持
- TCO： AWS PrivateLink 是一项完全托管的服务，因此所需的运营工作量更少
- 网络隔离：由于无法从 SaaS 提供商启动流量，因此可为 SaaS 消费者带来安全优势
- 网络隔离：SaaS 提供商可以获得安全优势，因为他们不会暴露整个子网或 VPC

以下是这种方法的缺点：

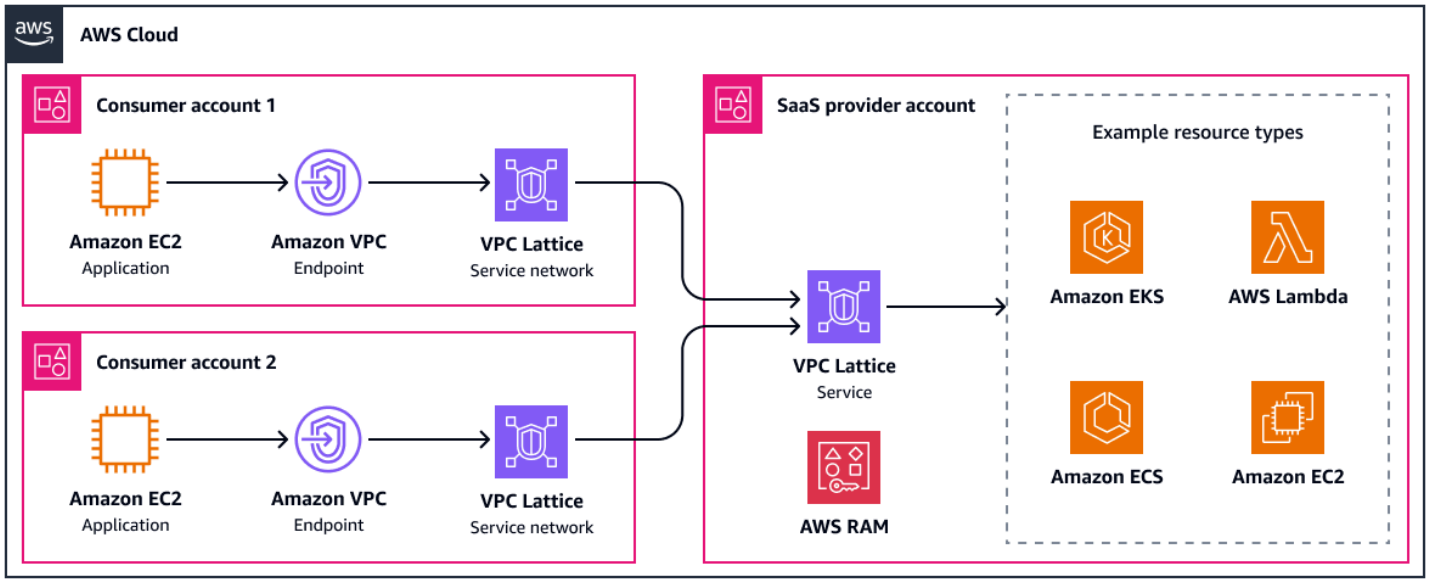
- 适应性：SaaS 提供商必须使用与消费者相同的可用区
- 适应性：仅支持客户端启动的连接，服务启动的通信需要资源 VPC 端点
- 适应性：Network Load Balancer 是唯一的直接集成 AWS PrivateLink

共享 Amazon VPC 莱迪思服务

要使用 Amazon VPC Lattice 作为 SaaS 应用程序的连接选项，您需要先创建一个或多个 VPC 莱迪思服务来代表您的 SaaS 应用程序组件。您可以配置侦听器和路由规则，将流量引导至您的后端目标，例如 Amazon EC2 实例、容器或 AWS Lambda 函数。有关更多信息，请参阅在 VPC莱迪思服务网络中连接Saas服务（AWS 博客文章）。从概念上讲，这与配置 Application Load Balancer 几乎相同。然后，您可以使用 AWS Resource Access Manager (AWS RAM) 与客户 AWS 账户 或组织安全地共享您的 SaaS 服务，指定他们拥有的权限。客户接受资源共享后，他们可以将您的SaaS服务与其现有或新创建的VPC Lattice服务网络关联以实现 service-to-service通信。

每个 VPC 莱迪思服务可支持高达 10 Gbps 的速度和每个可用区每秒 10,000 个请求。通过实施身份验证策略，您的客户可以精细控制哪些服务和资源可以访问 SaaS 应用程序。您可以使用资源网关来访问需要 TCP 连接的资源。例如，这可能是您管理的 Amazon EKS 集群，也可能是您的应用程序需要访问的客户管理的资源。有关为 SaaS 产品使用资源网关的更多信息，请参阅AWS 账户 使用对 VPC 资源的 AWS PrivateLink 支持将 SaaS 功能扩展到各处（AWS 博客文章）。

下图显示了VPC Lattice的高级配置以及一些示例集成。它使用客户管理的服务网络来访问SaaS应用程序。



这种方法的优点如下：

- 易于集成：无需更改路由表
- 易于集成：开箱即用的服务发现
- 可扩展性：它可以扩展到成千上万的 SaaS 用户
- 适应性：Support 支持重叠的 CIDR 范围
- 适应性：Support for IPv6
- 适应性：作为VPC莱迪 AWS 思服务与任何计算服务集成
- 总拥有成本：VPC Lattice 是一项完全托管的服务，因此所需的运营工作量更少
- TCO：内置负载均衡和高级流量路由
- 网络隔离：使用身份验证策略进行细粒度授权
- 网络隔离：由于无法从 SaaS 提供商启动流量，因此可为 SaaS 消费者带来安全优势
- 网络隔离：SaaS 提供商的安全优势，因为您不会暴露整个子网或 VPC

以下是这种方法的缺点：

- 适应性：仅支持客户端启动的连接，服务启动的通信需要资源网关
- 适应性：不支持跨区域

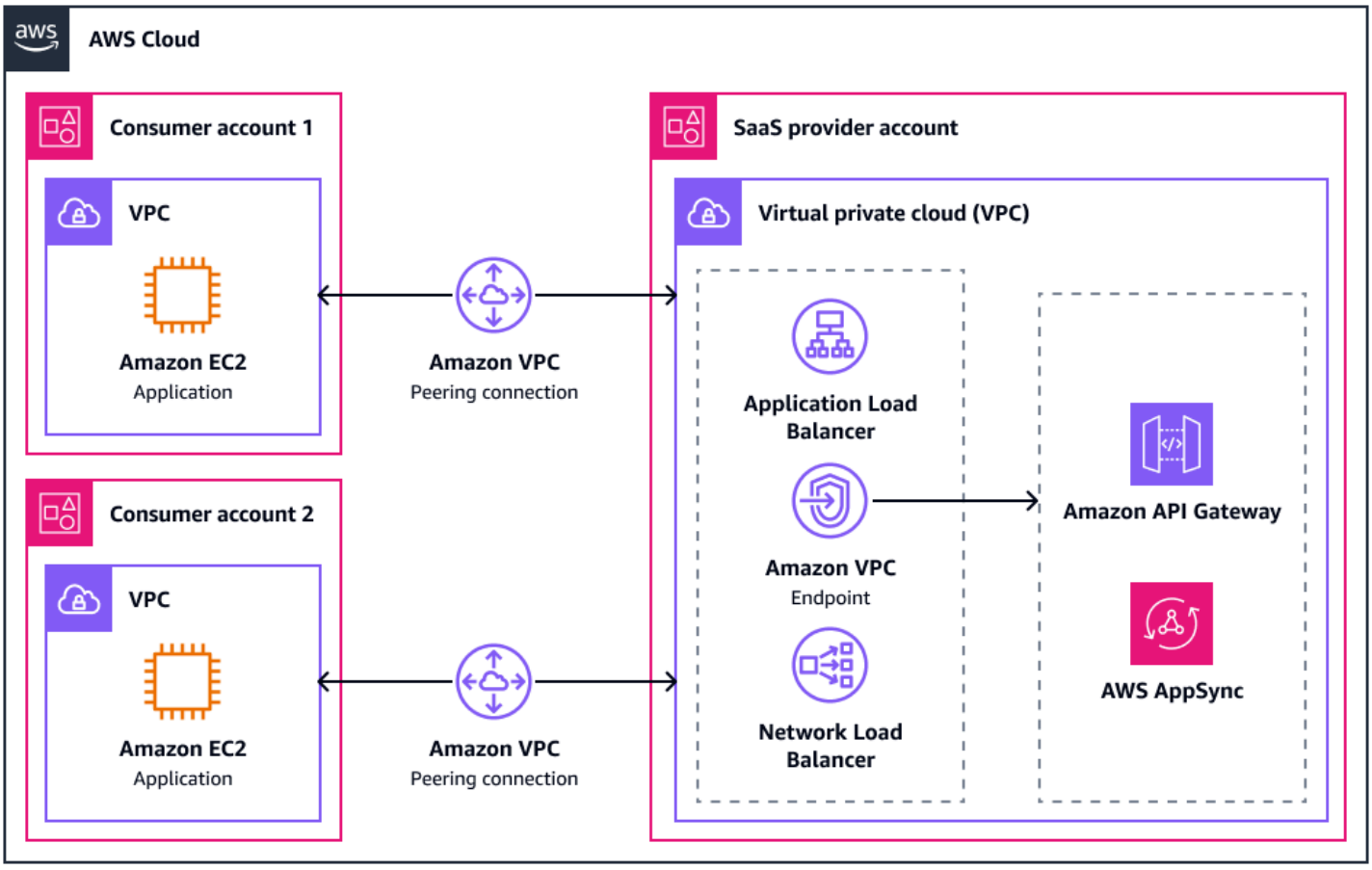
创建 VPC 对等连接

当您使用 VPC 对等连接将 SaaS 提供商的 VPC 与使用者的 VPC 连接时，双方都可以启动连接。这需要在两个账户中正确配置安全组、防火墙和网络访问控制列表 (NACLs)。否则，有害流量可能会通过对等连接进入网络。您可以使用安全组从对等设备引用安全组。 VPCs这可以帮助您控制对应用程序的访问权限，因为与允许名单 IP 地址相比，允许名单安全组提供了更明确、更精细的访问控制。

通过 VPC 对等互连，可以通过部署在 VPC 中的服务或资源访问 SaaS 产品。大多数 SaaS 应用程序都位于应用程序负载均衡器或网络负载均衡器之后。 AWS AppSync 私有 APIs或 Amazon API Gateway 私 APIs有版是 SaaS 应用程序的其他常见入口点，因为它们可以通过接口 VPC 终端节点通过对等连接成为目标。

建立对等连接后，必须更新两个账户 VPCs 中的路由表，将对等连接定义为相应 CIDR 范围的下一跳。由于管理多个对等连接很快就会变得过于复杂，因此仅推荐使用者较少的 SaaS 提供商使用此解决方案。

