

AWS 決策指南

選擇 AWS 分析服務



選擇 AWS 分析服務: AWS 決策指南

Copyright © 2026 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

決策指南	1
簡介	1
了解	2
考慮	5
選擇	11
使用	14
探索	22
文件歷史紀錄	24
.....	XXV

選擇 AWS 分析服務

採取第一步

用途	協助判斷哪些 AWS 分析服務最適合您的組織。
上次更新	2025 年 9 月 24 日
涵蓋的服務	<ul style="list-style-type: none">• Amazon Athena• AWS Clean Rooms• Amazon Data Firehose• Amazon DataZone• Amazon EMR• AWS Glue• Amazon Kinesis Data Streams• Amazon Managed Service for Apache Flink• Amazon Managed Streaming for Apache Kafka• Amazon Managed Workflows for Apache Airflow• Amazon OpenSearch Service• 快速• Amazon Redshift• Amazon S3• Amazon SageMaker

簡介

資料是現代業務的基礎。人員和應用程式需要安全地存取和分析來自全新且多樣化來源的資料。資料量也會不斷增加，這可能會導致組織難以擷取、儲存和分析所有必要的資料。

滿足這些挑戰意味著建立一個現代化的資料架構，以分解所有用於分析和洞察的資料孤島，包括第三方資料，並使組織中的每個人在一個地方能夠透過end-to-end管控來存取它。連接您的分析和機器學習 (ML) 系統以啟用預測分析也越來越重要。

此決策指南可協助您提出正確的問題，以在 AWS 服務上建置現代化資料架構。它說明如何細分資料孤島（透過連接您的資料湖和資料倉儲）、系統孤島（透過連接 ML 和分析）和人員孤島（透過將資料放在組織中每個人的手中）。

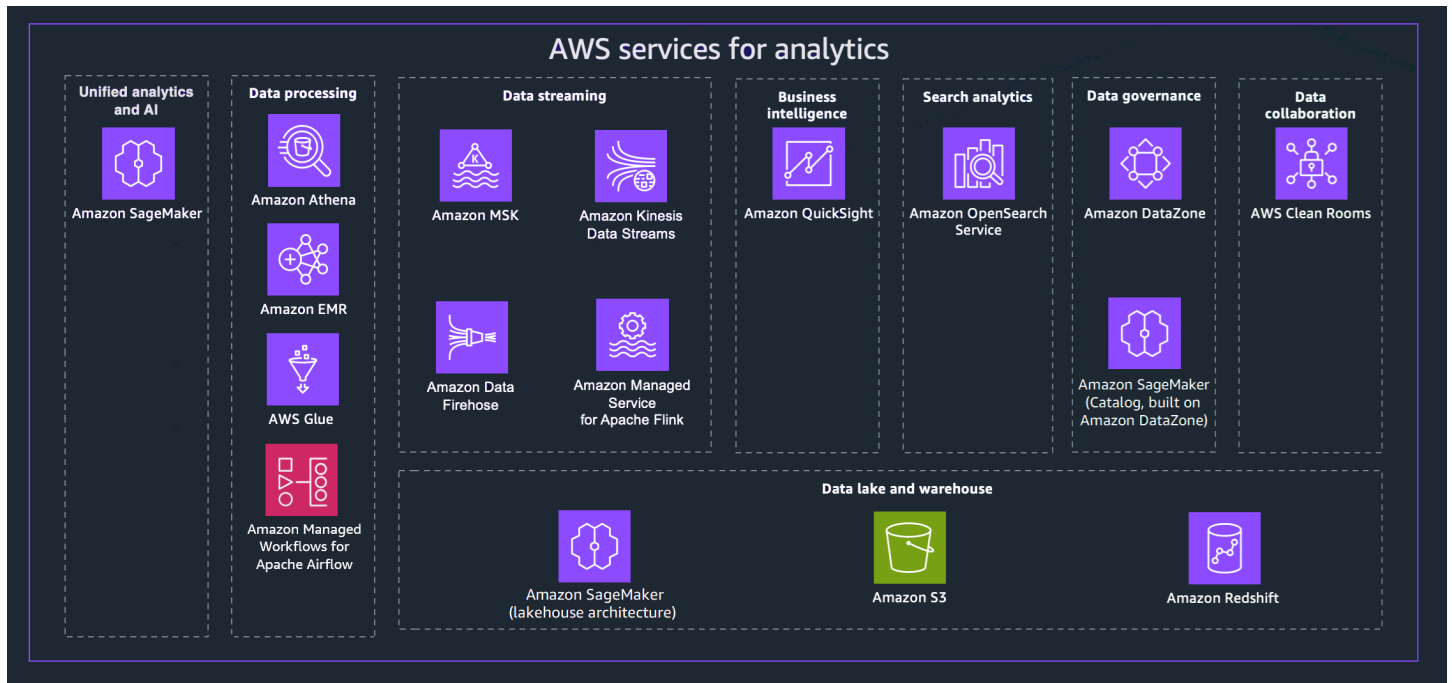
[這個八分鐘的練習來自 Sirish Chandrasekaran 和 Rick Sears 在 re : Invent 2024 的一小時簡報。它提供虛構公司 Maxdome 如何使用 SageMaker Unified Studio AI 和分析的概觀，這是新一代 Amazon SageMaker 的一部分，以解鎖資料洞見。](#)

