



生成式 AI 工作負載評估

AWS 方案指引



AWS 方案指引: 生成式 AI 工作負載評估

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon 的商標和商業外觀不得用於任何非 Amazon 的產品或服務，也不能以任何可能造成客戶混淆、任何貶低或使 Amazon 名譽受損的方式使用 Amazon 的商標和商業外觀。所有其他非 Amazon 擁有的商標均為其各自擁有者的財產，這些擁有者可能附屬於 Amazon，或與 Amazon 有合作關係，亦或受到 Amazon 贊助。

Table of Contents

簡介	1
本指南的目的	1
目標受眾和優勢	2
範圍	2
目標業務成果	4
評估考量事項和先決條件	6
從明確的使用案例開始	6
確保業務一致性	7
實作控管和監督	7
地址資料和技術先決條件	7
考慮運算資源需求	7
解決隱私權和安全性問題	7
及早吸引利益相關者	7
反覆運算和學習	8
生成式 AI 工作負載評估問卷	9
準備度	9
使用案例	11
架構	13
儲存	14
法規與合規	15
整合	15
測試	17
部署和自動化	18
資料策略	19
將評估洞見轉換為可行的結果	22
後續步驟	24
常見問答集	25
主要目標是什麼？	25
誰應該使用此評估？	25
什麼是關鍵元件？	25
這如何協助定義架構？	25
有哪些好處？	25
如何成功實作？	25
有哪些挑戰？	25

什麼是法規和合規要求？	26
利益相關者的角色是什麼？	26
如何衡量成功？	26
根據組織大小，方法有何不同？	26
資源	28
文件歷史紀錄	29
詞彙表	30
#	30
A	30
B	33
C	34
D	37
E	40
F	42
G	43
H	44
I	45
L	47
M	48
O	52
P	54
Q	56
R	57
S	59
T	62
U	63
V	64
W	64
Z	65
.....	lxvi

生成式 AI 工作負載評估

Tabby Ward 和 Deepak Dixit , Amazon Web Services (AWS)

2024 年 11 月 ([文件歷史記錄](#))

生成式 AI 工作負載評估是一種策略方法，旨在評估和改善組織建立或更新生成式 AI 工作負載的準備程度。此評估很重要，因為將生成式 AI 整合到業務營運中可能會大幅改變物件的運作方式，並可提供新的效率和功能。不過，若要成功採用生成式 AI，請務必徹底了解目前的系統，並制定明確的未來計畫。

生成式 AI 工作負載是指涉及使用人工智慧模型來建立新內容的運算任務，例如文字、影像、程式碼或其他資料類型。這些工作負載通常需要強大的運算能力、專業的硬體，例如 GPUs，以及用於訓練和推論的大型資料集。將生成式 AI 工作負載整合到操作中，有幾個挑戰：

- 基礎設施需求：佈建生成式 AI 模型所需的重要運算資源和專用硬體。
- 資料管理：在處理大型資料集時確保資料品質、隱私權和合規性。
- 技能差距：缺乏 AI 技術和模型部署方面的專業知識。
- 道德考量：解決 AI 產生內容中的偏差、公平性和透明度。
- 整合複雜性：將生成式 AI 無縫整合到現有的工作流程和舊版系統中。
- 成本管理：在潛在利益與高實作和操作成本之間取得平衡。

克服這些挑戰需要仔細規劃、投資基礎設施和人才，以及實施的策略方法。

本指南的目的

生成式 AI 迅速成為許多產業的重要元件。它提供轉型機會，但在整合、合規和可擴展性方面也帶來挑戰。由於技術基礎薄弱、變革阻力和資料品質問題，許多組織難以充分利用 AI。生成式 AI 工作負載評估透過識別現代化需求、定義實作範圍以及具有挑戰性的傳統系統和思維來解決這些挑戰。它還有助於確定最低可行產品 (MVPs)，並協助您開發目標解決方案架構，確保採用 AI 的結構化和策略方法。

本指南提供結構化的方法，協助組織應對採用生成式 AI 技術的複雜性。本指南不會從一開始就明確定義需求，而是協助：

- 識別組織內生成式 AI 的潛在使用案例。
- 評估組織是否準備好採用生成式 AI。

- 定義和精簡使用案例目標和延伸目標。
- 決定生成式 AI 實作的範圍和需求。
- 開發目標解決方案架構。

目標受眾和優勢

此評估專為想要評估生成式 AI 工作負載現代化技術層面的解決方案架構師、企業架構師和應用程式架構師而設計。對於想要衡量團隊整體整備度、資源分配和啟用需求的計畫和人事經理來說，這也很有價值。產業最佳實務強調全面評估的重要性，以確保 AI 採用的準備程度。這包括評估架構、儲存、合規、整合、測試、部署和自動化。

範圍

下列主題在生成式 AI 工作負載評估方法的範圍內：

- 目前的生成式 AI 技術和模型（例如，大型語言模型、影像產生模型）
- 使用生成式技術的窄版 AI 應用程式
- 生成式 AI 與現有系統和工作流程的整合
- 用於訓練和微調生成式 AI 模型的資料策略
- 目前生成式 AI 應用程式的道德考量和負責任的 AI 實務
- 在生產環境中測試和部署生成式 AI 的策略
- 生成式 AI 實作的安全性和隱私權考量
- 生成式 AI 工作負載的效能最佳化和可擴展性
- 各種產業中生成式 AI 的使用案例和應用程式
- 評估生成式 AI 輸出和品質保證程序

下列主題超出範圍：

- 人工智慧 (AGI) 和人工超智慧 (ASI) 案例
- AI 在目前生成模型之外的推測未來進展
- AI 中的量子運算應用程式
- 神經變形運算和大腦電腦界面
- AI 系統中的意識和自我意識

- 除了目前產生的 AI 應用程式之外，進階 AI 的長期社會影響
- 假設性未來 AI 技術的法規架構
- 哲學爭論機器中智慧和意識的本質
- AI 極端邊緣案例或高度投機的使用案例
- 專屬 AI 模型或架構的詳細技術規格

目標業務成果

生成式 AI 工作負載評估旨在提供數個目標成果，這對成功現代化生成式 AI 工作負載至關重要。這些結果可確保組織做好充分準備，以有效且有效率地整合 AI 技術。

對於每個目標成果，生成式 AI 工作負載評估著重於：

- **相互依存性**：識別並釐清結果與現代化程序的其他層面之間的任何相互依存性。這包括了解一個結果可能如何影響或受到其他人的影響，以確保現代化的整體方法。
- **利益相關者一致性**：概述使各種利益相關者與每個結果保持一致的策略。這涉及將每個結果的價值和影響傳達給不同的組織層級和部門，以促進接受和支援。
- **優先順序**：在識別多個使用案例或結果的情況下，提供根據業務影響、資源需求和策略一致性等因素排定優先順序的架構。
- **持續改進**：針對每個結果，建立持續評估和改進的機制。這可確保現代化工作保持適應性和回應不斷變化的技術環境和業務需求。

以下是每個目標成果的詳細討論：

目標架構

- **定義**：評估有助於為生成式 AI 工作負載定義清晰且可擴展的目標架構。
- **元件**：這包括選取適當的雲端服務、設計資料管道，以及確保系統互通性。
- **優點**：定義良好的架構支援可擴展性、可靠性和效能最佳化，並為現代化提供了堅實的基礎。

客戶整備

- **評估**：評估組織基礎設施、程序和文化的目前狀態，以確定是否準備好採用生成式 AI 現代化。
- **條件**：這涉及評估技術能力、資料品質和組織接受變革的意願。
- **結果**：識別差距和需要改進的領域，可確保組織準備好順利轉換到現代解決方案和技術。

使用案例目標和延伸目標

- **使用案例目標**為目標解決方案實作建立明確的目標，專注於特定的業務問題或機會。

生成式 AI 現代化環境中的使用案例目標是指組織透過實作生成式 AI 解決方案來實現的特定、可衡量的目標。這些目標通常與更廣泛的業務目標保持一致，並專注於解決組織內的特定挑戰或機會。使用案例目標的範例可能包括：

- 使用生成式 AI 支援的聊天機器人，將客戶服務回應時間縮短 50%。
- 透過生成式 AI 輔助程式碼分析，將程式碼檢閱效率提升 30%。
- 使用生成式 AI 模式辨識，將詐騙偵測準確度提高 25%。
- 延伸目標會定義遠大的目標，將生成式 AI 現代化在組織內可以實現的目標推向界限。
- 影響：同時設定可實現的目標和理想目標，有助於將生成式 AI 現代化計畫與策略業務目標保持一致，並鼓勵創新。

嘗試估算

- 目的：準確的工作量估算有助於資源規劃，並確保專案在預算內按時交付。
- 範圍：估計實作生成式 AI 現代化計畫所需的資源、時間和預算。
- 因素：考慮技術複雜性、整合挑戰和潛在風險。

啟用需求

- 訓練和開發：識別成功採用生成式 AI 現代化所需的技能和知識。
- 資源：判斷訓練計畫、研討會和其他啟用活動的需求。
- 成果：確保員工具備必要的技能，可增強生成式 AI 現代化計畫的有效性，並支援長期成功。

實作計畫

- 藍圖：制定詳細計畫，概述實現生成式 AI 現代化所需的步驟。
- 里程碑：定義關鍵里程碑和交付項目以追蹤進度。
- 優點：明確的實作計畫提供方向和責任，並促進生成式 AI 現代化的結構化方法。

評估考量事項和先決條件

從明確的使用案例開始

識別生成式 AI 可以解決的特定業務問題或機會。專注於符合策略業務目標並提供可衡量利益的使用案例。優先考慮以組織內常見挑戰為目標的使用案例，以確保解決方案架構可做為多個案例的模式。

以對潛在生成式 AI 應用程式的一般了解啟動評估程序是有益的，但不是強制性的。本指南隨附的[問卷](#)可因應各種程度的準備，從定義良好的使用案例的組織，到只有廣泛想法的組織。評估程序適用於：

- 精簡並釐清這些初始使用案例想法。
- 識別新的潛在使用案例。
- 為每個使用案例制定具體且可衡量的目標。
- 評估每個使用案例的可行性和潛在影響。

讓我們考慮一個假設性範例：金融服務公司決定探索生成式 AI 現代化。他們從改善客戶服務和詐騙偵測程序的廣泛概念開始。

- 初始評估：問卷有助於他們評估目前的系統、資料品質和組織準備程度，以採用生成式 AI。
- 使用案例精簡：透過評估程序，他們將初始想法精簡為兩個特定的使用案例：
 - 實作生成式 AI 支援的聊天機器人進行客戶查詢
 - 使用生成式 AI 進行即時交易詐騙偵測
- 目標設定：針對每個使用案例，它們會定義特定目標：
 - 在 6 個月內將客戶服務回應時間縮短 40%
 - 提升詐騙偵測準確度達 20%，並減少誤判達 15%
- 延伸目標：他們也會設定這些遠大的目標：
 - 透過 AI 協助回應實現 80% 的客戶滿意度
 - 開發可辨識新詐騙模式的預測性詐騙偵測模型
- MVP 定義：問卷可協助他們判斷每個使用案例的 MVP，專注於提供立即價值的基本功能。
- 目標架構：最後，他們會開發支援一個或兩個使用案例的目標架構，並確保可擴展性和與現有系統的整合。

確保業務一致性

使生成式 AI 計劃與整體業務策略和目標保持一致。對於每個使用案例，開發明確的價值主張，以示範生成式 AI 如何促進業務成長、效率或創新。建立指標來衡量生成式 AI 實作對關鍵績效指標 (KPIs) 的影響。

實作控管和監督

建立跨職能的轉向委員會來監督生成式 AI 計畫。制定負責任的 AI 使用政策和指導方針，解決道德考量和潛在的偏差。為生成式 AI 專案建立審核程序，以確保符合組織標準和法規要求。

地址資料和技術先決條件

評估並改善資料品質，並實作資料控管實務，以確保生成式 AI 模型的可靠輸入。制定資料策略，以解決生成式 AI 需求特定的資料收集、儲存和管理。評估和增強資料基礎設施，以支援生成式 AI 工作負載所需的資料量和速度。

考慮運算資源需求

評估目前的 IT 基礎設施，並識別生成式 AI 工作負載的運算容量差距。規劃可擴展的運算資源，並考慮雲端服務或內部部署高效能運算叢集等選項。最佳化資源配置，以平衡訓練和推論工作負載的效能和成本效益。

解決隱私權和安全性問題

實作強大的安全措施，以保護生成式 AI 訓練和操作中使用的敏感資料。處理個人資訊時，請確保遵循資料保護法規，例如一般資料保護法規 (GDPR) 或加州消費者隱私權法 (CCPA)。開發安全模型部署和監控的通訊協定，以防止未經授權的存取或濫用生成式 AI 功能。

及早吸引利益相關者

從一開始就讓關鍵利益相關者參與，以獲得領導層的認同和支援。清楚傳達現代化計畫的優勢和潛在影響，特別是針對生成式 AI 工作負載。提供訓練和資源，以協助利益相關者了解生成式 AI 技術及其影響。