下图显示了基本配置，其中包含一些可能的集成。 VPCs 有两个消费者账户与 SaaS 提供商账户中的 VPC 建立了对等连接。



这种方法的优点如下：

- 修复时间：通信没有单点故障
- 可扩展性：VPC 对等互连没有带宽限制
- TCO：在同一可用区内，对等连接或通过对等连接的流量不收取任何费用
- TCO：无需管理基础架构
- 适应性：Support for IPv6
- 适应性：支持区域间对等互连

以下是这种方法的缺点：

- 适应性：不支持传递路由
- 适应性：不支持重叠的 CIDR 范围
- 可扩展性：可扩展性有限（每个 VPC 最多 125 个对等连接）
- TCO：每增加一个对等连接，复杂性就会呈指数级增长
- TCO：管理路由表、对等连接本身、安全组规则和流量检查产生的开销
- 网络隔离：由于双方都处于暴露状态 VPCs ，因此需要严格的安全控制

VPCs 与... 连接 AWS Transit Gateway

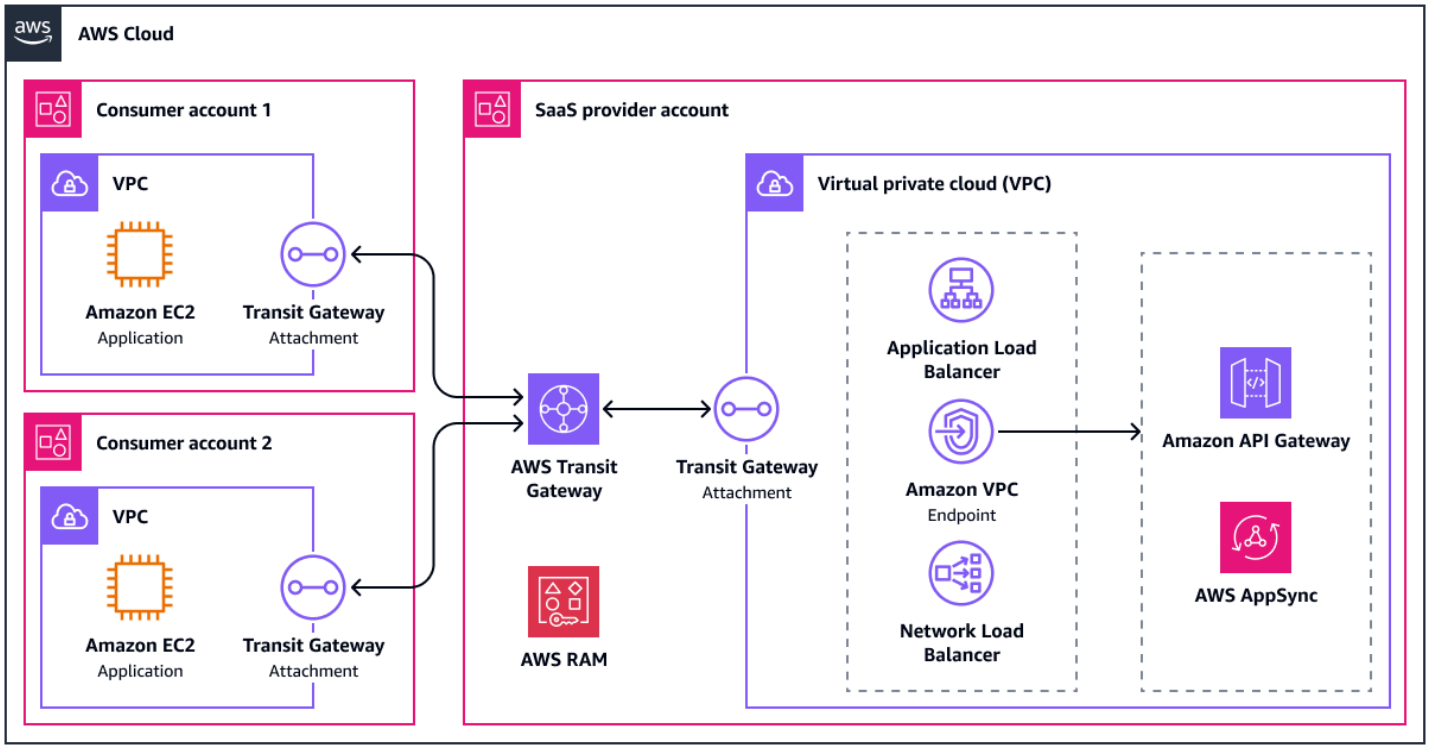
当您 VPCs 通过连接时 AWS Transit Gateway，它会创建 VPC 附件，并在每个可用区的子网中部署网络接口，用于路由进出该 VPC 的流量。建议在每个可用区中为 VPC 连接设置一个专用/28子网。有关更多信息，请参阅 Amazon VPC 传输网关设计最佳实践。 VPCs 需要更新路由表才能通过部署的网络接口发送流量，并需要相应地更新 Transit Gateway 路由表。在多租户配置中，您希望 SaaS 提供商的 VPC 具有通往所有消费 VPCs者的路由。消费者 VPCs应该只拥有通往 SaaS 提供商的 VPC 的路由。

Transit Gateway 的设计高度可用。它支持使用 VPC 流日志进行监控，Transit Gateway 连接的最大带宽为每个可用区 100 Gbps。与 VPC 对等连接一样，这种方法支持跨VPC安全组引用，从而简化了环境之间的访问控制。

使用 Transit Gateway 将消费者与你的 SaaS 产品联系起来，主要有两种选择。

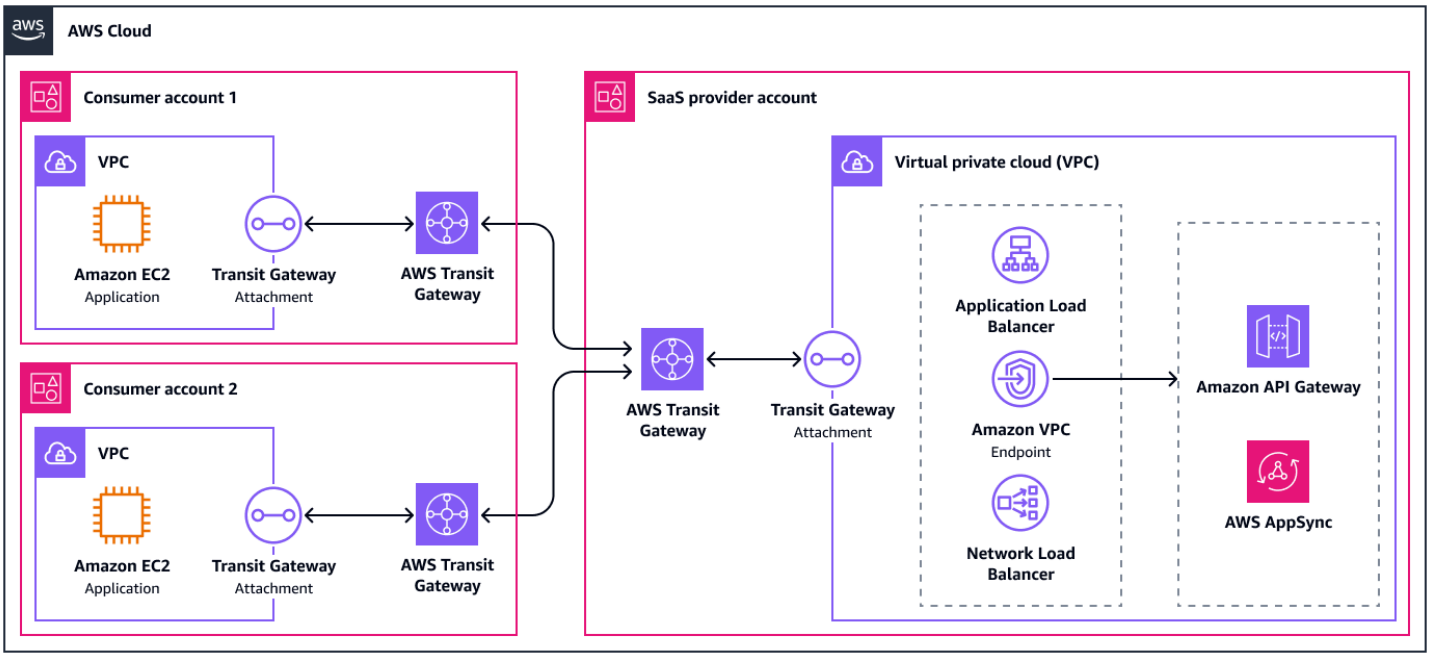
选项 1：使用内存

在第一个选项中，服务提供商使用 AWS Resource Access Manager (AWS RAM) 与消费者共享 Transit Gat eway。这允许使用者在自己的账户中部署 VPC 附件。下图从较高的层面上显示了此选项。



选项 2：对等传输网关

第二种选择是将您的公交网关与消费者账户中的公交网关对等。这为消费者提供了更大的灵活性，因为他们现在可以完全控制公交网关内的路由表。例如，他们可以在服务与其工作负载之间设置集中检查。此选项的一个缺点是仅支持传输网关之间的静态路由。下图从较高的层面上显示了此选项。



这种方法的优点如下：

- 可扩展性：支持多达 5,000 个附件
- 可扩展性：一站式管理和监控所有联网设备 VPCs
- 适应性：Transit Gateway 还可以连接到 VPNs Direct Connect 网关和第三方 SD-WAN 设备
- 适应性：灵活的架构，例如添加检查 VPC
- 适应性：Support 支持传递路由
- 适应性：能否对等区域内和区域间中转网关
- 适应性：Support for IPv6
- TCO： AWS Transit Gateway 是一项完全托管的服务，因此所需的运营工作量更少
- TCO：每增加一个公交网关连接，总拥有成本就会呈线性增长

以下是这种方法的缺点：

- 易于集成：路由配置需要高级网络知识
- 适应性：不支持重叠的 CIDR 范围
- TCO：管理路由表条目、安全组规则和流量检查产生的开销
- 安全：由于双方都处于危险之中 VPCs ，因此需要严格的安全控制