了解 AWS 分析服務

現代資料策略是以一組技術建置區塊建置而成，可協助您管理、存取、分析資料並採取行動。它還為您提供多個連接到資料來源的選項。現代資料策略應該讓您的團隊能夠：

- 使用您偏好的工具或技術
- 使用人工智慧 (AI) 協助尋找資料特定問題的答案
- 使用適當的安全與資料控管控制，管理可存取資料的人員
- 細分資料孤島，讓您同時獲得最佳的資料湖和專門建置的資料存放區
- 以低成本並以開放的標準型資料格式存放任意數量的資料
- 將您的資料湖、資料倉儲、操作資料庫、應用程式和聯合資料來源連接到一致的整體

AWS 提供各種服務，協助您實現現代資料策略。下圖說明本指南涵蓋的分析 AWS 服務。以下索引標籤提供其他詳細資訊。



Unified analytics and AI

新一代 [Amazon SageMaker](#) 結合了廣泛採用的 AWS 機器學習 (ML) 和分析功能，為分析和 AI 提供整合式體驗，讓您統一存取所有資料。使用 [Amazon SageMaker Unified Studio](#)，您可以與熟悉的模型開發工具、生成式 AI 應用程式開發、資料處理和 SQL 分析協作並更快速地建置，這些 AWS 工具全都由我們生成式 AI 軟體開發助理 Amazon Q Developer 加速。使用內建的控管功能，從資料湖、資料倉儲或第三方和聯合來源存取您的資料，以滿足企業安全需求。

Data processing

- [Amazon Athena](#) 可協助您分析存放在 Amazon S3 中的非結構化、半結構化和結構化資料。範例包括 CSV、JSON 或單欄資料格式，例如 Apache Parquet 和 Apache ORC。您可以使用 Athena 以透過 ANSI SQL 執行隨機操作查詢，而不需要將資料彙總或載入到 Athena。Athena 與 Quick AWS Glue Data Catalog、[和其他 AWS 服務](#) 整合。您也可以使用 [Trino](#) 大規模分析資料，而無需管理基礎設施，並使用 Apache Flink 和 Apache Spark 建置即時分析。
- [Amazon EMR](#) 是受管叢集平台，可簡化 Apache Hadoop 和 Apache Spark 等大數據架構的執行，AWS 以處理和分析大量資料。使用這些架構和相關的開放原始碼專案，可以處理用於分析用途和商業智慧工作負載的資料。Amazon EMR 也可讓您轉換大量資料，並將其移入和移出其他 AWS 資料存放區和資料庫，例如 Amazon S3。
- 使用 [AWS Glue](#)，您可以探索並連線到超過 100 個不同的資料來源，並在集中式資料目錄中管理您的資料。您可以視覺化地建立、執行和監控 ETL 管道，將資料載入資料湖。此外，您可以立即使用 Athena、Amazon EMR 和 Amazon Redshift Spectrum 搜尋和查詢編目的資料。

- [Amazon Managed Workflows for Apache Airflow \(MWAA\)](#) 是 Apache Airflow 的全受管實作，可讓您更輕鬆地在雲端中建立、排程和監控資料工作流程。MWAA 會自動擴展工作流程容量以符合您的需求，AWS 並與安全服務整合。您可以使用 MWAA 跨分析服務協調工作流程，包括資料處理、ETL 任務和機器學習管道。

Data streaming

- 透過 [Amazon Managed Streaming for Apache Kafka \(Amazon MSK\)](#)，您可以建置和執行使用 Apache Kafka 處理串流資料的應用程式。Amazon MSK 提供控制平面操作，例如用於建立、更新和刪除叢集的操作。它可以讓你使用 Apache Kafka 資料平面操作，如那些用於生產和使用數據。
- 使用 [Amazon Kinesis Data Streams](#)，您可以即時收集和處理大型資料記錄串流。使用的資料類型可包括 IT 基礎架構日誌資料、應用程式日誌、社交媒體、市場資料摘要和 web 點擊流資料。
- [Amazon Data Firehose](#) 是一項全受管服務，可將即時串流資料交付至目的地，例如 Amazon S3、Amazon Redshift、Amazon OpenSearch Service、Splunk 和 Apache Iceberg Tables。您也可以將資料傳送至支援的第三方服務提供者擁有的任何自訂 HTTP 端點或 HTTP 端點，包括 Datadog、Dynatrace、LogicMonitor、MongoDB、New Relic、Coralogix 和 Elastic。
- 透過 [Amazon Managed Service for Apache Flink](#)，您可以使用 Java、Scala、Python 或 SQL 來處理和分析串流資料。您可以針對串流來源和靜態來源撰寫和執行程式碼，以執行時間序列分析、饋送即時儀表板和指標。

Business intelligence

[快速](#)讓決策者有機會在互動式視覺化環境中探索和解釋資訊。在單一資料儀表板中，Quick 可以包含 AWS 資料、第三方資料、大數據、試算表資料、SaaS 資料、B2B 資料等。透過 Quick Q，您可以使用自然語言詢問有關資料的問題，並接收回應。例如，「加州最暢銷的類別是什麼？」