反覆運算和學習

採用增量方法，可讓您精簡目標解決方案。使用回饋迴圈來持續改善工作負載架構和程序。定期評估生成式 AI 實作的效能和影響，並根據實際結果和不斷變化的業務需求視需要調整策略。

生成式 AI 工作負載評估問卷

下列各節提供您可以用來評估組織生成式 AI 工作負載現代化不同層面的問題。此全面的問卷會評估組織在各個關鍵領域採用和實作生成式 AI 工作負載的準備程度，包括使用案例、架構、儲存、合規、整合、測試、部署和資料策略。透過解決生成式 AI 實作的關鍵層面，從技術基礎設施到法規考量，此問卷可協助您識別 AI 現代化旅程中的優勢、差距和機會。

章節：

- [準備度](#)
- [使用案例](#)
- [架構](#)
- [儲存](#)
- [法規與合規](#)
- [整合](#)
- [測試](#)
- [部署和自動化](#)
- [資料策略](#)

您也可以下載 Microsoft Excel 格式的問卷，並使用它來記錄您的資訊。



[下載問卷](#)

準備度

問題	回應範例
您是否有可用於這些工作負載 AWS 的帳戶？	是或否。
您是否有與 簽訂的現有企業協議 AWS？	是或否。
您目前雲端基礎設施處理生成式 AI 工作負載的可擴展性如何？	我們的雲端基礎設施具有高度可擴展性，具有適用於運算資源和分散式儲存系統的自動擴展

問題	回應範例
	功能，旨在有效地處理大規模生成式 AI 工作負載。
您是否有用於大規模預先處理和特徵工程的資料管道功能？	我們的資料管道使用分散式處理架構，例如 Apache Spark 進行大規模資料處理和特徵工程，同時支援批次和串流資料處理。
您是否有帳戶佈建和管理功能？	是或否。
您會如何描述組織的 AI 素養以及採用生成式 AI 技術的準備程度？	我們的組織已在 AI 教育計畫中投入大量資金，大多數技術人員也已完成基本的 AI/ML 訓練。組織具有創新文化，其中包含生成式 AI 等新技術。
組織中有哪些 AI/ML 專業知識，以及它如何分佈？	我們擁有與經驗豐富的資料科學家和 ML 工程師合作的專用 AI 卓越中心。我們提升跨不同業務單位的領域專家技能，以成為 AI-literate 並識別生成式 AI 使用案例。
您是否有表達雲端計畫目標、優點和成本的高階商業案例？	是或否。
將解決方案用於生產的時間表為何？	週、月等。
您的主要利益相關者（例如，CFO、CIT/CTO、COO）是否已做出資金承諾？	是或否。
如何確保符合生成式 AI 計畫中的資料保護法規？	我們有專門的合規團隊，與我們的 AI 團隊緊密合作。我們會定期進行隱私權影響評估、依設計原則實作資料保護，並維護所有生成式 AI 專案的詳細資料處理記錄。
您現有的系統與新的生成式 AI 技術整合有多成熟？	我們的 IT 架構是以微服務和 APIs 為基礎，允許靈活整合新的生成式 AI 技術。這些系統會標準化常用資料格式和通訊協定，以確保互通性。

問題	回應範例
您在操作 ML 模型方面有哪些經驗，以及這可能如何適用於生成式 AI 系統？	我們已建立 MLOps 實務，包括自動化模型部署管道、監控系統和 A/B 測試架構。這些實務正在調整，以處理大規模生成式 AI 模型的獨特需求。

使用案例

問題	回應範例
使用案例的主要目標或成功條件是什麼？	為了改善客戶支援回應時間、增加銷售轉換、增強產品建議。此外：改善使用者滿意度、任務完成率、回應品質等。
此使用案例如何符合您組織的策略目標？	這符合我們的策略目標，即透過減少客戶服務的回應時間來提高客戶滿意度。
使用案例的預期資料量或請求量是多少？	每秒 500 筆交易 (TPS)。
支援生成式 AI 工作負載需要哪些類型的資料來源？	內部結構化資料庫（客戶記錄、銷售資料等）；來自文件、電子郵件和社交媒體的非結構化文字資料；用於語音和影像辨識任務的音訊和影片檔案；來自 IoT 裝置和感應器的即時串流資料；用於擴充的公有資料集和 APIs。
您需要從這些來源更新或重新整理資料的頻率為何？	交易資料庫：近乎即時更新；文件儲存庫：每日批次更新；社交媒體摘要：每小時更新；IoT 感應器資料：持續即時串流；公有資料集：每月或每季更新。
生成式 AI 模型需要哪些資料格式做為輸入？	結構化資料：CSV、JSON 和 SQL 資料庫資料表；文字資料：純文字、PDF 和 HTML；影像資料：JPEG、PNG 和 TIFF；音訊資料：WAV 和 MP3；影片資料：MP4 和 AVI。

問題	回應範例
您對生成式 AI 工作負載的主要資料品質考量為何？	完整性：確保沒有關鍵欄位遺失；準確性：驗證資料正確性和消除錯誤；一致性：跨來源維持統一格式和值；及時性：確保資料是最新的以進行即時推論；相關性：確認資料符合特定的生成式 AI 任務。
什麼是關鍵效能需求（例如回應時間、輸送量、準確性）？	95% 準確度；回應時間 < 500 毫秒；每秒處理 1000 個請求的能力。高準確度 (95%+)、中等準確度 (80-90%)、盡最大努力等。
您是否有任何其他 KPIs 來衡量此使用案例的成功？	關鍵 KPIs 包括降低錯誤率、節省每筆交易的時間，以及客戶滿意度分數。
需要多少模型準確度，以及它如何與成本平衡？	具有中等成本的高準確度 (>90%)、具有低成本的中等準確度 (70-80%)，以此類推。
生成式 AI 解決方案的主要使用案例或案例是什麼？	客戶服務聊天機器人、內容產生、產品建議等。
生成式 AI 系統的目標使用者或角色是什麼？	客戶服務代理、行銷團隊、員工、最終使用者等。
預期的請求或使用者數量是多少？	每天 1,000 個請求；每月 10,000 個作用中使用者。
是否有任何特定的使用案例限制或要求？	即時回應、多語言支援、資料隱私權等。
您是否有任何用於開發和維護生成式 AI 解決方案的配置預算？	初始開發成本估計為 200,000 美元，年度維護成本為 50,000 美元。
此使用案例的預計投資報酬率 (ROI) 和償還期間為何？	三年內預期的投資報酬率為 150%，而回收期為 18 個月。
是否有任何應考慮的隱藏成本或潛在節省？	潛在的節省包括降低加班成本。隱藏成本可能涉及員工的額外訓練。

問題	回應範例
這個生成式 AI 解決方案的可擴展性和未來擴展可能性是什麼？	該解決方案旨在隨著我們的營運進行擴展，並可能在未來擴展到其他部門。
如何確保公平性並減少生成式 AI 模型中的偏差？	我們計劃透過各種資料收集、定期偏差稽核和偏差緩解技術的實作來緩解偏差。
您有哪些程序可以解決道德問題或意外後果？	我們將透過既定的 AI 事件回應計劃、定期的道德風險評估、員工的匿名報告系統、與外部道德專家的合作，以及根據意見回饋持續監控和調整部署的模型，來管理道德問題。
如何對組織中不同專案和部門的生成式 AI 工作負載評估進行優先順序和排序？	透過跨所有部門進行高階調查，以識別潛在的生成式 AI 使用案例，並根據三個關鍵條件進行評估：業務影響、技術可行性和道德考量。具有高潛在影響、較低的技術障礙和最低道德問題的專案會優先考慮。

架構

問題	回應範例
正在考慮哪種類型的生成式 AI 模型或架構？	轉換器、卷積神經網路 (CNN)、遞歸神經網路 (RNN)、決策樹等。
預期的資料和運算規模或數量是多少？	數百萬使用者、PB 的資料等。
訓練和推論的硬體需求（例如 CPUs 或 GPUs）有哪些？	高階 GPUs、CPU 叢集、雲端執行個體等。
生成式 AI 模型將如何隨著時間更新或重新訓練？	透過持續學習、定期重新訓練、手動更新等。
什麼是資料預先處理和特徵工程需求？	文字清理、影像增強、功能選擇等。
生成式 AI 系統會如何處理邊緣案例、極端值或低可信度輸入？	透過回復至人工監督、請求釐清等。

問題	回應範例
生成式 AI 應用程式的延遲要求是什麼？	即時、近乎即時、批次處理等。

儲存

問題	回應範例
訓練資料會存放在何處？	在雲端儲存體（例如 Amazon S3、檔案儲存體、區塊儲存體或物件儲存體）、內部部署儲存體等。
訓練資料和模型成品的儲存需求為何（例如，容量、耐用性、可用性）？	PB 級儲存、高耐用性 (99.999999999% 耐用性)、高可用性等。
訓練資料和模型成品的資料保留和備份需求是什麼？	x 年的資料保留、每日備份、異地備份等。
哪些檔案格式主要用於存放 AI 訓練資料集（例如 CSV、JSON、Parquet、HDF5）？	用於結構化資料的 Parquet 檔案，以及用於大型多維陣列和非結構化資料的 HDF5，例如影像和文字。我們使用特殊格式，例如 TFRecord，以在訓練期間最佳化資料載入。
如何組織訓練資料集：作為個別檔案、在資料庫中或使用特殊 AI 資料格式？	中小型資料集會儲存為物件儲存體中的個別 Parquet 檔案，以提供彈性。大型資料集會存放在分散式資料庫 (Cassandra) 中來處理擴展。
您是否特別針對生成式 AI 訓練資料使用任何資料壓縮或編碼技術？	對於表格式資料，我們使用 Parquet 中提供的字典編碼和位元封裝技術。對於影像，我們使用失真 JPEG 壓縮搭配針對模型最佳化的品質設定。
如何處理訓練資料集不同反覆運算的版本控制和儲存？這對您的整體儲存需求有何影響？	我們使用與 ML 平台整合的資料版本控制系統 (DVC)。

法規與合規

問題	回應範例
生成式 AI 解決方案 (例如 GDPR、HIPA A、PCI-DSS) 有哪些相關法規或合規要求？	用於處理個人資料的 GDPR、醫療保健資料的 HIPAA、付款資料的 PCI-DSS 等。
您的組織採用了哪些道德生成式 AI 指導方針或架構？	我們實作自己的負責任 AI 指導方針。所有生成式 AI 專案在核准和部署之前都會經過道德審查。
生成式 AI 系統的安全要求是什麼？	資料加密、安全網路通訊、定期安全稽核。
資料隱私權和保護有哪些要求？	資料匿名化、加密、存取控制等。
解決方案處理敏感或機密資料的要求是什麼？	嚴格的存取控制、資料遮罩、資料駐留要求等。
如何處理使用者身分驗證和授權？	透過使用 OAuth、API 金鑰、單一登入 (SSO) 和角色型存取控制 (RBAC)。
如何在生產環境中監控和管理解決方案？	透過使用 Prometheus 和 Datadog 等監控工具，記錄 ELK Stack、警示系統等工具。

整合

問題	回應範例
將生成式 AI 解決方案與現有系統或資料來源整合有哪些需求？	REST APIs、訊息佇列、資料庫連接器等。
如何擷取和預先處理生成式 AI 解決方案的資料？	透過使用批次處理、串流資料、資料轉換和功能工程。
如何耗用生成式 AI 解決方案的輸出，或與下游系統整合？	透過 API 端點、訊息佇列、資料庫更新等。

問題	回應範例
哪些事件驅動整合模式可用於生成式 AI 解決方案？	訊息佇列（例如 Amazon SQS、Apache Kafka、RabbitMQ）、pub/sub 系統、Webhook、事件串流平台。
哪些 API 型整合方法可用來將生成式 AI 解決方案與其他系統連線？	RESTful APIs、GraphQL APIs、SOAP APIs（適用於舊版系統）。
哪些微服務架構元件可用於生成式 AI 解決方案整合？	服務間通訊的服務網格、API 閘道、容器協同運作（例如 Kubernetes）。
如何為生成式 AI 解決方案實作混合整合？	透過結合事件驅動的即時更新模式、歷史資料的批次處理，以及外部系統整合 APIs。
生成式 AI 解決方案輸出如何與下游系統整合？	透過 API 端點、訊息佇列、資料庫更新、Webhook 和檔案匯出。
整合生成式 AI 解決方案時應考慮哪些安全措施？	身分驗證機制（例如 OAuth 或 JWT）、加密（傳輸中和靜態）、API 速率限制和存取控制清單 ACLs）。
您打算如何將 LlamaIndex 或 LangChain 等開放原始碼架構整合到現有的資料管道和生成式 AI 工作流程中？	我們計劃使用 LangChain 來建置複雜的生成式 AI 應用程式，特別是其代理程式和記憶體管理功能。我們的目標是讓 60% 的生成式 AI 專案在未來 6 個月內使用 LangChain。
如何確保所選開放原始碼架構與現有資料基礎設施之間的相容性？	我們正在建立專用整合團隊，以確保順暢的相容性。到了第三季度，我們的目標是擁有完全整合的管道，該管道使用 LlamaIndex 在我們目前的資料湖結構中實現有效的資料索引和擷取。
您計劃如何利用 LangChain 等架構的模組化元件進行快速原型設計和實驗？	我們正在設定沙盒環境，開發人員可以使用 LangChain 的元件快速建立原型。
在快速發展的開放原始碼架構中，您跟上更新和新功能的策略是什麼？	我們已指派一個團隊來監控 LangChain 和 LlamaIndex 的 GitHub 儲存庫和社群論壇。我們計劃每季評估和整合重大更新，專注於效能改善和新功能。

