使用者指南

AWS HealthOmics

版本 latest

AWS HealthOmics: 使用者指南

Table of Contents

什麼是 AWS HealthOmics？

AWS HealthOmics 是一項 AWS 服務，可協助生物資訊學家、研究人員和科學家等使用者儲存、查詢、分析和產生來自基因體和其他生物資料的洞見。它簡化並加速為研究和臨床組織儲存和分析基因體資訊的程序，並使科學探索和洞見產生更快。

HealthOmics 有三個主要元件。HealthOmics Storage 可協助您以每 GB 的低成本，有效率地存放和共用 PB 的基因體資料。HealthOmics Analytics 可簡化您準備基因體資料的方式，以進行多體學和多模態分析。HealthOmics Workflows 會自動為您的生物資訊學運算佈建和擴展基礎基礎設施。

主題

重要通知

HealthOmics 無法取代專業醫療建議、診斷或治療，也無法修復、治療、緩解、預防或診斷任何疾病或健康狀況。您負責將人工審核作為任何用途的一部分 AWS HealthOmics，包括與旨在告知臨床決策的任何第三方產品相關。

HealthOmics 僅用於傳輸、儲存、格式化或顯示資料，以及提供用於管理工作流程的基礎設施和組態支援。 AWS HealthOmics 並非直接執行變體呼叫或基因體分析和解釋。 AWS HealthOmics 並非用於解釋或分析臨床實驗室測試或其他裝置資料、結果和調查結果，也不能取代用於基因體分析的第三方工具。

HealthOmics 概念

本主題涵蓋 HealthOmics 特有的關鍵概念和術語的定義，以協助您了解 HealthOmics 使用本指南的術語。

主題

儲存

資料儲存體會分為序列存放區、用於基因體序列和相關資訊，以及用於所有參考基因體的參考存放區。下列術語說明 HealthOmics 特有的實作。

- 序列存放區 – 用於儲存基因體檔案的資料存放區。您可以在 HealthOmics 中擁有一或多個序列存放區。您可以在序列存放區上設定存取許可和 AWS KMS 加密，以控制誰可以存取資料。
- 讀取集 – 讀取集是基因體讀取的抽象，以 FASTQ、BAM 或 CRAM 格式儲存。讀取集可以匯入序列存放區，並以中繼資料標註。您可以使用屬性型存取控制 (ABAC) 將許可套用至讀取集。
- 參考 – 基因體參考會與讀取搭配使用，以識別特定讀取或一組讀取映射到的基因體。這些格式為 FASTA，並存放在參考存放區中。
- 參考存放區 – 用於儲存參考基因體的資料存放區。您可以在每個帳戶和區域中擁有單一參考存放區。

分析

您可以使用 HealthOmics Analytics 轉換和分析基因體資料。建立變體存放區或註釋存放區，以包含查詢的其他資訊。

- 變體存放區 – 以人口規模存放變體資料的資料存放區。變體存放區支援基因體變體呼叫格式 (gVCF) 和 VCF 輸入。
- 註釋存放區 – 代表註釋資料庫的資料存放區，例如來自 TSV/CSV、VCF 或一般功能格式 (GFF3) 檔案的資料存放區。註釋存放區會在匯入期間對應至與變體存放區相同的座標系統。

工作流程

透過 HealthOmics 工作流程，您可以處理和分析基因體資料。

- 工作流程 – 端對端程序的整體定義，包括參數和工具的參考。工作流程定義可以表示為 WDL、Nextflow 或 CWL。每個建立的工作流程都有唯一的識別符。

- 執行 – 工作流程的單一調用。個別執行會使用您定義的輸入資料，並產生輸出。每個建立的執行都有唯一的識別符。
- 任務 – 執行中的個別程序。HealthOmics Workflows 使用這些定義的運算規格來執行您的任務。每個任務都有唯一的識別符。
- 執行群組 – 一組執行，您可以設定最大 vCPU、最大持續時間或最大並行執行，以協助限制每次執行所使用的運算資源。您可以在執行群組中指定和設定執行的優先順序。例如，您可以指定在優先順序較低的執行之前執行高優先順序執行，以建立優先順序佇列。使用執行群組是選用的，而且每個執行群組都有唯一的識別符。

HealthOmics 功能

HealthOmics 提供下列功能。

- HealthOmics Storage – 可協助您以低成本、每 GB 的成本，有效率地存放和共用 PB 的原始基因體資料。
- HealthOmics Analytics — 簡化了如何準備基因體資料以進行多體學和多模態分析。
- HealthOmics 工作流程 — 為您的生物資訊工作流程自動佈建和擴展基礎基礎設施。

您可以獨立使用每個元件，或做為整合式end-to-end解決方案的一部分。

HealthOmics 為您提供下列優點。

- 安全地存放和合併基因體資料 — HealthOmics 整合其他 AWS 服務， AWS Lake Formation 例如 和 Amazon Athena。您可以安全地存放您的基因體資料，然後查詢或結合其與醫療歷史記錄資料，以獲得更好的診斷和個人化治療計劃。
- 保護患者隱私權 — HealthOmics 符合 HIPAA 資格。它還與 IAM 和 Amazon CloudWatch 整合，以便您可以控制和記錄資料存取，並追蹤資料在分析中如何使用。
- 專為擴展而建置：使用簡化的帳單和新的協同合作工具支援大型人口資料分析。
- 最大化效率 — 使用自動化工作流程和整合工具來簡化資料處理和分析。

您可以將 HealthOmics 用於下列生物醫學應用程式：

- 人口排序 — 一次查詢數千個基因組，以了解基因組變化如何映射到人口的表型。
- 臨床基因體 — 建置從排序器輸出到可報告資料的可重複基因體工作流程。您也可以最佳化高磁碟區輸送量，並設定高優先順序臨床樣本的運算需求，以減少周轉時間。

- 臨床試驗 — 將基因組分析整合到臨床試驗，以更好地了解新的候選藥物的有效性。透過長期節省成本和資料來源來簡化和加速臨床試驗，以符合管理機構的規定。
- 增強研究和創新 — 使用內建的資料列和資料欄型存取控制，簡化和控制匿名化基因體資料的儲存、存取和分析。

相關服務

下列 服務適用於 HealthOmics。

- Amazon Elastic Container Registry – 每個私有工作流程使用 Amazon ECR 映像 （ 在私有 Amazon ECR 儲存庫中 ） 來包含執行工作流程所需的所有可執行檔、程式庫和指令碼。
- Amazon Simple Storage Service – Amazon S3 為儲存和工作流程資料提供檔案儲存。
- AWS Lake Formation – Lake Formation 管理對 Analytics 資料存放區的資料存取。
- Amazon Athena – 使用 Athena 對變體存放區執行查詢。
- Amazon SageMaker AI – 使用 SageMaker AI 使用 Jupyter 筆記本執行 HealthOmics 任務。

AWS HealthOmics 的區域和端點

如需區域和端點的完整清單，請參閱 AWS 一般參考。

除了預設作用中 AWS 的區域之外，還需要啟用選擇加入區域。若要進一步了解如何啟用或停用區域，請參閱《帳戶管理指南》中的指定 AWS 您的帳戶可以使用的區域。 AWS

如何存取 HealthOmics

您可以使用 管理主控台、CLI、 SDKs或 API 來存取 AWS HealthOmics 功能。

- AWS 管理主控台 – 提供可用來存取 HealthOmics 的 Web 界面。
- AWS Command Line Interface (AWS CLI) – 為廣泛的 AWS 服務提供命令，包括 Windows AWS HealthOmics、macOS 和 Linux 支援 和 。如需安裝 的詳細資訊 AWS CLI，請參閱 AWS Command Line Interface。
- AWS SDKs – AWS 提供SDKs開發套件 （ 軟體開發套件 ），其中包含適用於各種程式設計語言和平台 （ 包括 Java、Python、Ruby、.NET、iOS 和 Android) 的程式庫和範本程式碼。SDKs提供以程式設計方式使用 HealthOmics 的便利方式。如需詳細資訊，請參閱 AWS SDK 開發人員中心。

- AWS API – 您可以使用 API 操作，以程式設計方式存取和管理 HealthOmics。如需詳細資訊，請參閱 HealthOmics API 參考。

進一步了解

從這些研討會和教學課程進一步了解 HealthOmics：

- HealthOmics 研討會 – HealthOmics 端對端研討會
- AWS 基因體資源 – 與基因體相關的公有 Amazon ECR 儲存庫
- Python 教學課程 – GitHub 上的 Jupyter 筆記本教學課程，涵蓋 HealthOmics 儲存、分析和工作流程

熟悉其他 HealthOmics 工具， AWS 提供：

- WDL linter – 適用於 WDL 的 HealthOmics linter
- Nextflow linter – 適用於 Nextflow 的 HealthOmics linter
- HealthOmics Amazon ECR 協助工具 – HealthOmics 的 Amazon ECR 協助工具
- GitHub 上的 HealthOmics 工具 – 使用 HealthOmics 的工具 (Transfer Manager、URI 剖析器、Omics rerun、Run Analyzer)。

設定 HealthOmics

若要設定 AWS HealthOmics，請註冊 AWS 帳戶、建立管理使用者，以及安全地管理其他使用者的存取權。

主題

註冊 AWS 帳戶

如果您沒有 AWS 帳戶，請完成下列步驟來建立一個。

註冊 AWS 帳戶

1. 開啟 https://portal.aws.amazon.com/billing/signup。
2. 請遵循線上指示進行。

   部分註冊程序需接收來電或簡訊，並在電話鍵盤輸入驗證碼。

   當您註冊 時 AWS 帳戶，AWS 帳戶根使用者會建立 。根使用者有權存取該帳戶中的所有 AWS 服務 和資源。作為安全最佳實務，請將管理存取權指派給使用者，並且僅使用根使用者來執行需要根使用者存取權的任務。

AWS 會在註冊程序完成後傳送確認電子郵件給您。您可以隨時登錄 https://aws.amazon.com/ 並選擇我的帳戶，以檢視您目前的帳戶活動並管理帳戶。

建立具有管理存取權的使用者

註冊 後 AWS 帳戶，請保護 AWS 帳戶根使用者、啟用 AWS IAM Identity Center和建立管理使用者，以免將根使用者用於日常任務。

保護您的 AWS 帳戶根使用者

1. 選擇根使用者並輸入 AWS 帳戶 您的電子郵件地址，以帳戶擁有者AWS Management Console身分登入 。在下一頁中，輸入您的密碼。

   如需使用根使用者登入的說明，請參閱 AWS 登入 使用者指南中的以根使用者身分登入。

2. 若要在您的根使用者帳戶上啟用多重要素驗證 (MFA)。

   如需說明，請參閱《IAM 使用者指南》中的為您的 AWS 帳戶 根使用者 （主控台） 啟用虛擬 MFA 裝置。

建立具有管理存取權的使用者

1. 啟用 IAM Identity Center。

   如需指示，請參閱《AWS IAM Identity Center 使用者指南》中的啟用 AWS IAM Identity Center。

2. 在 IAM Identity Center 中，將管理存取權授予使用者。

   如需使用 IAM Identity Center 目錄 做為身分來源的教學課程，請參閱AWS IAM Identity Center 《 使用者指南》中的使用預設值設定使用者存取權 IAM Identity Center 目錄。

以具有管理存取權的使用者身分登入

- 若要使用您的 IAM Identity Center 使用者簽署，請使用建立 IAM Identity Center 使用者時傳送至您電子郵件地址的簽署 URL。

  如需使用 IAM Identity Center 使用者登入的說明，請參閱AWS 登入 《 使用者指南》中的登入 AWS 存取入口網站。

指派存取權給其他使用者

1. 在 IAM Identity Center 中，建立一個許可集來遵循套用最低權限的最佳實務。

   如需指示，請參閱《AWS IAM Identity Center 使用者指南》中的建立許可集。

2. 將使用者指派至群組，然後對該群組指派單一登入存取權。

   如需指示，請參閱《AWS IAM Identity Center 使用者指南》中的新增群組。

建立 HealthOmics 的 IAM 許可

若要使用 HealthOmics，請設定下列 IAM 許可：

- 您帳戶中使用者存取 HealthOmics 的 IAM 身分型政策。
- HealthOmics 代表您存取資源的 IAM 服務角色。
- 其他 服務 （ 例如 Lake Formation 和 Amazon ECR) 中的許可，可供您的使用者和 HealthOmics 服務存取資源。

如需設定 HealthOmics IAM 許可的詳細資訊，請參閱 HealthOmics 的 IAM 許可。

與外部程式碼儲存庫連線

透過 AWS HealthOmics，您可以使用 Git 型儲存庫來管理工作流程 AWS CodeConnections。HealthOmics 使用此連線來存取您的原始碼儲存庫。

在使用外部程式碼儲存庫之前，請遵循設定連線指南開始使用 AWS CodeConnections。確認您已為 AWS 帳戶建立適當的 IAM 政策和許可。如需支援的 Git 提供者清單和詳細資訊，請參閱我可以為哪些第三方供應商建立連線？。

建立連線

若要與您偏好的儲存庫提供者建立連線，請遵循建立連線教學課程。

搭配 HealthOmics 使用 Amazon Q CLI

Amazon Q CLI 提供與 的自然語言互動 AWS HealthOmics，可讓您使用對話式命令執行複雜的基因體工作流程和分析任務。若要使用 Amazon Q CLI，請務必為 HealthOmics 和其他 服務 （ 例如 CloudWatch、Amazon ECR 或 Amazon S3) 設定 IAM 許可，讓 Amazon Q 存取其資源。

HealthOmics Agentic 生成式 AI 教學課程提供step-by-step指引，以設定內容檔案，並讓 Amazon Q CLI 能夠建立、執行和最佳化您的 AWS HealthOmics 工作流程。

HealthOmics 入門

若要開始使用 HealthOmics，請確定您已正確設定 HealthOmics 的 IAM 許可和角色。

在 HealthOmics 主控台中使用 Ready2Run 工作流程

下列練習示範如何使用 Ready2Run 工作流程。Ready2Run 工作流程已預先設定執行工作流程所需的參數和工具參考。工作流程發佈者提供範例資料，因此您不需要建立自己的資料。

1. 開啟 HealthOmics 主控台。
2. 選取左上方的導覽窗格 (≡)，然後選取 Ready2Run 工作流程。
3. 在 Ready2Run 工作流程頁面上，選擇ESMFold for up to 800 residues工作流程。主控台會開啟該工作流程的詳細資訊頁面。
4. 詳細資訊索引標籤提供工作流程的相關資訊。若要試用工作流程，請在頁面右上角選取開始執行。
5. 在指定執行詳細資訊頁面中，輸入執行名稱。
6. 輸入或選取執行輸出的 Amazon S3 位置。
7. 針對執行中繼資料保留模式，選擇是否要保留或移除 Runmeta 資料。
8. 在服務角色面板中，選擇建立並使用新的服務角色。
9. 選擇下一步。
10. 在新增參數值頁面上，選擇使用 Ready2Run 測試資料執行工作流程。
11. 選擇下一步。
12. 檢閱您的輸入，然後選擇開始執行。

Amazon Q CLI 的範例提示

Amazon Q CLI 可以使用 AWS HealthOmics 自然語言命令在 中執行基因體工作流程和分析任務。下列範例提示可讓您建立工作流程、管理執行和分析基因體資料。如需 HealthOmics 的詳細資訊和範例提示，請參閱 GitHub 上的 HealthOmics Agentic 生成式 AI 教學課程。

- 「建立將在 HealthOmics 上執行`main.wdl`的 WDL 1.1 工作流程檔案。工作流程會將參考基因體做為輸入和一組 fastq 檔案。它會使用 BWA 為參考基因體編製索引，然後將每對 fastq 檔案映射至參考。最後，將每個映射的 BAM 合併到單一 BAM 檔案，並輸出此檔案及其 bai 索引。」
- 「封裝工作流程並在 HealthOmics 中建立工作流程」

- 「更新 input.json 檔案以使用來自我的 Amazon S3 儲存貯體的真實檔案`omics-my-bucket-with-genome-data`」（提供特定的 Amazon S3 儲存貯體位置，或讓 Amazon Q 探索）
- 「在 Amazon ECR 儲存庫中尋找合適的容器並更新 input.json 以使用這些容器」
- 「尋找或建立適當的 IAM 角色以在執行工作流程時使用」
- 「為我的工作流程建立執行快取」
- 「在 HealthOmics 中執行工作流程」
- 「檢查執行的狀態」

> **Warning**
>
> 使用 Amazon Q CLI 時，請先檢閱所有產生的內容和建議的動作，再繼續。提供意見回饋以改善回應品質，並符合您工作流程的需求。如需詳細資訊，請參閱 Amazon Q 的安全考量和最佳實務。

HealthOmics 中的私有工作流程

當您想要建立自己的工作流程定義時，請使用私有工作流程。工作流程定義會指定工作流程的相關資訊，並定義工作流程任務。執行是工作流程的單一調用，而任務是執行中的單一程序。

HealthOmics 支援您在工作流程描述語言 (WDL)、常見工作流程語言 (CWL) 或 Nextflow 中建立的工作流程定義。

HealthOmics 工作流程提供下列選用功能：

- Run groups – 您可以將私有工作流程新增至執行群組，以控制運算用量。執行群組是工作流程執行的集合，可共用一組資源限制，例如並行執行上限和執行持續時間上限。您可以設定這些限制來控制執行群組使用的運算資源。
- Call caching – 您可以使用呼叫快取來儲存和重複使用任務輸出，進而縮短執行持續時間並節省運算成本。
- Sharing workflows – 您可以與相同 AWS 帳戶 區域中的其他 共用私有工作流程。
- Workflow versions – 您可以建立私有工作流程的版本。工作流程版本控制可讓使用者選擇何時開始使用更新的功能。工作流程版本不可變，並提供與工作流程相同的資料來源層級。

如需設定工作流程 IAM 許可的詳細資訊，請參閱 HealthOmics 的 IAM 許可。

如需如何使用 HealthOmics 私有工作流程的完整範例，請參閱 HealthOmics Github 教學課程或 HealthOmics 的 AWS 研討會端對端教學課程。

主題

- 在 HealthOmics 中建立私有工作流程
- 使用 README 檔案
- HealthOmics 中的工作流程版本控制
- 開始 HealthOmics 執行
- 在 HealthOmics 中刪除執行和執行群組
- 建立 HealthOmics 執行群組
- HealthOmics 執行的呼叫快取
- 共用 HealthOmics 工作流程

在 HealthOmics 中建立私有工作流程

私有工作流程取決於您在建立工作流程之前建立和設定的各種資源：

- Workflow definition file：以 WDL、 Nextflow或 撰寫的工作流程定義檔案CWL。工作流程定義會指定使用工作流程之執行的輸入和輸出。它還包含工作流程執行和執行任務的規格，包括運算和記憶體需求。工作流程定義檔案必須為 `.zip` 格式。如需詳細資訊，請參閱 HealthOmics 中的工作流程定義檔案。
- Parameter template file （選用）：以 寫入的參數範本檔案JSON。建立 檔案以定義執行參數，或 HealthOmics 會為您產生參數範本。如需詳細資訊，請參閱 HealthOmics 工作流程的參數範本檔案。
- Amazon ECR container images：建立工作流程的容器映像，並將其存放在私有 Amazon ECR 儲存庫中。
- Sentieon licenses （選用）：請求在私有工作流程中使用Sentieon軟體Sentieon的授權。