Search analytics

[Amazon OpenSearch Service](#) 會為您的 OpenSearch 叢集佈建所有資源並啟動它。它還會自動偵測和更換發生故障的 OpenSearch Service 節點，可降低與自我管理的基礎設施相關的營運成本。您可以使用 OpenSearch Service 直接查詢來分析 Amazon S3 和其他 AWS 服務中的資料。

Data governance

透過 [Amazon DataZone](#)，您可以使用精細控制來管理和控管對資料的存取。這些控制項有助於確保具有適當層級的權限和內容的存取。Amazon DataZone 透過整合資料管理服務來簡化您的架構，包括 Amazon Redshift、Athena、Quick AWS Glue、內部部署來源和第三方來源。

Data collaboration

[AWS Clean Rooms](#) 是一個安全的協作工作區，您可以在其中分析集合資料集，而無需提供原始資料的存取權。您可以選擇要與其他公司合作的合作夥伴、選取其資料集，以及為這些合作夥伴設定隱私權增強控制。當您執行查詢時，會從該資料的原始位置 AWS Clean Rooms 讀取資料，並套用內建分析規則，以協助您維持對該資料的控制。

Data lake and data warehouse

- [新一代 Amazon SageMaker](#) 與 Apache Iceberg 完全相容，可讓您跨 Amazon Simple Storage Service (Amazon S3) 資料湖和 Amazon Redshift 資料倉儲統一資料。這可在單一資料複本上建置分析和 AI 和機器學習 (ML) 應用程式。透過零 ETL 整合，您可以近乎即時地從操作來源串流資料、跨多個來源執行聯合查詢，以及使用 Apache Iceberg 相容工具存取資料。您可以透過定義在所有分析和 ML 工具和引擎中強制執行的精細許可來保護資料。
- [Amazon S3](#) 幾乎可以存放和保護您可以用於資料湖基礎的任何資料量和類型。Amazon S3 提供管理功能，讓您可以最佳化、組織和設定對資料的存取，從而滿足您的特定業務、組織和合規要求。Amazon S3 Tables 提供針對分析工作負載最佳化的 S3 儲存體。使用標準 SQL 陳述式，您可以使用支援 Iceberg 的查詢引擎來查詢資料表，例如 Athena、Amazon Redshift 和 Apache Spark。
- [Amazon Redshift](#) 是全受管的 PB 級資料倉儲服務。Amazon Redshift 可以連接到 Amazon SageMaker 中的資料湖，讓您可以在跨 Amazon Redshift 資料倉儲和 Amazon S3 資料湖的統一資料上使用其強大的 SQL 分析功能。您也可以 Amazon Redshift 中使用 Amazon Q，透過自然語言簡化 SQL 撰寫。

考慮 AWS 分析服務的條件

建置資料分析的原因有很多 AWS。您可能需要支援綠地或試行專案，作為雲端遷移之旅的第一步。或者，您可能在盡可能減少中斷的情況下遷移現有的工作負載。無論您的目標為何，下列考量事項都有助於您做出選擇。

Assess data sources and data types

分析可用的資料來源和資料類型，以全面了解資料多樣性、頻率和品質。了解處理和分析資料時的任何潛在挑戰。此分析至關重要，因為：

- 資料來源多樣化，來自各種系統、應用程式、裝置和外部平台。
- 資料來源具有唯一的結構、格式和資料更新頻率。分析這些來源有助於識別適當的資料收集方法和技術。

- 分析結構化、半結構化和非結構化資料等資料類型會決定適當的資料處理和儲存方法。
- 分析資料來源和類型有助於資料品質評估，協助您預測潛在的資料品質問題：遺失值、不一致或不正確。

Data processing requirements

判斷如何擷取、轉換、清理和準備分析資料的資料處理需求。主要考量事項包括：

- 資料轉換：判斷讓原始資料適合分析所需的特定轉換。這涉及資料彙總、標準化、篩選和擴充等任務。
- 資料清理：評估資料品質並定義處理遺漏、不準確或不一致資料的程序。實作資料清理技術，以確保高品質資料，以獲得可靠的洞見。
- 處理頻率：根據分析需求，判斷是否需要即時、近乎即時或批次處理。即時處理可立即獲得洞見，而批次處理可能足以進行定期分析。
- 可擴展性和輸送量：評估處理資料磁碟區、處理速度和並行資料請求數量的可擴展性要求。確保所選的處理方法可以適應未來的成長。
- 延遲：考慮資料處理可接受的延遲，以及從資料擷取到分析結果所花費的時間。這對即時或時間敏感的分析尤其重要。

Storage requirements

透過確定在整個分析管道中存放資料的方式和位置，來確定儲存需求。重要考量包括：

- 資料量：評估產生和收集的資料量，並預估未來的資料成長，以規劃足夠的儲存容量。
- 資料保留：定義應為歷史分析或合規目的保留資料的持續時間。決定適當的資料保留政策。
- 資料存取模式：了解如何存取和查詢資料，以選擇最適合的儲存解決方案。考慮讀取和寫入操作、資料存取頻率和資料地區性。
- 資料安全：透過評估加密選項、存取控制和資料保護機制來優先考慮資料安全，以保護敏感資訊。
- 成本最佳化：根據資料存取模式和用量選擇最具成本效益的儲存解決方案，以最佳化儲存成本。
- 與分析服務整合：確保所選儲存解決方案與管道中的資料處理和分析工具之間的無縫整合。