測試

問題	回應範例
測試要求是什麼（例如，單元測試、整合測試、end-to-end測試）？	個別元件的單元測試、與外部系統的整合測試、關鍵案例的end-to-end測試等。
如何確保生成式 AI 訓練在不同來源之間的資料品質和一致性？	我們透過自動化資料分析工具、定期資料稽核和集中式資料目錄來維護資料品質。我們已實作資料控管政策，以確保來源之間的一致性，並維護資料歷程。
如何評估和驗證生成式 AI 模型？	透過使用保留資料集、人工評估、A/B 測試等。
評估生成式 AI 模型的效能和準確性的條件是什麼？	精確度、召回、F1 分數、複雜度、人工評估等。
如何識別和處理邊緣案例和角落案例？	透過使用全方位的測試套件、人工評估、對手測試等。
如何測試生成式 AI 模型中的潛在偏差？	透過使用人口統計平等分析、平等機會測試、對手脫偏差技術、反事實測試等。
哪些指標將用於衡量模型輸出的公平性？	不同的影響比例、相等的奇數、人口統計平等、個別公平性指標等。
如何確保測試資料集中的多樣化表示以進行偏差偵測？	透過跨人口統計群組使用分層抽樣、與多樣性專家合作、使用合成資料填補差距等。
部署後將實作哪個程序來持續監控模型公平性？	定期公平性稽核、自動化偏差偵測系統、使用者意見回饋分析、定期重新訓練更新後的資料集等。
您要如何處理生成式 AI 模型中的交集偏差？	透過使用交集公平性分析、子組測試、與網域專家在交集方面的協作等。
如何測試模型在不同語言和文化環境中的效能？	透過使用多語言測試集、與文化專家的協作、當地公平性指標、跨文化比較研究等。

部署和自動化

問題	回應範例
擴展和負載平衡有哪些需求？	智慧請求路由；自動擴展系統；透過採用模型快取、延遲載入和分散式儲存系統等技術來最佳化快速冷啟動；設計系統來處理爆量、無法預測的流量模式。
更新和推出新版本有哪些要求？	藍/綠部署、金絲雀版本、滾動更新等。
災難復原和業務持續性有哪些要求？	備份和還原程序、容錯移轉機制、高可用性組態等。
自動化生成式 AI 模型的訓練、部署和管理有哪些需求？	自動化訓練管道、持續部署、自動擴展等。
當有新資料可用時，如何更新和重新訓練生成式 AI 模型？	透過定期重新訓練、增量學習、遷移學習等。
自動化監控和管理的需求是什麼？	自動化提醒、自動擴展、自我修復等。
對於生成式 AI 工作負載，您偏好的部署環境是什麼？	一種混合方法，使用 AWS 進行模型訓練，使用我們的內部部署基礎設施進行推論，以滿足資料駐留需求。
是否有任何您偏好用於生成式 AI 部署的特定雲端平台？	AWS 服務，特別是用於模型開發和部署的 Amazon SageMaker AI，以及用於基礎模型的 Amazon Bedrock。
您正在考慮將哪些容器化技術用於生成式 AI 工作負載？	我們希望標準化與 Kubernetes 協調的 Docker 容器，以確保混合環境中的可攜性和可擴展性。
在生成式 AI 管道中，您是否有任何偏好的 CI/CD 工具？	用於版本控制和 CI/CD 管道的 GitLab，與 Jenkins 整合，用於自動化測試和部署。
您考慮使用哪些協同運作工具來管理生成式 AI 工作流程？	工作流程協同運作的 Apache Airflow，特別是資料預先處理和模型訓練管道。

問題	回應範例
對於支援生成式 AI 工作負載的內部部署基礎設施，您是否有任何特定需求？	我們投資於 GPU 加速伺服器 and 高速聯網，以支援內部部署推論工作負載。
您打算如何管理跨不同環境的模型版本控制和部署？	我們計劃使用 MLflow 進行模型追蹤和版本控制，並將其與 Kubernetes 基礎設施整合，以跨環境無縫部署。
您正在考慮哪些用於生成式 AI 部署的監控和可觀測性工具？	用於指標收集的 Prometheus 和用於視覺化的 Grafana，以及用於模型特定監控的其他自訂記錄解決方案。
您如何解決混合部署模型中的資料移動和同步問題？	我們將使用在現場部署儲存與之間進行有效率 AWS DataSync 的資料傳輸 AWS，以及根據我們的訓練週期排程的自動同步任務。
您要針對跨不同環境的生成式 AI 部署實作哪些安全措施？	我們會將 IAM 用於雲端資源，與內部部署 Active Directory 整合，以實作 end-to-end 加密和網路分割，以保護資料流程。

資料策略

問題	回應範例
哪些特定資料類型對您的生成式 AI 工作負載至關重要，以及目前可存取這些類型的多少百分比？	客戶通話日誌和產品檢閱資料至關重要。目前，85% 的這些資料類型可供我們的生成式 AI 專案存取。
如何確保和測量資料的品質？	我們已實作資料品質指標，包括完整性、準確性、一致性和及時性。我們使用自動化工具定期評估這些指標，並擁有專用的團隊來清理資料和擴充資料。
有多少百分比的資料符合生成式 AI 使用的品質標準？	目前，78% 的資料符合我們的品質標準。透過改善資料清理程序，我們的目標是在未來 12 個月內達到 95%。

問題	回應範例
您打算如何在利益相關者之間建立對生成式 AI 中資料用量的信任？	我們正在實作 AI 道德委員會、提供 AI 決策的明確說明，以及每季進行 AI 稽核，以確保透明度和公平性。
資料來源和譜系的文件有多完整？	我們會維護詳細的資料目錄，其中包含所有資料來源的中繼資料，包括原始伺服器、更新頻率和用量。我們使用資料譜系工具來追蹤資料如何在系統中流動和轉換。
如何確保資料集的多樣性，以防止 AI 模型中的偏差？	我們主動從各種人口統計資訊中取得資料，並定期稽核資料集是否有代表性偏差。我們也使用合成資料產生技術來平衡代表性不足的類別。
關鍵生成式 AI 模型的資料重新整理率是多少，以及如何判斷此頻率？	每週重新整理關鍵模型。此頻率由 A/B 測試效能指標決定，我們的目標是在重新整理之間不會降解超過 2%。
您維護多少個關鍵資料集版本以及保留多久？	我們維護每個關鍵資料集的最後五個版本，每個版本的保留期間為 18 個月。
您的生成式 AI 計畫涉及多少個跨職能團隊，並且可以存取您的資料？	我們有三個跨職能團隊。每個團隊都包含資料科學家、網域專家、道德專家和商業分析師。
您有哪些資料控管政策和實務？	我們有一個跨功能資料管理委員會，負責監督我們的資料政策。我們已實作以角色為基礎的存取控制、資料分類機制和定期稽核，以確保符合我們的控管架構。
您採取哪些措施來確保資料隱私權、取得適當同意，以及維護機密性？	我們已實作符合 GDPR 和 CCPA 的完整資料隱私權架構。這包括取得資料用量的明確同意、實作資料匿名化技術，以及定期隱私權影響評估。
上個季度稽核了多少百分比的 AI 訓練資料集是否有偏差？	上個季度稽核了 70% 的 AI 訓練資料集是否有偏差。我們正在實作自動化偏差偵測工具，以達到 100% 每季稽核。

問題	回應範例
您目前的資料處理容量是多少，您預計未來生成式 AI 工作負載需要多少？	我們目前的容量為每天 10 TB。我們預計在一年內每天需要 30 TB，並且正在擴展我們的基礎設施以滿足此需求。
在資料隱私權與生成式 AI 模型的資料需求之間取得平衡的策略是什麼？	我們正在實作進階匿名化技術和合成資料產生。我們的目標是將 AI 的可用資料增加 40%，同時將明年的隱私權風險降低 60%。
準確標記機器學習 (ML) 資料集的百分比是多少？目標準確度是多少？	目前，85% 的 ML 資料集已正確標記。我們採用人工和自動化標記技術，以下一季 95% 的準確性率為目標。

將評估洞見轉換為可行的結果

本節提供分析問卷回應的架構，並使用這些洞見來塑造生成式 AI 現代化計畫的目標架構和其他關鍵交付項目。此架構填補了資料收集與實作之間的差距，並確保評估會直接通知並推動您的現代化策略。

目標架構定義：

- 使用問卷回應來通知雲端服務的選擇和資料管道的設計。
- 請確定架構設計支援可擴展性和互通性，如 指南中所強調。

客戶整備度評估：

- 分析與目前基礎設施、程序和組織文化相關的問題回應。
- 識別差距並建立解決差距的計畫。優先考慮對 MVP 成功至關重要的差距。

使用案例和延伸目標：

- 從問卷回應中擷取特定業務問題，以定義明確的使用案例目標。
- 設定延伸目標，以符合組織對生成式 AI 現代化的長期願景。

嘗試估計：

- 使用問卷資料來預估 MVP 和完整實作的資源、時間和預算。
- 建立以 MVP 開頭的分階段方法，並概述後續階段。

啟用需求：

- 根據問卷回應，找出技能差距和訓練需求。
- 制定訓練計畫，同時支援立即 MVP 需求和長期採用生成式 AI。

實作計畫：

- 建立以 MVP 開頭的完整藍圖，並概述完整生成式 AI 現代化的步驟。
- 定義實作每個階段的明確里程碑和交付項目。

實際步驟：

- 優先順序矩陣：建立將問卷回應映射至六個結果的矩陣，以協助排定特徵和工作的優先順序。
- 疊代方法：將 MVP 設計成一系列計劃版本中的第一個疊代，其中每個版本都朝向完整目標架構建置。
- 利益相關者一致性：使用問卷結果來調整 MVP 範圍的利益相關者，以及實現所有結果的分階段方法。
- 持續意見回饋迴圈：實作機制以在 MVP 部署後收集意見回饋，並使用洞察來精簡後續階段的計劃。
- 敏捷實作：採用敏捷的方法，允許彈性地解決一段時間內的所有結果，從 MVP 中最關鍵的結果開始。

後續步驟

完成生成式 AI 工作負載評估後，請遵循下列步驟：

1. 提供詳細的目標架構

- 目標：解決方案架構師會建立符合組織目標和評估結果的完整目標架構。
- 元件：此架構包含資料擷取、整合點和系統互通性的設計，以確保可擴展性、可靠性和效能最佳化。

2. 說明 AWS 服務 是否適合使用案例

- 服務映射：識別和映射 AWS 服務 最適合已識別使用案例的特定。
- 優點：強調這些服務如何處理特定業務需求、提高效率並提供可擴展性。

3. 提供具有優缺點的選用替代解決方案

- 替代方案：提供也可以符合組織需求的替代解決方案。
- 分析：透過考慮成本、複雜性和符合業務目標等因素，提供每個替代方案的優點和缺點的詳細分析。

4. 提供 的詳細價格估算 AWS 服務

- 成本分析：為提議的 提供詳細的成本估算 AWS 服務，包括潛在的使用案例和定價模型。
- 預算一致性：確保成本符合組織的預算限制條件，並清楚地了解財務影響。

5. 取得提議架構的意見回饋

- 利益相關者參與：與利益相關者互動，以呈現提議的架構並收集意見回饋。
- 反覆改進：使用意見回饋來精簡和改善解決方案，並確認解決方案符合所有利益相關者的需求和期望。

常見問答集

生成式 AI 工作負載評估的主要目標是什麼？

評估的主要目標是評估組織對現代化生成式 AI 工作負載、識別使用案例和開發目標解決方案架構的準備程度。它旨在定義現代化需求、確定實作範圍，並為成功的生成式 AI 現代化做好準備。

誰應該使用此評估？

此評估適用於想要評估生成式 AI 現代化技術層面的解決方案架構師、企業架構師和應用程式架構師。它也適用於計畫經理和人員經理，以衡量整體準備程度、資源分配和啟用需求。