或者，您可以在建立工作流程之前或之後，在工作流程定義上執行 linter。linter 主題說明 HealthOmics 中可用的 linter。

主題

- HealthOmics 中的工作流程定義檔案
- HealthOmics 工作流程的參數範本檔案
- 私有工作流程的 Amazon ECR 中的容器映像
- 請求私有工作流程的 Sentieon 授權
- HealthOmics 中的工作流程文字
- 建立或更新工作流程

HealthOmics 中的工作流程定義檔案

您可以使用工作流程定義來指定工作流程、執行和執行中任務的相關資訊。您可以使用工作流程定義語言，在一或多個檔案中建立工作流程定義。HealthOmics 支援以 WDL、Nextflow 或 CWL 撰寫的工作流程定義。如需這些語言的詳細資訊，請參閱 撰寫 HealthOmics 工作流程的工作流程定義

您可以在工作流程定義中指定下列類型的資訊：

- Language version – 工作流程定義的語言和版本。

- Compute and memory – 工作流程中任務的運算和記憶體需求。
- Inputs – 工作流程任務的輸入位置。如需詳細資訊，請參閱HealthOmics 執行輸入。
- Outputs – 儲存任務產生的輸出的位置。
- Task resources – 每個任務的運算和記憶體需求。
- Accelerators – 任務所需的其他資源，例如加速器。

主題

HealthOmics 工作流程定義需求

HealthOmics 工作流程定義檔案必須符合下列要求：

- 任務必須定義輸入/輸出參數、Amazon ECR 容器儲存庫，以及執行時間規格，例如記憶體或 CPU 配置。
- 確認您的 IAM 角色具有必要的許可。
  - 您的工作流程可以存取來自 AWS 資源的輸入資料，例如 Amazon S3。
  - 您的工作流程可以在需要時存取外部儲存庫服務。
- 宣告工作流程定義中的輸出檔案。若要將中繼執行檔案複製到輸出位置，請將它們宣告為工作流程輸出。
- 輸入和輸出位置必須與工作流程位於相同的區域。
- HealthOmics 儲存工作流程輸入必須處於 `ACTIVE` 狀態。HealthOmics 不會匯入`ARCHIVED`狀態為的輸入，導致工作流程失敗。如需 Amazon S3 物件輸入的資訊，請參閱 HealthOmics 執行輸入。
- 如果您的 ZIP 封存包含單一工作流程定義或名為「主要」的檔案，則工作流程main的位置是選用的。
  - 路徑範例： `workflow-definition/main-file.wdl`

- 從 Amazon S3 或本機磁碟機建立工作流程之前，請先建立工作流程定義檔案和任何相依項的 zip 封存檔，例如子工作流程。
- 建議您將工作流程中的 Amazon ECR 容器宣告為輸入參數，以驗證 Amazon ECR 許可。

其他 Nextflow 考量事項：

- /bin

  Nextflow 工作流程定義可能包含具有可執行指令碼的 /bin 資料夾。此路徑具有任務的唯讀和可執行存取權。依賴這些指令碼的任務應使用使用適當指令碼解譯器建置的容器。最佳實務是直接呼叫解譯器。例如：

  ```
  process my_bin_task {
     ...
     script:
        """
        python3 my_python_script.py
        """
  }
  ```

- includeConfig

  以 Nextflow 為基礎的工作流程定義可以包含 nextflow.config 檔案，以協助抽象參數定義或程序資源描述檔。若要支援在多個環境中開發和執行 Nextflow 管道，請使用您使用 includeConfig 指令新增至全域組態的 HealthOmics 特定組態。若要維持可攜性，請使用下列程式碼將工作流程設定為僅在 HealthOmics 上執行時包含 檔案：

  ```
  // at the end of the nextflow.config file
  if ("$AWS_WORKFLOW_RUN") {
      includeConfig 'conf/omics.config'
  }
  ```

- Reports

  HealthOmics 不支援引擎產生的 dag、追蹤和執行報告。您可以使用 GetRun 和 GetRunTask API 呼叫的組合，產生追蹤和執行報告的替代方案。

其他 CWL 考量事項：

- Container image uri interpolation

  HealthOmics 允許 DockerRequirement 的 dockerPull 屬性成為內嵌 javascript 表達式。DockerRequirement 例如：

  ```
  requirements:
    DockerRequirement:
      dockerPull: "$(inputs.container_image)"
  ```

  這可讓您指定容器映像 URIs做為工作流程的輸入參數。

- Javascript expressions

  Javascript 表達式必須`strict mode`合規。

- Operation process

  HealthOmics 不支援 CWL 操作程序。

HealthOmics 工作流程定義語言的版本支援

HealthOmics 支援以 Nextflow、WDL 或 CWL 撰寫的工作流程定義檔案。下列各節提供有關 HealthOmics 版本支援這些語言的資訊。

主題

- WDL 版本支援
- CWL 版本支援
- Nextflow 版本支援

WDL 版本支援

HealthOmics 支援版本 1.0、1.1 和 WDL 規格的開發版本。

每個 WDL 文件都必須包含版本陳述式，以指定其遵循的規格版本 （主要和次要）。如需版本的詳細資訊，請參閱 WDL 版本控制

WDL 規格的 1.0 和 1.1 版不支援 `Directory`類型。若要將 `Directory`類型用於輸入或輸出，請在檔案development的第一行將版本設定為 ：

```
version development  # first line of .wdl file
```

```
    ... remainder of the file ...
```

CWL 版本支援

HealthOmics 支援 CWL 語言的 1.0、1.1 和 1.2 版。

您可以在 CWL 工作流程定義檔案中指定語言版本。如需 CWL 的詳細資訊，請參閱 CWL 使用者指南

Nextflow 版本支援

HealthOmics 支援三個 Nextflow 穩定版本。Nextflow 通常會每六個月發行一個穩定的版本。HealthOmics 不支援每月的「邊緣」版本。

HealthOmics 支援每個版本中發行的功能，但不支援預覽功能。

支援的版本

HealthOmics 支援下列 Nextflow 版本：

- Nextflow v22.04.01 DSL 1 和 DSL 2
- Nextflow v23.10.0 DSL 2（預設）
- Nextflow v24.10.8 DSL 2

若要將工作流程遷移至最新支援的版本 (v24.10.8)，請遵循 Nextflow 升級指南。

從 Nextflow v23 遷移到 v24 時有一些重大變更，如 Nextflow 遷移指南的以下章節所述：

- 中斷 24.04 中的變更
- 中斷 24.10 版中的變更

偵測和處理 Nextflow 版本

HealthOmics 會偵測您指定的 DSL 版本和 Nextflow 版本。它會根據這些輸入自動決定要執行的最佳 Nextflow 版本。

DSL 版本

HealthOmics 會在工作流程定義檔案中偵測請求的 DSL 版本。例如，您可以指定：`nextflow.enable.dsl=2`。

HealthOmics 預設支援 DSL 2。如果您的工作流程定義檔案中指定 ，它會提供與 DSL 1 的回溯相容性。

- 如果您指定 DSL 2，HealthOmics 會執行 Nextflow v23.10.0，除非您指定 Nextflow v22.04.0 或 v24.10.8。
- 如果您指定 DSL 1，HealthOmics 會執行 Nextflow v22.04 DSL1 （唯一支援執行 DSL 1 的版本）。
- 如果您未指定 DSL 版本，或 HealthOmics 因任何原因無法剖析 DSL 資訊 （例如工作流程定義檔案中的語法錯誤），HealthOmics 會預設為 DSL 2 並執行 Nextflow v23.10.0。
- 若要將工作流程從 DSL 1 升級到 DSL 2，以利用最新的 Nextflow 版本和軟體功能，請參閱從 DSL 1 遷移。

Nextflow 版本

如果您提供此檔案，HealthOmics 會在 Nextflow 組態檔案 (nextflow.config) 中偵測請求的 Nextflow 版本。我們建議您在檔案結尾新增 `nextflowVersion`子句，以避免來自包含組態的任何意外覆寫。如需詳細資訊，請參閱 Nextflow 組態。

您可以使用下列語法指定 Nextflow 版本或一系列版本：

```
// exact match
manifest.nextflowVersion = '1.2.3'

// 1.2 or later (excluding 2 and later)
manifest.nextflowVersion = '1.2+'

// 1.2 or later
manifest.nextflowVersion = '>=1.2'

// any version in the range 1.2 to 1.5
manifest.nextflowVersion = '>=1.2, <=1.5'

// use the "!" prefix to stop execution if the current version
// doesn't match the required version.
manifest.nextflowVersion = '!>=1.2'
```

HealthOmics 會處理 Nextflow 版本資訊，如下所示：

- 如果您使用 = 來指定 HealthOmics 支援的確切版本，HealthOmics 會使用該版本。
- 如果您使用 !指定確切的版本或不支援的版本範圍，HealthOmics 會引發例外狀況並使執行失敗。如果您想要嚴格處理版本請求，請考慮使用此選項，如果請求包含不支援的版本，則快速失敗。

- 如果您指定某個版本範圍，HealthOmics 會使用該範圍中最新的支援版本，除非該範圍包含 v24.10.8。在此情況下，HealthOmics 會提供較舊版本的偏好設定。例如，如果範圍同時涵蓋 v23.10.0 和 v24.10.8，HealthOmics 會選擇 v23.10.0。
- 如果沒有請求的版本，或者請求的版本無效或因任何原因無法剖析：
  - 如果您指定 DSL 1，HealthOmics 會執行 Nextflow v22.04。
  - 否則，HealthOmics 會執行 Nextflow v23.10.0。

您可以擷取 HealthOmics 用於每次執行的 Nextflow 版本相關資訊：

- 執行日誌包含 HealthOmics 用於執行之實際 Nextflow 版本的相關資訊。
- HealthOmics 會在執行日誌中新增警告，如果沒有與您請求的版本直接相符，或者需要使用與您指定的版本不同的版本。
- 對 GetRun API 操作的回應包含欄位 (`engineVersion`)，其中包含 HealthOmics 用於執行的實際 Nextflow 版本。例如：

```
"engineVersion":"22.04.0"
```

HealthOmics 任務的運算和記憶體需求

HealthOmics 會在 omics 執行個體中執行您的私有工作流程任務。HealthOmics 提供各種執行個體類型，以容納不同類型的任務。每個執行個體類型都有固定的記憶體和 vCPU 組態 （以及加速運算執行個體類型的固定 GPU 組態）。使用 omics 執行個體的成本會根據執行個體類型而有所不同。如需詳細資訊，請參閱 HealthOmics 定價頁面。

對於工作流程中的任務，您可以在工作流程定義檔案中指定所需的記憶體和 vCPUs。當工作流程任務執行時，HealthOmics 會配置最小的 omics 執行個體，以容納請求的記憶體和 vCPUs。例如，如果任務需要 64 GiB 的記憶體和 8 vCPUs，HealthOmics 會選取 `omics.r.2xlarge`。

我們建議您檢閱執行個體類型，並設定請求vCPUs 和記憶體大小，以符合最符合您需求的執行個體。任務容器會使用您在工作流程定義檔案中指定的 vCPUs 數量和記憶體大小，即使執行個體類型具有額外的 vCPUs和記憶體。

下列清單包含有關 vCPU 和記憶體配置的其他資訊：

- 容器資源配置是硬性限制。如果任務耗盡記憶體或嘗試使用其他 vCPUs任務會產生錯誤日誌並結束。

- 如果您未指定任何運算或記憶體需求，HealthOmics 會選取omics.c.large並預設為具有 1 個 vCPU 和 1 GiB 記憶體的組態。
- 您可以請求的最小組態為 1 個 vCPU 和 1 GiB 的記憶體。
- 如果您指定的 vCPUs、記憶體或 GPUs 超過支援的執行個體類型，HealthOmics 會擲回錯誤訊息，且工作流程驗證失敗
- 如果您指定分數單位，HealthOmics 會四捨五入至最接近的整數。
- HealthOmics 會為管理和記錄代理程式保留少量記憶體 (5%)，因此任務中的應用程式不一定能使用完整的記憶體配置。
- HealthOmics 會比對執行個體類型，以符合您指定的運算和記憶體需求，並且可能會混合使用硬體世代。因此，相同任務的任務執行時間可能會有一些次要差異。

這些主題提供有關 HealthOmics 支援的執行個體類型的詳細資訊。

主題

> **Note**
>
> 對於標準、運算和記憶體最佳化執行個體，如果執行個體需要更高的輸送量，請增加執行個體頻寬大小。少於 16 個 vCPUs Amazon EC2 執行個體可能會發生輸送量爆增。如需 Amazon EC2 執行個體輸送量的詳細資訊，請參閱 Amazon EC2 可用的執行個體頻寬。

標準執行個體類型

對於標準執行個體類型，組態旨在平衡運算能力和記憶體。

HealthOmics 支援這些區域中的 32xlarge 和 48xlarge 執行個體：美國西部 （奧勒岡） 和美國東部 （維吉尼亞北部）。

執行個體	vCPUs數量	記憶體
omics.m.large	2	8 GiB
omics.m.xlarge	4	16 GiB
omics.m.2xlarge	8	32 GiB
omics.m.4xlarge	16	64 GiB
omics.m.8xlarge	32	128 GiB
omics.m.12xlarge	48	192 GiB
omics.m.16xlarge	64	256 GiB
omics.m.24xlarge	96	384 GiB
omics.m.32xlarge	128	512 GiB
omics.m.48xlarge	192	768 GiB

運算最佳化執行個體

對於運算最佳化執行個體類型，組態具有更多的運算能力和更少的記憶體。

HealthOmics 支援這些區域中的 32xlarge 和 48xlarge 執行個體：美國西部（奧勒岡）和美國東部（維吉尼亞北部）。

執行個體	vCPUs數量	記憶體
omics.c.large	2	4 GiB
omics.c.xlarge	4	8 GiB
omics.c.2xlarge	8	16 GiB
omics.c.4xlarge	16	32 GiB
omics.c.8xlarge	32	64 GiB

執行個體	vCPUs數量	記憶體
omics.c.12xlarge	48	96 GiB
omics.c.16xlarge	64	128 GiB
omics.c.24xlarge	96	192 GiB
omics.c.32xlarge	128	256 GiB
omics.c.48xlarge	192	384 GiB

記憶體最佳化執行個體

對於記憶體最佳化執行個體類型，組態具有較少的運算能力和更多記憶體。

HealthOmics 支援這些區域中的 32xlarge 和 48xlarge 執行個體：美國西部（奧勒岡）和美國東部（維吉尼亞北部）。

執行個體	vCPUs數量	記憶體
omics.r.large	2	16 GiB
omics.r.xlarge	4	32 GiB
omics.r.2xlarge	8	64 GiB
omics.r.4xlarge	16	128 GiB
omics.r.8xlarge	32	256 GiB
omics.r.12xlarge	48	384 GiB
omics.r.16xlarge	64	512 GiB
omics.r.24xlarge	96	768 GiB
omics.r.32xlarge	128	1024 GiB
omics.r.48xlarge	192	1536 GiB

加速運算執行個體

您可以選擇性地為工作流程中的每個任務指定 GPU 資源，以便 HealthOmics 為任務配置加速運算執行個體。如需如何在工作流程定義檔案中指定 GPU 資訊的資訊，請參閱 HealthOmics 工作流程定義中的任務加速器。

如果您指定支援多個執行個體類型的 GPU，HealthOmics 會根據可用性選取執行個體類型。如果兩種執行個體類型都可用，HealthOmics 會優先考慮成本較低的執行個體。

以色列（特拉維夫）區域不支援 G4 執行個體。亞太區域（新加坡）區域不支援 G5 執行個體。

主題

G6 和 G6e 執行個體類型

HealthOmics 支援下列 G6 加速運算執行個體組態。所有 omics.g6 執行個體都使用 Nvidia L4 或 Nvidia L4 A10G GPUs。

HealthOmics 支援這些區域中的 G6 和 G6e 執行個體：美國西部（奧勒岡）和美國東部（維吉尼亞北部）。

執行個體	vCPUs數量	記憶體	GPUs 數量	記憶體	
omics.g6.xlarge	4	16 GiB	1	48 GiB	
omics.g6.2xlarge	8	32 GiB	1	48 GiB	
omics.g6.4xlarge	16	64 GiB	1	48 GiB	
omics.g6.8xlarge	32	128 GiB	1	48 GiB	
omics.g6.12xlarge	48	192 GiB	4	192 GiB	

執行個體	vCPUs數量	記憶體	GPUs 數量	記憶體
omics.g6.16xlarge	64	256 GiB	1	48 GiB
omics.g6.24xlarge	96	192 GiB	4	192 GiB

所有 omics.g6e 執行個體都使用 Nvidia L40s GPUs。

執行個體	vCPUs數量	記憶體	GPUs 數量	記憶體
omics.g6e.xlarge	4	32 GiB	1	24 GiB
omics.g6e.2xlarge	8	64 GiB	1	24 GiB
omics.g6e.4xlarge	16	128 GiB	1	24 GiB
omics.g6e.8xlarge	32	256 GiB	1	24 GiB
omics.g6e.12xlarge	48	384 GiB	4	96 GiB
omics.g6e.16xlarge	64	512 GiB	1	96 GiB
omics.g6e.24xlarge	96	768 GiB	4	96 GiB

G4 和 G5 執行個體

HealthOmics 支援下列 G4 和 G5 加速運算執行個體組態。

所有 omics.g5 執行個體都使用 Nvidia L4 A10G、Nvidia Tesla A10G 或 Nvidia Tesla T4 A10G GPUs。

執行個體	vCPUs數量	記憶體	GPUs 數量	記憶體	
omics.g5.xlarge	4	16 GiB	1	24 GiB	
omics.g5.2xlarge	8	32 GiB	1	24 GiB	
omics.g5.4xlarge	16	64 GiB	1	24 GiB	
omics.g5.8xlarge	32	128 GiB	1	24 GiB	
omics.g5.12xlarge	48	192 GiB	4	96 GiB	
omics.g5.16xlarge	64	256 GiB	1	24 GiB	
omics.g5.24xlarge	96	384 GiB	4	96 GiB	

所有 omics.g4dn 執行個體都使用 Nvidia Tesla T4 或 Nvidia Tesla T4 A10G GPUs。

執行個體	vCPUs數量	記憶體	GPUs 數量	記憶體	
omics.g4dn.xlarge	4	16 GiB	1	16 GiB	
omics.g4dn.2xlarge	8	32 GiB	1	16 GiB	
omics.g4dn.4xlarge	16	64 GiB	1	16 GiB	

執行個體	vCPUs數量	記憶體	GPUs 數量	記憶體
omics.g4dn.8xlarge	32	128 GiB	1	16 GiB
omics.g4dn.12xlarge	48	192 GiB	4	64 GiB
omics.g4dn.16xlarge	64	256 GiB	1	24 GiB