Types of data

在決定收集和擷取資料的分析服務時，請考慮與組織需求和目標相關的各種資料類型。您可能需要考慮的常見資料類型包括：

- 交易資料：包括個別互動或交易的相關資訊，例如客戶購買、金融交易、線上訂單和使用者活動日誌。
- 檔案型資料：是指存放在檔案中的結構化或非結構化資料，例如日誌檔案、試算表、文件、影像、音訊檔案和影片檔案。Analytics 服務應支援擷取不同的檔案格式。
- 事件資料：擷取重大事件，例如使用者動作、系統事件、機器事件或業務事件。事件可以包含任何以高速抵達的資料，這些資料會擷取以進行串流或下游處理。

Operational considerations

營運責任由您和共同承擔 AWS，責任的劃分因不同層級的現代化而異。您可以選擇在上自我管理分析基礎設施，AWS 或利用多種無伺服器分析服務來減輕基礎設施管理負擔。

自我管理選項可讓使用者更好地控制基礎設施和組態，但他們需要更多的操作工作。

無伺服器選項可消除大部分的操作負擔，提供自動可擴展性、高可用性和強大的安全功能，讓使用者更專注於建置分析解決方案和推動洞察，而不是管理基礎設施和操作任務。考慮無伺服器分析解決方案的這些優點：

- 基礎設施抽象化：無伺服器服務抽象基礎設施管理，可讓使用者免於佈建、擴展和維護任務。會 AWS 處理這些操作層面，降低管理開銷。
- 自動擴展和效能：無伺服器服務會根據工作負載需求自動擴展資源，確保最佳效能，無需手動介入。
- 高可用性和災難復原：AWS 為無伺服器服務提供高可用性。AWS 管理資料備援、複寫和災難復原，以增強資料可用性和可靠性。
- 安全與合規：AWS 管理無伺服器服務的安全措施、資料加密和合規，遵守業界標準和最佳實務。
- 監控和記錄：AWS 提供無伺服器服務的內建監控、記錄和提醒功能。使用者可以透過 Amazon CloudWatch 存取詳細的指標和日誌。

Type of workload

建置現代分析管道時，決定要支援的工作負載類型對於有效滿足不同的分析需求至關重要。每種工作負載類型的考量重點包括：

批次工作負載

- 資料量和頻率：批次處理適用於具有定期更新的大量資料。

- 資料延遲：與即時處理相比，批次處理可能會延遲提供洞見。

互動式分析

- 資料查詢複雜性：互動式分析需要低延遲回應才能快速意見回饋。
- 資料視覺化：評估互動式資料視覺化工具的需求，讓商業使用者以視覺化方式探索資料。

串流工作負載

- 資料速度和磁碟區：串流工作負載需要即時處理才能處理高速資料。
- 資料視窗化：定義串流資料的資料視窗化和以時間為基礎的彙總，以擷取相關洞見。

Type of analysis needed

明確定義業務目標，以及您打算從分析衍生的洞察。不同類型的分析有不同的用途。例如：

- 描述性分析非常適合取得歷史概觀
- 診斷分析有助於了解過去事件背後的原因
- 預測分析預測未來成果
- 方案分析提供最佳動作的建議

將您的業務目標與相關類型的分析配對。以下是一些關鍵決策條件，可協助您選擇正確的分析類型：

- 資料可用性和品質：描述性和診斷分析依賴歷史資料，而預測和方案分析需要足夠的歷史資料和高品質資料來建立準確的模型。
- 資料量和複雜性：預測和方案分析需要大量的資料處理和運算資源。確保您的基礎設施和工具可以處理資料量和複雜性。
- 決策複雜性：如果決策涉及多個變數、限制條件和目標，則建議性分析可能更適合引導最佳動作。
- 風險承受能力：方案分析可能會提供建議，但伴隨著相關的不確定性。確保決策者了解與分析輸出相關的風險。

Evaluate scalability and performance

評估架構的可擴展性和效能需求。設計必須處理不斷增加的資料磁碟區、使用者需求和分析工作負載。要考慮的關鍵決策因素包括：

- 資料量和成長：評估目前的資料量並預測未來的成長。
- 資料速度和即時需求：判斷是否需要即時或近乎即時地處理和分析資料。
- 資料處理複雜性：分析資料處理和分析任務的複雜性。對於運算密集型任務，Amazon EMR 等服務可為大數據處理提供可擴展且受管的環境。
- 並行和使用者負載：考慮並行使用者的數量和系統上的使用者負載層級。
- 自動擴展功能：考慮提供自動擴展功能的服務，允許資源根據需求自動擴展或縮減。這可確保有效率的資源使用率和成本最佳化。
- 地理分佈：如果您的資料架構需要分佈在多個區域或位置，請考慮具有全域複寫和低延遲資料存取的服務。
- 成本效能權衡：平衡效能需求與成本考量。高效能的服務可能成本較高。
- 服務層級協議 (SLAs)：檢查 AWS 服務提供的 SLAs，以確保符合您的可擴展性和效能期望。