在场所运营的服务消费者

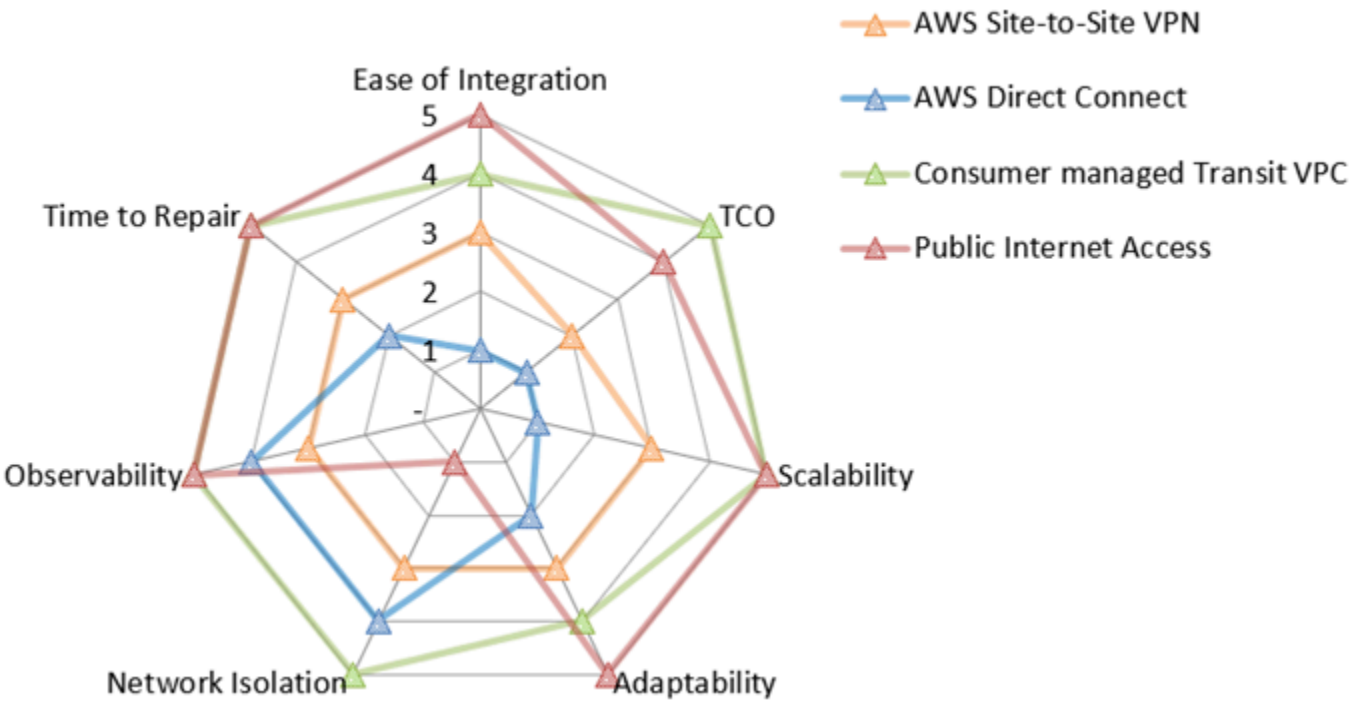
本节讨论本地数据中心内的 SaaS 工作负载之间的连接选项。 AWS Cloud 许多有本地需求的消费者，尤其是企业层面的消费者，将云视为其物理网络的延伸，他们希望将其反映在架构中。这意味着通过逻辑隧道甚至通过私有物理连接与云端SaaS产品的私有连接。其他消费者将接受通过公共互联网进行连接，本节也将对此进行讨论。

本节讨论以下网络访问方法：

- 与... 连接 AWS Site-to-Site VPN
- 与... 连接 AWS Direct Connect
- 连接传输 VPC 架构
- 通过公共互联网连接

以下网络价值图汇总了每个评估指标中每个选项的得分情况。有关评估指标的更多信息，请参阅本指南中的评估指标。在地图中，五表示最高分数，例如最低的 TCO、最佳的网络隔离或最低的修复时间。有关如何阅读此雷达图的更多信息，请参阅本指南网络价值地图中的。

> **Note**
>
> 不包括提供商管理的公交 VPC 选项，因为分数在很大程度上取决于正在运营的服务。



雷达图显示以下值。

评估指标	AWS Site-to-Site VPN	AWS Direct Connect	消费者管理的传输 VPC	公共互联网访问
易于集成	3	1	4	5
TCO	2	1	5	4
可扩展性	3	1	5	5
适应性	3	2	4	5
网络隔离	3	4	5	1
可观察性	3	4	5	5
是时候修理了	3	2	5	5

与... 连接 AWS Site-to-Site VPN

AWS Site-to-Site VPN连接可以在虚拟专用网关或传输网关上终止。虚拟专用网关是 VPN 连接 AWS 侧的 Site-to-Site VPN 终端节点，可以连接到单个 VPC。传输网关是一个交通枢纽，可用于互连多个 VPCs 本地网络。它也可以用作 VPN 连接 AWS 侧的 Site-to-Site VPN 端点。本节讨论这两个选项。

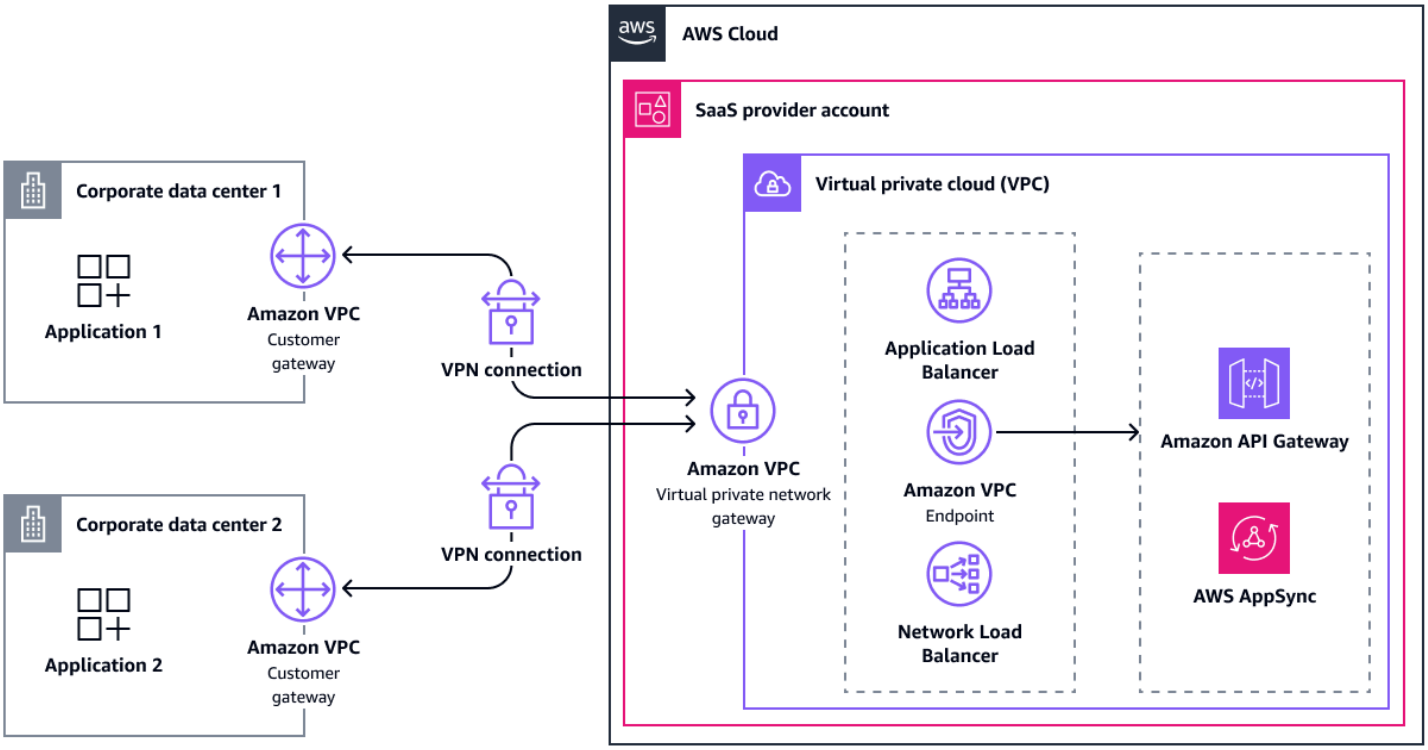
通过虚拟专用网关进行连接

创建虚拟私有网关后，将其连接到包含您的 SaaS 产品的 VPC。然后，您可以启用路由传播，将 VPN 路由传播到 VPC 路由表。这些路由可以是静态路由，也可以是 BGP 通告的动态路由。

为了实现高可用性， Site-to-SiteVPN 连接有两个 VPN 隧道，这些隧道终止在 AWS 侧面的两个可用区中。如果一条隧道不可用，则第二条隧道可以接管。单个隧道允许的最大带宽为 1.25 Gbps。由于虚拟专用网关不支持等价多路径路由 (ECMP)，因此您一次只能使用一条隧道。

要提高容错能力，您可以设置与第二个物理客户网关的第二个 VPN 连接。建立连接后，使用者可以访问 SaaS 提供商的 VPC 中的资源。

下图显示了这种架构。



这种方法的优点如下：

- 修复时间：托管故障转移到辅助 VPN 隧道

- 可观察性：使用网络合成监视器集成托管主动监控
- 易于集成：通过 BGP 支持动态路由
- 适应性：与大多数本地网络设备兼容
- 适应性：支持 IPv6
- TCO： AWS Site-to-Site VPN 是一项完全托管的服务，因此所需的运营工作量更少
- TCO：虚拟网关不收费，但每个网关上的两个公有 IPv4 地址都要收费
- 网络隔离：支持通过互联网进行安全的私人通信

以下是这种方法的缺点：

- 易于集成：消费者必须配置其客户网关
- 可扩展性：缺乏 ECMP 支持将每个虚拟网关的带宽限制为 1.25 Gbps
- 可扩展性：由于网络复杂性和运营开销增加，扩展受限
- 适应性：仅IPv6 支持 VPN 隧道的内部 IP 地址
- 适应性：没有传递路由
- TCO：为 SaaS 提供商维护、管理和配置大量 VPN 连接所需的运营开销