HealthOmics 工作流程定義中的任務輸出

您可以在工作流程定義中指定任務輸出。根據預設，HealthOmics 會在工作流程完成時捨棄所有中繼任務檔案。若要匯出中繼檔案，請將其定義為輸出。

如果您使用呼叫快取，HealthOmics 會將任務輸出儲存到快取，包括您定義為輸出的任何中繼檔案。

下列主題包含每個工作流程定義語言的任務定義範例。

主題

- WDL 的任務輸出
- Nextflow 的任務輸出
- CWL 的任務輸出

WDL 的任務輸出

對於以 WDL 撰寫的工作流程定義，請在最上層工作流程outputs區段中定義輸出。

HealthOmics

主題

- STDOUT 的任務輸出
- STDERR 的任務輸出
- 任務輸出至檔案
- 檔案陣列的任務輸出

STDOUT 的任務輸出

此範例會建立名為 的任務SayHello，將 STDOUT 內容回應至任務輸出檔案。WDL stdout函數會擷取檔案 中的 STDOUT 內容 （在此範例中為 Hello World！ 輸入字串）stdout_file。

由於 HealthOmics 會為所有 STDOUT 內容建立日誌，因此輸出也會出現在 CloudWatch Logs 中，以及任務的其他 STDERR 記錄資訊。

```
version 1.0
 workflow HelloWorld {
    input {
        String message = "Hello, World!"
        String ubuntu_container = "123456789012.dkr.ecr.us-east-1.amazonaws.com/
dockerhub/library/ubuntu:20.04"
    }

    call SayHello {
        input:
            message = message,
            container = ubuntu_container
    }

    output {
        File stdout_file = SayHello.stdout_file
    }
}

task SayHello {
    input {
        String message
        String container
    }

    command <<<
        echo "~{message}"
        echo "Current date: $(date)"
        echo "This message was printed to STDOUT"
    >>>

    runtime {
        docker: container
        cpu: 1
        memory: "2 GB"
```

```
    }

    output {
        File stdout_file = stdout()
    }
}
```

STDERR 的任務輸出

此範例會建立名為 的任務SayHello，將 STDERR 內容回應至任務輸出檔案。WDL stderr函數會擷取檔案 中的 STDERR 內容 （在此範例中為 Hello World！ 輸入字串）stderr_file。

由於 HealthOmics 會為所有 STDERR 內容建立日誌，因此輸出會出現在 CloudWatch Logs 中，以及任務的其他 STDERR 記錄資訊。

```
version 1.0
 workflow HelloWorld {
    input {
        String message = "Hello, World!"
        String ubuntu_container = "123456789012.dkr.ecr.us-east-1.amazonaws.com/
dockerhub/library/ubuntu:20.04"
    }

    call SayHello {
        input:
            message = message,
            container = ubuntu_container
    }

    output {
        File stderr_file = SayHello.stderr_file
    }
}

task SayHello {
    input {
        String message
        String container
    }

    command <<<
        echo "~{message}" >&2
        echo "Current date: $(date)" >&2
```

```
        echo "This message was printed to STDERR" >&2
    >>>

    runtime {
        docker: container
        cpu: 1
        memory: "2 GB"
    }

    output {
        File stderr_file = stderr()
    }
}
```

任務輸出至檔案

在此範例中，SayHello 任務會建立兩個檔案 (message.txt 和 info.txt)，並明確將這些檔案宣告為具名輸出 (message_file 和 info_file)。

```
version 1.0
workflow HelloWorld {
    input {
        String message = "Hello, World!"
        String ubuntu_container = "123456789012.dkr.ecr.us-east-1.amazonaws.com/
dockerhub/library/ubuntu:20.04"
    }

    call SayHello {
        input:
            message = message,
            container = ubuntu_container
    }

    output {
        File message_file = SayHello.message_file
        File info_file = SayHello.info_file
    }
}

task SayHello {
    input {
        String message
        String container
```

```
    }

    command <<<
        # Create message file
        echo "~{message}" > message.txt

        # Create info file with date and additional information
        echo "Current date: $(date)" > info.txt
        echo "This message was saved to a file" >> info.txt
    >>>

    runtime {
        docker: container
        cpu: 1
        memory: "2 GB"
    }

    output {
        File message_file = "message.txt"
        File info_file = "info.txt"
    }
}
```

檔案陣列的任務輸出

在此範例中，GenerateGreetings任務會產生一組檔案做為任務輸出。任務會為輸入陣列 的每個成員動態產生一個問候語檔案names。由於檔案名稱在執行時間之前是未知的，因此輸出定義會使用 WDL glob() 函數來輸出符合模式 的所有檔案*_greeting.txt。

```
version 1.0
 workflow HelloArray {
    input {
        Array[String] names = ["World", "Friend", "Developer"]
        String ubuntu_container = "123456789012.dkr.ecr.us-east-1.amazonaws.com/
dockerhub/library/ubuntu:20.04"
    }

    call GenerateGreetings {
        input:
            names = names,
            container = ubuntu_container
    }
```

```
    output {
        Array[File] greeting_files = GenerateGreetings.greeting_files
    }
}

task GenerateGreetings {
    input {
        Array[String] names
        String container
    }

    command  <<<
        # Create a greeting file for each name
        for name in ~{sep=" " names}; do
            echo "Hello, $name!" > ${name}_greeting.txt
        done
    >>>

    runtime {
        docker: container
        cpu: 1
        memory: "2 GB"
    }

    output {
        Array[File] greeting_files = glob("*_greeting.txt")
    }
 }
```

Nextflow 的任務輸出

對於以 Nextflow 編寫的工作流程定義，請定義 publishDir 指令，將任務內容匯出到您的輸出 Amazon S3 儲存貯體。將 publishDir 值設定為 `/mnt/workflow/pubdir`。

若要讓 HealthOmics 將檔案匯出至 Amazon S3，檔案必須位於此目錄中。

如果任務產生一組輸出檔案，做為後續任務的輸入，建議您將這些檔案分組在 目錄中，並將該目錄發出為任務輸出。列舉每個個別檔案可能會導致基礎檔案系統中的 I/O 瓶頸。例如：

```
process my_task {
      ...
      // recommended
      output "output-folder/", emit: output
```

```
    // not recommended
    // output "output-folder/**", emit: output
    ...
}
```

CWL 的任務輸出

對於以 CWL 編寫的工作流程定義，您可以使用任務指定CommandLineTool任務輸出。下列各節顯示定義不同輸出類型的CommandLineTool任務範例。

主題

- STDOUT 的任務輸出
- STDERR 的任務輸出
- 任務輸出至檔案
- 檔案陣列的任務輸出

STDOUT 的任務輸出

此範例會建立將 STDOUT 內容回應至名為 的文字輸出檔案CommandLineTool的任務output.txt。例如，如果您提供下列輸入，則產生的任務輸出在 output.txt 檔案中為 Hello World！。

```
{
    "message": "Hello World!"
}
```

outputs 指令指定輸出名稱為 ，example_out其類型為 stdout。若要讓下游任務消耗此任務的輸出，它會將輸出稱為 example_out。

由於 HealthOmics 會為所有 STDERR 和 STDOUT 內容建立日誌，因此輸出也會出現在 CloudWatch Logs 中，以及任務的其他 STDERR 日誌記錄資訊。

```
cwlVersion: v1.2
class: CommandLineTool
baseCommand: echo
stdout: output.txt
inputs:
  message:
    type: string
    inputBinding:
```

```
      position: 1
outputs:
  example_out:
    type: stdout

requirements:
    DockerRequirement:
        dockerPull: 123456789012.dkr.ecr.us-east-1.amazonaws.com/dockerhub/library/
ubuntu:20.04
    ResourceRequirement:
        ramMin: 2048
        coresMin: 1
```

STDERR 的任務輸出

此範例會建立將 STDERR 內容回應至名為 的文字輸出檔案CommandLineTool的任務stderr.txt。任務會修改 ，baseCommand讓 echo寫入 STDERR （ 而非 STDOUT)。

outputs 指令指定輸出名稱為 ，stderr_out其類型為 stderr。

由於 HealthOmics 會為所有 STDERR 和 STDOUT 內容建立日誌，因此輸出會出現在 CloudWatch Logs 中，以及任務的其他 STDERR 日誌記錄資訊。

```
cwlVersion: v1.2
class: CommandLineTool
baseCommand: [bash, -c]
stderr: stderr.txt
inputs:
  message:
    type: string
    inputBinding:
      position: 1
      shellQuote: true
      valueFrom: "echo $(self) >&2"
outputs:
  stderr_out:
    type: stderr

requirements:
    DockerRequirement:
        dockerPull: 123456789012.dkr.ecr.us-east-1.amazonaws.com/dockerhub/library/
ubuntu:20.04
    ResourceRequirement:
```

```
        ramMin: 2048
        coresMin: 1
```

任務輸出至檔案

此範例會建立從輸入檔案建立壓縮 tar 封存CommandLineTool的任務。您提供封存的名稱做為輸入參數 (archive_name)。

outputs 指令指定archive_file輸出類型為 File，並使用輸入參數的參考archive_name繫結至輸出檔案。

```
cwlVersion: v1.2
class: CommandLineTool
baseCommand: [tar, cfz]
inputs:
  archive_name:
    type: string
    inputBinding:
      position: 1
  input_files:
    type: File[]
    inputBinding:
      position: 2

outputs:
  archive_file:
    type: File
    outputBinding:
      glob: "$(inputs.archive_name)"

requirements:
    DockerRequirement:
        dockerPull: 123456789012.dkr.ecr.us-east-1.amazonaws.com/dockerhub/library/
ubuntu:20.04
    ResourceRequirement:
        ramMin: 2048
        coresMin: 1
```

檔案陣列的任務輸出

在此範例中，CommandLineTool任務會使用 touch命令建立檔案陣列。命令會使用files-to-create輸入參數中的字串來命名檔案。命令會輸出一組檔案。陣列包含工作目錄中符合 glob 模式的任何檔案。此範例使用符合所有檔案的萬用字元模式 ("*")。

```
cwlVersion: v1.2
class: CommandLineTool
baseCommand: touch
inputs:
  files-to-create:
    type:
      type: array
      items: string
    inputBinding:
      position: 1
outputs:
  output-files:
    type:
      type: array
      items: File
    outputBinding:
      glob: "*"

requirements:
    DockerRequirement:
        dockerPull: 123456789012.dkr.ecr.us-east-1.amazonaws.com/dockerhub/library/
ubuntu:20.04
    ResourceRequirement:
        ramMin: 2048
        coresMin: 1
```

HealthOmics 工作流程定義中的任務資源

在工作流程定義中，為每個任務定義下列項目：

- 任務的容器映像。如需詳細資訊，請參閱私有工作流程的 Amazon ECR 中的容器映像。
- 任務所需的 CPUs 和記憶體數量。如需詳細資訊，請參閱HealthOmics 任務的運算和記憶體需求。

HealthOmics 會忽略任何每個任務的儲存規格。HealthOmics 提供執行中所有任務都可以存取的執行儲存體。如需詳細資訊，請參閱在 HealthOmics 工作流程中執行儲存類型。

WDL

```
task my_task {
   runtime {
      container: "<aws-account-id>.dkr.ecr.<aws-region>.amazonaws.com/<image-name>"
```

```
      cpu: 2
      memory: "4 GB"
   }
   ...
}
```

對於 WDL 工作流程，HealthOmics 會針對因服務錯誤而失敗的任務嘗試最多兩次重試 (API 請求傳回 5XX HTTP 狀態碼）。如需任務重試的詳細資訊，請參閱 任務重試。

您可以在 WDL 定義檔案中指定任務的下列組態，以選擇退出重試行為：

```
runtime {
   preemptible: 0
}
```

NextFlow

```
process my_task {
   container "<aws-account-id>.dkr.ecr.<aws-region>.amazonaws.com/<image-name>"
   cpus 2
   memory "4 GiB"
   ...
}
```

CWL

```
cwlVersion: v1.2
class: CommandLineTool
requirements:
    DockerRequirement:
        dockerPull: "<aws-account-id>.dkr.ecr.<aws-region>.amazonaws.com/<image-
name>"
    ResourceRequirement:
        coresMax: 2
        ramMax: 4000 # specified in mebibytes
```

HealthOmics 工作流程定義中的任務加速器

在工作流程定義中，您可以選擇為任務指定 GPU 加速器規格。HealthOmics 支援下列加速器規格值，以及支援的執行個體類型：

加速器規格	運作狀態執行個體類型				
nvidia-tesla-t4	G4				
nvidia-tesla-t4-a10g	G4 和 G5				
nvidia-tesla-a10g	G5				
nvidia-l4-a10g	G5 和 G6				
nvidia-l4	G6				
nvidia-l40s	G6e				

如果您指定的加速器類型支援多個執行個體類型，HealthOmics 會根據可用容量選取執行個體類型。如果兩種執行個體類型都可用，HealthOmics 會優先考慮成本較低的執行個體。

如需執行個體類型的詳細資訊，請參閱 加速運算執行個體。

在下列範例中，工作流程定義指定 `nvidia-l4`作為加速器：

WDL

```
task my_task {
 runtime {
    ...
    acceleratorCount: 1
    acceleratorType: "nvidia-l4"
 }
 ...
}
```

NextFlow

```
process my_task {
```

```
 ...
 accelerator 1, type: "nvidia-l4"
 ...
}
```

CWL

```
cwlVersion: v1.2
class: CommandLineTool
requirements:
  ...
  cwltool:CUDARequirement:
      cudaDeviceCountMin: 1
      cudaComputeCapability: "nvidia-l4"
      cudaVersionMin: "1.0"
```

撰寫 HealthOmics 工作流程的工作流程定義

HealthOmics 支援以 WDL、Nextflow 或 CWL 撰寫的工作流程定義。若要進一步了解這些工作流程語言，請參閱 WDL、Nextflow 或 CWL 的規格。

HealthOmics 支援三種工作流程定義語言的版本管理。如需詳細資訊，請參閱HealthOmics 工作流程定義語言的版本支援 。

主題

- 在 WDL 中撰寫工作流程
- 在 Nextflow 中撰寫工作流程
- 在 CWL 中撰寫工作流程
- 工作流程定義範例
- WDL 工作流程定義範例

在 WDL 中撰寫工作流程

下表顯示 WDL 中的輸入如何對應至相符的基本類型或複雜的 JSON 類型。類型強制受到限制，且盡可能明確類型。

基本類型

WDL 類型	JSON 類型	範例 WDL	範例 JSON 金鑰和值	備註
Boolean	boolean	Boolean b	"b": true	值必須是小寫且沒有引號。
Int	integer	Int i	"i": 7	必須取消引號。
Float	number	Float f	"f": 42.2	必須取消引號。
String	string	String s	"s": "characters"	做為 URI 的 JSON 字串必須映射至要匯入的 WDL 檔案。
File	string	File f	"f": "s3://amzn-s3-demo-bucket1/path/to/file"	只要為工作流程提供的 IAM 角色具有這些物件的讀取存取權，Amazon S3 和 HealthOmics 儲存 URIs 就會匯入。不支援其他 URI 結構描述（例如 file://、https://和 ftp://）。URI 必須指定 物件。它不能是目錄，表示它不能以 結尾/。
Directory	string	Directory d	"d": "s3://bucket/path/"	Directory 類型不包含在 WDL 1.0 或 1.1 中，因此您需

WDL 類型	JSON 類型	範例 WDL	範例 JSON 金鑰和值	備註
				要將 version development 新增至 WDL 檔案的 標頭。URI 必須是 Amazon S3 URI，且字首結尾為 '/'。目錄的所有內容都會以單一下載方式遞迴複製到工作流程。Directory 應該只包含與工作流程相關的檔案。

WDL 中的複雜類型是由基本類型組成的資料結構。清單等資料結構將轉換為陣列。

複雜類型

WDL 類型	JSON 類型	範例 WDL	範例 JSON 金鑰和值	備註
Array	array	Array[Int] nums	“nums": [1, 2, 3]	陣列的成員必須遵循 WDL 陣列類型的格式。
Pair	object	Pair[String, Int] str_to_i	“str_to_i": {"left": "0", "right": 1}	配對的每個值都必須使用其相符 WDL 類型的 JSON 格式。
Map	object	Map[Int, String]	"int_to_string": { 2:	映射中的每個項目都必須使用其

WDL 類型	JSON 類型	範例 WDL	範例 JSON 金鑰和值	備註
		`int_to_string`	`"hello", 1: "goodbye" }`	相符 WDL 類型的 JSON 格式。
`Struct`	`object`	`struct SampleBamAndIndex { String sample_name File bam File bam_index } SampleBamAndIndex b_and_i`	`"b_and_i": { "sample_name": "NA12878", "bam": "s3://amzn-s3-demo-bucket1/NA12878.bam", "bam_index": "s3://amzn-s3-demo-bucket1/NA12878.bam.bai" }`	結構成員的名稱必須與 JSON 物件金鑰的名稱完全相符。每個值必須使用相符 WDL 類型的 JSON 格式。
`Object`	N/A	N/A	N/A	WDL `Object`類型已過期，`Struct`在所有情況下都應該由 取代。

HealthOmics 工作流程引擎不支援合格或命名空間的輸入參數。WDL 語言不會指定合格參數的處理及其對 WDL 參數的映射，而且可能模棱兩可。基於這些原因，最佳實務是宣告頂層（主要）工作流程定義檔案中的所有輸入參數，並使用標準 WDL 機制將其傳遞至子工作流程呼叫。

在 Nextflow 中撰寫工作流程

HealthOmics 支援 Nextflow DSL1 和 DSL2。如需詳細資訊，請參閱Nextflow 版本支援。

Nextflow DSL2 是以 Groovy 程式設計語言為基礎，因此參數是動態的，而且類型強制可以使用與 Groovy 相同的規則。輸入 JSON 提供的參數和值可在工作流程的參數 (`params`) 映射中使用。

> **Note**
>
> HealthOmics 支援 Nextflow 版本 v23.10 （但不支援 v22.04) 的 `nf-schema`和 `nf-validation` 外掛程式。

下列資訊與將這些外掛程式與 Nextflow v23.10 工作流程搭配使用相關：

- HealthOmics 會預先安裝 nf-schema@2.3.0 和 nf-validation@1.1.1 外掛程式。HealthOmics 會忽略您在 `nextflow.config` 檔案中指定的任何其他外掛程式版本。
- 您無法在工作流程執行期間擷取其他外掛程式。
- 在 Nextflow v24.04 及更高版本中，`nf-validation`外掛程式會重新命名為 `nf-schema`。如需詳細資訊，請參閱 Nextflow GitHub 儲存庫中的 nf-schema。

使用 Amazon S3 或 HealthOmics URI 建構 Nextflow 檔案或路徑物件時，只要授予讀取存取權，就會將相符的物件提供給工作流程。允許 Amazon S3 URIs 使用字首或目錄。如需範例，請參閱 Amazon S3 輸入參數格式。

HealthOmics 支援在 Amazon S3 URIs 或 HealthOmics 儲存 URIs 中使用 glob 模式。在工作流程定義中使用 Glob 模式來建立 `path`或 `file`頻道。