Data governance

資料控管是您需要實作的一組程序、政策和控制項，以確保資料資產的有效管理、品質、安全性和合規性。要考慮的關鍵決策點包括：

- 資料保留政策：根據法規要求和業務需求定義資料保留政策，並在不再需要時建立安全資料處理的程序。
- 稽核追蹤和記錄：決定記錄和稽核機制，以監控資料存取和用量。實作全面的稽核線索，以追蹤資料變更、存取嘗試和使用者活動，以進行合規和安全性監控。
- 合規要求：了解適用於您組織的產業特定和地理資料合規法規。確保資料架構符合這些法規和指導方針。
- 資料分類：根據資料的敏感度分類資料，並為每個資料類別定義適當的安全控制。
- 災難復原和業務連續性：規劃災難復原和業務連續性，以確保發生意外事件或系統故障時的資料可用性和彈性。
- 第三方資料共用：如果與第三方實體共用資料，請實作安全的資料共用通訊協定和協議，以保護資料機密性並防止資料濫用。

Security

分析管道中資料的安全性涉及保護管道每個階段的資料，以確保其機密性、完整性和可用性。要考慮的關鍵決策點包括：

- 存取控制和授權：實作強大的身分驗證和授權通訊協定，以確保只有獲得授權的使用者才能存取特定資料資源。
- 資料加密：針對存放在資料庫、資料湖中的資料，以及在架構不同元件之間的資料移動期間，選擇適當的加密方法。
- 資料遮罩和匿名化：考慮需要資料遮罩或匿名化來保護敏感資料，例如 PII 或敏感商業資料，同時允許某些分析程序繼續。
- 安全資料整合：建立安全的資料整合實務，以確保資料在架構的不同元件之間安全地流動，避免資料在資料移動期間洩漏或未經授權的存取。
- 網路隔離：考慮支援 [Amazon VPC 端點](#) 的服務，以避免將資源暴露到公有網際網路。

Plan for integration and data flows

定義分析管道各種元件之間的整合點和資料流程，以確保無縫的資料流程和互通性。要考慮的關鍵決策點包括：

- 資料來源整合：識別將從中收集資料的資料來源，例如資料庫、應用程式、檔案或外部 APIs。決定資料擷取方法（批次、即時、事件型），以最低的延遲有效率地將資料帶入管道。
- 資料轉換：判斷準備資料以供分析所需的轉換。決定工具和程序，以在資料流經管道時清除、彙總、標準化或充實資料。
- 資料移動架構：為管道元件之間的資料移動選擇適當的架構。根據即時需求和資料量，考慮批次處理、串流處理或兩者的組合。
- 資料複寫和同步：決定資料複寫和同步機制，讓所有元件的資料保持 up-to-date 狀態。根據資料新鮮度需求，考慮即時複寫解決方案或定期資料同步。
- 資料品質和驗證：實作資料品質檢查和驗證步驟，以確保資料在管道中移動時的完整性。決定資料驗證失敗時要採取的動作，例如提醒或錯誤處理。
- 資料安全和加密：決定如何在傳輸和靜態期間保護資料。考慮根據資料敏感度所需的安全層級，決定加密方法來保護整個管道的敏感資料。
- 可擴展性和彈性：確保資料流程設計允許水平可擴展性，並且可以處理增加的資料量和流量。

Architect for cost optimization

在上建置您的分析管道 AWS 可提供各種成本最佳化機會。為了確保成本效益，請考慮下列策略：

- 資源大小和選擇：根據實際工作負載需求調整資源大小。選擇符合工作負載效能需求的 AWS 服務和執行個體類型，同時避免過度佈建。
- 自動擴展：針對遇到不同工作負載的服務實作自動擴展。自動擴展會根據需求動態調整執行個體數量，在低流量期間降低成本。
- Spot 執行個體：將 Amazon EC2 Spot 執行個體用於非關鍵和容錯的工作負載。相較於隨需執行個體，Spot 執行個體可以大幅降低成本。
- 預留執行個體：考慮購買 AWS 預留執行個體，在具有可預測用量的穩定工作負載上，相較於隨需定價，可大幅節省成本。
- 資料儲存分層：根據資料存取頻率使用不同的儲存類別，以最佳化資料儲存成本。
- 資料生命週期政策：建立資料生命週期政策，根據資料的存留期和使用模式自動移動或刪除資料。這有助於管理儲存成本，並讓資料儲存與其值保持一致。

選擇 AWS 分析服務

現在您知道評估分析需求的條件，您就可以選擇適合您的組織需求的 AWS 分析服務。下表將一組服務與常見功能和業務目標保持一致。

類別	它針對什麼進行最佳化？	服務
統一分析和 AI	分析和 AI 開發 針對使用單一開發環境進行最佳化，Amazon SageMaker Unified Studio 可存取資料、分析和 AI 功能。	Amazon SageMaker
資料處理	互動式分析 針對執行即時資料分析和探勘進行最佳化，可讓使用者以互動方式查詢和視覺化資料。	Amazon Athena
	大數據處理	Amazon EMR