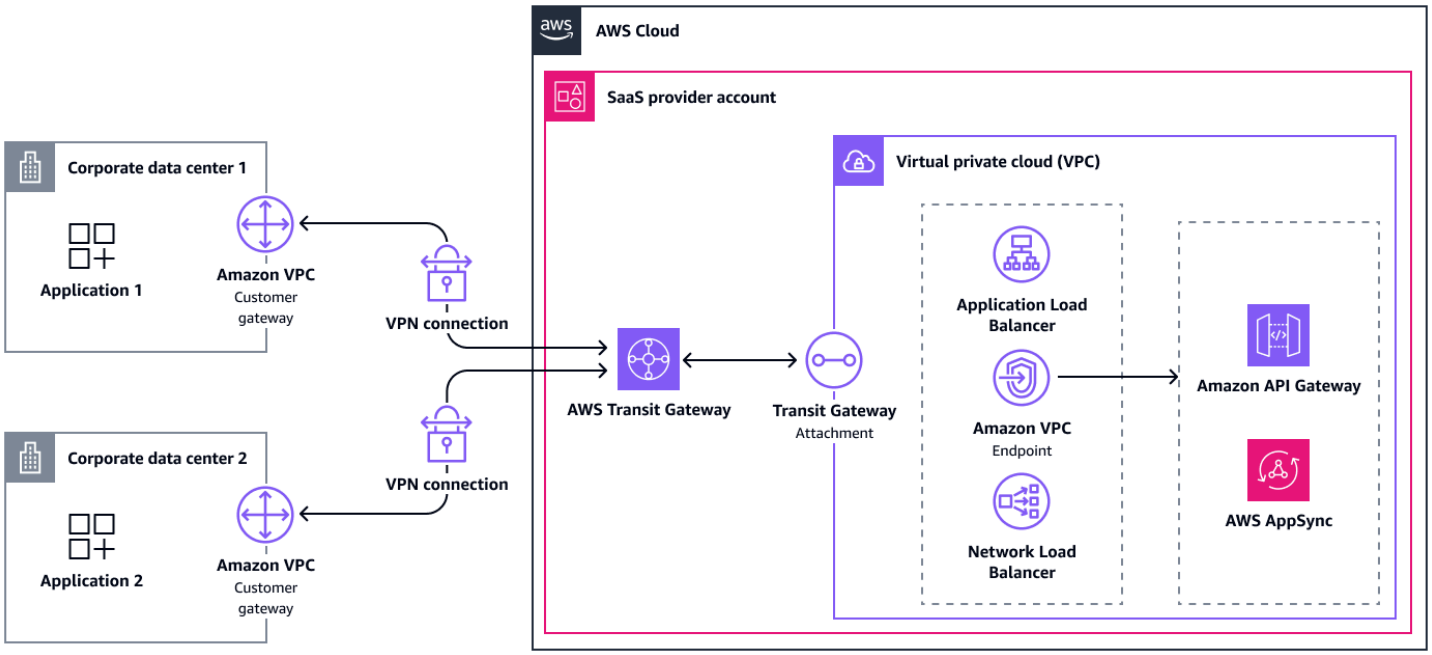
通过中转网关进行连接

通过中转网关进行的连接类似于虚拟网关。但是，有一些区别需要记住。

首先，可以在传输网关路由表中自动传播 VPN 连接的路由，但您必须手动将路由添加到连接 VPCs的路由中。

与虚拟网关相比，Transit Gateway 支持 ECMP。如果客户网关支持 ECMP，则它可以使用两条隧道来实现 2.5 Gbps 的最大总吞吐量。您可以在同一个本地网络与传输网关之间建立多个连接。使用这种方法，您最多可以将每个连接的最大带宽增加 2.5 Gbps。

下图显示了这种架构。



这种方法的优点如下：

- 修复时间：托管故障转移到辅助 VPN 隧道
- 可观察性：使用网络合成监视器集成托管主动监控
- 易于集成：通过 BGP 支持动态路由
- 可扩展性：ECMP 支持允许扩展 VPN 吞吐量以满足较大的带宽需求
- 可扩展性：单个传输网关支持大量 VPN 连接（最多可达 5,000 个）
- 可扩展性：一站式管理和监控所有 VPN 连接
- 适应性：与大多数本地网络设备兼容
- 适应性：支持 IPv6
- 适应性：继承灵活性 AWS Transit Gateway
- TCO： AWS Transit Gateway 是一项完全托管的服务，因此所需的运营工作量更少
- TCO：虚拟网关不收费，但每个网关上的两个公有 IPv4 地址都要收费
- 网络隔离：支持通过互联网进行安全的私人通信

以下是这种方法的缺点：

- 易于集成：消费者必须配置其客户网关
- 可扩展性：由于网络复杂性和运营开销增加，扩展受限

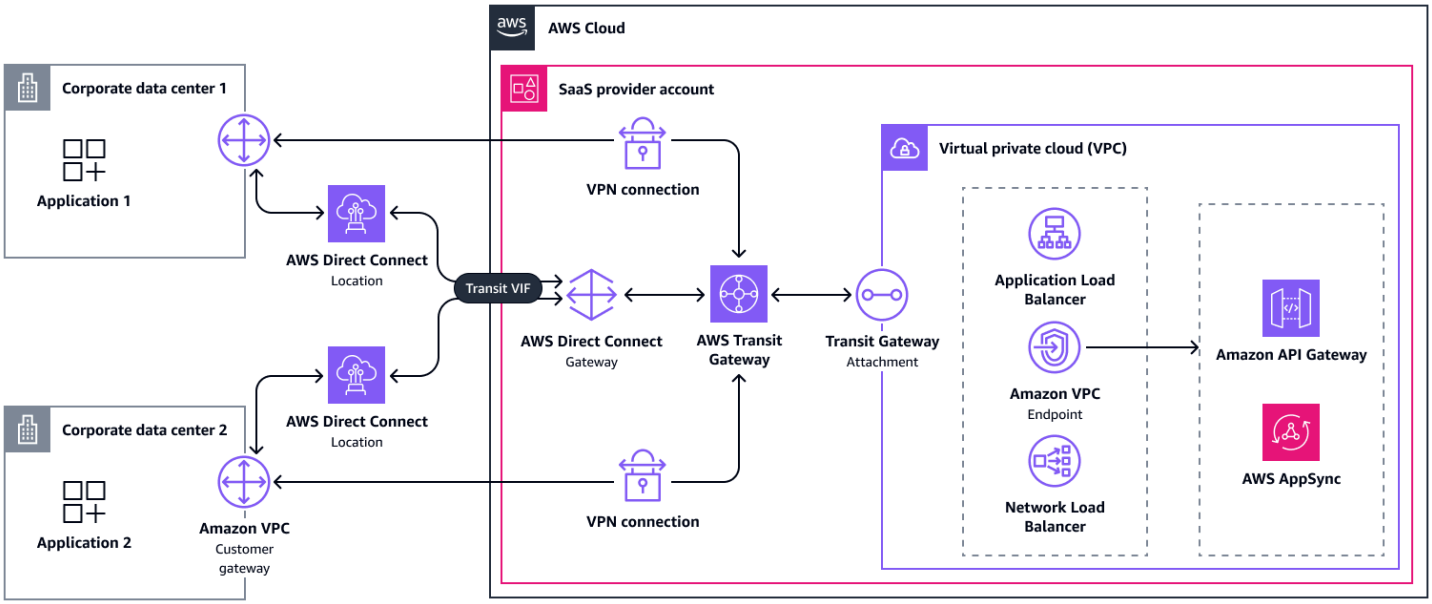
- 适应性：仅IPv6 支持 VPN 隧道的内部 IP 地址
- TCO：为 SaaS 提供商维护、管理和配置大量 VPN 连接所需的运营开销
- TCO：使用时需额外收费 AWS Transit Gateway
- TCO：管理公交网关路由表更加复杂

与... 连接 AWS Direct Connect

AWS Direct Connect通过标准的以太网光纤电缆将您的内部网络链接到某个 Direct Connect 位置。与其他架构选项不同，无法在几分钟内建立专用连接。相反，如果满足所有要求，则此过程最多可能需要几天的时间。否则，可能需要更长的时间。因此，我们建议您联系您的 AWS 客户团队或寻求 AWS 支持 有关此方法的帮助。或者，您可以选择由 AWS 合作伙伴提供并与其他客户共享的托管连接。无论如何，架构都是一样的。您 Direct Connect 之所以选择，是因为它可以减少延迟、提高带宽或符合监管要求。

要使用该 Direct Connect 连接，消费者必须创建公共、私有或传输虚拟接口。有不同的架构选项可供选择。将多个本地位置连接到的最灵活的方法 AWS Cloud 是连接到Direct Connect 网关的传输虚拟接口。 Direct Connect 网关是一个全球性的逻辑组件，允许服务提供商将多达六个传输网关连接到网关。此外，您最多可以将 30 个虚拟接口连接到网关。为了实现扩展，您可以创建其他 Direct Connect 网关。如前所述，在 SaaS 提供商账户中 VPCs，传输网关随后会附加到。

消费者可以使用一到四个 Direct Connect 连接从总共一两个Direct Connect 位置进行连接，具体取决于所需的弹性级别。有关更多信息，请参阅配置 Direct Connect 以实现最大弹性。通过 Internet 建立的 AWS Site-to-Site VPN 连接也可以作为低成本的 Direct Connect 连接备用路径。支持的 Direct Connect 专用连接可用于MACsec加密第 2 层上该 Direct Connect 位置和数据中心之间的链路。为了提高数据的机密性，通常会有 Site-to-Site VPN 连接。使用普通的 Site-to-Site VPN 连接，可以在传输网关上终止 VPN 连接。下图显示了这种架构。



这种方法的优点如下：

- 可观察性：使用网络合成监视器集成托管主动监控
- 可扩展性：Support 支持更高的带宽吞吐量
- 适应性：支持 IPv6
- TCO：有可能减少数据传输
- 总拥有成本：一致的网络体验
- 网络隔离：可满足监管要求的私有连接

以下是这种方法的缺点：

- 易于集成：设置所需的时间和手动操作
- 可扩展性：由于有多个配额需要跟踪，因此可扩展性有限，超过数十个 Direct Connect 连接
- 适应性：配置选项取决于可用位置 Direct Connect
- TCO：定期 Direct Connect 维护可能会导致停机，需要采取措施

连接传输 VPC 架构

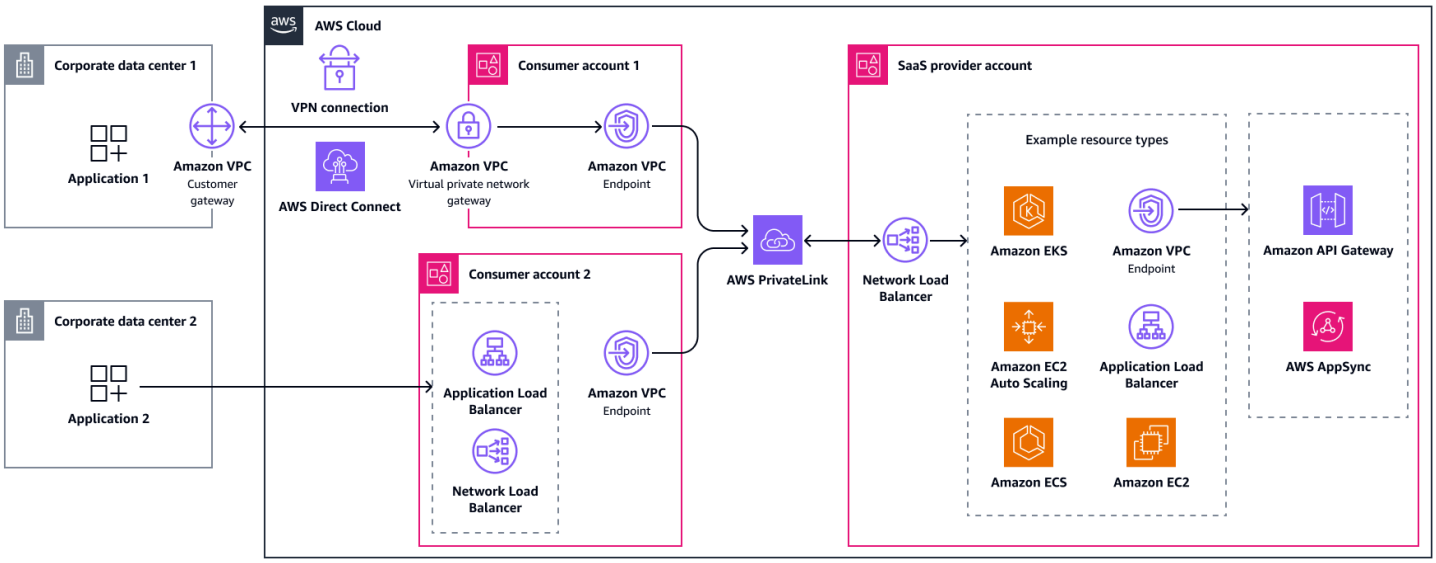
Transit VPC 是一种架构选项，它为消费者提供了连接方式的灵活性 AWS，它允许 SaaS 提供商通过统一访问其服务而受益 AWS PrivateLink。使用者从内部连接到仅包含入口点（例如虚拟私有网关）和