對於以 Nextflow 撰寫的工作流程，請定義 publishDir 指令，將任務內容匯出至輸出 Amazon S3 儲存貯體。如下列範例所示，將 publishDir 值設定為 `/mnt/workflow/pubdir`。若要將檔案匯出至 Amazon S3，檔案必須位於此目錄中。

```
 nextflow.enable.dsl=2

workflow {
  CramToBamTask(params.ref_fasta, params.ref_fasta_index, params.ref_dict,
 params.input_cram, params.sample_name)
  ValidateSamFile(CramToBamTask.out.outputBam)
}

process CramToBamTask {
  container "<account>.dkr.ecr.us-west-2.amazonaws.com/genomes-in-the-cloud"

  publishDir "/mnt/workflow/pubdir"
```

```
  input:
      path ref_fasta
      path ref_fasta_index
      path ref_dict
      path input_cram
      val sample_name

  output:
      path "${sample_name}.bam", emit: outputBam
      path "${sample_name}.bai", emit: outputBai

  script:
  """
      set -eo pipefail

      samtools view -h -T $ref_fasta $input_cram |
      samtools view -b -o ${sample_name}.bam -
      samtools index -b ${sample_name}.bam
      mv ${sample_name}.bam.bai ${sample_name}.bai
  """
}

process ValidateSamFile {
  container "<account>.dkr.ecr.us-west-2.amazonaws.com/genomes-in-the-cloud"

  publishDir "/mnt/workflow/pubdir"

  input:
      file input_bam

  output:
      path "validation_report"

  script:
  """
      java -Xmx3G -jar /usr/gitc/picard.jar \
      ValidateSamFile \
      INPUT=${input_bam} \
      OUTPUT=validation_report \
      MODE=SUMMARY \
      IS_BISULFITE_SEQUENCED=false
  """
```

```
}
```

在 CWL 中撰寫工作流程

以通用工作流程語言或 CWL 編寫的工作流程提供與以 WDL 和 Nextflow 編寫的工作流程類似的功能。您可以使用 Amazon S3 或 HealthOmics 儲存 URIs做為輸入參數。

如果您在子工作流程的 secondaryFile 中定義輸入，請在主要工作流程中新增相同的定義。

HealthOmics 工作流程不支援操作程序。若要進一步了解 CWL 工作流程中的操作程序，請參閱 CWL 文件。

若要將現有的 CWL 工作流程定義檔案轉換為使用 HealthOmics，請進行下列變更：

- 將所有 Docker 容器 URIs取代為 Amazon ECR URIs。
- 請確保在主要工作流程中將所有工作流程檔案宣告為輸入，並明確定義所有變數。
- 確保所有 JavaScript 程式碼都是嚴格模式投訴。

應該為每個使用的容器定義 CWL 工作流程。不建議使用固定的 Amazon ECR URI 硬式編碼 dockerPull 項目。

以下是以 CWL 撰寫的工作流程範例。

```
cwlVersion: v1.2
class: Workflow

inputs:
in_file:
  type: File
  secondaryFiles: [.fai]

out_filename: string
docker_image: string


outputs:
copied_file:
  type: File
  outputSource: copy_step/copied_file
```

```
steps:
copy_step:
  in:
    in_file: in_file
    out_filename: out_filename
    docker_image: docker_image
  out: [copied_file]
  run: copy.cwl
```

下列檔案定義 copy.cwl任務。

```
cwlVersion: v1.2
class: CommandLineTool
baseCommand: cp

inputs:
in_file:
  type: File
  secondaryFiles: [.fai]
  inputBinding:
    position: 1

out_filename:
  type: string
  inputBinding:
    position: 2
docker_image:
  type: string

outputs:
copied_file:
  type: File
  outputBinding:
      glob: $(inputs.out_filename)

requirements:
InlineJavascriptRequirement: {}
DockerRequirement:
  dockerPull: "$(inputs.docker_image)"
```

以下是以 CWL 撰寫且具有 GPU 需求的工作流程範例。

```
cwlVersion: v1.2
class: CommandLineTool
baseCommand: ["/bin/bash", "docm_haplotypeCaller.sh"]
$namespaces:
cwltool: http://commonwl.org/cwltool#
requirements:
cwltool:CUDARequirement:
  cudaDeviceCountMin: 1
  cudaComputeCapability: "nvidia-tesla-t4"
  cudaVersionMin: "1.0"
InlineJavascriptRequirement: {}
InitialWorkDirRequirement:
  listing:
  - entryname: 'docm_haplotypeCaller.sh'
    entry: |
            nvidia-smi --query-gpu=gpu_name,gpu_bus_id,vbios_version --format=csv

inputs: []
outputs: []
```

工作流程定義範例

下列範例顯示 WDL、Nextflow 和 CWL 中的相同工作流程定義。

WDL

```
version 1.1

task my_task {
   runtime { ... }
   inputs {
       File input_file
       String name
       Int threshold
   }

   command <<<
   my_tool --name ~{name} --threshold ~{threshold} ~{input_file}
   >>>

   output {
       File results = "results.txt"
   }
```

```
}

workflow my_workflow {
   inputs {
       File input_file
       String name
       Int threshold = 50
   }

   call my_task {
       input:
          input_file = input_file,
          name = name,
          threshold = threshold
   }
   outputs {
       File results = my_task.results
   }
}
```

Nextflow

```
nextflow.enable.dsl = 2

params.input_file = null
params.name = null
params.threshold = 50

process my_task {
   // <directives>

   input:
     path input_file
     val name
     val threshold

   output:
     path 'results.txt', emit: results

   script:
     """
     my_tool --name ${name} --threshold ${threshold} ${input_file}
     """
```

```
}

workflow MY_WORKFLOW {
   my_task(
       params.input_file,
       params.name,
       params.threshold
   )
}

workflow {
   MY_WORKFLOW()
}
```

CWL

```
cwlVersion: v1.2
class: Workflow

requirements:
    InlineJavascriptRequirement: {}

inputs:
   input_file: File
   name: string
   threshold: int

outputs:
    result:
        type: ...
        outputSource: ...

steps:
    my_task:
        run:
            class: CommandLineTool
            baseCommand: my_tool
            requirements:
                ...
```

```
            inputs:
                name:
                    type: string
                    inputBinding:
                        prefix: "--name"
                threshold:
                    type: int
                    inputBinding:
                        prefix: "--threshold"
                input_file:
                    type: File
                    inputBinding: {}
            outputs:
                results:
                    type: File
                    outputBinding:
                        glob: results.txt
```

WDL 工作流程定義範例

下列範例顯示用於在 WDL CRAMBAM中從 轉換為 的私有工作流程定義。CRAM 工作流程的 BAM定義了兩個任務，並使用來自 genomes-in-the-cloud容器的工具，如範例所示，可公開取得。

下列範例示範如何將 Amazon ECR 容器包含為參數。這可讓 HealthOmics 在啟動執行之前驗證容器的存取許可。

```
{
     ...
     "gotc_docker":"<account_id>.dkr.ecr.<region>.amazonaws.com/genomes-in-the-
cloud:2.4.7-1603303710"
  }
```

下列範例顯示當檔案位於 Amazon S3 儲存貯體時，如何指定要在執行中使用的檔案。

```
{
      "input_cram": "s3://amzn-s3-demo-bucket1/inputs/NA12878.cram",
      "ref_dict": "s3://amzn-s3-demo-bucket1/inputs/Homo_sapiens_assembly38.dict",
      "ref_fasta": "s3://amzn-s3-demo-bucket1/inputs/Homo_sapiens_assembly38.fasta",
      "ref_fasta_index": "s3://amzn-s3-demo-bucket1/inputs/
Homo_sapiens_assembly38.fasta.fai",
      "sample_name": "NA12878"
```

```
  }
```

如果您想要從序列存放區指定檔案，請指出 ，如下列範例所示，使用序列存放區的 URI。

```
{
      "input_cram": "omics://429915189008.storage.us-west-2.amazonaws.com/111122223333/
readSet/4500843795/source1",
      "ref_dict": "s3://amzn-s3-demo-bucket1/inputs/Homo_sapiens_assembly38.dict",
      "ref_fasta": "s3://amzn-s3-demo-bucket1/inputs/Homo_sapiens_assembly38.fasta",
      "ref_fasta_index": "s3://amzn-s3-demo-bucket1/inputs/
Homo_sapiens_assembly38.fasta.fai",
      "sample_name": "NA12878"
  }
```

然後，您可以在 WDL 中定義工作流程，如下所示。

```
 version 1.0
  workflow CramToBamFlow {
      input {
          File ref_fasta
          File ref_fasta_index
          File ref_dict
          File input_cram
          String sample_name
          String gotc_docker = "<account>.dkr.ecr.us-west-2.amazonaws.com/genomes-in-
the-
  cloud:latest"
      }
      #Converts CRAM to SAM to BAM and makes BAI.
      call CramToBamTask{
           input:
              ref_fasta = ref_fasta,
              ref_fasta_index = ref_fasta_index,
              ref_dict = ref_dict,
              input_cram = input_cram,
              sample_name = sample_name,
              docker_image = gotc_docker,
       }
       #Validates Bam.
       call ValidateSamFile{
          input:
             input_bam = CramToBamTask.outputBam,
             docker_image = gotc_docker,
```

```
      }
      #Outputs Bam, Bai, and validation report to the FireCloud data model.
      output {
          File outputBam = CramToBamTask.outputBam
          File outputBai = CramToBamTask.outputBai
          File validation_report = ValidateSamFile.report
       }
}
#Task definitions.
task CramToBamTask {
    input {
       # Command parameters
       File ref_fasta
       File ref_fasta_index
       File ref_dict
       File input_cram
       String sample_name
       # Runtime parameters
       String docker_image
    }
   #Calls samtools view to do the conversion.
   command {
       set -eo pipefail

       samtools view -h -T ~{ref_fasta} ~{input_cram} |
       samtools view -b -o ~{sample_name}.bam -
       samtools index -b ~{sample_name}.bam
       mv ~{sample_name}.bam.bai ~{sample_name}.bai
    }

    #Runtime attributes:
    runtime {
        docker: docker_image
    }

    #Outputs a BAM and BAI with the same sample name
     output {
         File outputBam = "~{sample_name}.bam"
         File outputBai = "~{sample_name}.bai"
    }
}

#Validates BAM output to ensure it wasn't corrupted during the file conversion.
task ValidateSamFile {
```

```
    input {
       File input_bam
       Int machine_mem_size = 4
       String docker_image
    }
    String output_name = basename(input_bam, ".bam") + ".validation_report"
    Int command_mem_size = machine_mem_size - 1
    command {
        java -Xmx~{command_mem_size}G -jar /usr/gitc/picard.jar \
        ValidateSamFile \
        INPUT=~{input_bam} \
        OUTPUT=~{output_name} \
        MODE=SUMMARY \
        IS_BISULFITE_SEQUENCED=false
     }
     runtime {
     docker: docker_image
     }
    #A text file is generated that lists errors or warnings that apply.
     output {
         File report = "~{output_name}"
     }
 }
```

HealthOmics 工作流程的參數範本檔案

參數範本定義工作流程的輸入參數。您可以定義輸入參數，讓您的工作流程更具彈性和多樣化。例如，您可以為參考基因體檔案的 Amazon S3 位置定義參數。參數範本可以透過 Git 型儲存庫服務或本機磁碟機提供。然後，使用者可以使用各種資料集執行工作流程。

您可以為工作流程建立參數範本，或者 HealthOmics 可以為您產生參數範本。

參數範本是 JSON 檔案。在 檔案中，每個輸入參數都是具名物件，必須符合工作流程輸入的名稱。當您開始執行時，如果您未提供所有必要參數的值，則執行會失敗。

輸入參數物件包含下列屬性：

- description – 此必要屬性是主控台顯示在開始執行頁面中的字串。此描述也會保留為執行中繼資料。
- optional – 此選用屬性指出輸入參數是否為選用。如果您未指定 optional 欄位，則需要輸入參數。

下列參數範本範例示範如何指定輸入參數。

```
{
  "myRequiredParameter1": {
     "description": "this parameter is required",
  },
  "myRequiredParameter2": {
     "description": "this parameter is also required",
     "optional": false
  },
  "myOptionalParameter": {
     "description": "this parameter is optional",
     "optional": true
  }
}
```

產生參數範本

HealthOmics 透過剖析工作流程定義來偵測輸入參數來產生參數範本。如果您提供工作流程的參數範本檔案，檔案中的參數會覆寫工作流程定義中偵測到的參數。

CWL、WDL 和 Nextflow 引擎的剖析邏輯之間有些微差異，如以下各節所述。

主題

- CWL 的參數偵測
- WDL 的參數偵測
- Nextflow 的參數偵測

CWL 的參數偵測

在 CWL 工作流程引擎中，剖析邏輯會做出下列假設：

- 任何 null 支援的類型都會標示為選用的輸入參數。
- 任何非 Null 支援的類型都會標示為必要的輸入參數。
- 任何具有預設值的參數都會標示為選用的輸入參數。
- 描述會從`main`工作流程定義的 `label`區段擷取。如果`label`未指定 ，則描述將為空白 （ 空字串 ）。

下表顯示 CWL 插補範例。對於每個範例，參數名稱為 x。如果需要 參數，您必須提供 參數的值。如果 參數是選用的，則不需要提供值。

此資料表顯示基本類型的 CWL 插補範例。

輸入	輸入/輸出範例	必要
`x:` `  type: int`	1 或 2 或 ...	是

`x:`
`  type: int`
`  default: 2`
	預設值為 2。有效輸入為 1 或 2 或 ...	否

`x:` `  type: int?`	有效輸入為無或 1 或 2 或 ...	否

`x:`
`  type: int?`
`  default: 2`
	預設值為 2。有效輸入為無或 1 或 2 或 ...	否

下表顯示複雜類型的 CWL 插補範例。複雜類型是基本類型的集合。

輸入	輸入/輸出範例	必要

`x:`
`  type: array`
`  items: int`
	【】 或 【1 , 2 , 3】	是

`x:`
`  type: array?`
`  items: int`
	無 或 【】 或 【1 , 2 , 3】	否

`x:`
`  type: array`
`  items: int?`
	【】 或 【無、3、無】	是

輸入	輸入/輸出範例	必要

`x:`
`  type: array?`
`  items: int?`
	【無】 或 無 或 【1，2，3】 或 【無，3】 但 【】	否

WDL 的參數偵測

在 WDL 工作流程引擎中，剖析邏輯會做出下列假設：

- 任何 null 支援的類型都會標示為選用的輸入參數。
- 對於非 Null 支援的類型：
  - 任何指派常值或表達式的輸入變數都會標示為選用參數。例如：

```
 Int x = 2
Float f0 = 1.0 + f1
```

  - 如果未將值或表達式指派給輸入參數，則會將其標記為必要參數。
- 描述是從`main`工作流程定義`parameter_meta`中的 擷取。如果`parameter_meta`未指定 ，則描述將為空白 （ 空字串 ）。如需詳細資訊，請參閱參數中繼資料的 WDL 規格。

下表顯示 WDL 插補範例。對於每個範例，參數名稱為 x。如果需要 參數，您必須提供 參數的值。如果 參數是選用的，則不需要提供值。

此資料表顯示基本類型的 WDL 插補範例。

輸入	輸入/輸出範例	必要
Int x	1 或 2 或 ...	是
Int x = 2	2	否
Int x = 1+2	3	否
Int x = y+z	y+z	否
Int？ x	無 或 1 或 2 或 ...	是

輸入	輸入/輸出範例	必要
Int？ x = 2	無或 2	否
Int？ x = 1+2	無或 3	否
Int？ x = y+z	無或 y+z	否

下表顯示複雜類型的 WDL 插補範例。複雜類型是基本類型的集合。

輸入	輸入/輸出範例	必要
Array【Int】 x	【1，2，3】 或 【】	是
Array【Int】+ x	【1】，但不是 【】	是
Array【Int】？ x	無 或 【】 或 【1，2，3】	否
Array【Int？】 x	【】 或 【無、3、無】	是
Array【Int ？】=？ x	【無】 或 無 或 【1，2，3】 或 【無，3】 但 【】	否
結構範例 {String a， Int y} 稍後在輸入中：mySample 範例	`String a = mySample.a Int y = mySample.y`	是
結構範例 {String a， Int y}	`if (defined( mySample)) {`	否

輸入	輸入/輸出範例	必要		
稍後在輸入中：Sample？mySample	`String a = mySample.a Int y = mySample.y }`			

Nextflow 的參數偵測

對於 Nextflow，HealthOmics 會剖析 `nextflow_schema.json` 檔案來產生參數範本。如果工作流程定義不包含結構描述檔案，HealthOmics 會剖析主要工作流程定義檔案。

主題

- 剖析結構描述檔案
- 剖析主要檔案
- 巢狀參數
- Nextflow 插補的範例

剖析結構描述檔案

若要讓剖析正常運作，請確定結構描述檔案符合下列要求：

- 結構描述檔案名為 `nextflow_schema.json`，且位於與主要工作流程檔案相同的目錄中。
- 結構描述檔案是有效的 JSON，如下列任一結構描述所定義：
  - https：//json-schema.org/draft/2020-12/schema。
  - https：//json-schema.org/draft-07/schema。

HealthOmics 會剖析`nextflow_schema.json`檔案以產生參數範本：

- 擷取結構描述中properties定義的所有 。
- description 如果 屬性可用，請包含 屬性。
- 根據 屬性required的 欄位，識別每個參數是否為選用或必要。

下列範例顯示定義檔案和產生的參數檔案。

```
{
    "$schema": "https://json-schema.org/draft/2020-12/schema",
    "type": "object",
    "$defs": {
        "input_options": {
            "title": "Input options",
            "type": "object",
            "required": ["input_file"],
            "properties": {
                "input_file": {
                    "type": "string",
                    "format": "file-path",
                    "pattern": "^s3://[a-z0-9.-]{3,63}(?:/\\S*)?$",
                    "description": "description for input_file"
                },
                "input_num": {
                    "type": "integer",
                    "default": 42,
                    "description": "description for input_num"
                }
            }
        },
        "output_options": {
            "title": "Output options",
            "type": "object",
            "required": ["output_dir"],
            "properties": {
                "output_dir": {
                    "type": "string",
                    "format": "file-path",
                    "description": "description for output_dir",
                }
            }
        }
    },
    "properties": {
        "ungrouped_input_bool": {
            "type": "boolean",
            "default": true
        }
    },
    "required": ["ungrouped_input_bool"],
    "allOf": [
```

```
        { "$ref": "#/$defs/input_options" },
        { "$ref": "#/$defs/output_options" }
    ]
}
```

產生的參數範本：

```
{
    "input_file": {
        "description": "description for input_file",
        "optional": False
    },
    "input_num": {
        "description": "description for input_num",
        "optional": True
    },
    "output_dir": {
        "description": "description for output_dir",
        "optional": False
    },
    "ungrouped_input_bool": {
        "description": None,
        "optional": False
    }
}
```

剖析主要檔案

如果工作流程定義不包含`nextflow_schema.json`檔案，HealthOmics 會剖析主要工作流程定義檔案。

HealthOmics 會分析主要工作流程定義檔案和 `nextflow.config` 檔案中找到的`params`表達式。`params` 具有預設值的所有 都會標示為選用。

若要讓剖析正常運作，請注意下列需求：

- HealthOmics 只會剖析主要工作流程定義檔案。為了確保擷取所有參數，建議您params將所有參數全部連接到任何子模組和匯入的工作流程。
- 組態檔案是選用的。如果您定義一個，請命名它`nextflow.config`，並將其放在與主要工作流程定義檔案相同的目錄中。

下列範例顯示定義檔案和產生的參數範本。

```
params.input_file = "default.txt"
params.threads = 4
params.memory = "8GB"

workflow {
    if (params.version) {
        println "Using version: ${params.version}"
    }
}
```

產生的參數範本：

```
{
    "input_file": {
        "description": None,
        "optional": True
    },
    "threads": {
        "description": None,
        "optional": True
    },
    "memory": {
        "description": None,
        "optional": True
    },
    "version": {
        "description": None,
        "optional": False
    }
}
```