類別	它針對什麼進行最佳化？	服務
	<p>針對處理、移動和轉換大量資料進行最佳化。</p>	
	<p>資料型錄</p> <p>針對提供可用資料、其結構、特性和關係的詳細資訊進行最佳化。</p>	<p>AWS Glue</p>
	<p>工作流程協同運作</p> <p>針對使用 Apache Airflow 建立、排程和監控資料工作流程進行最佳化，以協調分析程序和 ETL 任務。</p>	<p>Amazon MWAA</p>
資料串流	<p>串流資料的 Apache Kafka 處理</p> <p>針對使用 Apache Kafka 資料平面操作和執行 Apache Kafka 的開放原始碼版本進行最佳化。</p>	<p>Amazon MSK</p>
	<p>即時處理</p> <p>針對快速且持續的資料擷取和彙總進行最佳化，包括 IT 基礎設施日誌資料、應用程式日誌、社交媒體、市場資料饋送和 Web Clickstream 資料。</p>	<p>Amazon Kinesis Data Streams</p>

類別	它針對什麼進行最佳化？	服務
	<p>即時串流資料交付</p> <p>針對將即時串流資料交付至目的地進行最佳化，例如 Amazon S3、Amazon Redshift、OpenSearch Service、Splunk、Apache Iceberg Tables，以及支援的第三方服務供應商擁有的任何自訂 HTTP 端點或 HTTP 端點。</p>	Amazon Data Firehose
	<p>建置 Apache Flink 應用程式</p> <p>針對使用 Java、Scala、Python 或 SQL 來處理和分析串流資料進行最佳化。</p>	Amazon Managed Service for Apache Flink
商業智慧	<p>儀表板和視覺化</p> <p>針對視覺化呈現複雜的資料集進行最佳化，並提供資料的自然語言查詢。</p>	快速
搜尋分析	<p>受管 OpenSearch 叢集</p> <p>針對日誌分析、即時應用程式監控和點擊串流分析進行最佳化。</p>	Amazon OpenSearch Service
資料控管	<p>管理資料存取</p> <p>已針對資料在其整個生命週期中設定適當的管理、可用性、可用性、完整性和安全性進行最佳化。</p>	Amazon DataZone

類別	它針對什麼進行最佳化？	服務
資料協作	<p>保護資料無塵室</p> <p>已針對與其他公司協作進行最佳化，無需共用原始基礎資料。</p>	AWS Clean Rooms
資料湖和倉儲	<p>統一存取資料湖和資料倉儲</p> <p>以湖房架構為基礎，以最佳化 Amazon S3 資料湖、Amazon Redshift 資料倉儲、操作資料庫，以及第三方和聯合資料來源之間的資料存取。</p>	Amazon SageMaker
	<p>資料湖的物件儲存</p> <p>針對提供資料湖基礎進行最佳化，具有幾乎無限的可擴展性和高耐用性。</p>	Amazon S3
	<p>資料倉儲</p> <p>已針對集中儲存、組織和擷取組織中各種來源的大量結構化和有時半結構化資料進行最佳化。</p>	Amazon Redshift

使用 AWS 分析服務

您現在應該清楚了解業務目標，以及您將要擷取和分析的資料量和速度，以開始建置資料管道。

為了探索如何使用和進一步了解每個可用的服務，我們提供了探索每個服務運作方式的途徑。下列各節提供深入文件、實作教學和資源的連結，協助您從基本使用開始使用到更進階的深入探討。

Amazon Athena

- [Amazon Athena 入門](#)

了解如何使用 Amazon Athena 查詢資料，並根據存放在 Amazon S3 中的範例資料建立資料表、查詢資料表，以及檢查查詢的結果。

[教學課程入門](#)

- Athena 上的 Apache Spark 入門

在 Athena 主控台中使用簡化的筆記本體驗，使用 Python 或 Athena 筆記本 APIs 開發 Apache Spark 應用程式。

[教學課程入門](#)

- 使用 Amazon SageMaker Lakehouse 架構來編目和管理 Athena 聯合查詢

了解如何透過 Amazon SageMaker 中的資料湖中心，對存放在 Amazon Redshift、DynamoDB 和 Snowflake 中的資料進行連線、控管和執行聯合查詢。

[閱讀部落格](#)

- 使用 Athena 分析 Amazon S3 中的資料

探索如何在 Elastic Load Balancer 的日誌上使用 Athena，並以預先定義的格式產生為文字檔案。我們向您展示如何建立資料表、以 Athena 使用的格式分割資料、將其轉換為 Parquet，以及比較查詢效能。

[閱讀部落格文章](#)

AWS Clean Rooms

- 設定 AWS Clean Rooms

了解如何 AWS Clean Rooms 在您的帳戶中設定 AWS。

[閱讀指南](#)

- 在上使用 AWS 實體解析來解鎖跨多方資料集的資料洞見，AWS Clean Rooms 而不共用基礎資料

了解如何使用準備和比對來協助改善與協作者的資料比對。

[閱讀部落格文章](#)

- 差異隱私權如何協助釋放洞見，而不會在個別層級洩漏資料

了解 AWS Clean Rooms 差異隱私權如何簡化套用差異隱私權，並協助保護使用者的隱私權。

[閱讀部落格](#)

Amazon Data Firehose

- 教學課程：從主控台建立 Firehose 串流

了解如何使用 AWS 管理主控台 或 AWS 開發套件來建立 Firehose 串流到您選擇的目的地。

[閱讀指南](#)