接口 VPC 终端节点（一种 AWS PrivateLink 资源）的中转 VPC 终端节点。公交 VPCs 应归SaaS提供商或消费者所有。本节讨论这两个选项。

您可以创建与本地数据中心兼容的 CIDR 范围的传输 VPC 和子网。如果用户需要私有连接，则可以通过 AWS Direct Connect 或连接到该 VPC AWS Site-to-Site VPN。您还可以使用指向 VPC 终端节点的 Application Load Balancer 或 Network Load Balancer 配置从公共互联网访问传输账户。

消费者管理的传输 VPC

在这种方法中，SaaS提供商将公交的管理留给 VPCs 了消费者。从技术角度来看，SaaS提供商的架构与 AWS Cloud 通过连接消费者时的架构相同 AWS PrivateLink。从销售和产品的角度来看，这需要付出额外的努力，因为有些消费者 AWS 账户 还没有。他们可能对开设和操作账户犹豫不决。SaaS 提供商应就如何创建 AWS 账户 和连接其本地数据中心向其消费者提供指导。下图显示了公共和私人接入的组合，其中消费者拥有公交系统 VPCs。



这种方法的优点如下：

- 该修复了：运营开销在很大程度上转移给了 SaaS 消费者
- 适应性：SaaS 消费者可以从不同的访问选项中进行选择
- 适应性：即使在使用 Site-to-Site VPN 时也没有 CIDR 范围冲突 Direct Connect
- 所有指标：服务提供商继承权益 AWS PrivateLink

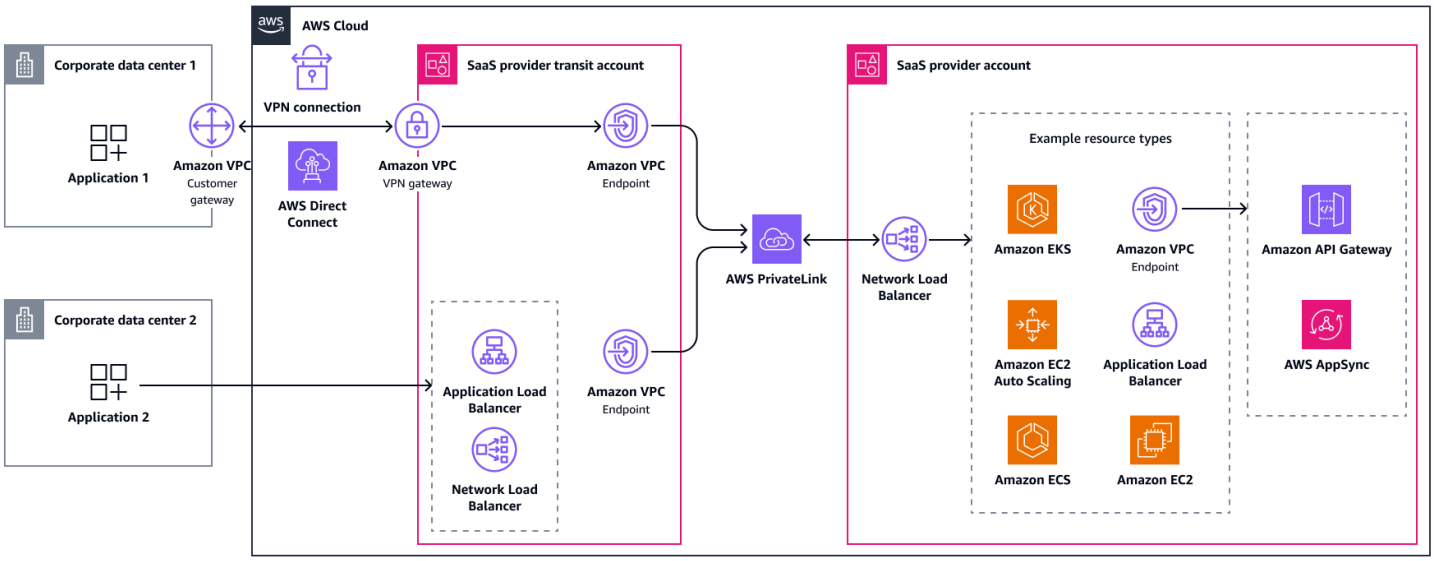
以下是这种方法的缺点：

- 易于集成：SaaS 消费者至少需要一个 AWS 账户

- TCO：传输 VPC 是一种架构，而不是完全托管的服务，因此需要更多的运营工作

提供商管理的传输 VPC

这种方法使用相同的技术，但账户界限和职责会发生变化。在这里，SaaS提供商拥有公交 VPCs，最好与SaaS产品分开存放。这种脱钩降低了成本，降低了风险，并允许公交账户独立扩展。对于需要高度隔离的环境，您可以使用子网或为每个使用者创建单独的中转 VPC，从而在租户之间建立额外的隔离。然后，使用者可以选择如何连接到传输 VPC。这种方法为扩大整个潜在市场提供了更多选择，但由于需要操作和监控额外的架构组件，它对SaaS提供商来说具有更高的总体拥有成本。



这种方法的优点如下：

- 适应性：SaaS 消费者可以从不同的访问选项中进行选择
- 适应性：SaaS 消费者不必拥有 AWS 账户
- 适应性：即使在使用 Site-to-Site VPN 时也没有 CIDR 范围冲突 Direct Connect

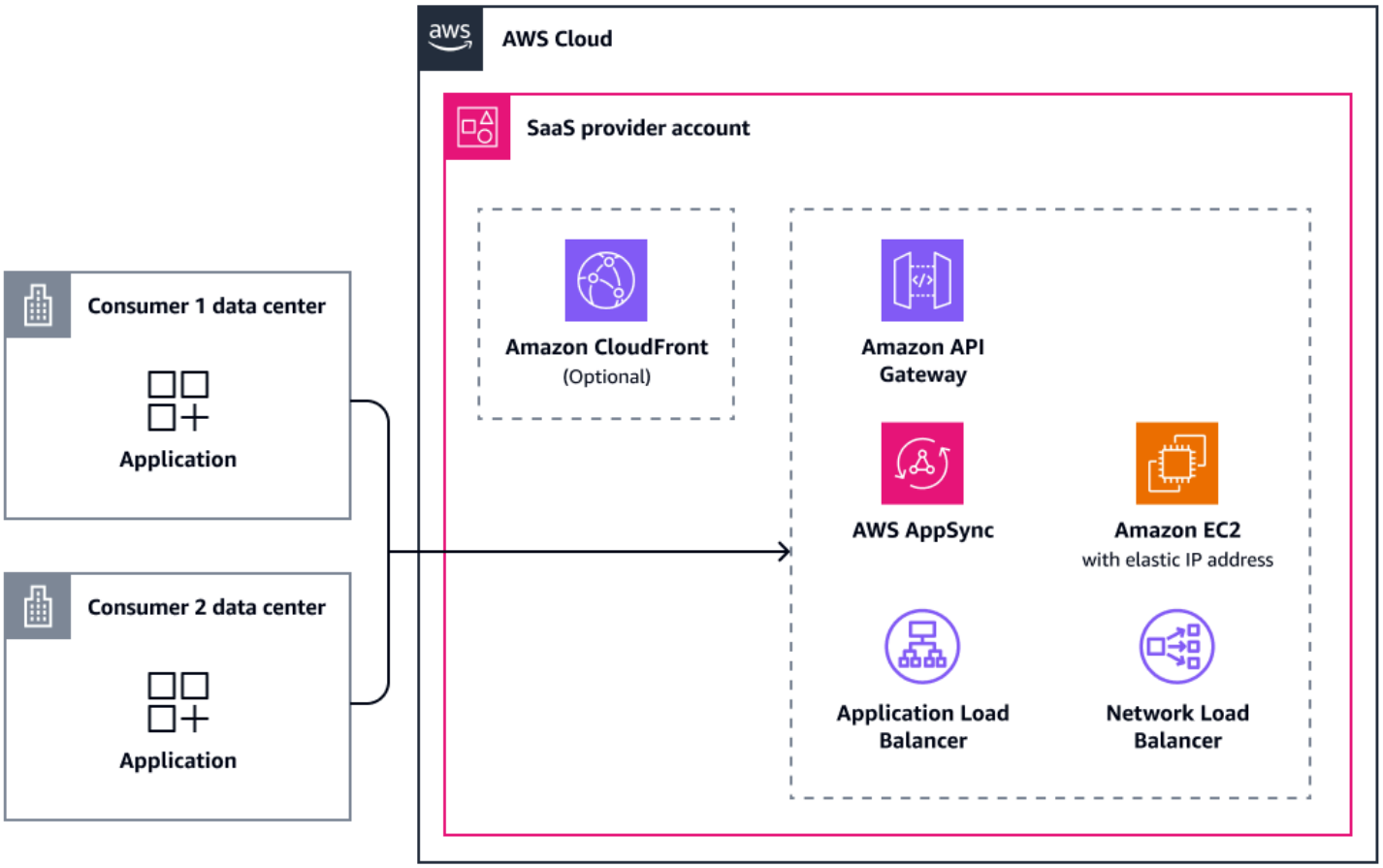
以下是这种方法的缺点：

- TCO：传输 VPC 是一种架构，而不是完全托管的服务，因此需要更多的运营工作
- 总拥有成本：SaaS 提供商需要操作和监控其他架构组件

通过公共互联网连接

公共互联网接入也是提供SaaS产品访问权限的有效选择，尽管它不提供传统意义上的私有连接。一些消费者可能仍然更喜欢公共访问方法，因为它不需要在他们和SaaS提供商之间建立额外的网络基础架构。它降低了复杂性、成本和集成时间，以换取更大的攻击面。强大的身份验证和授权机制可以帮助缓解更高的威胁级别，并且您应始终对流量进行加密。在这种情况下，仍然建议您增加一层安全保护，例如使用AWS WAF。

这种场景中的架构很简单。消费者通过互联网连接到公共主机（SaaS 提供商）。该应用程序可以直接托管在具有弹性 IP 地址的公共亚马逊弹性计算云 (Amazon EC2) 实例上。首选选项是将其托管在 Application Load Balancer 或类似服务后面。为了提高性能和缓存静态资产，您可以使用内容分发网络，例如 Amazon CloudFront。要在两个全球静态 Anycast IP 地址上以最小的延迟为应用程序提供服务，您可以放置AWS Global Accelerator在 Amazon EC2 实例、Network Load Balancer 或 Application Load Balancast 或 Application Load Balancer 的前面。此外 CloudFront，应用程序负载均衡器和 Amazon API Gateway 都与 AWS WAF集成。 AWS AppSync下图概述了公共互联网接入连接选项。