對於 nextflow.config 中定義的預設值，HealthOmics 會收集在 中宣告的`params`指派和參數`params {}`，如下列範例所示。在指派陳述式中， `params` 必須出現在陳述式的左側。

```
params.alpha = "alpha"
params.beta = "beta"

params {
    gamma = "gamma"
    delta = "delta"
```

```
}

env {
   // ignored, as this assignment isn't in the params block
   VERSION = "TEST"
}

// ignored, as params is not on the left side
interpolated_image = "${params.cli_image}"
```

產生的參數範本：

```
{
    // other params in your main workflow defintion
    "alpha": {
        "description": None,
        "optional": True
    },
    "beta": {
        "description": None,
        "optional": True
    },
    "gamma": {
        "description": None,
        "optional": True
    },
    "delta": {
        "description": None,
        "optional": True
    }
}
```

巢狀參數

nextflow_schema.json 和 nextflow.config 允許巢狀參數。不過，HealthOmics 參數範本只需要最上層參數。如果您的工作流程使用巢狀參數，您必須提供 JSON 物件做為該參數的輸入。

結構描述檔案中的巢狀參數

HealthOmics 會在剖析nextflow_schema.json檔案params時略過巢狀。例如，如果您定義下列nextflow_schema.json檔案：

```
{
```

```
    "properties": {
        "input": {
            "properties": {
                "input_file": { ... },
                "input_num": { ... }
            }
        },
        "input_bool": { ... }
    }
}
```

HealthOmics 會在產生參數範本input_num時忽略 input_file和 ：

```
{
    "input": {
        "description": None,
        "optional": True
    },
    "input_bool": {
        "description": None,
        "optional": True
    }
}
```

當您執行此工作流程時，HealthOmics 預期會有類似下列input.json的檔案：

```
{
   "input": {
       "input_file": "s3://bucket/obj",
       "input_num": 2
   },
   "input_bool": false
}
```

組態檔案中的巢狀參數

HealthOmics 不會收集nextflow.config檔案中params的巢狀，並在剖析期間略過它們。例如，如果您定義下列nextflow.config檔案：

```
params.alpha = "alpha"
  params.nested.beta = "beta"
```

```
  params {
      gamma = "gamma"
      group {
          delta = "delta"
      }
  }
```

HealthOmics 會在產生參數範本`params.group.delta`時忽略 `params.nested.beta`和 ：

```
{
    "alpha": {
        "description": None,
        "optional": True
    },
    "gamma": {
        "description": None,
        "optional": True
    }
}
```

Nextflow 插補的範例

下表顯示主檔案中參數的 Nextflow 插補範例。

參數	必要
params.input_file	是
params.input_file = "s3：//bucket/data.json"	否
params.nested.input_file	N/A
params.nested.input_file = "s3：//bucket/data.json"	N/A

下表顯示 `nextflow.config` 檔案中參數的 Nextflow 插補範例。

參數	必要
`params.input_file = "s3://bucket/data.json"`	否
`params { input_file = "s3://bucket/data.json" }`	否
`params { nested { input_file = "s3://bucket/data.json" } }`	N/A
`input_file = params.input_file`	N/A

私有工作流程的 Amazon ECR 中的容器映像

在建立私有工作流程之前，您可以為工作流程建立容器映像。您可以將映像上傳至 Amazon Elastic Container Registry (Amazon ECR) 中的私有映像儲存庫。當您執行工作流程時，HealthOmics 服務會存取您提供的容器。

容器映像 Amazon ECR 儲存庫必須與呼叫 服務的帳戶位於相同的 AWS 區域。只要來源映像儲存庫提供適當的許可，不同的 AWS 帳戶 就可以擁有容器映像。如需詳細資訊，請參閱共用工作流程的 Amazon Elastic Container Registry 儲存庫政策。

建議您將 Amazon ECR 容器映像 URIs定義為工作流程中的參數，以便在執行開始之前驗證存取權。變更區域參數也可讓您更輕鬆地在新區域中執行工作流程。

> **Note**
>
> HealthOmics 不支援 ARM 容器，也不支援存取公有儲存庫。

如需設定 HealthOmics 存取 Amazon ECR 的 IAM 許可的詳細資訊，請參閱 資源許可。

主題

Amazon ECR 容器映像的一般考量

- 架構

  HealthOmics 支援 x86_64 容器。如果您的本機機器以 ARM 為基礎，例如 Apple Mac)，請使用如下所示的命令來建置 x86_64 容器映像：

  ```
  docker build --platform amd64 -t my_tool:latest .
  ```

- 進入點和 shell

  HealthOmics 工作流程引擎會將 bash 指令碼做為命令覆寫注入工作流程任務所使用的容器映像。因此，應該在沒有指定 ENTRYPOINT 的情況下建置容器映像，使得 bash shell 為預設值。

- 掛載路徑

  共用檔案系統掛載到位於 /tmp 的容器任務。此位置的容器映像內建的任何資料或工具都會遭到覆寫。

  透過位於 /mnt/workflow 的唯讀掛載，工作流程定義可供任務使用。

- 映像大小

  如需容器映像大小上限HealthOmics 工作流程固定大小配額，請參閱 。

HealthOmics 工作流程的環境變數

HealthOmics 提供環境變數，其中包含容器中執行之工作流程的相關資訊。您可以在工作流程任務的邏輯中使用這些變數的值。

所有 HealthOmics 工作流程變數都以 `AWS_WORKFLOW_`字首開頭。此字首是受保護的環境變數字首。請勿在工作流程容器中將此字首用於您自己的變數。

HealthOmics 提供下列工作流程環境變數：

AWS_REGION

此變數是容器執行所在的區域。

AWS_WORKFLOW_RUN

此變數是目前執行的名稱。

AWS_WORKFLOW_RUN_ID

此變數是目前執行的執行識別符。

AWS_WORKFLOW_RUN_UUID

此變數是目前執行的執行 UUID。

AWS_WORKFLOW_TASK

此變數是目前任務的名稱。

AWS_WORKFLOW_TASK_ID

此變數是目前任務的任務識別符。

AWS_WORKFLOW_TASK_UUID

此變數是目前任務的任務 UUID。

下列範例顯示每個環境變數的典型值：

```
AWS Region: us-east-1
Workflow Run: arn:aws:omics:us-east-1:123456789012:run/6470304
Workflow Run ID: 6470304
Workflow Run UUID: f4d9ed47-192e-760e-f3a8-13afedbd4937
Workflow Task: arn:aws:omics:us-east-1:123456789012:task/4192063
Workflow Task ID: 4192063
Workflow Task UUID: f0c9ed49-652c-4a38-7646-60ad835e0a2e
```

在 Amazon ECR 容器映像中使用 Java

如果工作流程任務使用 Java 應用程式，例如 GATK，請考慮容器的下列記憶體需求：

- Java 應用程式使用堆疊記憶體和堆積記憶體。根據預設，堆積記憶體上限是容器中可用記憶體總數的百分比。此預設值取決於特定的 JVM 分佈和 JVM 版本，因此請參閱 JVM 的相關文件，或使用 Java 命令列選項 （ 例如 `-Xmx`) 明確設定堆積記憶體上限。
- 請勿將堆積記憶體上限設定為容器記憶體配置的 100%，因為 JVM 堆疊也需要記憶體。JVM 垃圾收集器和容器內執行的任何其他作業系統程序也需要記憶體。
- 有些 Java 應用程式，例如 GATK，可以使用原生方法叫用或其他最佳化，例如記憶體映射檔案。這些技術需要執行「關閉堆積」的記憶體配置，這些配置不受 JVM 最大堆積參數控制。

  如果您知道 （ 或懷疑 ） Java 應用程式配置堆積外記憶體，請確定任務記憶體配置包含堆積外記憶體需求。

  如果這些堆積外配置導致容器記憶體不足，您通常不會看到 Java OutOfMemory錯誤，因為 JVM 無法控制此記憶體。

將任務輸入新增至 ECR 容器映像

將執行工作流程任務所需的所有可執行檔、程式庫和指令碼新增至用於執行任務的 Amazon ECR 映像。

最佳實務是避免使用任務容器映像外部的指令碼、二進位檔和程式庫。這在使用將`bin`目錄做為`nf-core`工作流程套件一部分的工作流程時特別重要。雖然此目錄可供工作流程任務使用，但會掛載為唯讀目錄。此目錄中的必要資源應複製到任務映像中，並在執行時間或建置用於任務的容器映像時提供。

HealthOmics 工作流程固定大小配額 如需 HealthOmics 支援的容器映像大小上限，請參閱 。

請求私有工作流程的 Sentieon 授權

如果您的私有工作流程使用 Sentieon 軟體，您需要 Senieon 授權。請依照下列步驟請求和設定 Sentieon 軟體的授權：

- 請求 Sentieon 授權
  - 傳送電子郵件至 Sentieon 支援群組 (support@sentieon.com) 以請求軟體授權。
    - 在電子郵件中提供 AWS 正式使用者 ID。
    - 請依照這些指示尋找您的 AWS 正式使用者 ID。
- 更新您的 HealthOmics 服務角色，以授予其存取您區域中的 Sentieon 授權伺服器代理和 Sentieon Omics 儲存貯體。下列範例會授予 中的存取權`us-east-1`。如有必要，請將此文字取代為您的 區域。

JSON

```
{
          "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObjectAcl",
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::omics-ap-us-east-1/*",
                "arn:aws:s3:::sentieon-omics-license-us-east-1/*"
            ]
        }
    ]
}
```

- 產生 AWS 支援案例以存取 Sentieon 授權伺服器代理。
  - 若要建立支援案例，請導覽至 https：//support.console.aws.amazon.com。
  - 在支援案例中提供您的 AWS 帳戶 和 區域。您的帳戶會新增至授權伺服器代理的允許清單。
- 使用 Sentieon 容器和 Sentieon 授權指令碼建置您的私有工作流程。
  - 如需在私有工作流程中使用 Sentieon 工具的其他說明，請參閱 GitHub 中的 Sentieon-Amazon-Omics。
- Sentieon 軟體版本 202112.07 及更高版本支援 HealthOmics 授權伺服器代理。若要使用早於 202112.07 的 Sentieon 軟體版本，請聯絡 Sentieon 支援。

HealthOmics 中的工作流程文字

建立工作流程之後，建議您先在工作流程上執行 linter，再開始第一次執行。linter 偵測到可能導致執行失敗的錯誤。

對於 WDL，HealthOmics 會在您建立工作流程時自動執行 linter。回應的 `statusMessage` 欄位提供 linter 輸出get-workflow。使用下列 CLI 命令擷取狀態輸出 （ 使用您建立之 WDL 工作流程的工作流程 ID）：

```
aws omics get-workflow
   –id 123456
```

```
  –query 'statusMessage'
```

HealthOmics 提供 linter，您可以在建立工作流程之前在工作流程棄用上執行。在您要遷移至 HealthOmics 的現有管道上執行這些文字。

- WDL – 執行 WDL linter 的公有 Amazon ECR 映像。

- Nextflow – 用於執行 Nextflow Linter 規則的公有 Amazon ECR 映像。您可以從 GitHub 存取此 linter 的原始碼。
- CWL – 無法使用

建立或更新工作流程

若要建立私有工作流程，您需要：

- Workflow definition file：以 WDL、 Nextflow或 撰寫的工作流程定義檔案CWL。工作流程定義會指定使用工作流程之執行的輸入和輸出。它還包含工作流程執行和執行任務的規格，包括運算和記憶體需求。工作流程定義檔案必須為 `.zip` 格式。如需詳細資訊，請參閱 HealthOmics 中的工作流程定義檔案。
- Parameter template file （選用）：以 寫入的參數範本檔案JSON。建立 檔案以定義執行參數，或 HealthOmics 會為您產生參數範本。如需詳細資訊，請參閱 HealthOmics 工作流程的參數範本檔案。
- Amazon ECR container images：建立工作流程的容器映像，並將其存放在私有 Amazon ECR 儲存庫中。
- Sentieon licenses （選用）：請求在私有工作流程中使用Sentieon軟體Sentieon的授權。

對於大於 4 MiB （壓縮） 的工作流程定義檔案，請在工作流程建立期間選擇下列其中一個選項：

- 上傳至 Amazon Simple Storage Service 資料夾並指定位置。
- 上傳至外部儲存庫 （大小上限為 1 GiB)，並指定儲存庫詳細資訊。

建立工作流程之後，您可以使用 `UpdateWorkflow`操作更新下列工作流程資訊：

- 名稱
- 描述

- 預設儲存類型
- 預設儲存容量 （ 使用工作流程 ID)
- README.md 檔案

若要變更工作流程中的其他資訊，請建立新的工作流程或工作流程版本。

使用工作流程版本控制來組織和建構您的工作流程。版本也可協助您管理反覆工作流程更新的簡介。如需有關版本的詳細資訊，請參閱建立工作流程版本。

主題

- 建立私有工作流程
- 更新私有工作流程
- 刪除私有工作流程
- 驗證工作流程狀態
- 從工作流程定義參考基因體檔案

建立私有工作流程

使用 HealthOmics 主控台、 AWS CLI 命令或其中一個 AWS SDKs建立工作流程。

> **Note**
>
> 請勿在工作流程名稱中包含任何個人身分識別資訊 (PII)。這些名稱會顯示在 CloudWatch 日誌中。

當您建立工作流程時，HealthOmics 會將通用唯一識別碼 (UUID) 指派給工作流程。工作流程 UUID 是全域唯一識別符 (guid)，在工作流程和工作流程版本中是唯一的。基於資料來源目的，我們建議您使用工作流程 UUID 來唯一識別工作流程。

主題

- 使用主控台建立工作流程
- 使用 CLI 建立工作流程
- 使用 SDK 建立工作流程

使用主控台建立工作流程

建立工作流程的步驟

1. 開啟 HealthOmics 主控台。
2. 選取左上方的導覽窗格 (≡)，然後選取私有工作流程。
3. 在私有工作流程頁面上，選擇建立工作流程。
4. 在定義工作流程頁面上，提供下列資訊：
    1. 工作流程名稱：此工作流程的獨特名稱。建議您設定工作流程名稱，以在 AWS HealthOmics 主控台和 CloudWatch 日誌中組織您的執行。
    2. 描述 （選用 ）：此工作流程的描述。
5. 在工作流程定義面板中，提供下列資訊：
    1. 工作流程語言 （選用 ）：選取工作流程的規格語言。否則，HealthOmics 會從工作流程定義判斷語言。
    2. 對於工作流程定義來源，選擇從 Git 型儲存庫、Amazon S3 位置或從本機磁碟機匯入定義資料夾。
        a. 對於從儲存庫服務匯入：

        > **Note**
        > HealthOmics 支援 GitHub、、GitLab、Bitbucket、 的公有和私有儲存庫GitHub self-managedGitLab self-managed。

        i. 選擇連線，將您的 AWS 資源連線至外部儲存庫。若要建立連線，請參閱 與外部程式碼儲存庫連線。

        > **Note**
        > TLV 區域中的客戶需要在 IAD(us-east-1) 區域中建立連線，才能建立工作流程。

        ii. 在完整儲存庫 ID 中，將您的儲存庫 ID 輸入為 user-name/repo-name。確認您有權存取此儲存庫中的檔案。
        iii. 在來源參考 （選用 ） 中，輸入儲存庫來源參考 （分支、標籤或遞交 ID)。如果未指定來源參考，HealthOmics 會使用預設分支。

iv. 在排除檔案模式中，輸入檔案模式以排除特定資料夾、檔案或副檔名。這有助於在匯入儲存庫檔案時管理資料大小。最多有 50 種模式，而且 patter 必須遵循 glob 模式語法。例如：

A. `tests/`

B. `*.jpeg`

C. `large_data.zip`

b. 對於從 S3 選取定義資料夾：

i. 輸入包含壓縮工作流程定義資料夾的 Amazon S3 位置。Amazon S3 儲存貯體必須與工作流程位於相同的區域。

ii. 如果您的帳戶沒有 Amazon S3 儲存貯體，請在 S3 儲存貯體擁有者 AWS 的帳戶 ID 中輸入儲存貯體擁有者的帳戶 ID。 S3 需要此資訊，以便 HealthOmics 可以驗證儲存貯體擁有權。

c. 對於從本機來源選取定義資料夾：

i. 輸入壓縮工作流程定義資料夾的本機磁碟機位置。

3. 主要工作流程定義檔案路徑 （選用）：輸入從壓縮工作流程定義資料夾或儲存庫到`main`檔案的檔案路徑。如果工作流程定義資料夾中只有一個檔案，或主要檔案名為「主要」，則不需要此參數。

6. 在 README 檔案 （選用） 面板中，提供下列資訊：

1. 選取 README 檔案的來源。

a. 對於從儲存庫服務匯入，在 README 檔案路徑中，輸入儲存庫中 README 檔案的路徑。

b. 針對從 S3 選取檔案，在 S3 的 README 檔案中，輸入 README 檔案的 Amazon S3 URI。

c. 對於從本機來源選取檔案：從本機來源在 README 檔案中，從本機來源上傳 README 檔案。

2. 在 README 檔案路徑中，輸入來源中 README 檔案的路徑。

7. 在預設執行儲存組態面板中，為使用此工作流程的執行提供預設執行儲存類型和容量：

1. 執行儲存體類型：選擇是否使用靜態或動態儲存體做為暫時執行儲存體的預設值。預設為靜態儲存。

2. 執行儲存容量 （選用）：對於靜態執行儲存類型，您可以輸入此工作流程所需的預設執行儲存量。此參數的預設值為 1200 GiB。您可以在開始執行時覆寫這些預設值。

8. 標籤 （選用）：您最多可以將 50 個標籤與此工作流程建立關聯。

9. 選擇下一步。
10. 在新增工作流程參數（選用）頁面上，選取參數來源：
    1. 對於從工作流程定義檔案剖析，HealthOmics 會自動從工作流程定義檔案剖析工作流程參數。
    2. 對於從 Git 儲存庫提供參數範本，請使用從儲存庫到參數範本檔案的路徑。
    3. 對於從本機來源選取 JSON 檔案，請從指定參數的本機來源上傳JSON檔案。
    4. 對於手動輸入工作流程參數，請手動輸入參數名稱和描述。
11. 在參數預覽面板中，您可以檢閱或變更此工作流程版本的參數。如果您還原JSON檔案，則會遺失您所做的任何本機變更。
12. 選擇下一步。
13. 檢閱工作流程組態，然後選擇建立工作流程。

使用 CLI 建立工作流程

定義工作流程和參數之後，您可以使用 CLI 建立工作流程，如下所示。

```
aws omics create-workflow  \
  --name "my_workflow"   \
  --definition-zip fileb://my-definition.zip \
  --parameter-template file://my-parameter-template.json
```

如果您的工作流程定義檔案位於 Amazon S3 資料夾中，請使用 `definition-uri` 參數輸入位置，而非 `definition-zip`。如需詳細資訊，請參閱 AWS HealthOmics API 參考中的 CreateWorkflow。

`create-workflow` 請求會以下列方式回應：

```
{
  "arn": "arn:aws:omics:us-west-2:....",
  "id": "1234567",
  "status": "CREATING",
  "tags": {
      "resourceArn": "arn:aws:omics:us-west-2:...."
  },
  "uuid": "64c9a39e-8302-cc45-0262-2ea7116d854f"
}
```