評估中評估哪些關鍵元件？

評估涵蓋整體整備、使用案例、架構、儲存、法規和合規、整合、測試、部署自動化和資料策略。這些元件對於判斷生成式 AI 現代化採用的技術和組織準備程度至關重要。

評估如何協助定義目標架構？

評估提供結構化方法來評估目前的系統並識別改善項目。它可協助您選擇適當的技術和設計符合業務目標和使用案例需求的可擴展架構。

執行生成式 AI 工作負載評估有何好處？

優點包括提高效率、改善決策、合規保證、創新促進和可擴展性準備。評估建立生成式 AI 現代化的策略方法，並在降低風險的同時最大化潛在效益。

組織如何確保評估後成功實作？

組織應制定明確的實作計畫，其中包含定義的里程碑、及早吸引利益相關者參與，並採用反覆方法。建立卓越中心 (CoE) 並專注於人才培養也是建議的最佳實務。

組織在評估期間可能會面臨哪些挑戰？

挑戰可能包括對變革的阻力、資料品質問題和合規複雜性。解決這些挑戰需要培養創新文化、確保資料準備度，以及實作強大的安全措施。

評估如何解決法規和合規要求？

評估會評估目前的合規措施並識別差距。它可確保目標解決方案遵守相關法規和資料隱私權法律，並納入安全最佳實務來保護敏感資訊。

利益相關者參與在評估過程中扮演什麼角色？

利益相關者參與對於獲得接受、使現代化計劃與業務目標保持一致，以及確保成功實作至關重要。早期參與和明確傳達好處是克服阻力和培養支援的關鍵。

評估後，組織如何衡量其生成式 AI 現代化計劃的成功？

您可以使用符合業務目標的關鍵績效指標 (KPIs) 來衡量成功。定期監控和評估這些指標有助於引導決策，並向利益相關者展示生成式 AI 現代化的價值。

對於大小不一的組織（小型、中型或企業）或產業，評估方法有何不同？

小型組織：

- 全面評估的資源和專業知識可能有限
- 可能專注於特定的高影響使用案例，而不是整個企業的採用
- 可能更依賴第三方工具和服務進行評估
- 評估程序可能較不正式且更靈活

中型組織：

- 通常擁有專用 IT 或資料團隊，但可能缺乏專業 AI 專業知識
- 可能採取分階段方法，從關鍵部門的試驗專案開始
- 需要平衡創新與現有系統和程序
- 評估可能涉及跨職能團隊

企業組織：

- 通常具有專用 AI/ML 團隊和更多資源以進行全面評估

- 需要考慮與現有企業系統的複雜整合
- 可能需要考慮產業特定的法規要求
- 評估通常涉及正式的控管程序

資源

- [上的生成式 AI AWS](#)
- [AWS 提供新的人工智慧、機器學習和生成式 AI 指南，以規劃您的 AI 策略](#) (AWS 部落格文章)
- [在上建置生成式 AI 應用程式的最佳實務 AWS](#)(AWS 部落格文章)
- [上的生成式 AI 應用程式建置器 AWS](#) (AWS 解決方案程式庫)
- [生成式 AI 功能 AWS](#) (安全參考架構)
- [AWS 生成式 AI 最佳實務架構](#) (AWS Audit Manager 使用者指南)
- [選擇生成式 AI 服務](#) (AWS 決策指南)
- [什麼是 Amazon Bedrock ?](#) (Amazon Bedrock 使用者指南)
- [什麼是 Amazon SageMaker AI ?](#) (Amazon SageMaker AI 開發人員指南)

文件歷史紀錄

下表描述了本指南的重大變更。如果您想收到有關未來更新的通知，可以訂閱 [RSS 摘要](#)。

變更	描述	日期
初次出版	—	2024 年 11 月 6 日

AWS 規範性指引詞彙表

以下是 AWS Prescriptive Guidance 提供的策略、指南和模式中常用的術語。若要建議項目，請使用詞彙表末尾的提供意見回饋連結。

數字

7 R

將應用程式移至雲端的七種常見遷移策略。這些策略以 Gartner 在 2011 年確定的 5 R 為基礎，包括以下內容：

- 重構/重新架構 – 充分利用雲端原生功能來移動應用程式並修改其架構，以提高敏捷性、效能和可擴展性。這通常涉及移植作業系統和資料庫。範例：將您的現場部署 Oracle 資料庫遷移至 Amazon Aurora PostgreSQL 相容版本。
- 平台轉換 (隨即重塑) – 將應用程式移至雲端，並引入一定程度的優化以利用雲端功能。範例：將您的現場部署 Oracle 資料庫遷移至 中的 Amazon Relational Database Service (Amazon RDS) for Oracle AWS 雲端。
- 重新購買 (捨棄再購買) – 切換至不同的產品，通常從傳統授權移至 SaaS 模型。範例：將您的客戶關係管理 (CRM) 系統遷移至 Salesforce.com。
- 主機轉換 (隨即轉移) – 將應用程式移至雲端，而不進行任何變更以利用雲端功能。範例：將您的現場部署 Oracle 資料庫遷移至 中 EC2 執行個體上的 Oracle AWS 雲端。
- 重新放置 (虛擬機器監視器等級隨即轉移) – 將基礎設施移至雲端，無需購買新硬體、重寫應用程式或修改現有操作。您可以將伺服器從內部部署平台遷移到相同平台的雲端服務。範例：將 Microsoft Hyper-V 應用程式遷移至 AWS。
- 保留 (重新檢視) – 將應用程式保留在來源環境中。其中可能包括需要重要重構的應用程式，且您希望將該工作延遲到以後，以及您想要保留的舊版應用程式，因為沒有業務理由來進行遷移。
- 淘汰 – 解除委任或移除來源環境中不再需要的應用程式。

A

ABAC

請參閱 [屬性型存取控制](#)。

抽象服務

請參閱 [受管服務](#)。

ACID

請參閱 [原子性、一致性、隔離性、持久性](#)。

主動-主動式遷移

一種資料庫遷移方法，其中來源和目標資料庫保持同步 (透過使用雙向複寫工具或雙重寫入操作)，且兩個資料庫都在遷移期間處理來自連接應用程式的交易。此方法支援小型、受控制批次的遷移，而不需要一次性切換。它更靈活，但比 [主動-被動遷移](#) 需要更多的工作。

主動-被動式遷移

一種資料庫遷移方法，其中來源和目標資料庫保持同步，但只有來源資料庫會在資料複寫至目標資料庫時處理來自連線應用程式的交易。目標資料庫在遷移期間不接受任何交易。

彙總函數

在一組資料列上操作並計算群組單一傳回值的 SQL 函數。彙總函數的範例包括 SUM 和 MAX。

AI

請參閱 [人工智慧](#)。

AIOps

請參閱 [人工智慧操作](#)。

匿名化

永久刪除資料集中個人資訊的程序。匿名化有助於保護個人隱私權。匿名資料不再被視為個人資料。

反模式

經常用於重複性問題的解決方案，其中解決方案具有反效益、無效，或比替代解決方案更有效。

應用程式控制

一種安全方法，僅允許使用核准的應用程式，以協助保護系統免受惡意軟體攻擊。

應用程式組合

有關組織使用的每個應用程式的詳細資訊的集合，包括建置和維護應用程式的成本及其商業價值。此資訊是 [產品組合探索和分析程序](#) 的關鍵，有助於識別要遷移、現代化和優化的應用程式並排定其優先順序。

人工智慧 (AI)

電腦科學領域，致力於使用運算技術來執行通常與人類相關的認知功能，例如學習、解決問題和識別模式。如需詳細資訊，請參閱[什麼是人工智慧？](#)

人工智慧操作 (AIOps)

使用機器學習技術解決操作問題、減少操作事件和人工干預以及提高服務品質的程序。如需有關如何在 AWS 遷移策略中使用 AIOps 的詳細資訊，請參閱[操作整合指南](#)。

非對稱加密

一種加密演算法，它使用一對金鑰：一個用於加密的公有金鑰和一個用於解密的私有金鑰。您可以共用公有金鑰，因為它不用於解密，但對私有金鑰存取應受到高度限制。

原子性、一致性、隔離性、持久性 (ACID)

一組軟體屬性，即使在出現錯誤、電源故障或其他問題的情況下，也能確保資料庫的資料有效性和操作可靠性。

屬性型存取控制 (ABAC)

根據使用者屬性 (例如部門、工作職責和團隊名稱) 建立精細許可的實務。如需詳細資訊，請參閱《AWS Identity and Access Management (IAM) 文件》中的[ABAC for AWS](#)。

授權資料來源

您存放主要版本資料的位置，被視為最可靠的資訊來源。您可以將授權資料來源中的資料複製到其他位置，以處理或修改資料，例如匿名、修訂或假名化資料。

可用區域

中的不同位置 AWS 區域，可隔離其他可用區域中的故障，並提供相同區域中其他可用區域的低成本、低延遲網路連線。

AWS 雲端採用架構 (AWS CAF)

的指導方針和最佳實務架構 AWS，可協助組織制定高效且有效的計劃，以成功地移至雲端。AWS CAF 將指導方針組織到六個重點領域：業務、人員、治理、平台、安全和營運。業務、人員和控管層面著重於業務技能和程序；平台、安全和操作層面著重於技術技能和程序。例如，人員層面針對處理人力資源 (HR)、人員配備功能和人員管理的利害關係人。因此，AWS CAF 為人員開發、訓練和通訊提供指引，協助組織做好成功採用雲端的準備。如需詳細資訊，請參閱[AWS CAF 網站](#)和[AWS CAF 白皮書](#)。

AWS 工作負載資格架構 (AWS WQF)

一種工具，可評估資料庫遷移工作負載、建議遷移策略，並提供工作預估值。AWS WQF 隨附於 AWS Schema Conversion Tool (AWS SCT)。它會分析資料庫結構描述和程式碼物件、應用程式程式碼、相依性和效能特性，並提供評估報告。

B

錯誤的機器人

旨在中斷或傷害個人或組織的[機器人](#)。

BCP

請參閱[業務持續性規劃](#)。

行為圖

資源行為的統一互動式檢視，以及一段時間後的互動。您可以將行為圖與 Amazon Detective 搭配使用來檢查失敗的登入嘗試、可疑的 API 呼叫和類似動作。如需詳細資訊，請參閱偵測文件中的[行為圖中的資料](#)。

大端序系統

首先儲存最高有效位元組的系統。另請參閱 [Endianness](#)。

二進制分類

預測二進制結果的過程 (兩個可能的類別之一)。例如，ML 模型可能需要預測諸如「此電子郵件是否是垃圾郵件？」等問題 或「產品是書還是汽車？」

Bloom 篩選條件

一種機率性、記憶體高效的資料結構，用於測試元素是否為集的成員。

藍/綠部署

一種部署策略，您可以在其中建立兩個不同但相同的環境。您可以在一個環境（藍色）中執行目前的應用程式版本，並在另一個環境（綠色）中執行新的應用程式版本。此策略可協助您快速復原，並將影響降至最低。

機器人

透過網際網路執行自動化任務並模擬人類活動或互動的軟體應用程式。有些機器人有用或有益，例如在網際網路上編製資訊索引的 Web 爬蟲程式。有些其他機器人稱為惡意機器人，旨在中斷或傷害個人或組織。

殭屍網路

受到[惡意軟體](#)感染且由單一方控制的[機器人](#)網路，稱為機器人繼承器或機器人運算子。殭屍網路是擴展機器人及其影響的最佳已知機制。

分支

程式碼儲存庫包含的區域。儲存庫中建立的第一個分支是主要分支。您可以從現有分支建立新分支，然後在新分支中開發功能或修正錯誤。您建立用來建立功能的分支通常稱為功能分支。當準備好發佈功能時，可以將功能分支合併回主要分支。如需詳細資訊，請參閱[關於分支](#) (GitHub 文件)。

碎片存取

在特殊情況下，並透過核准的程序，讓使用者快速取得他們通常無權存取 AWS 帳戶 之 的存取權。如需詳細資訊，請參閱 Well-Architected 指南中的 AWS [實作碎片程序](#) 指標。

棕地策略

環境中的現有基礎設施。對系統架構採用棕地策略時，可以根據目前系統和基礎設施的限制來設計架構。如果正在擴展現有基礎設施，則可能會混合棕地和[綠地](#)策略。

緩衝快取

儲存最常存取資料的記憶體區域。

業務能力

業務如何創造價值 (例如，銷售、客戶服務或營銷)。業務能力可驅動微服務架構和開發決策。如需詳細資訊，請參閱在 [AWS 上執行容器化微服務](#) 白皮書的 [圍繞業務能力進行組織](#) 部分。

業務連續性規劃 (BCP)

一種解決破壞性事件 (如大規模遷移) 對營運的潛在影響並使業務能夠快速恢復營運的計畫。

C

CAF

請參閱[AWS 雲端採用架構](#)。

Canary 部署

版本對最終使用者的緩慢和增量版本。當您有信心時，您可以部署新版本並完全取代目前的版本。

CCoE

請參閱 [Cloud Center of Excellence](#)。

CDC

請參閱[變更資料擷取](#)。

變更資料擷取 (CDC)

追蹤對資料來源 (例如資料庫表格) 的變更並記錄有關變更的中繼資料的程序。您可以將 CDC 用於各種用途，例如稽核或複寫目標系統中的變更以保持同步。

混沌工程

故意引入故障或破壞性事件，以測試系統的彈性。您可以使用 [AWS Fault Injection Service \(AWS FIS\)](#) 執行實驗，為您的 AWS 工作負載帶來壓力，並評估其回應。