下表描述了此场景支持的协议和集成。

服务或资源	IPv6	AWS WAF 整合	可以是全球加速器端点
Amazon CloudFront	支持	支持	不支持
Amazon API Gateway	支持	支持	不支持
AWS AppSync	部分支持	支持	不支持
具有弹性 IP 地址的 Amazon EC2	支持	不支持	支持
应用程序负载均衡器	支持	支持	支持
网络负载均衡器	支持	不支持	支持

这种方法的优点如下：

- 易于集成：简单性和可访问性
- 可扩展性：无限扩展
- 适应性：不可能发生 CIDR 范围冲突
- 适应性：支持 CloudFront

以下是这种方法的缺点：

- 网络隔离：没有私有连接
- 网络隔离：需要采取强有力的安全措施

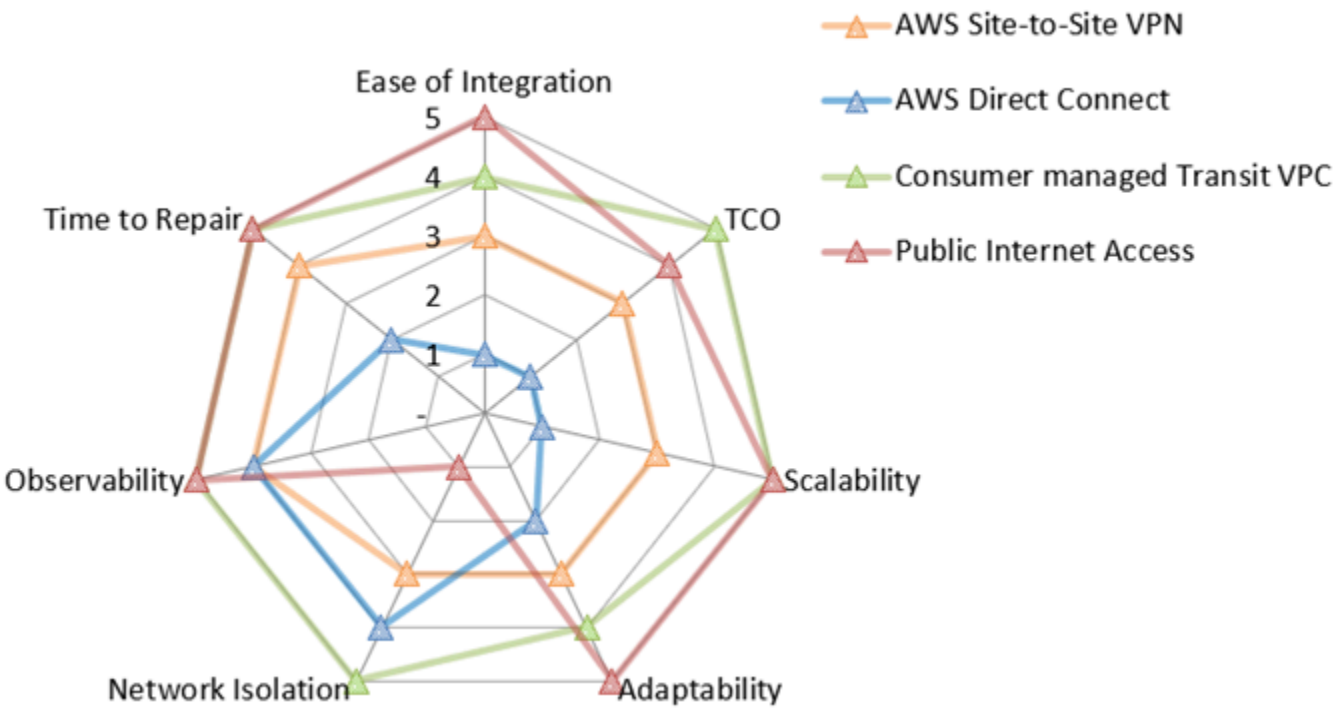
其他优点和缺点也适用，具体取决于您选择的服务。

在其他云服务提供商上运营的 SaaS 消费者

此场景描述了其他云服务提供商（CSPs）上面向消费者的解决方案。这种情况与本地数据中心的连接有一些共同点。实际上，本地环境的所有连接选项对其他环境的消费者同样有效 CSPs，有些甚至可以与之建立私 AWS Direct Connect 有连接 CSPs。大多数都 CSPs 提供有关如何 AWS Cloud 通过 AWS Site-to-Site VPN 或连接到的文档和支持 AWS Direct Connect。

在选择 Site-to-Site VPN 时，消费者可以从各自的 CSP 的托管网关或类似资源中受益。消费者不一定要自己设置它们，就像在本地场景中一样。这会影响 Site-to-Site VPN 的某些指标，例如修复时间和可观察性的缩短。这是因为现在连接的两端都已被管理。

以下网络价值图汇总了每个评估指标中每个选项的得分情况。尽管 Site-to-Site VPN的值不同，但它与本地连接的网络价值图非常相似。有关评估指标的更多信息，请参阅本指南评估指标中的。在地图中，五表示最高分数，例如最低的 TCO、最佳的网络隔离或最低的修复时间。有关如何阅读此雷达图的更多信息，请参阅本指南网络价值地图中的。



雷达图显示以下值。

评估指标	AWS Site-to-Site VPN	AWS Direct Connect	消费者管理的传输 VPC	公共互联网访问
易于集成	3	1	4	5
TCO	3	1	5	4
可扩展性	3	1	5	5
适应性	3	2	4	5

网络隔离	3	4	5	1
可观察性	4	4	5	5
是时候修理了	4	2	5	5

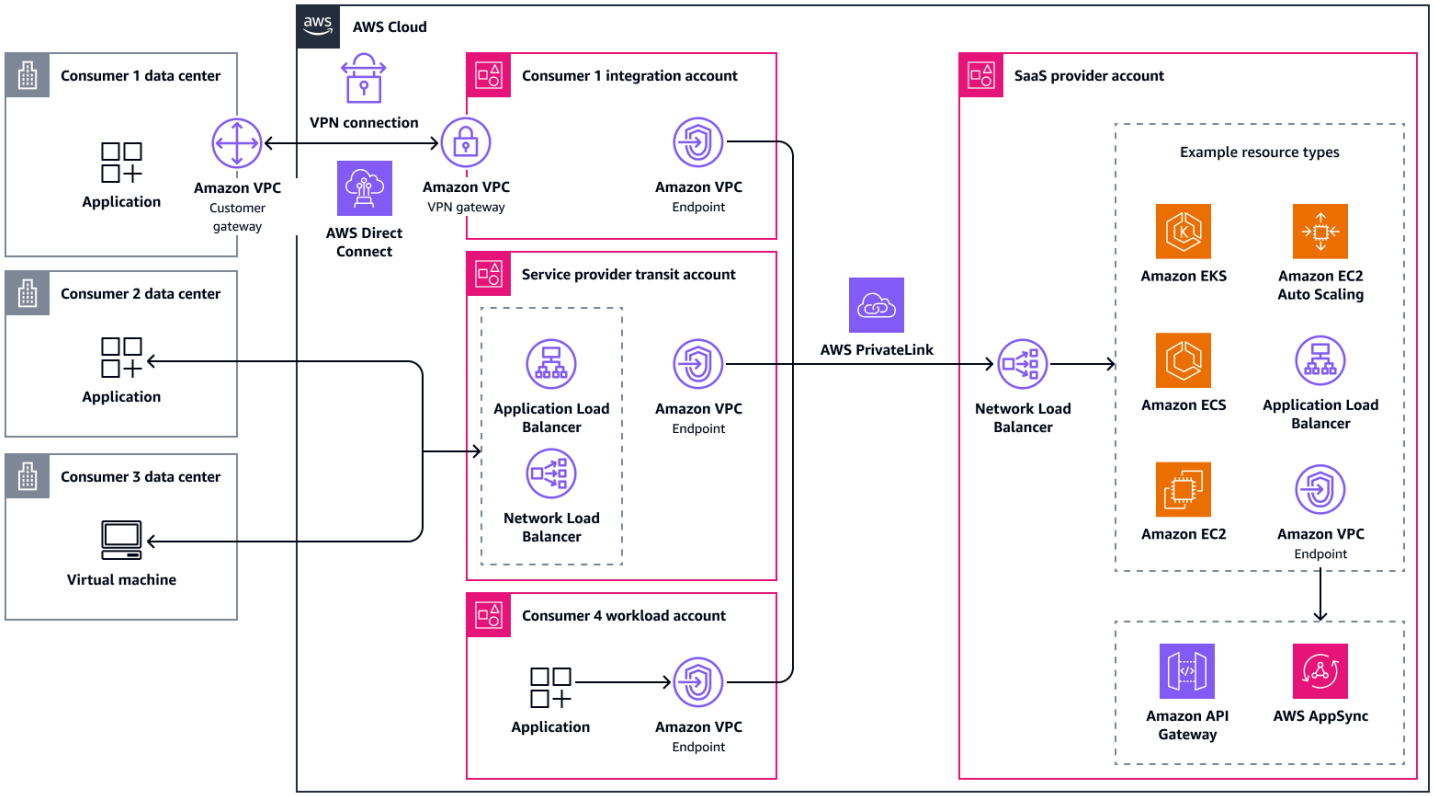
支持混合环境

消费者通常来自不同的环境，每个环境都有自己的技术和安全限制。有些客户可能完全在本地数据中心运营，这些数据中心需要通过互联网或专用网络链路进行安全连接。其他人可能已经在其中运行工作负载， AWS 并期望使用低延迟的私有网络路径。第三组可能依赖于其他群体 CSPs，其中连接必须桥接不同的云网络。

无论如何，您都应该以标准化网络访问您的 SaaS 应用程序为目标，以简化架构并降低运营复杂性。前面介绍的两种方法—— 公共互联网接入和传输 VPCs ——在这些场景中效果很好。公共互联网接入为您的客户提供了最快的入门路径，只需最少的设置即可。Transit VPCs 提供更受控制和私密的访问，通常使用 AWS PrivateLink。

在设计 SaaS 产品时，您可以采用单一网络接入模式，也可以将多种方法组合成分层产品。例如，您可以为优先考虑便捷连接和快速入门的客户提供公共访问部署层，也可以为有严格合规或安全控制要求的客户提供私有访问部署层。这些等级的成本、性能和风险状况各不相同。也可以将这两种方法组合到一个架构中。在这种情况下，请确保采取强有力的安全措施，以便公共和私有路径保持隔离。

下图显示了一种混合接入方法，在这种方法中，消费者可以选择从其数据中心或 CSP 进行私密连接，也可以公开连接，也可以直接通过 AWS PrivateLink （ 如果他们有工作负载 AWS Cloud ）进行连接。