建立工作流程時要使用的選用參數

您可以在建立工作流程時指定任何選用參數。如需詳細資訊，請參閱 AWS HealthOmics API 參考中的 CreateWorkflow。

如果您包含多個工作流程定義檔案，請使用 `main` 參數來指定哪個檔案是工作流程的主要定義檔案。

如果您已將工作流程定義檔案上傳至 Amazon S3 資料夾，請使用 `definition-uri` 參數指定位置，如下列範例所示。如果您的帳戶未擁有 Amazon S3 儲存貯體，請提供擁有者的 AWS 帳戶 ID。

```
aws omics create-workflow  \
  --name Test  \
  --main multi_workflow/workflow2.wdl  \
  --definition-uri s3://omics-bucket/workflow-definition/  \
  --owner-id  123456789012  \
  --parameter-template file://params_sample_description.json
```

您可以指定預設執行儲存類型 (DYNAMIC 或 STATIC)，並執行儲存容量 （靜態儲存需要）。如需執行儲存體類型的詳細資訊，請參閱 在 HealthOmics 工作流程中執行儲存類型。

```
aws omics create-workflow  \
  --name my_workflow   \
  --definition-zip fileb://my-definition.zip \
  --parameter-template file://my-parameter-template.json   \
  --storage-type 'STATIC'  \
  --storage-capacity 1200  \
```

使用 加速器參數來建立在加速運算執行個體上執行的工作流程。下列範例示範如何使用 `--accelerators` 參數。

```
aws omics create-workflow --name workflow name \
   --definition-uri s3://amzn-s3-demo-bucket1/GPUWorkflow.zip \
   --accelerators GPU
```

使用 SDK 建立工作流程

您可以使用其中一個 SDKs建立工作流程。下列範例示範如何使用 Python SDK 建立工作流程

```
import boto3

omics = boto3.client('omics')
```

```
with open('definition.zip', 'rb') as f:
   definition = f.read()

response = omics.create_workflow(
   name='my_workflow',
   definitionZip=definition,
   parameterTemplate={ ... }
)
```

更新私有工作流程

您可以使用 HealthOmics 主控台、 AWS CLI 命令或其中一個 AWS SDKs來更新工作流程。

> **Note**
>
> 請勿在工作流程名稱中包含任何個人身分識別資訊 (PII)。這些名稱會顯示在 CloudWatch 日誌中。

主題

- 使用主控台更新工作流程
- 使用 CLI 更新工作流程
- 使用 SDK 更新工作流程

使用主控台更新工作流程

更新工作流程的步驟

1. 開啟 HealthOmics 主控台。
2. 選取左上方的導覽窗格 (≡)，然後選取私有工作流程。
3. 在私有工作流程頁面上，選擇要更新的工作流程。
4. 在工作流程頁面上：
   - 如果工作流程有版本，請確定您選取預設版本。
   - 從動作清單中選擇編輯。
5. 在編輯工作流程頁面上，您可以變更下列任何值：
   - 工作流程名稱。

- 工作流程描述。
- 工作流程的預設執行儲存類型。
- 預設執行儲存容量 （如果執行儲存類型是靜態儲存）。如需預設執行儲存組態的詳細資訊，請參閱 使用主控台建立工作流程。

6. 選擇儲存變更以套用變更。

使用 CLI 更新工作流程

如下列範例所示，您可以更新工作流程名稱和描述。您也可以變更預設執行儲存類型 (STATIC 或 DYNAMIC)，並執行儲存容量 （靜態儲存類型）。如需執行儲存類型的詳細資訊，請參閱 在 HealthOmics 工作流程中執行儲存類型。

```
aws omics update-workflow
  --id 1234567
  --name my_workflow
  --description "updated workflow"
  --storage-type 'STATIC'
  --storage-capacity 1200
```

您不會收到對update-workflow請求的回應。

使用 SDK 更新工作流程

您可以使用其中一個 SDKs更新工作流程。

下列範例示範如何使用 Python SDK 更新工作流程

```
import boto3

omics = boto3.client('omics')

response = omics.update_workflow(
   name='my_workflow',
   description='updated workflow'
)
```

刪除私有工作流程

當您不再需要工作流程時，您可以使用 HealthOmics 主控台、 AWS CLI 命令或其中一個 AWS SDKs 將其刪除。您可以刪除符合下列條件的工作流程：

- 其狀態為 ACTIVE 或 FAILED。
- 它沒有作用中的共享。
- 您已刪除所有工作流程版本。

刪除工作流程不會影響任何使用工作流程的持續執行。

主題

- 使用主控台刪除工作流程
- 使用 CLI 刪除工作流程
- 使用 SDK 刪除工作流程

使用主控台刪除工作流程

刪除工作流程

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇私有工作流程。
3. 在私有工作流程頁面上，選擇要刪除的工作流程。
4. 在工作流程頁面上，從動作清單中選擇刪除。
5. 在刪除工作流程模態中，輸入「確認」以確認刪除。
6. 選擇 刪除。

使用 CLI 刪除工作流程

下列範例示範如何使用 AWS CLI 命令來刪除工作流程。若要執行範例，*workflow id*請以您要刪除的工作流程 ID 取代 。

```
aws omics delete-workflow
  --id workflow id
```

HealthOmics 不會傳送回應給`delete-workflow`請求。

使用 SDK 刪除工作流程

您可以使用其中一個 SDKs刪除工作流程。

下列範例示範如何使用 Python SDK 刪除工作流程。

```
import boto3

omics = boto3.client('omics')

response = omics.delete_workflow(
   id='1234567'
)
```

驗證工作流程狀態

建立工作流程後，您可以使用 get-workflow 驗證狀態並檢視工作流程的其他詳細資訊，如下所示。

```
aws omics get-workflow --id 1234567
```

回應包含工作流程詳細資訊，包括狀態，如下所示。

```
{
    "arn": "arn:aws:omics:us-west-2:....",
    "creationTime": "2022-07-06T00:27:05.542459"
    "id": "1234567",
    "engine": "WDL",
    "status": "ACTIVE",
    "type": "PRIVATE",
    "main": "workflow-crambam.wdl",
    "name": "workflow_name",
    "storageType": "STATIC",
    "storageCapacity": "1200",
    "uuid": "64c9a39e-8302-cc45-0262-2ea7116d854f"
  }
```

在狀態轉換為 之後，您可以使用此工作流程開始執行ACTIVE。

從工作流程定義參考基因體檔案

HealthOmics 參考存放區物件可以使用如下的 URI 來參考。在指示*reference ID*的地方使用您自己的 *reference store ID*、 *account ID*和 。

```
omics://account ID.storage.us-west-2.amazonaws.com/reference store id/reference/id
```

有些工作流程需要參考基因體的 SOURCE和 INDEX 檔案。先前的 URI 是預設的短格式，並將預設為 SOURCE 檔案。若要指定任一檔案，您可以使用長 URI 表單，如下所示。

```
omics://account ID.storage.us-west-2.amazonaws.com/reference store id/reference/id/
source
omics://account ID.storage.us-west-2.amazonaws.com/reference store id/reference/id/
index
```

使用序列讀取集會有類似的模式，如下所示。

```
aws omics create-workflow \
     --name workflow name \
     --main sample workflow.wdl \
     --definition-uri omics://account ID.storage.us-
west-2.amazonaws.com/sequence_store_id/readSet/id \
     --parameter-template file://parameters_sample_description.json
```

有些讀取集，例如以 FASTQ 為基礎的讀取集，可以包含配對的讀取。在下列範例中，它們稱為 SOURCE1 和 SOURCE2。BAM 和 CRAM 等格式只會有 SOURCE1 檔案。有些讀取集將包含 INDEX 檔案，例如 bai或 crai 檔案。上述 URI 是預設的簡短格式，預設為 SOURCE1 檔案。若要指定確切的檔案或索引，您可以使用長 URI 表單，如下所示。

```
omics://123456789012.storage.us-west-2.amazonaws.com/<sequence_store_id>/readSet/<id>/
source1
omics://123456789012.storage.us-west-2.amazonaws.com/<sequence_store_id>/readSet/<id>/
source2
omics://123456789012.storage.us-west-2.amazonaws.com/<sequence_store_id>/readSet/<id>/
index
```

以下是使用兩個 Omics 儲存 URIs 的輸入 JSON 檔案範例。

```
{
   "input_fasta": "omics://123456789012.storage.us-west-2.amazonaws.com/
<reference_store_id>/reference/<id>",
   "input_cram": "omics://123456789012.storage.us-west-2.amazonaws.com/
<sequence_store_id>/readSet/<id>"
}
```

將 `--inputs file://<input_file.json>`新增至您的開始執行請求 AWS CLI ，以參考 中的輸入 JSON 檔案。

使用 README 檔案

上傳包含工作流程指示、圖表和必要資訊的 README.md 檔案。每個工作流程版本都支援一個 README 檔案，可以隨時更新。

README 要求包括：

- README 檔案必須是 Markdown (.md) 格式
- 檔案大小上限：500 KiB

主題

使用現有的 README

從 Git 儲存庫匯出READMEs 包含通常無法在儲存庫外部運作的相對連結。HealthOmics Git 整合會自動將這些連結轉換為絕對連結，以便在主控台中正確轉譯，無需手動更新 URL。

對於從 Amazon S3 或本機磁碟機匯入READMEs，映像和連結必須使用公有 URLs或更新其相對路徑，以進行適當的轉譯。

> **Note**
>
> 必須公開託管映像，才能在 HealthOmics 主控台中顯示。存放在 GitHub Enterprise Server或 GitLab Self-Managed 儲存庫中的影像無法轉譯。

轉譯條件

HealthOmics 主控台會使用絕對路徑插入可公開存取的影像和連結。若要從私有儲存庫轉譯 URLs，使用者必須能夠存取儲存庫。對於使用自訂網域的 GitHub Enterprise Server或 GitLab Self-Managed 儲存庫，HealthOmics 無法解析存放在這些私有儲存庫中的相對連結或轉譯映像。

下表顯示 AWS 主控台 README 檢視支援的 Markdown 元素。

Element	AWS 主控台
Alerts (提醒)	是，但沒有圖示
徽章	是
基本文字格式	是
程式碼區塊	是，但沒有語法反白和複製按鈕功能
可摺疊區段	是
標題	是
影像格式	是
影像 （ 可點選 ）	是
換行	是
美人魚圖	只有 可以開啟圖形、移動圖形位置和複製程式碼
引號	是
下標和上標	是
資料表	是，但不支援文字對齊
文字對齊	是

HealthOmics 中的工作流程版本控制

如果您需要變更工作流程，您可以建立新的工作流程或新的工作流程版本。版本不可變，但允許的組態變更不會影響執行邏輯。

工作流程版本提供下列優點：

- 版本會形成相關工作流程的邏輯群組。您可以將使用者定義的名稱新增至每個工作流程版本，以便更輕鬆地管理它們 （ 特別是具有大量版本的工作流程 ）。

- 您可以同時執行多個版本的工作流程。
- 工作流程的所有版本都會共用相同的工作流程 ID 和基本 ARN，這可以簡化修改工作流程後的管道管理。
- 工作流程版本提供與工作流程相同的資料來源層級。版本不可變，且 HealthOmics 會為每個工作流程版本建立唯一的 ARN。版本 ARN 包含工作流程 ID 和版本名稱，如下列範例所示：

  ```
  arn:aws:omics:us-west-2:123456789012:workflow/1234567/version/
  myUniqueVersionName
  ```
- 如果您擁有共享工作流程，則可以在不中斷訂閱者的情況下更新工作流程（他們可以繼續使用先前的版本）。訂閱者可以存取所有工作流程版本。如果您建立新的版本，則不需要重新共用工作流程。
- 啟動工作流程執行時，您可以指定工作流程版本。
  - 使用者可以選擇維持生產執行的穩定版本，並嘗試測試執行的最新版本。
  - 如果使用者遇到新版本的問題，則可以還原至工作流程的先前版本。
  - 共用工作流程的訂閱者可以選擇要使用的版本。

主題

預設工作流程版本

在您建立一或多個版本的工作流程後，HealthOmics 會將原始工作流程視為預設版本。當您開始執行時，您可以選擇指定執行的工作流程版本。如果您在開始執行時未指定版本，HealthOmics 會使用預設版本。

在 主控台中，HealthOmics 會使用預設版本標籤來指出原始工作流程。主控台只會在您建立一或多個工作流程版本後使用此標籤。原始工作流程一律維持預設版本。您無法將任何其他版本指派為預設值。

如果有與工作流程相關聯的其他版本，則無法刪除工作流程的預設版本。如需詳細資訊，請參閱刪除私有工作流程。

建立工作流程版本

當您建立新的工作流程版本時，您需要指定新版本的組態值。它不會從工作流程繼承任何組態值。

當您建立版本時，請提供此工作流程唯一的版本名稱。HealthOmics 建立版本後，您無法變更名稱。

版本名稱必須以字母或數字開頭，並可包含大小寫字母、數字、連字號、句點和底線。長度上限為 64 個字元。例如，您可以使用簡單的命名機制，例如 version1、Version2、Version3。您也可以將工作流程版本與您自己的內部版本控制慣例進行比對，例如 2.7.0、2.7.1、2.7.2。

或者，使用版本描述欄位來新增有關此版本的備註。例如：Fix for syntax error in workflow definition。

> **Note**
>
> 請勿在版本名稱中包含任何個人身分識別資訊 (PII)。版本名稱會出現在工作流程版本 ARN 中。

HealthOmics 會將唯一的 ARN 指派給工作流程版本。根據工作流程 ID 和版本名稱的組合，ARN 是唯一的。

> **Warning**
>
> 刪除工作流程版本後，HealthOmics 可讓您重複使用不同工作流程版本的版本名稱。最佳實務是不要重複使用版本名稱。如果您重複使用名稱，工作流程和每個版本都有唯一的 UUID，可用於原始伺服器。

主題

使用主控台建立工作流程版本

建立工作流程的步驟

1. 開啟 HealthOmics 主控台。
2. 選取左上方的導覽窗格 (≡)，然後選取私有工作流程。
3. 在私有工作流程頁面上，選擇新版本的工作流程。

4. 在工作流程詳細資訊頁面上，選擇建立新版本。
5. 在建立版本頁面上，提供下列資訊：
   1. 版本名稱：輸入工作流程中唯一工作流程版本的名稱。
   2. 版本描述（選用）：您可以使用描述欄位來新增有關此版本的備註。
6. 在工作流程定義面板中，提供下列資訊：
   1. 工作流程語言（選用）：選取工作流程版本的規格語言。否則，HealthOmics 會從工作流程定義判斷語言。
   2. 對於工作流程定義來源，選擇從 Git 型儲存庫、Amazon S3 位置或從本機磁碟機匯入定義資料夾。
      a. 對於從儲存庫服務匯入：

> **Note**
>
> HealthOmics 支援 GitHub、、GitLab、Bitbucket、 的公有和私有儲存庫GitHub self-managedGitLab self-managed。

      i. 選擇連線，將您的 AWS 資源連線至外部儲存庫。若要建立連線，請參閱 與外部程式碼儲存庫連線。

> **Note**
>
> TLV 區域中的客戶需要在 IAD(us-east-1) 區域中建立連線，才能建立工作流程。

      ii. 在完整儲存庫 ID 中，將您的儲存庫 ID 輸入為 user-name/repo-name。確認您有權存取此儲存庫中的檔案。
      iii. 在來源參考（選用）中，輸入儲存庫來源參考（分支、標籤或遞交 ID)。如果未指定來源參考，HealthOmics 會使用預設分支。
      iv. 在排除檔案模式中，輸入檔案模式以排除特定資料夾、檔案或副檔名。這有助於在匯入儲存庫檔案時管理資料大小。最多有 50 個模式，而且 修補程式必須遵循 glob 模式語法。例如：

         A. `tests/`

         B. `*.jpeg`

         C. `large_data.zip`

b. 對於從 S3 選取定義資料夾：

i. 輸入包含壓縮工作流程定義資料夾的 Amazon S3 位置。Amazon S3 儲存貯體必須與工作流程位於相同的區域。

ii. 如果您的帳戶沒有 Amazon S3 儲存貯體，請在 S3 儲存貯體擁有者 AWS 的帳戶 ID 中輸入儲存貯體擁有者的帳戶 ID。 S3 需要此資訊，以便 HealthOmics 可以驗證儲存貯體擁有權。

c. 對於從本機來源選取定義資料夾：

i. 輸入壓縮工作流程定義資料夾的本機磁碟機位置。

3. 主要工作流程定義檔案路徑 （選用）：輸入從壓縮工作流程定義資料夾或儲存庫到`main`檔案的檔案路徑。如果工作流程定義資料夾中只有一個檔案，或主要檔案名為「主要」，則不需要此參數。

7. 在預設執行儲存組態面板中，為使用此工作流程的執行提供預設執行儲存類型和容量：

1. 執行儲存體類型：選擇是否使用靜態或動態儲存體做為暫時執行儲存體的預設值。預設為靜態儲存。

2. 執行儲存容量 （選用）：對於靜態執行儲存類型，您可以輸入此工作流程所需的預設執行儲存量。此參數的預設值為 1200 GiB。您可以在開始執行時覆寫這些預設值。

8. 標籤 （選用）：您最多可以將 50 個標籤與此工作流程版本建立關聯。

9. 選擇下一步。

10. 在新增工作流程參數 （選用） 頁面上，選取參數來源：

1. 對於從工作流程定義檔案剖析，HealthOmics 會自動從工作流程定義檔案剖析工作流程參數。

2. 對於從 Git 儲存庫提供參數範本，請使用從儲存庫到參數範本檔案的路徑。

3. 對於從本機來源選取 JSON 檔案，請從指定參數的本機來源上傳JSON檔案。

4. 對於手動輸入工作流程參數，手動輸入參數名稱和描述。

11. 在參數預覽面板中，您可以檢閱或變更此工作流程版本的參數。如果您還原JSON檔案，則會遺失您所做的任何本機變更。

12. 選擇下一步。

13. 檢閱版本組態，然後選擇建立版本。

建立版本時，主控台會返回工作流程詳細資訊頁面，並在工作流程和版本資料表中顯示新版本。

使用 CLI 建立工作流程版本

您可以使用 `CreateWorkflowVersion` API 操作建立工作流程版本。針對選用參數，HealthOmics 會使用下列預設值：

參數	預設
引擎	從工作流程定義決定
儲存體類型	STATIC
儲存容量（用於靜態儲存）	1200 GiB
主要	根據工作流程定義資料夾的內容決定。如需詳細資訊，請參閱HealthOmics 工作流程定義需求。
加速器	無
標籤	無

下列 CLI 範例會建立工作流程版本，並以靜態儲存做為預設執行儲存：

```
aws omics create-workflow-version \
--workflow-id 1234567  \
--version-name "my_version" \
--engine WDL \
--definition-zip fileb://workflow-crambam.zip \
--description "my version description" \
--main file://workflow-params.json \
--parameter-template file://workflow-params.json \
--storage-type='STATIC'   \
--storage-capacity 1200   \
--tags example123=string  \
--accelerators GPU
```

如果您的工作流程定義檔案位於 Amazon S3 資料夾，請使用 `definition-uri` 參數輸入位置，而非 `definition-zip`。如需詳細資訊，請參閱 AWS HealthOmics API 參考中的 CreateWorkflowVersion。