CI/CD

請參閱[持續整合和持續交付](#)。

分類

有助於產生預測的分類程序。用於分類問題的 ML 模型可預測離散值。離散值永遠彼此不同。例如，模型可能需要評估影像中是否有汽車。

用戶端加密

在目標 AWS 服務接收資料之前，在本機加密資料。

雲端卓越中心 (CCoE)

一個多學科團隊，可推動整個組織的雲端採用工作，包括開發雲端最佳實務、調動資源、制定遷移時間表以及領導組織進行大規模轉型。如需詳細資訊，請參閱 AWS 雲端企業策略部落格上的 [CCoE 文章](#)。

雲端運算

通常用於遠端資料儲存和 IoT 裝置管理的雲端技術。雲端運算通常連接到[邊緣運算](#)技術。

雲端操作模型

在 IT 組織中，用於建置、成熟和最佳化一或多個雲端環境的操作模型。如需詳細資訊，請參閱[建置您的雲端操作模型](#)。

採用雲端階段

組織在遷移至時通常會經歷的四個階段 AWS 雲端：

- 專案 – 執行一些與雲端相關的專案以進行概念驗證和學習用途
- 基礎 – 進行基礎投資以擴展雲端採用 (例如，建立登陸區域、定義 CCoE、建立營運模型)

- 遷移 – 遷移個別應用程式
- 重塑 – 優化產品和服務，並在雲端中創新

部落格文章中的 Stephen Orban 定義了這些階段：AWS 雲端 企業策略部落格上的[邁向雲端優先之旅和採用階段](#)。如需有關它們如何與 AWS 遷移策略關聯的資訊，請參閱[遷移整備指南](#)。

CMDB

請參閱[組態管理資料庫](#)。

程式碼儲存庫

透過版本控制程序來儲存及更新原始程式碼和其他資產 (例如文件、範例和指令碼) 的位置。常見的雲端儲存庫包括 GitHub 或 Bitbucket Cloud。程式碼的每個版本都稱為分支。在微服務結構中，每個儲存庫都專用於單個功能。單一 CI/CD 管道可以使用多個儲存庫。

冷快取

一種緩衝快取，它是空的、未填充的，或者包含過時或不相關的資料。這會影響效能，因為資料庫執行個體必須從主記憶體或磁碟讀取，這比從緩衝快取讀取更慢。

冷資料

很少存取且通常是歷史資料的資料。查詢這類資料時，通常可接受慢查詢。將此資料移至效能較低且成本較低的儲存層或類別，可以降低成本。

電腦視覺 (CV)

AI 欄位^{???}，使用機器學習從數位影像和影片等視覺化格式分析和擷取資訊。例如，Amazon SageMaker AI 提供 CV 的影像處理演算法。

組態偏離

對於工作負載，組態會從預期狀態變更。這可能會導致工作負載變得不合規，而且通常是漸進和無意的。

組態管理資料庫 (CMDB)

儲存和管理有關資料庫及其 IT 環境的資訊的儲存庫，同時包括硬體和軟體元件及其組態。您通常在遷移的產品組合探索和分析階段使用 CMDB 中的資料。

一致性套件

您可以組合的 AWS Config 規則和修補動作集合，以自訂您的合規和安全檢查。您可以使用 YAML 範本，將一致性套件部署為 AWS 帳戶 和 區域中或整個組織的單一實體。如需詳細資訊，請參閱 AWS Config 文件中的[一致性套件](#)。

持續整合和持續交付 (CI/CD)

自動化軟體發程序的來源、建置、測試、暫存和生產階段的程序。CI/CD 通常被描述為管道。CI/CD 可協助您將程序自動化、提升生產力、改善程式碼品質以及加快交付速度。如需詳細資訊，請參閱[持續交付的優點](#)。CD 也可表示持續部署。如需詳細資訊，請參閱[持續交付與持續部署](#)。

CV

請參閱[電腦視覺](#)。

D

靜態資料

網路中靜止的資料，例如儲存中的資料。

資料分類

根據重要性和敏感性來識別和分類網路資料的程序。它是所有網路安全風險管理策略的關鍵組成部分，因為它可以協助您確定適當的資料保護和保留控制。資料分類是 AWS Well-Architected Framework 中安全支柱的元件。如需詳細資訊，請參閱[資料分類](#)。

資料偏離

生產資料與用於訓練 ML 模型的資料之間有意義的變化，或輸入資料隨時間有意義的變更。資料偏離可以降低 ML 模型預測的整體品質、準確性和公平性。

傳輸中的資料

在您的網路中主動移動的資料，例如在網路資源之間移動。

資料網格

架構架構，提供分散式、分散式資料擁有權與集中式管理。

資料最小化

僅收集和處理嚴格必要資料的原則。在中實作資料最小化 AWS 雲端可以降低隱私權風險、成本和分析碳足跡。

資料周邊

AWS 環境中的一組預防性防護機制，可協助確保只有信任的身分才能從預期的網路存取信任的資源。如需詳細資訊，請參閱[在上建置資料周邊 AWS](#)。

資料預先處理

將原始資料轉換成 ML 模型可輕鬆剖析的格式。預處理資料可能意味著移除某些欄或列，並解決遺失、不一致或重複的值。

資料來源

在整個生命週期中追蹤資料的原始伺服器 and 歷史記錄的程序，例如資料的產生、傳輸和儲存方式。

資料主體

正在收集和處理其資料的個人。

資料倉儲

支援商業智慧的資料管理系統，例如分析。資料倉儲通常包含大量歷史資料，通常用於查詢和分析。

資料庫定義語言 (DDL)

用於建立或修改資料庫中資料表和物件之結構的陳述式或命令。

資料庫處理語言 (DML)

用於修改 (插入、更新和刪除) 資料庫中資訊的陳述式或命令。

DDL

請參閱[資料庫定義語言](#)。

深度整體

結合多個深度學習模型進行預測。可以使用深度整體來獲得更準確的預測或估計預測中的不確定性。

深度學習

一個機器學習子領域，它使用多層人工神經網路來識別感興趣的輸入資料與目標變數之間的對應關係。

深度防禦

這是一種資訊安全方法，其中一系列的安全機制和控制項會在整個電腦網路中精心分層，以保護網路和其中資料的機密性、完整性和可用性。當您在上採用此策略時 AWS，您可以在 AWS Organizations 結構的不同層新增多個控制項，以協助保護資源。例如，defense-in-depth 方法可能會結合多重重要素驗證、網路分割和加密。

委派的管理員

在中 AWS Organizations，相容的服務可以註冊 AWS 成員帳戶，以管理組織的帳戶和管理該服務的許可。此帳戶稱為該服務的委派管理員。如需詳細資訊和相容服務清單，請參閱 AWS Organizations 文件中的[可搭配 AWS Organizations運作的服務](#)。

部署

在目標環境中提供應用程式、新功能或程式碼修正的程序。部署涉及在程式碼庫中實作變更，然後在應用程式環境中建置和執行該程式碼庫。

開發環境

請參閱[環境](#)。

偵測性控制

一種安全控制，用於在事件發生後偵測、記錄和提醒。這些控制是第二道防線，提醒您注意繞過現有預防性控制的安全事件。如需詳細資訊，請參閱在 AWS 上實作安全控制中的[偵測性控制](#)。

開發值串流映射 (DVSM)

一種程序，用於識別對軟體開發生命週期中的速度和品質造成負面影響的限制並排定優先順序。DVSM 擴展了最初專為精簡製造實務設計的價值串流映射程序。它著重於透過軟體開發程序建立和移動價值所需的步驟和團隊。

數位分身

真實世界系統的虛擬呈現，例如建築物、工廠、工業設備或生產線。數位分身支援預測性維護、遠端監控和生產最佳化。

維度資料表

在[星星結構描述](#)中，較小的資料表包含有關事實資料表中量化資料的資料屬性。維度資料表屬性通常是文字欄位或離散數字，其行為類似於文字。這些屬性通常用於查詢限制、篩選和結果集標記。

災難

防止工作負載或系統在其主要部署位置實現其業務目標的事件。這些事件可能是自然災難、技術故障或人為動作的結果，例如意外設定錯誤或惡意軟體攻擊。

災難復原 (DR)

您用來將[災難](#)造成的停機時間和資料遺失降至最低的策略和程序。如需詳細資訊，請參閱 AWS Well-Architected Framework 中的[上工作負載的災難復原 AWS：雲端中的復原](#)。

DML

請參閱[資料庫處理語言](#)。

領域驅動的設計

一種開發複雜軟體系統的方法，它會將其元件與每個元件所服務的不斷發展的領域或核心業務目標相關聯。Eric Evans 在其著作 *Domain-Driven Design: Tackling Complexity in the Heart of Software* (Boston: Addison-Wesley Professional, 2003) 中介紹了這一概念。如需有關如何將領域驅動的設計與 strangler fig 模式搭配使用的資訊，請參閱[使用容器和 Amazon API Gateway 逐步現代化舊版 Microsoft ASP.NET \(ASMX\) Web 服務](#)。

DR

請參閱[災難復原](#)。

偏離偵測

追蹤與基準組態的偏差。例如，您可以使用 AWS CloudFormation 來偵測系統資源中的偏離，也可以使用 AWS Control Tower 來[偵測登陸區域中可能影響控管要求合規性的變更](#)。<https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/using-cfn-stack-drift.html>

DVSM

請參閱[開發值串流映射](#)。

E

EDA

請參閱[探索性資料分析](#)。

EDI

請參閱[電子資料交換](#)。

邊緣運算

提升 IoT 網路邊緣智慧型裝置運算能力的技術。與[雲端運算](#)相比，邊緣運算可以減少通訊延遲並改善回應時間。

電子資料交換 (EDI)

在組織之間自動交換商業文件。如需詳細資訊，請參閱[什麼是電子資料交換](#)。

加密

將人類可讀取的純文字資料轉換為加密文字的運算程序。

加密金鑰

由加密演算法產生的隨機位元的加密字串。金鑰長度可能有所不同，每個金鑰的設計都是不可預測且唯一的。

端序

位元組在電腦記憶體中的儲存順序。大端序系統首先儲存最高有效位元組。小端序系統首先儲存最低有效位元組。

端點

請參閱 [服務端點](#)。

端點服務

您可以在虛擬私有雲端 (VPC) 中託管以與其他使用者共用的服務。您可以使用 [建立端點服務](#)，AWS PrivateLink 並將許可授予其他 AWS 帳戶 或 AWS Identity and Access Management (IAM) 委託人。這些帳戶或主體可以透過建立介面 VPC 端點私下連接至您的端點服務。如需詳細資訊，請參閱 Amazon Virtual Private Cloud (Amazon VPC) 文件中的 [建立端點服務](#)。

企業資源規劃 (ERP)

一種系統，可自動化和管理企業的關鍵業務流程（例如會計、[MES](#) 和專案管理）。

信封加密

使用另一個加密金鑰對某個加密金鑰進行加密的程序。如需詳細資訊，請參閱 AWS Key Management Service (AWS KMS) 文件中的 [信封加密](#)。

環境

執行中應用程式的執行個體。以下是雲端運算中常見的環境類型：

- 開發環境 – 執行中應用程式的執行個體，只有負責維護應用程式的核心團隊才能使用。開發環境用來測試變更，然後再將開發環境提升到較高的環境。此類型的環境有時稱為測試環境。
- 較低的環境 – 應用程式的所有開發環境，例如用於初始建置和測試的開發環境。
- 生產環境 – 最終使用者可以存取的執行中應用程式的執行個體。在 CI/CD 管道中，生產環境是最後一個部署環境。
- 較高的環境 – 核心開發團隊以外的使用者可存取的所有環境。這可能包括生產環境、生產前環境以及用於使用者接受度測試的環境。

epic

在敏捷方法中，有助於組織工作並排定工作優先順序的功能類別。epic 提供要求和實作任務的高層級描述。例如，AWS CAF 安全概念包括身分和存取管理、偵測控制、基礎設施安全、資料保護和事件回應。如需有關 AWS 遷移策略中的 Epic 的詳細資訊，請參閱[計畫實作指南](#)。

ERP

請參閱[企業資源規劃](#)。

探索性資料分析 (EDA)

分析資料集以了解其主要特性的過程。您收集或彙總資料，然後執行初步調查以尋找模式、偵測異常並檢查假設。透過計算摘要統計並建立資料可視化來執行 EDA。

F

事實資料表

[星狀結構描述](#)中的中央資料表。它存放有關業務操作的量化資料。一般而言，事實資料表包含兩種類型的資料欄：包含度量的資料，以及包含維度資料表外部索引鍵的資料欄。

快速失敗

一種使用頻繁和增量測試來縮短開發生命週期的理念。這是敏捷方法的關鍵部分。

故障隔離界限

在中 AWS 雲端，像是可用區域 AWS 區域、控制平面或資料平面等邊界會限制故障的影響，並有助於改善工作負載的彈性。如需詳細資訊，請參閱[AWS 故障隔離界限](#)。