AWS Cloud
Consumer 1 data center
Application
Amazon VPC
Customer gateway
VPN connection
Amazon VPC
VPN gateway
AWS Direct Connect
Consumer 1 integration account
Amazon VPC
Endpoint
Consumer 2 data center
Application
Consumer 3 data center
Virtual machine
Service provider transit account
Application Load Balancer
Network Load Balancer
Amazon VPC
Endpoint
Consumer 4 workload account
Application
Amazon VPC
Endpoint
AWS PrivateLink
SaaS provider account
Network Load Balancer
Example resource types
Amazon EKS
Amazon EC2 Auto Scaling
Amazon ECS
Application Load Balancer
Amazon EC2
Amazon VPC
Endpoint
Amazon API Gateway
AWS AppSync

中 SaaS 产品的高级网络访问场景 AWS Cloud

本中 SaaS 产品的网络访问场景 AWS Cloud节中讨论的架构应该可以帮助您找到适合大多数用例的解决方案。但是，有些场景有特定的技术要求。许多都超出了本指南的范围。

本节讨论以下高级技术要求和注意事项：

- 双向通信
- TCP、UDP 和专有协议

双向通信

在某些情况下，应用程序需要双向流量才能按预期运行。常见的用例是 Webhook 或通知服务。通常，您可以通过在服务器和客户端之间建立 WebSocket 连接来实现此目的。此连接保持 TCP 会话处于打开状态，并允许两个参与者通过该连接发送流量。本指南中讨论的大多数服务都是原生支持的 WebSocket，包括网络负载均衡器、应用程序负载均衡器、Amazon API Gateway 和 AWS AppSync（通过私有实时终端节点）。 AWS PrivateLink

在其他情况下，SaaS 提供商端的应用程序可能需要访问消费者端的资源，例如数据库。当您通过双向通道（例如连接）进行 AWS Site-to-Site VPN 连接时，这不是问题。

另一方面， AWS PrivateLink Elastic Load Balancing 仅支持单向流量。如果您使用这些服务，则必须为从 SaaS 产品发起的流量设置另一条网络路径。例如，这可能是朝相反方向移动的附加 AWS PrivateLink 连接。

TCP、UDP 和专有协议

许多应用程序都通过 HTTP 或 HTTPS 提供服务，但并非所有应用程序都通过。有些协议可能在 TCP 之外使用其他第 7 层协议，例如消息队列遥测支持 (MQTT)。其他人甚至可能使用UDP为消费者提供服务。在极少数情况下，服务使用必须在数据包内传输的专有协议（第 3 层）。对于这些场景，了解哪些服务支持您的 SaaS 产品非常重要。

对于第 3 层服务，您可以使用 AWS PrivateLink 和网络负载均衡器，两者都支持所有 TCP 和 UDP 流量。

对于第 7 层服务，应用程序负载均衡器和亚马逊 CloudFront 支持 HTTP WebSocket、HTTPS 和谷歌远程过程调用 (gRPC)。同样，Amazon API Gateway AWS AppSync 均支持 HTTP、HTTPS 和 WebSocket。亚马逊 CloudFront 是目前唯一支持 HTTP/3 的服务。

您可以使用 Amazon VPC Lattice 连接第 7 层应用程序和第 3 层资源。它支持 HTTP、HTTPS、gRPC、TCP 和 TLS 直通。

如果应用程序只能通过第 3 层提供流量，则必须使用核心 AWS 网络服务，例如、 AWS Transit Gateway AWS Direct Connect AWS Site-to-Site VPN、和 VPC 对等连接。然后，流量应直接从SaaS使用者路由到SaaS产品的计算层。

中网络访问的反模式 AWS Cloud

反模式是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。本节中提到的设计选项通常有效，但它们有明显的缺点。如果可能，应避免使用它们，因为有更好的替代品可供选择。

本节讨论以下反模式和挑战：

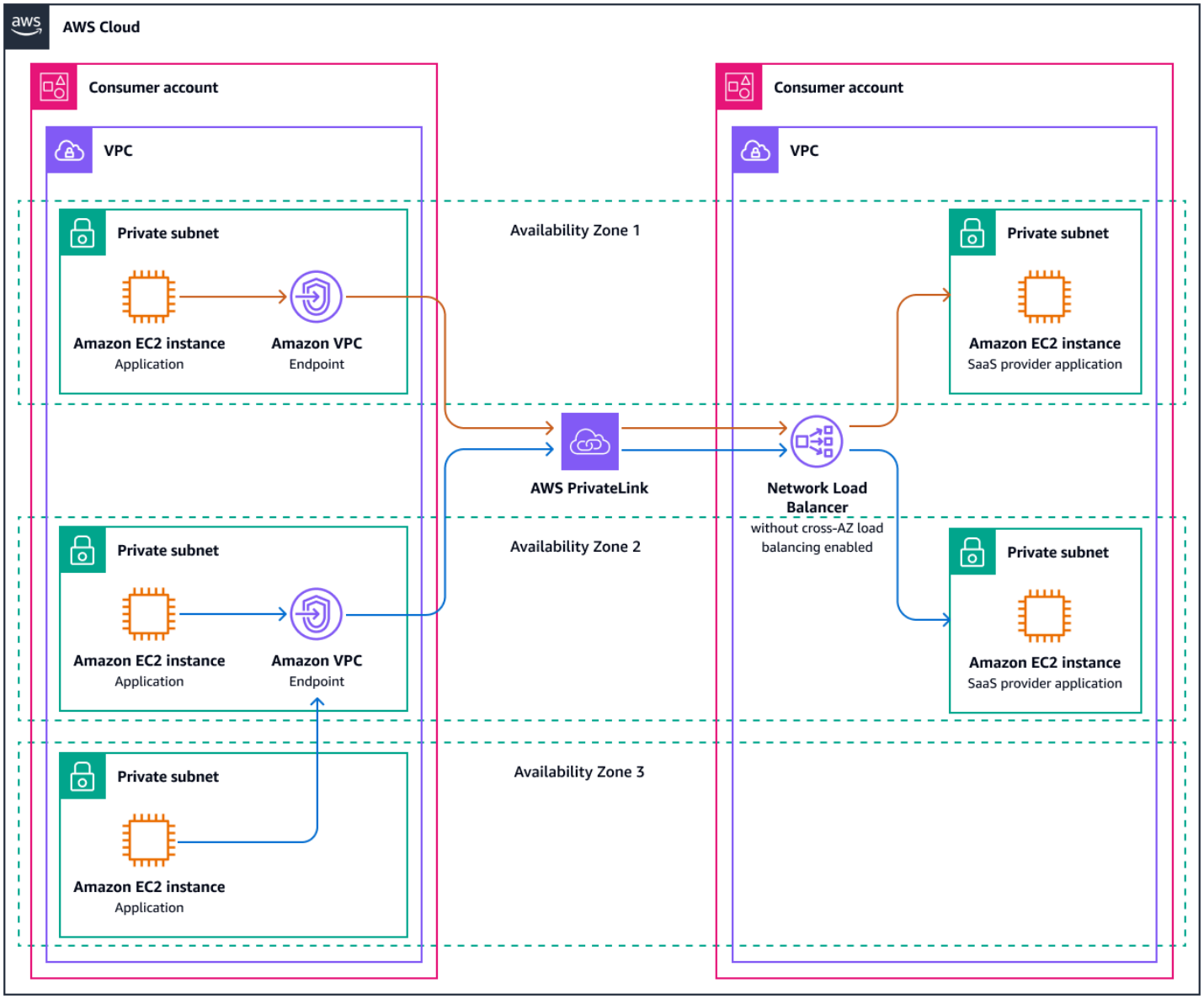
- 可用区与不匹配 AWS PrivateLink
- AWS Site-to-Site VPN 之间的连接 AWS 账户

可用区与不匹配 AWS PrivateLink

通过提供对应用程序的访问权限时 AWS PrivateLink，SaaS 用户只能在部署应用程序的可用区中创建接口 VPC 终端节点。例如，如果应用程序部署在use1-az1和中use1-az2，则使用者无法在中部署 VPC 终端节点use1-az3。我们建议您在每个可用区部署 SaaS 产品。大多数 AWS 区域 都有三个可用区，但有些还有更多。有关完整列表，请参阅地区和可用区。在选择可用区时，请考虑可用区的数量 AWS 区域。

> **Note**
>
> 可用区域名称不同于可用区 IDs。有关更多信息，请参阅您的 IDs AWS 资源的可用区。

如果 SaaS 提供商选择不在所有可用区域进行部署，则会产生一些后果。假设 SaaS 产品部署在use1-az1和中use1-az2，但消费者正在使用所有三个可用区，包括use1-az3。接口 VPC 终端节点部署在use1-az1和中的消费者端use1-az2，现在中的应用程序use1-az3需要访问其中一个终端节点。首先，必须允许流量从不匹配的可用区中的子网进入相应的 VPC 终端节点。消费者可以决定使用区域 AWS PrivateLink DNS 名称，该名称可以解析到任一 VPC 终端节点，并在两者之间均匀分配流量。或者，消费者可以选择将流量直接发送到终端节点，例如use1-az2。这导致67%的流量来自提供商一方，33%进入提供商。use1-az2 use1-az1下图描述了这种情况。



由于消费者数量众多，流量分布不均匀，工作负载可能会在一个可用区遇到容量问题，而另一个可用区则出现容量不足。为了解决这个问题，SaaS 提供商可以决定通过在 Network Load Balancer 上启用跨区域负载平衡来均衡其侧面的流量。这会产生额外费用。

如果服务提供商只匹配了一个可用区，则所有流量都将通过单个终端节点进入。这造成了更大的不平衡。因此，SaaS产品对消费者的可用性不再高。对于消费者来说，应用程序是否通过他们自己没有使用的其他可用区提供服务并不重要。在最坏的情况下，SaaS 提供商可能无法为不使用任何相同可用区的消费者提供服务。

在极少数情况下，SaaS 提供商无法在所有可用区中配置其应用程序，也可以仅在缺少的可用区中创建子网，然后将服务扩展到那些空的可用区。然后，跨区域负载平衡可以将传入流量分配到其他可用区域中的实际应用程序终端节点。

AWS Site-to-Site VPN 之间的连接 AWS 账户

从本地环境迁移到云端的公司有时会尝试提升和转移整个网络。这可能会导致问题，因为本地网络和云网络实践之间存在显著差异。如果不发生这种思维方式的转变，就会发生诸如从一个 VPC 到另一个 VPC 的 AWS Site-to-Site VPN 连接之类的事情。这种方法无法利用中专门构建的网络服务 AWS Cloud，这些服务可以简化管理并提高性能。适应云原生设计有助于减少运营开销，并在两者之间 VPCs实现更可靠、更可扩展的连接。

如果您正在考虑以SaaS提供商的身份提供此连接选项，请问自己或消费者为什么 AWS Site-to-Site VPN 要使用。然后，从这些要求向后推进，找到更好的连接选项。本指南的 “比较服务功能” 部分包含一个矩阵，您可以使用该矩阵来帮助确定选项。然后，您可以仔细阅读本指南的相关部分，找到适合您的用例的架构方法。