您會收到以下對`create-workflow-version`請求的回應。

```
{
  "workflowId": "1234567",
  "versionName": "my_version",
  "arn": "arn:aws:omics:us-west-2:123456789012:workflow/1234567/version/3",
  "status": "ACTIVE",
  "tags": {
    "environment": "production",
    "owner": "team-alpha"
  },
  "uuid": "0ac9a563-355c-fc7a-1b47-a115167af8a2"
}
```

使用 SDK 建立工作流程版本

您可以使用其中一個 SDKs建立工作流程。

下列範例示範如何使用 Python SDK 建立工作流程版本

```
import boto3

omics = boto3.client('omics')

with open('definition.zip', 'rb') as f:
   definition = f.read()

response = omics.create_workflow_version(
   workflowId='1234567',
   versionName='my_version',
   requestId='my_request_1'
   definitionZip=definition,
   parameterTemplate={ ... }
)
```

驗證工作流程版本的狀態

建立工作流程版本後，您可以使用 get-workflow-version 驗證狀態並檢視工作流程的其他詳細資訊，如下所示。

```
aws omics get-workflow-version
--workflow-id 9876543
--version-name "my_version"
```

回應會提供您工作流程的詳細資訊，包括狀態，如下所示。

```
{
  "workflowId": "1234567",
  "versionName": "3.0.0",
  "arn": "arn:aws:omics:us-west-2:123456789012:workflow/1234567/version/3.0.0",
  "status": "ACTIVE",
  "description": ...
  "uuid": "0ac9a563-355c-fc7a-1b47-a115167af8a2"
}
```

在使用此工作流程版本開始執行之前，狀態必須轉換為 ACTIVE。

更新工作流程版本

您可以更新私有工作流程版本的描述和預設執行儲存組態。若要變更工作流程版本中的任何其他資訊，請建立新的版本。

主題

- 使用主控台更新工作流程版本
- 使用 CLI 更新工作流程版本
- 使用 SDK 更新工作流程版本

使用主控台更新工作流程版本

更新工作流程

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇私有工作流程。
3. 在私有工作流程頁面上，選擇工作流程。
4. 在工作流程頁面上，選擇要更新的工作流程版本，然後從動作清單中選擇編輯。
   - 如果您選擇預設版本，主控台會開啟編輯工作流程頁面。如需詳細資訊，請參閱更新私有工作流程。
   - 如果您選擇使用者定義的版本，主控台會開啟編輯版本頁面。
5. 在編輯版本頁面上，提供下列資訊
   - 版本描述 （選用） - 此版本的描述。

6. 在預設執行儲存組態面板中，為使用此工作流程版本的執行提供下列預設值。您可以在開始執行時覆寫預設值：
   - 針對執行儲存類型，選取靜態或動態。
   - 針對靜態執行儲存，選取使用此工作流程版本之執行的預設執行儲存容量。此參數的預設值為 1200 GiB。
7. 選擇儲存變更。

主控台會返回工作流程詳細資訊頁面，並顯示具有更新工作流程版本的頁面橫幅。

使用 CLI 更新工作流程版本

您可以使用下列 CLI 命令更新工作流程版本的參數。工作流程 ID 和版本名稱的組合可唯一識別版本。

```
aws omics update-workflow-version
--workflow-id 1234567
--version-name "my_version"
--storage-type 'STATIC'
--storage-capacity 2400
--description "version description"
```

您不會收到對update-workflow-version請求的回應。

使用 SDK 更新工作流程版本

您可以使用其中一個 SDKs更新工作流程版本。下列 python SDK 範例示範如何更新工作流程版本的儲存類型和描述。

```
import boto3

omics = boto3.client('omics')

response = omics.update_workflow_version(
   workflowID=1234567,
   versionName='3.0.0',
   storageType='DYNAMIC',
   description='new version description'
)
```

刪除工作流程版本

您可以使用主控台、CLI 或其中一個 SDKs 來刪除使用者定義的工作流程版本。刪除工作流程版本不會影響任何使用工作流程版本的持續執行。

您無法刪除 預設工作流程版本。您刪除所有使用者定義的版本，然後刪除工作流程。

主題

- 使用主控台刪除工作流程版本
- 使用 CLI 刪除工作流程版本
- 使用 SDK 刪除工作流程版本

使用主控台刪除工作流程版本

刪除工作流程版本

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇私有工作流程。
3. 在私有工作流程頁面上，選擇工作流程。
4. 在工作流程頁面上，選擇要刪除的工作流程版本，然後從動作清單中選擇刪除。
5. 在刪除工作流程版本模態中，輸入「確認」以確認刪除。
6. 選擇 刪除。

主控台會顯示頁面橫幅，其中包含已刪除的工作流程版本。

使用 CLI 刪除工作流程版本

您可以使用下列 CLI 命令刪除使用者定義的工作流程版本。工作流程 ID 和版本名稱的組合可唯一識別版本。

```
aws omics delete-workflow-version
--workflow-id 9876543
--version-name "my_version"
```

您不會收到對`delete-workflow-version`請求的回應。

使用 SDK 刪除工作流程版本

您可以使用其中一個 SDKs刪除工作流程。

下列範例示範如何使用 Python SDK 刪除工作流程。

```
import boto3

omics = boto3.client('omics')

response = omics.delete_workflow_version(
   workflowID=1234567,
   versionName='3.0.0'
)
```

開始 HealthOmics 執行

建立工作流程之後，您可以使用工作流程開始執行。

當您開始執行時，HealthOmics 會配置暫時執行儲存體，供工作流程引擎在執行期間使用。為了確保資料隔離和安全性，HealthOmics 會在每次執行開始時佈建儲存體，並在執行結束時取消佈建儲存體。

HealthOmics 提供與工作流程執行和任務相關的數個配額。預設值是刻意保守的，可協助您避免意外的成本超支。您可以請求提高這些配額。如需詳細資訊，請參閱HealthOmics 服務配額。

當您開始執行時，HealthOmics 會指派執行 ID 和執行 Uuid 給執行。帳戶中的執行具有唯一的執行 IDs。不過，HealthOmics 會重複使用已刪除IDs，因此執行和已刪除的執行可以有相同的執行 ID。此外，共用工作流程的執行 ID 與您帳戶中的執行 ID 相同，很少見。

run uuid 是全域唯一識別符 (guid)，可用來識別跨 帳戶的執行，或區分您帳戶中具有相同執行 ID 的兩個執行。

> **Note**
>
> 基於資料來源目的，我們建議您使用 run uuid 來唯一識別執行。run uuid 也是連結至內部實驗室資訊管理系統 (LIMs) 或範例追蹤系統的最佳識別符。

主題

- 在 HealthOmics 工作流程中執行儲存類型

- HealthOmics 執行的執行保留模式
- HealthOmics 執行輸入
- 在 HealthOmics 中開始執行
- 在 HealthOmics 工作流程中執行生命週期
- HealthOmics 執行輸出
- 執行失敗原因
- HealthOmics 執行中的任務生命週期
- 私有 HealthOmics 工作流程的執行最佳化

在 HealthOmics 工作流程中執行儲存類型

當您開始執行時，HealthOmics 會配置暫時執行儲存體，供工作流程引擎在執行期間使用。HealthOmics 提供暫時執行儲存做為檔案系統。

對於指定的工作流程或工作流程執行，您可以選擇動態或靜態執行儲存。根據預設，HealthOmics 會提供靜態執行儲存體。

> **Note**
>
> 執行儲存體用量會對您的帳戶產生費用。如需靜態和動態執行儲存的定價資訊，請參閱 HealthOmics 定價。

下列各節提供決定要使用的執行儲存類型時所要考慮的資訊。

動態執行儲存

我們建議對大多數執行使用動態執行儲存，包括需要更快開始時間的執行、您事先不知道儲存需求的執行，以及反覆開發測試週期。

您不需要預估執行所需的儲存體或輸送量。HealthOmics 會根據執行期間的檔案系統使用率，動態擴展或縮減儲存體大小。HealthOmics 也會根據工作流程的需求動態擴展輸送量。執行永遠不會因為檔案系統錯誤的儲存空間不足而失敗。

動態執行儲存提供比靜態執行儲存更快的佈建/取消佈建時間。更快速的設定是大多數工作流程的優點，也是開發/測試週期期間的優點。

執行完成後（成功路徑或失敗路徑），getRun API 操作會在 storageCapacity 欄位中傳回執行所使用的最大儲存體。您也可以在位於日誌群組的執行資訊omics清單日誌中找到此資訊。對於在 2 小時內完成的動態儲存執行，最大儲存值可能無法使用。

對於動態執行儲存，執行會佈建使用 NFS 通訊協定的檔案系統。NFS 將 CREATE、DELETE 和 RENAME 檔案操作視為非等冪性，有時可能會導致程式碼必須正常處理的這些操作的競爭條件。例如，如果程式碼嘗試刪除不存在的檔案，則不應失敗。採用動態執行儲存之前，建議您調整工作流程程式碼，使其對非等冪檔案操作具有彈性。請參閱 安全處理非等冪操作的程式碼範例。

安全處理非等冪操作的程式碼範例

下列 python 範例示範如何在檔案不存在時刪除檔案而不失敗。

```
import os
import errno

def remove_file(file_path):
    try:
        os.remove(file_path)
    except OSError as e:
        # If the error is "No such file or directory", ignore it (or log it)
        if e.errno != errno.ENOENT:
            # Otherwise, raise the error
            raise

# Example usage
remove_file("myfile")
```

下列範例使用 Bash shell。若要安全地移除檔案，即使檔案不存在，請使用：

```
rm -f my_file
```

若要安全地移動（重新命名）檔案，只有在檔案old_name存在於目前目錄中時，才執行移動命令。

```
[ -f old_name ] && mv old_name new_name
```

若要建立目錄，請使用下列命令：

```
mkdir -p mydir/subdir/
```

靜態執行儲存

對於靜態執行儲存，執行會佈建使用 Lustre 通訊協定的檔案系統。根據預設，此通訊協定對非等冪檔案操作具有彈性。您不需要調整工作流程程式碼來處理非等冪檔案操作。

HealthOmics 會配置固定數量的執行儲存體。您可以在開始執行時指定此值。如果您未指定值，預設執行儲存體為 1200 GiB。當您在 StartRun API 請求中指定儲存體大小的值時，系統會將該值四捨五入到最接近的 1200 GiB 倍數。如果該儲存體大小不可用，則會四捨五入至最接近的 2400 GiB 倍數。

對於靜態執行儲存，HealthOmics 會佈建下列輸送量值：

- 每個 TiB 佈建儲存容量的基準輸送量為 200 MB/s。
- 每個佈建 TiB 儲存容量的爆量輸送量高達 1300 MB/s。

如果指定的儲存體大小太低，則執行會失敗，並因檔案系統錯誤導致儲存不足。靜態執行儲存非常適合具有已知儲存需求的可預測工作流程。

靜態執行儲存體適用於具有高任務並行的大型高載工作負載（例如，平行處理的大量 RNASeq 樣本）。它提供比動態執行儲存更高的每個 GiB 檔案系統輸送量和更低的每 GiB 成本。

計算所需的靜態執行儲存體

當工作流程使用靜態執行儲存體（相較於動態執行儲存體）時，需要額外的容量，因為基本檔案系統安裝會使用 7% 的靜態檔案系統容量。

如果您執行動態執行儲存工作流程來測量執行使用的最大儲存體，請使用下列計算來判斷所需的靜態儲存體數量下限：

```
static storage required =
       maximum storage in GiB used by the dynamic run storage
         + (total static file system size in GiB * 0.07)
```

例如：

```
    Maximum storage measured from a dynamic run storage workflow run:  500GiB
      File system size: 1200GiB
      7% of the file system size:  84GiB
      500 + 84 = 584GiB of static run storage required for this run.
```

因此，1200GiB（靜態執行儲存體的最小容量）足以進行此執行。

HealthOmics 執行的執行保留模式

執行完成後，HealthOmics 會將執行中繼資料封存至 CloudWatch。根據預設，除非您變更 CloudWatch 保留政策，否則 CloudWatch 會無限期保留執行資料。執行輸出也會儲存在 Amazon S3 中，直到您將其刪除為止。

其中一個可調整的 HealthOmics 服務配額 是區域中maximum number of runs (active and inactive)的 。HealthOmics 會將執行中繼資料保留最多此數量的執行，以供主控台和 API 操作 (ListRuns 和 GetRun) 使用。當您開始執行時，您可以設定執行保留模式參數，以指出執行的保留行為。參數支援值 REMOVE 和 RETAIN。

對於將保留模式設定為 REMOVE 的新執行，如果 HealthOmics 嘗試在儲存執行數量上限之後新增執行，則會自動移除已設定 REMOVE 模式之最舊執行的中繼資料。此移除不會影響存放在 CloudWatch 或 Amazon S3 中的資料。

RETAIN 是執行保留模式的預設值。對於在此模式下執行的 ，系統不會刪除執行中繼資料。如果 HealthOmics 達到執行次數上限，全部都設定為 RETAIN，您將無法建立其他執行，直到您刪除一些執行為止。

如果您打算同時執行超過執行數量上限的批次，請務必將執行保留模式設定為 REMOVE。否則，當 HealthOmics 嘗試在最大值之後啟動下一次執行時，批次會失敗。

使用 REMOVE 保留模式的其他考量事項：

- 當您第一次開始使用 REMOVE 做為保留模式時，請考慮刪除一或多個使用 RETAIN 模式的執行，以釋放插槽。當您開始額外的 REMOVE 執行時，自動移除會接管，因此有足夠的插槽可供新執行使用。
- 如果您想要重新執行封存的執行（或一組執行），請使用 HealthOmics 重新執行 CLI 工具。如需如何使用此工具的詳細資訊和範例，請參閱 HealthOmics 工具 GitHub 儲存庫中的 Omics 重新執行。HealthOmics GitHub
- 我們建議您為每個執行設定唯一的名稱。HealthOmics 移除執行後，您就無法使用主控台或 API 來尋找執行名稱或執行 ID。不過，您可以使用 CloudWatch 來搜尋執行名稱，因此請使用唯一名稱來取得最佳的搜尋結果。
- 您可以使用 CloudWatch start-query命令來取得封存執行的相關資訊。如果執行名稱不是唯一的，查詢可能會傳回多個資訊清單。開始時間和結束時間參數會定義搜尋的時間範圍。

```
aws logs start-query  \
```

```
    --log-group-name "/aws/omics/WorkflowLog" \
    --query-string 'filter @logStream like "manifest" and @message like "myRunName"'
\
    --end-time <END-EPOCH-TIME> --start-time <START-EPOCH-TIME>
```

start-query 命令會傳回查詢 ID。將查詢 ID 傳遞至 get-query-results命令會傳回查詢結果。

```
aws logs get-query-results --query-id QueryId
```

HealthOmics 執行輸入

如果工作流程定義指定工作流程或工作流程任務的輸入檔案，HealthOmics 會將檔案分階段到工作流程執行專用的暫存磁碟區。這些輸入檔案為唯讀，可防止任務修改工作流程中其他任務的潛在輸入。對於目錄匯入，目錄也是唯讀的。

許多基因體應用程式假設索引檔案與序列檔案（例如bam檔案的配套bai檔案）位於同一位置。若要包含索引檔案，請在工作流程定義中將其指定為任務輸入。

主題

- 管理執行參數大小
- Amazon S3 輸入參數格式
- Amazon S3 輸入封存狀態

管理執行參數大小

當您開始執行時，您可以在執行參數 JSON 物件或檔案中指定執行輸入。您可以為工作流程指定最多 50 KB 的執行參數。您可以使用下列技術來維持在此大小限制內：

- 使用目錄匯入

  若要指定大量輸入檔案，請指定一個參數做為包含所有檔案的 Amazon S3 位置，而不是為每個檔案位置指定參數。如需詳細資訊，請參閱下一個主題 (Amazon S3 輸入參數格式）。
- 使用範例工作表

  範例工作表是 CSV 或 TSV 檔案，其中包含一個用於 fastq.gz 地址的資料欄（或兩個用於配對讀取），以及範例名稱等中繼資料的其他資料欄。您可以將範例工作表指定為執行輸入參數，而不是每個輸入檔案的參數。

您的工作流程會定義範例工作表如何映射至工作流程中的資料結構。雖然您可以在 WDL 和 CWL 中撰寫範例工作表的程式碼，但它們在 NextFlow 中更為常見。如需範例，請參閱 nf-core GitHub 網站上的範例工作表。

Amazon S3 輸入參數格式

對於接受 Amazon S3 位置的輸入參數， 參數可以指定一個檔案或整個檔案目錄的位置。使用目錄有下列優點：

- 便利性 – 您可以將目錄名稱指定為 參數。您不會列出每個檔案名稱。
- 精簡性 – 輸入參數檔案大小上限為 50 KB。如果您提供輸入檔案名稱的長清單，則可以超過此上限。

Amazon S3 是平面物件儲存系統，因此不支援目錄。您可以將檔案分組為「目錄」，方法是為每個檔案提供相同的物件金鑰字首。如需 Amazon S3 物件金鑰字首的詳細資訊，請參閱使用字首組織物件。

HealthOmics 會解譯輸入參數值，如下所示：

- 如果 Amazon S3 位置結尾不是正斜線或使用 glob 模式，HealthOmics 預期參數值是某個 Amazon S3 物件的索引鍵。

  例如，您可以指定 `s3://myfiles/runs/inputs/a/file1.fastq`輸入 file1.fastq
- 如果 Amazon S3 位置以正斜線結尾，HealthOmics 會將參數值解譯為 Amazon S3 字首。它會載入具有該字首的所有 Amazon S3 物件。

  例如，您可以指定 `s3://myfiles/runs/inputs/a/`載入其金鑰以此字首開頭的所有物件。
- 對於 Nextflow，HealthOmics 支援輸入參數中 Amazon S3 URIs 的 glob 模式。

  例如，您可以指定 `“s3://myfiles/runs/inputs/a/*.gz”`輸入其金鑰以此字首開頭的所有 .gz 檔案。

Amazon S3 輸入中雙斜線的語言特定處理

HealthOmics 會在 Amazon S3 URIs 中處理雙斜線時保留每個工作流程引擎的原生引擎行為，因此您在將工作流程遷移至 HealthOmics 時不需要對工作流程進行任何變更。下列各節說明每個引擎如何處理各種案例。

WDL

如果輸入參數在 URI 中間或結尾包含雙斜線，則 WDL 引擎會保留雙斜線。

輸入參數	預期位置	
s3：//myfiles/runs/inputs//file1.fastq	s3：//myfiles/runs/inputs//file1.fastq	
s3：//myfiles/runs/inputs//	s3：//myfiles/runs/inputs//	

下一個流程

如果輸入參數在 URI 中間包含雙斜線，則 Nextflow 引擎會保留雙斜線。對於 URI 結尾的雙斜線，Nextflow 引擎會將其解析為單一斜線。

輸入參數	預期位置	
s3：//myfiles/runs/inputs//file1.fastq	s3：//myfiles/runs/inputs//file1.fastq	
s3：//myfiles//runs/inputs//*.gz	s3：//myfiles//runs/inputs//*.gz	
s3：//myfiles//runs/inputs//	s3：//myfiles//runs/inputs/	