功能分支

請參閱[分支](#)。

特徵

用來進行預測的輸入資料。例如，在製造環境中，特徵可能是定期從製造生產線擷取的影像。

功能重要性

特徵對於模型的預測有多重要。這通常表示為可以透過各種技術來計算的數值得分，例如 Shapley Additive Explanations (SHAP) 和積分梯度。如需詳細資訊，請參閱[機器學習模型可解譯性 AWS](#)。

特徵轉換

優化 ML 程序的資料，包括使用其他來源豐富資料、調整值、或從單一資料欄位擷取多組資訊。這可讓 ML 模型從資料中受益。例如，如果將「2021-05-27 00:15:37」日期劃分為「2021」、「五月」、「週四」和「15」，則可以協助學習演算法學習與不同資料元件相關聯的細微模式。

少量擷取提示

在要求 [LLM](#) 執行類似的任務之前，提供少量示範任務和所需輸出的範例給 LLM。此技術是內容內學習的應用程式，其中模型會從內嵌在提示中的範例 (快照) 中學習。對於需要特定格式、推理或網域知識的任務，少量的提示非常有效。另請參閱[零鏡頭提示](#)。

FGAC

請參閱[精細存取控制](#)。

精細存取控制 (FGAC)

使用多個條件來允許或拒絕存取請求。

閃切遷移

一種資料庫遷移方法，透過[變更資料擷取](#)使用連續資料複寫，以盡可能在最短的時間內遷移資料，而不是使用分階段方法。目標是將停機時間降至最低。

FM

請參閱[基礎模型](#)。

基礎模型 (FM)

大型深度學習神經網路，已針對廣義和未標記資料的大量資料集進行訓練。FMs 能夠執行各種一般任務，例如了解語言、產生文字和影像，以及以自然語言交談。如需詳細資訊，請參閱[什麼是基礎模型](#)。

G

生成式 AI

已針對大量資料進行訓練的 [AI](#) 模型子集，可使用簡單的文字提示建立新的內容和成品，例如影像、影片、文字和音訊。如需詳細資訊，請參閱[什麼是生成式 AI](#)。

地理封鎖

請參閱[地理限制](#)。

地理限制 (地理封鎖)

Amazon CloudFront 中的選項，可防止特定國家/地區的使用者存取內容分發。您可以使用允許清單或封鎖清單來指定核准和禁止的國家/地區。如需詳細資訊，請參閱 CloudFront 文件中的[限制內容的地理分佈](#)。

Gitflow 工作流程

這是一種方法，其中較低和較高環境在原始碼儲存庫中使用不同分支。Gitflow 工作流程會被視為舊版，而以[幹線為基礎的工作流程](#)是現代、偏好的方法。

黃金影像

系統或軟體的快照，做為部署該系統或軟體新執行個體的範本。例如，在製造中，黃金映像可用於在多個裝置上佈建軟體，並有助於提高裝置製造操作的速度、可擴展性和生產力。

綠地策略

新環境中缺乏現有基礎設施。對系統架構採用綠地策略時，可以選擇所有新技術，而不會限制與現有基礎設施的相容性，也稱為[棕地](#)。如果正在擴展現有基礎設施，則可能會混合棕地和綠地策略。

防護機制

有助於跨組織單位 (OU) 來管控資源、政策和合規的高層級規則。預防性防護機制會強制執行政策，以確保符合合規標準。透過使用服務控制政策和 IAM 許可界限來將其實施。偵測性防護機制可偵測政策違規和合規問題，並產生提醒以便修正。它們是透過使用 AWS Config、AWS Security Hub、Amazon GuardDuty、Amazon Inspector、AWS Trusted Advisor 和自訂 AWS Lambda 檢查來實施。

H

HA

請參閱[高可用性](#)。

異質資料庫遷移

將來源資料庫遷移至使用不同資料庫引擎的目標資料庫 (例如，Oracle 至 Amazon Aurora)。異質遷移通常是重新架構工作的一部分，而轉換結構描述可能是一項複雜任務。[AWS 提供有助於結構描述轉換的 AWS SCT](#)。

高可用性 (HA)

在遇到挑戰或災難時，工作負載能夠在不介入的情況下持續運作。HA 系統的設計目的是自動容錯移轉、持續提供高品質的效能，以及處理不同的負載和故障，並將效能影響降至最低。

歷史現代化

一種方法，用於現代化和升級操作技術 (OT) 系統，以更好地滿足製造業的需求。歷史資料是一種資料庫，用於從工廠中的各種來源收集和存放資料。

保留資料

從用於訓練機器學習模型的資料集中保留的部分歷史標記資料。您可以使用保留資料，透過比較模型預測與保留資料來評估模型效能。

異質資料庫遷移

將您的來源資料庫遷移至共用相同資料庫引擎的目標資料庫 (例如，Microsoft SQL Server 至 Amazon RDS for SQL Server)。同質遷移通常是主機轉換或平台轉換工作的一部分。您可以使用原生資料庫公用程式來遷移結構描述。

熱資料

經常存取的資料，例如即時資料或最近的轉譯資料。此資料通常需要高效能儲存層或類別，才能提供快速的查詢回應。

修補程序

緊急修正生產環境中的關鍵問題。由於其緊迫性，通常會在典型 DevOps 發行工作流程之外執行修補程式。

超級護理期間

在切換後，遷移團隊在雲端管理和監控遷移的應用程式以解決任何問題的時段。通常，此期間的長度為 1-4 天。在超級護理期間結束時，遷移團隊通常會將應用程式的責任轉移給雲端營運團隊。

I

IaC

將[基礎設施視為程式碼](#)。

身分型政策

連接至一或多個 IAM 主體的政策，可定義其在 AWS 雲端環境中的許可。

閒置應用程式

90 天期間 CPU 和記憶體平均使用率在 5% 至 20% 之間的應用程式。在遷移專案中，通常會淘汰這些應用程式或將其保留在內部部署。

IloT

請參閱[工業物聯網](#)。

不可變的基礎設施

為生產工作負載部署新基礎設施的模型，而不是更新、修補或修改現有的基礎設施。不可變基礎設施本質上比[可變基礎設施](#)更一致、可靠且可預測。如需詳細資訊，請參閱 AWS Well-Architected Framework [中的使用不可變基礎設施的部署](#)最佳實務。

傳入 (輸入) VPC

在 AWS 多帳戶架構中，接受、檢查和路由來自應用程式外部之網路連線的 VPC。[AWS 安全參考架構](#)建議您使用傳入、傳出和檢查 VPC 來設定網路帳戶，以保護應用程式與更廣泛的網際網路之間的雙向介面。

增量遷移

一種切換策略，您可以在其中將應用程式分成小部分遷移，而不是執行單一、完整的切換。例如，您最初可能只將一些微服務或使用者移至新系統。確認所有項目都正常運作之後，您可以逐步移動其他微服務或使用者，直到可以解除委任舊式系統。此策略可降低與大型遷移關聯的風險。

工業 4.0

2016 年 [Klaus Schwab](#) 推出的術語，透過連線能力、即時資料、自動化、分析和 AI/ML 的進展，指製造程序的現代化。

基礎設施

應用程式環境中包含的所有資源和資產。

基礎設施即程式碼 (IaC)

透過一組組態檔案來佈建和管理應用程式基礎設施的程序。IaC 旨在協助您集中管理基礎設施，標準化資源並快速擴展，以便新環境可重複、可靠且一致。

工業物聯網 (IIoT)

在製造業、能源、汽車、醫療保健、生命科學和農業等產業領域使用網際網路連線的感測器和裝置。如需詳細資訊，請參閱[建立工業物聯網 \(IIoT\) 數位轉型策略](#)。

檢查 VPC

在 AWS 多帳戶架構中，集中式 VPC，可管理 VPCs 之間（在相同或不同的 AWS 區域）、網際網路和內部部署網路之間的網路流量檢查。[AWS 安全參考架構](#)建議您使用傳入、傳出和檢查 VPC 來設定網路帳戶，以保護應用程式與更廣泛的網際網路之間的雙向介面。

物聯網 (IoT)

具有內嵌式感測器或處理器的相連實體物體網路，其透過網際網路或本地通訊網路與其他裝置和系統進行通訊。如需詳細資訊，請參閱[什麼是 IoT？](#)

可解釋性

機器學習模型的一個特徵，描述了人類能夠理解模型的預測如何依賴於其輸入的程度。如需詳細資訊，請參閱[的機器學習模型可解釋性 AWS](#)。

IoT

請參閱[物聯網](#)。

IT 資訊庫 (ITIL)

一組用於交付 IT 服務並使這些服務與業務需求保持一致的最佳實務。ITIL 為 ITSM 提供了基礎。

IT 服務管理 (ITSM)

與組織的設計、實作、管理和支援 IT 服務關聯的活動。如需有關將雲端操作與 ITSM 工具整合的資訊，請參閱[操作整合指南](#)。

ITIL

請參閱[IT 資訊庫](#)。

ITSM

請參閱[IT 服務管理](#)。

L

標籤型存取控制 (LBAC)

強制存取控制 (MAC) 的實作，其中使用者和資料本身都會獲得明確指派的安全標籤值。使用者安全標籤和資料安全標籤之間的交集會決定使用者可以看到哪些資料列和資料欄。

登陸區域

登陸區域是架構良好的多帳戶 AWS 環境，可擴展且安全。這是一個起點，您的組織可以從此起點快速啟動和部署工作負載與應用程式，並對其安全和基礎設施環境充滿信心。如需有關登陸區域的詳細資訊，請參閱[設定安全且可擴展的多帳戶 AWS 環境](#)。

大型語言模型 (LLM)

預先訓練大量資料的深度學習 [AI](#) 模型。LLM 可以執行多個任務，例如回答問題、摘要文件、將文字翻譯成其他語言，以及完成句子。如需詳細資訊，請參閱[什麼是 LLMs](#)。

大型遷移

遷移 300 部或更多伺服器。

LBAC

請參閱[標籤型存取控制](#)。

最低權限

授予執行任務所需之最低許可的安全最佳實務。如需詳細資訊，請參閱 IAM 文件中的[套用最低權限許可](#)。

隨即轉移

請參閱 [7 個 R](#)。

小端序系統

首先儲存最低有效位元組的系統。另請參閱 [Endianness](#)。

LLM

請參閱[大型語言模型](#)。

較低的環境

請參閱 [環境](#)。

M

機器學習 (ML)

一種使用演算法和技術進行模式識別和學習的人工智慧。機器學習會進行分析並從記錄的資料 (例如物聯網 (IoT) 資料) 中學習，以根據模式產生統計模型。如需詳細資訊，請參閱[機器學習](#)。

主要分支

請參閱[分支](#)。

惡意軟體

旨在危及電腦安全或隱私權的軟體。惡意軟體可能會中斷電腦系統、洩露敏感資訊，或取得未經授權的存取。惡意軟體的範例包括病毒、蠕蟲、勒索軟體、特洛伊木馬程式、間諜軟體和鍵盤記錄器。

受管服務

AWS 服務會 AWS 操作基礎設施層、作業系統和平台，而您會存取端點來存放和擷取資料。Amazon Simple Storage Service (Amazon S3) 和 Amazon DynamoDB 是受管服務的範例。這些也稱為抽象服務。

製造執行系統 (MES)

一種軟體系統，用於追蹤、監控、記錄和控制生產程序，將原物料轉換為現場成品。

MAP

請參閱[遷移加速計劃](#)。

機制

建立工具、推動工具採用，然後檢查結果以進行調整的完整程序。機制是在操作時強化和改善自身的循環。如需詳細資訊，請參閱 AWS Well-Architected Framework 中的[建置機制](#)。

成員帳戶

除了屬於組織一部分的管理帳戶 AWS 帳戶 之外的所有 AWS Organizations。一個帳戶一次只能是一個組織的成員。

製造執行系統

請參閱[製造執行系統](#)。

訊息佇列遙測傳輸 (MQTT)

根據[發佈/訂閱](#)模式的輕量型machine-to-machine(M2M) 通訊協定，適用於資源受限的 [IoT](#) 裝置。

微服務

一種小型的獨立服務，它可透過定義明確的 API 進行通訊，通常由小型獨立團隊擁有。例如，保險系統可能包含對應至業務能力 (例如銷售或行銷) 或子領域 (例如購買、索賠或分析) 的微服務。微服務的優點包括靈活性、彈性擴展、輕鬆部署、可重複使用的程式碼和適應力。如需詳細資訊，請參閱[使用無 AWS 伺服器服務整合微服務](#)。

微服務架構

一種使用獨立元件來建置應用程式的方法，這些元件會以微服務形式執行每個應用程式程序。這些微服務會使用輕量型 API，透過明確定義的介面進行通訊。此架構中的每個微服務都可以進行更新、部署和擴展，以滿足應用程式特定功能的需求。如需詳細資訊，請參閱[在上實作微服務 AWS](#)。

Migration Acceleration Program (MAP)

此 AWS 計畫提供諮詢支援、訓練和服務，以協助組織建立強大的營運基礎，以移至雲端，並協助抵銷遷移的初始成本。MAP 包括用於有條不紊地執行舊式遷移的遷移方法以及一組用於自動化和加速常見遷移案例的工具。