- 將資料傳送至 Firehose 串流

了解如何使用不同的資料來源將資料傳送至 Firehose 串流。

[閱讀指南](#)

- 在 Firehose 中轉換來源資料

了解如何叫用 Lambda 函數來轉換傳入的來源資料，並將轉換的資料傳遞至目的地。

[閱讀指南](#)

Amazon DataZone

- Amazon DataZone 入門

了解如何建立 Amazon DataZone 根網域、取得資料入口網站 URL、逐步解說資料生產者和資料消費者的基本 Amazon DataZone 工作流程。

[教學課程入門](#)

- 宣布新一代 Amazon SageMaker 和 Amazon DataZone 中資料歷程的一般可用性

了解 Amazon DataZone 如何使用自動譜系擷取，專注於自動收集和映射來自 AWS Glue 和 Amazon Redshift 的譜系資訊。

[閱讀部落格](#)

Amazon EMR

- Amazon EMR 入門

了解如何使用 Spark 啟動範例叢集，以及如何執行存放在 Amazon S3 儲存貯體中的簡單 PySpark 指令碼。

[教學課程入門](#)

- Amazon EMR on Amazon EKS 入門

我們向您展示如何在虛擬叢集上部署 Spark 應用程式，以開始使用 Amazon EMR on Amazon EKS。

[探索指南](#)

- EMR Serverless 入門

探索 Amazon EMR Serverless 如何提供無伺服器執行期環境，以簡化使用最新開放原始碼架構的分析應用程式的操作。

[教學課程入門](#)

AWS Glue

- 入門 AWS Glue DataBrew

了解如何建立您的第一個 DataBrew 專案。您可以載入範例資料集、在該資料集上執行轉換、建置配方來擷取這些轉換，以及執行任務以將轉換後的資料寫入 Amazon S3。

[教學課程入門](#)

- 使用 轉換資料 AWS Glue DataBrew

了解 AWS Glue DataBrew 是一種視覺化資料準備工具，可讓資料分析師和資料科學家輕鬆清理和標準化資料，以準備資料進行分析和機器學習。了解如何使用 建構 ETL 程序 AWS Glue DataBrew。

[實驗室入門](#)

- AWS Glue DataBrew 沉浸日

探索如何使用 AWS Glue DataBrew 來清理和標準化分析和機器學習的資料。

[研討會入門](#)

- 開始使用 AWS Glue Data Catalog

了解如何建立您的第一個 AWS Glue Data Catalog，這會使用 Amazon S3 儲存貯體做為資料來源。

[教學課程入門](#)

- 中的資料目錄和爬蟲程式 AWS Glue

探索如何使用 Data Catalog 中的資訊來建立和監控 ETL 任務。

[探索指南](#)

Amazon Kinesis Data Streams

- Amazon Kinesis Data Streams 入門教學課程

了解如何處理和分析即時股票資料。

[開始使用教學課程](#)

- 使用 Amazon Kinesis Data Streams 進行即時分析的架構模式，第 1 部分

了解兩種使用案例的常見架構模式：時間序列資料分析和事件驅動微服務。

[閱讀部落格](#)

- 使用 Amazon Kinesis Data Streams 進行即時分析的架構模式，第 2 部分

在三種案例中了解使用 Kinesis Data Streams 的 AI 應用程式：即時生成商業智慧、即時建議系統，以及物聯網資料串流和推論。

[閱讀部落格](#)

Amazon Managed Service for Apache Flink

- 什麼是 Amazon Managed Service for Apache Flink？

了解 Amazon Managed Service for Apache Flink 的基本概念。

[探索指南](#)

- [Amazon Managed Service for Apache Flink 研討會](#)

在本研討會中，您將了解如何使用 Amazon Managed Service for Apache Flink 部署、操作和擴展 Flink 應用程式。

[參加虛擬研討會](#)

Amazon MSK

- [Amazon MSK 入門](#)

了解如何建立 Amazon MSK 叢集、產生和使用資料，以及使用 指標監控叢集的運作狀態。

[指南入門](#)

- [Amazon MSK 研討會](#)

透過此實作 Amazon MSK 研討會深入探索。

[研討會入門](#)

Amazon MWAA

- [Amazon MWAA 入門](#)

了解如何建立您的第一個 MWAA 環境、將 DAG 上傳至 Amazon S3，以及執行您的第一個工作流程。

[教學課程入門](#)

- [使用 Amazon MWAA 建置資料管道](#)

了解如何建置end-to-end資料管道，以協調 Glue、EMR 和 Redshift 等其他 AWS 分析服務。此部落格文章探索使用 MWAA 和 Cosmos 協調 dbt Core 任務的簡化、組態驅動方法，以及在 Amazon Redshift 上執行轉換的任務。

[閱讀部落格文章](#)

- [Amazon MWAA 研討會](#)

探索實作練習，了解如何部署、設定和使用 Amazon MWAA 進行資料工作流程協同運作。

[研討會入門](#)

- Amazon MWAA 的最佳實務

了解在分析工作流程中使用 Amazon MWAA 的架構模式和最佳實務。

[閱讀指南](#)

OpenSearch Service

- OpenSearch Service 入門

了解如何使用 Amazon OpenSearch Service 建立和設定測試網域。

[教學課程入門](#)