后续步骤

本指南描述了不同场景下的各种网络访问方法，并描述了每种架构的优缺点。你应该明白为什么选择网络接入方法不应该纯粹是技术方面的讨论。业务和技术之间的协调至关重要。以下后续步骤和建议可通过评估当前能力、分析市场需求和实施治理控制来帮助您评估和标准化您的网络架构策略。

本节包含以下主题：

- 评估当前架构和功能
- 市场和客户分析
- 战略调整
- 标准化
- Governance
- 重复

评估当前架构和功能

根据相关数据源审查当前的网络架构，例如本指南中的自我评估框架、当前的监管要求和市场现状（包括客户和竞争分析）。例如，可以考虑使用 Well-A AWS rchitected Framework，该框架基于数十年来大规模运行生产系统的经验。 AWS Cloud

查看任何潜在的例外情况、一次性产品和历史产品决策。保持好奇，挑战他们，不要自动假设他们的有效性。几年前的客户要求可能不再有效。具有挑战性的假设为简化和降低架构的复杂性创造了机会。

简而言之，记录观察结果，以便组织中的不同角色可以访问和理解它们。捕捉当前状态与目标状态的不同之处、目标状态是什么、影响以及何时进行观测。记录这些信息有助于您的组织根据最新数据做出决策。

市场和客户分析

收集对市场趋势的见解。当前，消费者访问像你这样的SaaS产品的首选方式是什么？ 您还能在客户所在的地方与他们会面吗？ 客户群组或行为是否发生了变化？ 您的高管们是否将这艘船引向了新的市场、具有特定监管要求的地理位置或新的客户等级？ 您的业务或运营模式是否发生了变化？ 例如，您是否正在考虑为服务贴上白标？ 您的增长计划是否包括与合作伙伴合作，以便在客户与这些合作伙伴建立联系时向他们提供您的服务？

战略调整

当您了解当前的能力、当前的架构、市场和客户后，请召开战略协调会议。与相关的产品、业务和技术利益相关者一起质疑哪些要求仍然有效，哪些新要求需要考虑。通过删除不再需要的要求来寻找降低复杂性的机会。这不是委员会的设计；工程团队需要准备并拥有实际的架构和实施细节。但是，这次会议应该阐明为什么这是一组可以最大限度地为客户和组织带来好处的要求。

标准化

为了吸引客户，让每个人自由选择如何连接到您的服务可能很诱人。毕竟，任何解决方案在技术上都可能奏效，而且您可能还拥有管理和操作所有这些解决方案的专业知识和资源。在某种程度上，这可以很好地发挥作用，但是随着业务的扩展，它变得难以管理。您的可观测性堆栈需要支持来自多个解决方案的指标，而您的网站可靠性工程师也需要能够理解这些指标。每种连接方法都需要 up-to-date文档。需要根据您提供的每种访问方法对应用程序中的重大更改进行评估。您需要为每种访问方法编写和维护自动化和基础设施即代码 (IaC)。必须权衡不对服务访问进行标准化所产生的额外开销与您希望为客户提供的灵活性。

如果您需要北极星来指导您的决策，我们建议您进行标准化。客户与您提供的服务互动方式的标准化通常是您可以采取的最有影响力的措施来改善整个组织的许多成功指标。标准化使产品团队更容易了解您的服务的成本结构，并做出数据驱动的产品决策。在根据预定义标准开发、推出和运行的环境中，运营团队可以更轻松地解决问题并自动执行部分故障排除过程。它可以帮助您检测异常、意外行为或恶意行为者的行为。标准化还可以减少技术债务。工程团队测试和推出生产变更所需的周期更短。它还可以加快您的上市速度，提高自助服务入职成功率，并降低监管风险。

因此，我们建议您同时查看当今可能存在的任何一次性措施。量化您为支持现有客户所花费的运营周期数。将您的结果与历史数据进行比较，并评估您当前的方法是否适用于未来几年。每当需要偏离标准时，都要质疑这些要求背后的要求。评估影响，在眼前收益与长期承诺之间取得平衡。

如果定制不可避免但与您的标准相冲突，请考虑采用责任共担模式。在此模型中，您的产品在很大程度上不受所要求的更改的影响，并且自定义是在极简主义的专用环境中进行的。有关示例，请参阅一连接传输 VPC 架构节。

Governance

为了遵守监管要求和您自己的内部标准，管理至关重要。有了适当的治理，您就可以控制在何处以及如何执行标准。您还可以建立控制措施，以检测与标准的差异，并告知资源所有者必要的纠正措施。AWS Organizations、AWS Config、AWS CloudTrail、和AWS Control Tower是许多可以帮助你在 AWS 服务 中管理和控制工作负载的工具中的一小部分 AWS Cloud。

重复

利用从最初的工作中吸取的教训，建立一个轻量级、可重复的流程，以便在将来保持一致。定义您需要从哪些角色那里输入数据、频率、数据的准确性、如何共享数据以及谁将对数据采取行动。

资源

AWS 文档

- 在 AWS Cloud（AWS 规范性指南）中集成第三方服务
- 多租户 SaaS 授权和 API 访问控制（AWS 规范性指南）
- 在@@ 单个控制平面上管理多个 SaaS 产品的租户（AWS 规范性指南）
- 什么是 AWS Direct Connect？（Direct Connect 文档）
- 什么是 AWS PrivateLink？（亚马逊 VPC 文档）
- 什么是 AWS Site-to-Site VPN？（AWS Site-to-Site VPN 文档）
- 什么是 AWS Transit Gateway？（亚马逊 VPC 文档）
- 什么是 VPC 对等互连？（亚马逊 VPC 文档）

其他 AWS 资源

- 亚马逊 Virtual Private Cloud 连接选项（AWS 白皮书）
- AWS re: Invent 2021-如何为您的 AWS 工作负载选择合适的负载均衡器 () YouTube
- 什么是 SaaS？（AWS 网站）
- AWS SaaS 工厂计划（AWS Partner 计划）
- 多租户架构指南 AWS（AWS 解决方案库）

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9月12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：

- Refactor/re-architect — 充分利用云原生功能来提高敏捷性、性能和可扩展性，从而移动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 数据库迁移到亚马逊 Aurora PostgreSQL-Compatible 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：将您的客户关系管理 (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，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

A2A () Agent-to-Agent

一种支持任务委托和状态转移的代理到代理协作的状态协议。

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。

座席

一种能够使用工具自主推理、计划和采取行动来实现目标的人工智能系统。

特工行动

在生产环境中大规模构建、测试、部署和运行 AI 代理的操作实践。

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？

人工智能运营（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 AWS 迁移策略中使用 AIOps 的更多信息，请参阅运营集成指南。

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。

blue/green 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，请参阅指南中的 “实施破碎玻璃程序” AWS Well-Architected 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。

公民开发者

使用无code/low代码平台创建 AI 应用程序但没有专业技术技能的企业用户。

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，创建登录区、定义 CCoE、建立运营模型）
- 迁移 - 迁移单个应用程序
- Re-invention — 优化产品和服务，在云端进行创新

Stephen Orban在 AWS Cloud 企业战略博客的博客文章《走向之旅 Cloud-First 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）

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《 AWS Well-Architected 框架》中安全支柱的组成部分。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

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。

深度防御

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，深度防御方法可能将多因素身份验证、网络分段和加密结合起来。

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。

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《工作负载灾难恢复 AWS： AWS Well-Architected 框架中的云端恢复》。

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。埃里克·埃文斯（Eric Evans）在他的《Domain-Driven 设计：解决软件核心的复杂性》（波士顿： Addison-Wesley 专业版，2003年）一书中介绍了这个概念。有关如何使用带有 strangler fig 模式的域驱动设计的信息，请参阅使用容器和 Amazon API Gateway 逐步实现传统微软 ASP.NET (ASMX) 网络服务的现代化。

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。 Big-endian 系统首先存储最重要的字节。 Little-endian 系统首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和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。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。 Few-shot 对于需要特定格式、推理或领域知识的任务，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，它已使用海量的通用和未标注数据集进行训练。FM 能够执行各种常规任务，例如理解语言、生成文本和图像以及使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。

FM 网关

一种集中式中介，用于控制和规范对基础模型的访问。也称为 LLM 网关。

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

一种高级规则，用于跨组织单位（OU）管理资源、策略和合规性。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。

护栏 (AI)

用于过滤、验证和限制代理输入和输出的安全机制，有助于确保负责任和安全的 AI 行为。

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。

人机在圈 (HitL)

一种工作流程模式，其中代理执行在关键决策点暂停以供人工审查和批准。

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

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。

IIoT

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，请参阅框架中的使用不可变基础架构部署最佳实践。 AWS Well-Architected

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 安全参考架构建议使用入站、出站和检查 VPC 设置网络账户，保护应用程序与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）在2016年推出，指的是通过连接性、实时数据、自动化、分析和的进步实现制造流程的现代化。 AI/ML

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（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）数字化转型策略。

检查 VPC

在 AWS 多账户架构中，一种集中式 VPC，用于管理 VPC（相同或不同 AWS 区域）、互联网和本地网络之间的网络流量检查。AWS 安全参考架构建议使用入站、出站和检查 VPC 设置网络账户，保护应用程序与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 LLM。

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、记录和控制在车间将原材料转化为成品的生产过程。

MAP

请参阅迁移加速计划。

MCP

参见模型上下文协议。

模型上下文协议（MCP）

一种用于代理与工具通信的无状态协议。

MCP 服务器

一种通过模型上下文协议公开一个或多个工具的服务。

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 AWS Well-Architected 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publish/subscribe模式的轻量级机器对机器 (M2M) 通信协议，适用于资源受限的物联网设备。

微服务

一种小型独立服务，通过明确定义的 API 进行通信，通常由小型独立团队拥有。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 API 通过明确定义的接口进行通信。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

Cross-functional 通过自动化、敏捷的方法简化工作负载迁移的团队。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：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，该 AWS Well-Architected 框架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机器对机器 (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 框架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 安全参考架构建议使用入站、出站和检查 VPC 设置网络账户，保护应用程序与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

私有托管区就是一个容器，其中包含的信息说明您希望 Amazon Route 53 如何响应一个或多个 VPC 中的某个域及其子域的 DNS 查询。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 AWS Organizations的组织中所有账户的权限。SCP 为管理员可以委托给用户或角色的操作定义了防护机制或设定了限制。您可以将 SCP 用作允许列表或拒绝列表，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。

暗影人工智能

在组织内受管控渠道之外构建或使用的未经授权的 AI 应用程序。

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 模式

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 逐步实现传统微软 ASP.NET (ASMX) 网络服务的现代化。

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。你可以使用 Amazon S CloudWatch ynthetics 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

Key-value 对充当用于组织 AWS 资源的元数据。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。

工具

代理可以调用以在外部系统中执行操作的函数或 API。

中转网关

中转网关是网络中转中心，您可用它来互连 VPC 和本地网络。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。

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

两个 VPC 之间的连接，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。