大規模遷移

將大部分應用程式組合依波次移至雲端的程序，在每個波次中，都會以更快的速度移動更多應用程式。此階段使用從早期階段學到的最佳實務和經驗教訓來實作團隊、工具和流程的遷移工廠，以透過自動化和敏捷交付簡化工作負載的遷移。這是[AWS 遷移策略](#)的第三階段。

遷移工廠

可透過自動化、敏捷的方法簡化工作負載遷移的跨職能團隊。遷移工廠團隊通常包括營運、業務分析師和擁有者、遷移工程師、開發人員以及從事 Sprint 工作的 DevOps 專業人員。20% 至 50% 之間的企業應用程式組合包含可透過工廠方法優化的重複模式。如需詳細資訊，請參閱此內容集中的[遷移工廠的討論](#)和[雲端遷移工廠指南](#)。

遷移中繼資料

有關完成遷移所需的應用程式和伺服器的資訊。每種遷移模式都需要一組不同的遷移中繼資料。遷移中繼資料的範例包括目標子網路、安全群組和 AWS 帳戶。

遷移模式

可重複的遷移任務，詳細描述遷移策略、遷移目的地以及所使用的遷移應用程式或服務。範例：使用 AWS Application Migration Service 重新託管遷移至 Amazon EC2。

遷移組合評定 (MPA)

線上工具，提供驗證商業案例以遷移至的資訊 AWS 雲端。MPA 提供詳細的組合評定 (伺服器適當規模、定價、總體擁有成本比較、遷移成本分析) 以及遷移規劃 (應用程式資料分析和資料收集、應用程式分組、遷移優先順序，以及波次規劃)。[MPA 工具](#) (需要登入) 可供所有 AWS 顧問和 APN 合作夥伴顧問免費使用。

遷移準備程度評定 (MRA)

使用 AWS CAF 取得組織雲端整備狀態的洞見、識別優缺點，以及建立行動計劃以消除已識別差距的程序。如需詳細資訊，請參閱[遷移準備程度指南](#)。MRA 是 [AWS 遷移策略](#) 的第一階段。

遷移策略

用來將工作負載遷移至的方法 AWS 雲端。如需詳細資訊，請參閱本詞彙表中的 [7 個 Rs](#) 項目，並請參閱[動員您的組織以加速大規模遷移](#)。

機器學習 (ML)

請參閱[機器學習](#)。

現代化

將過時的 (舊版或單一) 應用程式及其基礎架構轉換為雲端中靈活、富有彈性且高度可用的系統，以降低成本、提高效率並充分利用創新。如需詳細資訊，請參閱 [《》中的現代化應用程式的策略 AWS 雲端](#)。

現代化準備程度評定

這項評估可協助判斷組織應用程式的現代化準備程度；識別優點、風險和相依性；並確定組織能夠在多大程度上支援這些應用程式的未來狀態。評定的結果就是目標架構的藍圖、詳細說明現代化程序的開發階段和里程碑的路線圖、以及解決已發現的差距之行動計畫。如需詳細資訊，請參閱 [《》中的評估應用程式的現代化準備 AWS 雲端](#) 程度。

單一應用程式 (單一)

透過緊密結合的程序作為單一服務執行的應用程式。單一應用程式有幾個缺點。如果一個應用程式功能遇到需求激增，則必須擴展整個架構。當程式碼庫增長時，新增或改進單一應用程式的功能也會變得更加複雜。若要解決這些問題，可以使用微服務架構。如需詳細資訊，請參閱[將單一體系分解為微服務](#)。

MPA

請參閱[遷移產品組合評估](#)。

MQTT

請參閱[訊息佇列遙測傳輸](#)。

多類別分類

一個有助於產生多類別預測的過程 (預測兩個以上的結果之一)。例如，機器學習模型可能會詢問「此產品是書籍、汽車還是電話？」或者「這個客戶對哪種產品類別最感興趣？」

可變基礎設施

更新和修改生產工作負載現有基礎設施的模型。為了提高一致性、可靠性和可預測性，AWS Well-Architected Framework 建議使用[不可變的基礎設施](#)作為最佳實務。

O

OAC

請參閱[原始存取控制](#)。

OAI

請參閱[原始存取身分](#)。

OCM

請參閱[組織變更管理](#)。

離線遷移

一種遷移方法，可在遷移過程中刪除來源工作負載。此方法涉及延長停機時間，通常用於小型非關鍵工作負載。

OI

請參閱[操作整合](#)。

OLA

請參閱[操作層級協議](#)。

線上遷移

一種遷移方法，無需離線即可將來源工作負載複製到目標系統。連接至工作負載的應用程式可在遷移期間繼續運作。此方法涉及零至最短停機時間，通常用於關鍵的生產工作負載。

OPC-UA

請參閱[開啟程序通訊 - 統一架構](#)。

開放程序通訊 - 統一架構 (OPC-UA)

用於工業自動化的machine-to-machine(M2M) 通訊協定。OPC-UA 提供資料加密、身分驗證和授權機制的互通性標準。

操作水準協議 (OLA)

一份協議，闡明 IT 職能群組承諾向彼此提供的內容，以支援服務水準協議 (SLA)。

操作整備審查 (ORR)

問題和相關最佳實務的檢查清單，可協助您了解、評估、預防或減少事件和可能失敗的範圍。如需詳細資訊，請參閱 AWS Well-Architected Framework 中的 [操作準備審查 \(ORR\)](#)。

操作技術 (OT)

使用實體環境控制工業操作、設備和基礎設施的硬體和軟體系統。在製造業中，整合 OT 和資訊技術 (IT) 系統是 [工業 4.0](#) 轉型的關鍵重點。

操作整合 (OI)

在雲端中將操作現代化的程序，其中包括準備程度規劃、自動化和整合。如需詳細資訊，請參閱 [操作整合指南](#)。

組織追蹤

建立的線索 AWS CloudTrail 會記錄 AWS 帳戶 組織中所有 的所有事件 AWS Organizations。在屬於組織的每個 AWS 帳戶 中建立此追蹤，它會跟蹤每個帳戶中的活動。如需詳細資訊，請參閱 CloudTrail 文件中的 [建立組織追蹤](#)。

組織變更管理 (OCM)

用於從人員、文化和領導力層面管理重大、顛覆性業務轉型的架構。OCM 透過加速變更採用、解決過渡問題，以及推動文化和組織變更，協助組織為新系統和策略做好準備，並轉移至新系統和策略。在 AWS 遷移策略中，此架構稱為人員加速，因為雲端採用專案所需的變更速度。如需詳細資訊，請參閱 [OCM 指南](#)。

原始存取控制 (OAC)

CloudFront 中的增強型選項，用於限制存取以保護 Amazon Simple Storage Service (Amazon S3) 內容。OAC 支援所有 S3 儲存貯體中的所有伺服器端加密 AWS KMS (SSE-KMS) AWS 區域，以及對 S3 儲存貯體的動態PUT和DELETE請求。

原始存取身分 (OAI)

CloudFront 中的一個選項，用於限制存取以保護 Amazon S3 內容。當您使用 OAI 時，CloudFront 會建立一個可供 Amazon S3 進行驗證的主體。經驗證的主體只能透過特定 CloudFront 分發來存取 S3 儲存貯體中的內容。另請參閱 [OAC](#)，它可提供更精細且增強的存取控制。

ORR

請參閱 [操作整備審核](#)。

OT

請參閱[操作技術](#)。

傳出 (輸出) VPC

在 AWS 多帳戶架構中，處理從應用程式內啟動之網路連線的 VPC。[AWS 安全參考架構](#)建議您使用傳入、傳出和檢查 VPC 來設定網路帳戶，以保護應用程式與更廣泛的網際網路之間的雙向介面。

P

許可界限

附接至 IAM 主體的 IAM 管理政策，可設定使用者或角色擁有的最大許可。如需詳細資訊，請參閱 IAM 文件中的[許可界限](#)。

個人身分識別資訊 (PII)

當直接檢視或與其他相關資料配對時，可用來合理推斷個人身分的資訊。PII 的範例包括名稱、地址和聯絡資訊。

PII

請參閱[個人身分識別資訊](#)。

手冊

一組預先定義的步驟，可擷取與遷移關聯的工作，例如在雲端中提供核心操作功能。手冊可以採用指令碼、自動化執行手冊或操作現代化環境所需的程序或步驟摘要的形式。

PLC

請參閱[可程式設計邏輯控制器](#)。

PLM

請參閱[產品生命週期管理](#)。

政策

可定義許可的物件（請參閱[身分型政策](#)）、指定存取條件（請參閱[資源型政策](#)），或定義組織中所有帳戶的最大許可 AWS Organizations（請參閱[服務控制政策](#)）。

混合持久性

根據資料存取模式和其他需求，獨立選擇微服務的資料儲存技術。如果您的微服務具有相同的資料儲存技術，則其可能會遇到實作挑戰或效能不佳。如果微服務使用最適合其需求的資料儲存，則可以更輕鬆地實作並達到更好的效能和可擴展性。如需詳細資訊，請參閱[在微服務中啟用資料持久性](#)。

組合評定

探索、分析應用程式組合並排定其優先順序以規劃遷移的程序。如需詳細資訊，請參閱[評估遷移準備程度](#)。

述詞

傳回 true 或的查詢條件 false，通常位於 WHERE 子句中。

述詞下推

一種資料庫查詢最佳化技術，可在傳輸前篩選查詢中的資料。這可減少必須從關聯式資料庫擷取和處理的資料量，並改善查詢效能。

預防性控制

旨在防止事件發生的安全控制。這些控制是第一道防線，可協助防止對網路的未經授權存取或不必要變更。如需詳細資訊，請參閱在 AWS 上實作安全控制中的[預防性控制](#)。

委託人

中可執行動作和存取資源 AWS 的實體。此實體通常是 AWS 帳戶、IAM 角色或使用者的根使用者。如需詳細資訊，請參閱 IAM 文件中[角色術語和概念](#)中的主體。

設計隱私權

透過整個開發程序將隱私權納入考量的系統工程方法。

私有託管區域

一種容器，它包含有關您希望 Amazon Route 53 如何回應一個或多個 VPC 內的域及其子域之 DNS 查詢的資訊。如需詳細資訊，請參閱 Route 53 文件中的[使用私有託管區域](#)。

主動控制

旨在防止部署不合規資源的[安全控制](#)。這些控制項會在佈建資源之前對其進行掃描。如果資源不符合控制項，則不會佈建。如需詳細資訊，請參閱 AWS Control Tower 文件中的[控制項參考指南](#)，並參閱實作安全[控制項中的主動](#)控制項。 AWS

產品生命週期管理 (PLM)

管理產品整個生命週期的資料和程序，從設計、開發和啟動，到成長和成熟，再到拒絕和移除。

生產環境

請參閱 [環境](#)。

可程式設計邏輯控制器 (PLC)

在製造中，高度可靠、可調整的電腦，可監控機器並自動化製造程序。

提示鏈結

使用一個 [LLM](#) 提示的輸出做為下一個提示的輸入，以產生更好的回應。此技術用於將複雜任務分解為子任務，或反覆精簡或展開初步回應。它有助於提高模型回應的準確性和相關性，並允許更精細、個人化的結果。

擬匿名化

將資料集中的個人識別符取代為預留位置值的程序。假名化有助於保護個人隱私權。假名化資料仍被視為個人資料。

發佈/訂閱 (pub/sub)

一種模式，可啟用微服務之間的非同步通訊，以提高可擴展性和回應能力。例如，在微服務型 [MES](#) 中，微服務可以將事件訊息發佈到其他微服務可訂閱的頻道。系統可以新增新的微服務，而無需變更發佈服務。

Q

查詢計劃

一系列步驟，如指示，用於存取 SQL 關聯式資料庫系統中的資料。

查詢計劃迴歸

在資料庫服務優化工具選擇的計畫比對資料庫環境進行指定的變更之前的計畫不太理想時。這可能因為對統計資料、限制條件、環境設定、查詢參數繫結的變更以及資料庫引擎的更新所導致。

R

RACI 矩陣

請參閱[負責、負責、諮詢、告知 \(RACI\)](#)。

RAG

請參閱[擷取增強產生](#)。

勒索軟體

一種惡意軟體，旨在阻止對計算機系統或資料的存取，直到付款為止。

RASCI 矩陣

請參閱[負責、負責、諮詢、告知 \(RACI\)](#)。

RCAC

請參閱[資料列和資料欄存取控制](#)。

僅供讀取複本

用於唯讀用途的資料庫複本。您可以將查詢路由至僅供讀取複本以減少主資料庫的負載。

重新架構師

請參閱[7 個 R](#)。

復原點目標 (RPO)

自上次資料復原點以來可接受的時間上限。這會決定最後一個復原點與服務中斷之間可接受的資料遺失。

復原時間目標 (RTO)

服務中斷與服務還原之間的可接受延遲上限。

重構

請參閱[7 個 R](#)。

區域

地理區域中的 AWS 資源集合。每個 AWS 區域 都獨立於其他 ，以提供容錯能力、穩定性和彈性。如需詳細資訊，請參閱[指定 AWS 區域 您的帳戶可以使用哪些](#)。