- 使用 OpenSearch Service 和 OpenSearch Dashboards 視覺化客戶支援呼叫

探索以下情況的完整演練：企業會收到一些客戶支援呼叫，並想要分析它們。每個呼叫的主題為何？有多少是正面的？有多少是負面的？主管人員如何搜尋或檢閱這些呼叫的文字記錄？

[教學課程入門](#)

- Amazon OpenSearch Serverless 研討會入門

了解如何在主控台中設定新的 Amazon OpenSearch Serverless AWS 網域。探索可用的不同類型的搜尋查詢，並設計吸引人的視覺化效果，並了解如何根據指派的使用者權限來保護您的網域和文件。

[研討會入門](#)

- 成本最佳化向量資料庫：Amazon OpenSearch Service 量化技術簡介

了解 OpenSearch Service 如何支援純量和產品量化技術，以最佳化記憶體用量並降低營運成本。

[閱讀部落格文章](#)

Quick

- 快速資料分析入門

了解如何建立您的第一個分析。使用範例資料建立簡單或更進階的分析。或者您可以連線至您自己的資料來建立分析。

[探索指南](#)

- 使用 Quick 視覺化

探索商業智慧 (BI) 和資料視覺化的技術面 AWS。了解如何將儀表板嵌入應用程式和網站，並安全地管理存取和許可。

[課程入門](#)

- 快速研討會

使用研討會開始您的快速旅程

[研討會入門](#)

Amazon Redshift

- Amazon Redshift Serverless 入門

了解 Amazon Redshift Serverless 的基本流程，以建立無伺服器資源、連線至 Amazon Redshift Serverless、載入範例資料，然後對資料執行查詢。

[探索指南](#)

- Amazon Redshift 深入探討研討會

探索一系列有助於使用者開始使用 Amazon Redshift 平台的練習。

[研討會入門](#)

Amazon S3

- Amazon S3 入門

了解如何建立您的第一個 DataBrew 專案。您可以載入範例資料集、在該資料集上執行轉換、建置配方來擷取這些轉換，以及執行任務以將轉換後的資料寫入 Amazon S3。

[指南入門](#)

Amazon SageMaker

- SageMaker 入門

了解如何建立專案、新增成員，以及使用範例 JupyterLab 筆記本開始建置。

[閱讀指南](#)

- 新一代 Amazon SageMaker 簡介：所有資料、分析和 AI 的中心

了解如何開始使用資料處理、模型開發和生成式 AI 應用程式開發。

[閱讀部落格](#)

- 什麼是 SageMaker Unified Studio？

了解 SageMaker Unified Studio 的功能，以及如何在 Amazon SageMaker 時存取這些功能。

[閱讀指南](#)

- 開始使用 Amazon SageMaker 的 Lakehouse 架構

了解如何在 Amazon SageMaker 中建立專案，以及瀏覽、上傳和查詢業務使用案例的資料。

[閱讀指南](#)

- Amazon SageMaker 湖房架構中的資料連線

了解 Lakehouse 架構如何提供統一方法來管理 AWS 服務和企業應用程式之間的資料連線。

[閱讀指南](#)

- 使用 SageMaker Lakehouse 架構來編目和管理 Athena 聯合查詢

了解如何針對 Amazon SageMaker 專案，在 Amazon Redshift、DynamoDB 和 Snowflake 中存放的資料上連接、管理和執行聯合查詢。

[閱讀部落格](#)

探索使用 AWS 分析服務的方法

Editable architecture diagrams

參考架構圖

探索架構圖表，協助您開發、擴展和測試您的分析解決方案 AWS。

[探索分析參考架構](#)

Ready-to-use code

精選解決方案

在 上使用 Apache Druid 進行可擴展分析 AWS

部署建置 AWS 的程式碼，協助您設定、操作和管理 Apache Druid on AWS，這是一種經濟實惠、高可用性、彈性且容錯的託管環境。

[探索此解決方案](#)

AWS 解決方案

探索由 建置的預先設定、可部署的解決方案及其實作指南 AWS。

[探索所有 AWS 安全性、身分和管理解決方案](#)

Documentation

分析白皮書

探索白皮書，進一步了解選擇、實作和使用最適合您組織的分析服務時的洞見和最佳實務。

[探索分析白皮書](#)

AWS 大數據部落格

探索解決特定大數據使用案例的部落格文章。

[探索 AWS 大數據部落格](#)

文件歷史紀錄

下表說明此決策指南的重要變更。如需有關本指南更新的通知，您可以訂閱 RSS 摘要。

變更	描述	日期
re : Invent 更新	更新決策指南中的連結，並新增 Amazon Managed Workflows for Apache Airflow。	2025 年 9 月 24 日
re : Invent 更新	在整個決策指南中更新了對 Amazon SageMaker、Amazon SageMaker Unified Studio（不再預覽）和 Amazon SageMaker Lakehouse 的參考。	2025 年 9 月 9 日
re : Invent 更新	新增 SageMaker AI Unified Studio 和 AWS Clean Rooms。使用新的 AI 功能更新整個文件。	2025 年 2 月 20 日
初次出版	指南首先發佈。	2023 年 11 月 17 日

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。