迴歸

預測數值的 ML 技術。例如，為了解決「這房子會賣什麼價格？」的問題 ML 模型可以使用線性迴歸模型，根據已知的房屋事實 (例如，平方英尺) 來預測房屋的銷售價格。

重新託管

請參閱 [7 個 R](#)。

版本

在部署程序中，它是將變更提升至生產環境的動作。

重新放置

請參閱 [7 個 R](#)。

Replatform

請參閱 [7 個 R](#)。

回購

請參閱 [7 個 R](#)。

彈性

應用程式抵禦中斷或從中斷中復原的能力。[在中規劃彈性時，高可用性和災難復原](#)是常見的考量 AWS 雲端。如需詳細資訊，請參閱[AWS 雲端 彈性](#)。

資源型政策

附接至資源的政策，例如 Amazon S3 儲存貯體、端點或加密金鑰。這種類型的政策會指定允許存取哪些主體、支援的動作以及必須滿足的任何其他條件。

負責者、當責者、事先諮詢者和事後告知者 (RACI) 矩陣

定義所有涉及遷移活動和雲端操作之各方的角色和責任的矩陣。矩陣名稱衍生自矩陣中定義的責任類型：負責人 (R)、責任 (A)、已諮詢 (C) 和知情 (I)。支援 (S) 類型為選用。如果您包含支援，則矩陣稱為 RASCI 矩陣，如果您排除它，則稱為 RACI 矩陣。

回應性控制

一種安全控制，旨在驅動不良事件或偏離安全基準的補救措施。如需詳細資訊，請參閱在 AWS 上實作安全控制中的[回應性控制](#)。

保留

請參閱 [7 個 R](#)。

淘汰

請參閱 [7 Rs](#)。

檢索增強生成 (RAG)

[一種生成式 AI](#) 技術，其中 [LLM](#) 會在產生回應之前參考訓練資料來源以外的授權資料來源。例如，RAG 模型可能會對組織的知識庫或自訂資料執行語意搜尋。如需詳細資訊，請參閱 [什麼是 RAG](#)。

輪換

定期更新 [秘密](#) 的程序，讓攻擊者更難存取登入資料。

資料列和資料欄存取控制 (RCAC)

使用已定義存取規則的基本、彈性 SQL 表達式。RCAC 包含資料列許可和資料欄遮罩。

RPO

請參閱 [復原點目標](#)。

RTO

請參閱 [復原時間目標](#)。

執行手冊

執行特定任務所需的一組手動或自動程序。這些通常是為了簡化重複性操作或錯誤率較高的程序而建置。

S

SAML 2.0

許多身分提供者 (IdP) 使用的開放標準。此功能可啟用聯合單一登入 (SSO)，讓使用者可以登入 AWS Management Console 或呼叫 AWS API 操作，而無需為您組織中的每個人在 IAM 中建立使用者。如需有關以 SAML 2.0 為基礎的聯合詳細資訊，請參閱 IAM 文件中的 [關於以 SAML 2.0 為基礎的聯合](#)。

SCADA

請參閱 [監督控制和資料擷取](#)。

SCP

請參閱 [服務控制政策](#)。

秘密

您以加密形式存放的 AWS Secrets Manager 機密或限制資訊，例如密碼或使用者的登入資料。它由秘密值及其中繼資料組成。秘密值可以是二進位、單一字串或多個字串。如需詳細資訊，請參閱 [Secrets Manager 文件中的 Secrets Manager 秘密中的什麼內容？](#)。

依設計的安全性

透過整個開發程序將安全性納入考量的系統工程方法。

安全控制

一種技術或管理防護機制，它可預防、偵測或降低威脅行為者利用安全漏洞的能力。安全控制有四種主要類型：[預防性](#)、[偵測性](#)、[回應性](#)和[主動性](#)。

安全強化

減少受攻擊面以使其更能抵抗攻擊的過程。這可能包括一些動作，例如移除不再需要的資源、實作授予最低權限的安全最佳實務、或停用組態檔案中不必要的功能。

安全資訊與事件管理 (SIEM) 系統

結合安全資訊管理 (SIM) 和安全事件管理 (SEM) 系統的工具與服務。SIEM 系統會收集、監控和分析來自伺服器、網路、裝置和其他來源的資料，以偵測威脅和安全漏洞，並產生提醒。

安全回應自動化

預先定義和程式設計的動作，旨在自動回應或修復安全事件。這些自動化可做為[偵測或回應式](#)安全控制，協助您實作 AWS 安全最佳實務。自動化回應動作的範例包括修改 VPC 安全群組、修補 Amazon EC2 執行個體或輪換登入資料。

伺服器端加密

由 AWS 服務接收資料的在其目的地加密資料。

服務控制政策 (SCP)

為 AWS Organizations 中的組織的所有帳戶提供集中控制許可的政策。SCP 會定義防護機制或設定管理員可委派給使用者或角色的動作限制。您可以使用 SCP 作為允許清單或拒絕清單，以指定允許或禁止哪些服務或動作。如需詳細資訊，請參閱 AWS Organizations 文件中的[服務控制政策](#)。

服務端點

的進入點 URL AWS 服務。您可以使用端點，透過程式設計方式連接至目標服務。如需詳細資訊，請參閱 AWS 一般參考中的 [AWS 服務端點](#)。

服務水準協議 (SLA)

一份協議，闡明 IT 團隊承諾向客戶提供的服務，例如服務正常執行時間和效能。

服務層級指標 (SLI)

服務效能方面的測量，例如其錯誤率、可用性或輸送量。

服務層級目標 (SLO)

代表服務運作狀態的目標指標，由[服務層級指標](#)測量。

共同責任模式

描述您與共同 AWS 承擔雲端安全與合規責任的模型。AWS 負責雲端的安全，而負責雲端的安全。如需詳細資訊，請參閱[共同責任模式](#)。

SIEM

請參閱[安全資訊和事件管理系統](#)。

單一故障點 (SPOF)

應用程式的單一關鍵元件故障，可能會中斷系統。

SLA

請參閱[服務層級協議](#)。

SLI

請參閱[服務層級指標](#)。

SLO

請參閱[服務層級目標](#)。

先拆分後播種模型

擴展和加速現代化專案的模式。定義新功能和產品版本時，核心團隊會進行拆分以建立新的產品團隊。這有助於擴展組織的能力和服務，提高開發人員生產力，並支援快速創新。如需詳細資訊，請參閱[中的階段式應用程式現代化方法 AWS 雲端](#)。

SPOF

請參閱[單一故障點](#)。

星狀結構描述

使用一個大型事實資料表來存放交易或測量資料的資料庫組織結構，並使用一或多個較小的維度資料表來存放資料屬性。此結構旨在用於[資料倉儲](#)或商業智慧用途。

Strangler Fig 模式

一種現代化單一系統的方法，它會逐步重寫和取代系統功能，直到舊式系統停止使用為止。此模式源自無花果藤，它長成一棵馴化樹並最終戰勝且取代了其宿主。該模式由 [Martin Fowler 引入](#)，作為重寫單一系統時管理風險的方式。如需有關如何套用此模式的範例，請參閱 [使用容器和 Amazon API Gateway 逐步現代化舊版 Microsoft ASP.NET \(ASMX\) Web 服務](#)。

子網

您 VPC 中的 IP 地址範圍。子網必須位於單一可用區域。

監控控制和資料擷取 (SCADA)

在製造中，使用硬體和軟體來監控實體資產和生產操作的系統。

對稱加密

使用相同金鑰來加密及解密資料的加密演算法。

合成測試

以模擬使用者互動的方式測試系統，以偵測潛在問題或監控效能。您可以使用 [Amazon CloudWatch Synthetics](#) 來建立這些測試。

系統提示

一種向 [LLM](#) 提供內容、指示或指導方針以指示其行為的技術。系統提示有助於設定內容，並建立與使用者互動的規則。

T

標籤

做為中繼資料以組織 AWS 資源的鍵值對。標籤可協助您管理、識別、組織、搜尋及篩選資源。如需詳細資訊，請參閱 [標記您的 AWS 資源](#)。

目標變數

您嘗試在受監督的 ML 中預測的值。這也被稱為結果變數。例如，在製造設定中，目標變數可能是產品瑕疵。

任務清單

用於透過執行手冊追蹤進度的工具。任務清單包含執行手冊的概觀以及要完成的一般任務清單。對於每個一般任務，它包括所需的預估時間量、擁有者和進度。

測試環境

請參閱 [環境](#)。

訓練

為 ML 模型提供資料以供學習。訓練資料必須包含正確答案。學習演算法會在訓練資料中尋找將輸入資料屬性映射至目標的模式 (您想要預測的答案)。它會輸出擷取這些模式的 ML 模型。可以使用 ML 模型，來預測您不知道的目標新資料。

傳輸閘道

可以用於互連 VPC 和內部部署網路的網路傳輸中樞。如需詳細資訊，請參閱 AWS Transit Gateway 文件中的 [什麼是傳輸閘道](#)。

主幹型工作流程

這是一種方法，開發人員可在功能分支中本地建置和測試功能，然後將這些變更合併到主要分支中。然後，主要分支會依序建置到開發環境、生產前環境和生產環境中。

受信任的存取權

將許可授予您指定的服務，以代表您在組織中 AWS Organizations 及其帳戶中執行任務。受信任的服務會在需要該角色時，在每個帳戶中建立服務連結角色，以便為您執行管理工作。如需詳細資訊，請參閱文件中的 AWS Organizations [搭配使用 AWS Organizations 與其他 AWS 服務](#)。

調校

變更訓練程序的各個層面，以提高 ML 模型的準確性。例如，可以透過產生標籤集、新增標籤、然後在不同的設定下多次重複這些步驟來訓練 ML 模型，以優化模型。

雙比薩團隊

兩個比薩就能吃飽的小型 DevOps 團隊。雙披薩團隊規模可確保軟體開發中的最佳協作。

U

不確定性

這是一個概念，指的是不精確、不完整或未知的資訊，其可能會破壞預測性 ML 模型的可靠性。有兩種類型的不確定性：認知不確定性是由有限的、不完整的資料引起的，而隨機不確定性是由資料中固有的噪聲和隨機性引起的。如需詳細資訊，請參閱 [量化深度學習系統的不確定性](#) 指南。

未區分的任務

也稱為繁重工作，這是建立和操作應用程式的必要工作，但不為最終使用者提供直接價值或提供競爭優勢。未區分任務的範例包括採購、維護和容量規劃。

較高的環境

請參閱 [環境](#)。

V

清空

一種資料庫維護操作，涉及增量更新後的清理工作，以回收儲存並提升效能。

版本控制

追蹤變更的程序和工具，例如儲存庫中原始程式碼的變更。

VPC 對等互連

兩個 VPC 之間的連線，可讓您使用私有 IP 地址路由流量。如需詳細資訊，請參閱 Amazon VPC 文件中的 [什麼是 VPC 對等互連](#)。

漏洞

危害系統安全性的軟體或硬體瑕疵。

W

暖快取

包含經常存取的目前相關資料的緩衝快取。資料庫執行個體可以從緩衝快取讀取，這比從主記憶體或磁碟讀取更快。

暖資料

不常存取的資料。查詢這類資料時，通常可接受中等緩慢的查詢。

視窗函數

SQL 函數，對與目前記錄在某種程度上相關的資料列群組執行計算。視窗函數適用於處理任務，例如根據目前資料列的相對位置計算移動平均值或存取資料列的值。

工作負載

提供商業價值的資源和程式碼集合，例如面向客戶的應用程式或後端流程。

工作串流

遷移專案中負責一組特定任務的功能群組。每個工作串流都是獨立的，但支援專案中的其他工作串流。例如，組合工作串流負責排定應用程式、波次規劃和收集遷移中繼資料的優先順序。組合工作串流將這些資產交付至遷移工作串流，然後再遷移伺服器 and 應用程式。

WORM

請參閱[寫入一次，多次讀取](#)。

WQF

請參閱[AWS 工作負載資格架構](#)。

寫入一次，讀取許多 (WORM)

儲存模型，可一次性寫入資料，並防止刪除或修改資料。授權使用者可以視需要多次讀取資料，但無法變更資料。此資料儲存基礎設施被視為[不可變](#)。

Z

零時差入侵

利用[零時差漏洞](#)的攻擊，通常是惡意軟體。

零時差漏洞

生產系統中未緩解的瑕疵或漏洞。威脅行為者可以使用這種類型的漏洞來攻擊系統。開發人員經常因為攻擊而意識到漏洞。

零鏡頭提示

提供 [LLM](#) 執行任務的指示，但沒有可協助引導任務的範例 (快照)。LLM 必須使用其預先訓練的知識來處理任務。零鏡頭提示的有效性取決於任務的複雜性和提示的品質。另請參閱[少量擷取提示](#)。

殭屍應用程式

CPU 和記憶體平均使用率低於 5% 的應用程式。在遷移專案中，通常會淘汰這些應用程式。

本文為英文版的機器翻譯版本，如內容有任何歧義或不一致之處，概以英文版為準。