CWL

如果輸入參數在 URI 中間或結尾包含雙斜線，CWL 引擎會保留雙斜線。

輸入參數	預期位置	
s3：//myfiles//runs/inputs//file1.fastq	s3：//myfiles//runs/inputs//file1.fastq	
s3：//myfiles//runs/inputs//	s3：//myfiles//runs/inputs//	

Amazon S3 輸入封存狀態

HealthOmics 可以即時擷取 Amazon S3 S3 物件。對於處於下列封存儲存狀態的物件，要提供給 HealthOmics 的restore物件：

- Amazon S3 Glacier 中的 Flexible Retrieval 或 Deep Archive 儲存類別。
- 智慧型分層中的封存存取或 Deep Archive Access 層。

如需有關還原物件的資訊，請參閱《Amazon S3 使用者指南》中的還原封存的物件。

在 HealthOmics 中開始執行

開始執行時，您可以設定執行儲存類型和儲存量（用於靜態儲存）。如需其他資訊，請參閱 在 HealthOmics 工作流程中執行儲存類型。

您也可以設定執行優先順序。優先順序如何影響執行取決於執行是否與執行群組相關聯。如需其他資訊，請參閱 執行優先順序。

如果您已建立一或多個工作流程版本，您可以在開始執行時指定版本。如果您未指定版本，HealthOmics 會啟動預設工作流程版本。

指定輸出檔案的 Amazon S3 位置。如果您同時執行大量工作流程，請針對每個工作流程使用個別的 Amazon S3 輸出 URIs，以避免儲存貯體調節。如需詳細資訊，請參閱 Amazon S3 使用者指南中的使用字首組織物件，以及最佳化 Amazon S3 效能白皮書中的水平擴展儲存連線。

> **Note**
>
> 您可以在開始執行時指定 IAM 服務角色。或者，主控台可以為您建立服務角色。如需詳細資訊，請參閱的服務角色 AWS HealthOmics。

主題

HealthOmics 執行參數

當您開始執行時，您可以在執行參數 JSON 檔案中指定執行輸入，也可以內嵌輸入參數值。如需管理執行參數 JSON 檔案大小的資訊，請參閱 管理執行參數大小。

HealthOmics 支援下列參數值的 JSON 類型。

JSON 類型	範例索引鍵和值	備註
boolean	"b"：true	值不在引號中，且全部小寫。
integer	「i」：7	值不在引號中。
number	"f"：42.3	值不在引號中。
string	"s"："字元"	值以引號表示。針對文字值和 URIs使用字串類型。URI 目標必須是預期的輸入類型。
陣列	"a"：【1，2，3】	值不在引號中。陣列成員必須各自具有輸入參數定義的類型。
object	"o"：{"left"："a"， "right"：1}	在 WDL 中，物件映射至 WDL 配對、映射或結構

使用主控台啟動執行

啟動工作流程執行

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇執行。
3. 在執行頁面上，選擇開始執行。
4. 在執行詳細資訊面板中，提供下列資訊
   - 工作流程來源 - 選擇擁有的工作流程或共享工作流程。
   - 工作流程 ID - 與此執行相關聯的工作流程 ID。

- 工作流程版本 （選用） - 選取要用於此執行的工作流程版本。如果您未選取版本，則執行會使用工作流程預設版本。
- 執行名稱 - 此執行的獨特名稱。
- 執行優先順序 （選用） - 此執行的優先順序。較高的數字會指定較高的優先順序，而最高優先順序的任務會先執行。
- 執行儲存類型 - 在此指定儲存類型，以覆寫為工作流程指定的預設執行儲存類型。靜態儲存會配置執行的固定儲存量。動態儲存會視需要擴展和縮減執行中每個任務的規模。
- 執行儲存容量 - 針對靜態執行儲存，指定執行所需的儲存量。此項目會覆寫為工作流程指定的預設執行儲存量。
- 選取 S3 輸出目的地 - 將儲存執行輸出的 S3 位置。
- 輸出儲存貯體擁有者的帳戶 ID （選用） - 如果您的帳戶不擁有輸出儲存貯體，請輸入儲存貯體擁有者的 AWS 帳戶 ID。需要此資訊，以便 HealthOmics 可以驗證儲存貯體擁有權。
- 執行中繼資料保留模式 - 選擇是否要保留所有執行的中繼資料，或在您的帳戶達到執行次數上限時，讓系統移除最舊的執行中繼資料。如需詳細資訊，請參閱HealthOmics 執行的執行保留模式。

5. 在服務角色下，您可以使用現有的服務角色或建立新的服務角色。
6. （選用） 對於標籤，您最多可以將 50 個標籤指派給執行。
7. 選擇下一步。
8. 在新增參數值頁面上，提供執行參數。您可以上傳指定參數的 JSON 檔案，或手動輸入值。
9. 選擇下一步。
10. 在執行群組面板中，您可以選擇指定此執行的執行群組。如需詳細資訊，請參閱建立 HealthOmics 執行群組。
11. 在執行快取面板中，您可以選擇為此執行指定執行快取。如需詳細資訊，請參閱使用主控台設定具有執行快取的執行。
12. 選擇 Review and start run (檢閱並開始執行)。
13. 檢閱執行組態後，選擇開始執行。

使用 API 開始執行

使用啟動執行 API 操作搭配您建立的 IAM 角色和 Amazon S3 儲存貯體。此範例會將保留模式設定為 REMOVE。如需保留模式的詳細資訊，請參閱 HealthOmics 執行的執行保留模式。

```
aws omics start-run
```

```
     --workflow-id workflow id \
     --role-arn arn:aws:iam::1234567892012:role/service-role/
OmicsWorkflow-20221004T164236 \
     --name workflow name \
     --retention-mode REMOVE
```

為了回應，您會取得下列輸出。對執行uuid是唯一的，並且 與 outputUri可用於追蹤輸出資料寫入的位置。

```
{
    "arn": "arn:aws:omics:us-west-2:....:run/1234567",
    "id": "123456789",
    "uuid":"96c57683-74bf-9d6d-ae7e-f09b097db14a",
    "outputUri":"s3://bucket/folder/8405154/96c57683-74bf-9d6d-ae7e-f09b097db14a"
    "status": "PENDING"
}
```

如果工作流程的參數範本宣告任何必要的參數，您可以在啟動工作流程執行時提供輸入的本機 JSON 檔案。JSON 檔案包含每個輸入參數的確切名稱，以及 參數的值。

將 `--parameters file://<input_file.json>`新增至您的`start-run`請求 AWS CLI ，以參考中的輸入 JSON 檔案。如需執行參數的詳細資訊，請參閱 HealthOmics 執行輸入。

您可以指定執行的工作流程版本。

```
aws omics start-run
     --workflow-id workflow id \
      ...
     --workflow-version-name '1.2.1'
```

您可以覆寫工作流程中指定的預設執行儲存體類型。

```
aws omics start-run
     --workflow-id workflow id \
      ...
     --storage-type STATIC
     --storage-capacity 2400
```

您也可以使用啟動 API 搭配 GPU 工作流程 ID，如下所示。

```
aws omics start-run
```

```
     --workflow-id workflow id \
     --role-arn arn:aws:iam::1234567892012:role/service-role/
OmicsWorkflow-20221004T164236 \
     --name GPUTestRunModel \
     --output-uri s3://amzn-s3-demo-bucket1
```

取得工作流程執行的相關資訊

您可以在回應中使用 ID 搭配 get-run API 來檢查執行的狀態，如下所示。

```
aws omics get-run --id run id
```

此 API 操作的回應會告訴您工作流程執行的狀態。可能的狀態為 PENDING、RUNNING、 STARTING和 COMPLETED。當執行為 時COMPLETED，您可以在輸出 Amazon S3 儲存貯outfile.txt體中找到名為 的輸出檔案，位於以執行 ID 命名的資料夾中。

get-run API 操作也會傳回其他詳細資訊，例如工作流程是 Ready2Run還是 PRIVATE、工作流程引擎和加速器詳細資訊。下列範例顯示私有工作流程執行的 get-run 回應，如使用 GPU 加速器且未指派給執行的標籤的 WDL 中所述。

```
{
    "arn": "arn:aws:omics:us-west-2:123456789012:run/7830534",
    "id": "7830534",
    "uuid":"96c57683-74bf-9d6d-ae7e-f09b097db14a",
    "outputUri":"s3://bucket/folder/8405154/96c57683-74bf-9d6d-ae7e-f09b097db14a"
    "status": "COMPLETED",
    "workflowId": "4074992",
    "workflowType": "PRIVATE",
    "workflowVersionName": "3.0.0",
    "roleArn": "arn:aws:iam::123456789012:role/service-role/
OmicsWorkflow-20221004T164236",
    "name": "RunGroupMaxGpuTest",
    "runGroupId": "9938959",
    "digest":
 "sha256:a23a6fc54040d36784206234c02147302ab8658bed89860a86976048f6cad5ac",
    "accelerators": "GPU",
    "outputUri": "s3://amzn-s3-demo-bucket1",
    "startedBy": "arn:aws:sts::123456789012:assumed-role/Admin/<role_name>",
    "creationTime": "2023-04-07T16:44:22.262471+00:00",
    "startTime": "2023-04-07T16:56:12.504000+00:00",
    "stopTime": "2023-04-07T17:22:29.908813+00:00",
    "tags": {}
```

```
}
```

您可以使用 list-runs API 操作查看所有執行的狀態，如下所示。

```
aws omics list-runs
```

若要查看特定執行完成的所有任務，請使用 list-run-tasks API。

```
aws omics list-run-tasks --id task ID
```

若要取得任何特定任務的詳細資訊，請使用 get-run-task API。

```
aws omics get-run-task --id <run_id> --task-id task ID
```

執行完成後，中繼資料會傳送至串流 下的 CloudWatch**manifest/run/<run ID>/<run UUID>**。

以下是資訊清單的範例。

```
{
    "arn": "arn:aws:omics:us-east-1:123456789012:run/1695324",
    "creationTime": "2022-08-24T19:53:55.284Z",
    "resourceDigests": {
      "s3://omics-data/broad-references/hg38/v0/Homo_sapiens_assembly38.dict":
 "etag:3884c62eb0e53fa92459ed9bff133ae6",
      "s3://omics-data/broad-references/hg38/v0/Homo_sapiens_assembly38.fasta":
 "etag:e307d81c605fb91b7720a08f00276842-388",
      "s3://omics-data/broad-references/hg38/v0/Homo_sapiens_assembly38.fasta.fai":
 "etag:f76371b113734a56cde236bc0372de0a",
      "s3://omics-data/intervals/hg38-mjs-whole-chr.500M.intervals":
 "etag:27fdd1341246896721ec49a46a575334",
      "s3://omics-data/workflow-input-lists/dragen-gvcf-list.txt":
 "etag:e22f5aeed0b350a66696d8ffae453227"
    },
    "digest":
 "sha256:a5baaff84dd54085eb03f78766b0a367e93439486bc3f67de42bb38b93304964",
    "engine": "WDL",
    "main": "gatk4-basic-joint-genotyping-v2.wdl",
    "name": "1044-gvcfs",
    "outputUri": "s3://omics-data/workflow-output",
    "parameters": {
      "callset_name": "cohort",
      "input_gvcf_uris": "s3://omics-data/workflow-input-lists/dragen-gvcf-list.txt",
```

```
      "interval_list": "s3://omics-data/intervals/hg38-mjs-whole-chr.500M.intervals",
      "ref_dict": "s3://omics-data/broad-references/hg38/v0/
Homo_sapiens_assembly38.dict",
      "ref_fasta": "s3://omics-data/broad-references/hg38/v0/
Homo_sapiens_assembly38.fasta",
      "ref_fasta_index": "s3://omics-data/broad-references/hg38/v0/
Homo_sapiens_assembly38.fasta.fai"
    },
    "roleArn": "arn:aws:iam::123456789012:role/OmicsServiceRole",
    "startedBy": "arn:aws:sts::123456789012:assumed-role/admin/ahenroid-Isengard",
    "startTime": "2022-08-24T20:08:22.582Z",
    "status": "COMPLETED",
    "stopTime": "2022-08-24T20:08:22.582Z",
    "storageCapacity": 9600,
    "uuid": "a3b0ca7e-9597-4ecc-94a4-6ed45481aeab",
    "workflow": "arn:aws:omics:us-east-1:123456789012:workflow/1558364",
    "workflowType": "PRIVATE"
  },
  {
    "arn": "arn:aws:omics:us-east-1:123456789012:task/1245938",
    "cpus": 16,
    "creationTime": "2022-08-24T20:06:32.971290",
    "image": "123456789012.dkr.ecr.us-west-2.amazonaws.com/gatk",
    "imageDigest":
 "sha256:8051adab0ff725e7e9c2af5997680346f3c3799b2df3785dd51d4abdd3da747b",
    "memory": 32,
    "name": "geno-123",
    "run": "arn:aws:omics:us-east-1:123456789012:run/1695324",
    "startTime": "2022-08-24T20:08:22.278Z",
    "status": "SUCCESS",
    "stopTime": "2022-08-24T20:08:22.278Z",
    "uuid": "44c1a30a-4eee-426d-88ea-1af403858f76"
  },
  ...
```


如果 CloudWatch 日誌中沒有執行中繼資料，則不會將其刪除。您也可以使用執行 ID，使用 CLI 工具重新執行工作流程執行。進一步了解並從 HealthOmics Tool GitHub 儲存庫下載工具。

重新執行工作流程執行

下列範例示範如何使用 rerun工具重新執行執行。您需要可從 CloudWatch 日誌擷取的執行 ID。

```
omics-rerun 9876543 --name workflow name --retention-mode REMOVE
```

如果執行存在於 CloudWatch 中，您會收到類似以下的回應。

```
Original request:
{
  "workflowId": "9679729",
  "roleArn": "arn:aws:iam::123456789012:role/DemoRole",
  "name": "sample_rerun",
  "parameters": {
    "image": "123456789012.dkr.ecr.us-west-2.amazonaws.com/default:latest",
    "file1": "omics://123456789012.storage.us-west-2.amazonaws.com/8647780323/
readSet/6389608538"
  },
  "outputUri": "s3://workflow-output-bcf2fcb1"
}
StartRun request:
{
  "workflowId": "9679729",
  "roleArn": "arn:aws:iam::123456789012:role/DemoRole",
  "name": "new test",
  "parameters": {
    "image": "123456789012.dkr.ecr.us-west-2.amazonaws.com/default:latest",
    "file1": "omics://123456789012.storage.us-west-2.amazonaws.com/8647780323/
readSet/6389608538"
  },
  "outputUri": "s3://workflow-output-bcf2fcb1"
}
StartRun response:
{
  "arn": "arn:aws:omics:us-west-2:123456789012:run/9171779",
  "id": "9171779",
  "status": "PENDING",
  "tags": {}
}
```

如果工作流程不再存在，您會收到錯誤訊息。

在 HealthOmics 工作流程中執行生命週期

您可以透過監控執行狀態來追蹤執行的進度。HealthOmics 會在執行進行生命週期時更新執行狀態。

您可以使用下列任一方法擷取執行狀態：

- HealthOmics 主控台會在Runs頁面上顯示每個執行的狀態。

- GetRun API 操作會傳回目前的執行狀態。
- 您可以使用 EventBridge 事件來監控執行狀態。如需詳細資訊，請參閱搭配 使用 EventBridge AWS HealthOmics。

主題

執行狀態值

當您開始執行時，HealthOmics 會將執行狀態設定為 Pending。隨著執行的生命週期進行，HealthOmics 會更新狀態值，以反映其目前的進度。

> **Note**
>
> 您在執行中以外的任何執行狀態期間不會產生費用。如需詳細資訊，請參閱下節。

HealthOmics 支援下列執行狀態值：

待定

執行在佇列中，正在等待啟動。執行通常會在啟動前維持在待定狀態一小段時間。

- 如果您同時提交多個任務，則執行可以保留在待定中更長的時間。
- 在您的 帳戶達到並行執行數目上限後，執行會保持在待定狀態。
- 如果執行是達到其任何資源最大值的執行群組的一部分，則執行會保持在待定中。
- 您可以調整執行優先順序，讓特定排入佇列的執行在其他人之前開始。如需執行優先順序的詳細資訊，請參閱 執行優先順序。

啟動

HealthOmics 會建立執行並佈建執行所需的資源 （例如暫時執行儲存體和引擎節點）。

- HealthOmics 會在執行開始時佈建暫時執行儲存體，並在執行停止時取消佈建執行儲存體。

執行中

在匯入程序、每個任務的處理和匯出程序期間，執行會保持在執行中狀態。

- HealthOmics 會將輸入檔案匯入臨時執行儲存檔案系統。輸入檔案是唯讀的，可防止任務修改工作流程中其他任務的輸入。
- 在檔案匯出期間，HealthOmics 會將輸出檔案從執行儲存檔案系統匯出至 S3 位置。
- HealthOmics 會在執行狀態執行時，將執行日誌和任務日誌即時交付至 CloudWatch。如需詳細資訊，請參閱CloudWatch 中的日誌 。

正在停止

匯出程序完成後，執行會轉換為停止狀態。

- HealthOmics 會取消佈建所有資源 （包括執行儲存檔案系統和引擎節點）。

已完成

HealthOmics 完成資源取消佈建後，執行會轉換為已完成。

- HealthOmics 已完成所有執行任務，並無錯誤地匯出輸出資料。
- 執行輸出可在指定的 Amazon S3 URI 輸出位置使用。對於 WDL 和 CWL，HealthOmics 會產生執行輸出摘要檔案，提供 的相關資訊HealthOmics 執行輸出。
- 最終執行資訊清單日誌和引擎日誌 （如適用） 可在 CloudWatch 中使用。
- 對於支援任務重試的執行，狀態為已完成的執行可以包含一或多個失敗的任務。只要每個失敗的任務都重試成功，HealthOmics 就會將執行轉換為已完成。HealthOmics 會為每個重試指派新的任務 ID，因此執行會包含失敗嘗試和完成嘗試的任務 IDs。

失敗

HealthOmics 遇到一或多個錯誤，無法完成所有執行任務。

- 失敗的執行會在 HealthOmics 取消佈建資源時轉換到停止狀態。

已取消

使用者啟動取消執行的請求。

- HealthOmics 會停止任何執行中的任務，並取消佈建所有資源。
- 當使用者取消執行時，HealthOmics 不會匯出任何執行輸出資料。您無法存取任何已取消執行的中繼檔案。
- 您的帳戶會針對取消之前在執行狀態期間執行的任務和資源產生費用。
- 如果您取消處於待定或開始狀態的執行，則不會產生任何費用。

任務重試

如果任務在執行期間失敗，HealthOmics 會在下列情況下再次重試任務：

- 對於 WDL 工作流程，當任務因服務錯誤 (5XX HTTP 狀態碼） 失敗時，HealthOmics 支援任務重試。

  根據預設，HealthOmics 最多會嘗試兩次失敗任務的重試。您可以設定 WDL 定義檔案來選擇退出任務重試。如需範例組態，請參閱 HealthOmics 工作流程定義中的任務資源。
- 對於 Nextflow 工作流程，您可以在工作流程棄用中設定任務的重試條件。
- 如果執行中的每個任務最終都完成，即使他們需要重試，HealthOmics 也會將執行轉換為已完成。
- HealthOmics 會為每個重試指派新的任務 ID，因此執行會包含失敗嘗試和完成嘗試的任務 IDs。

執行狀態的定價影響

當執行狀態為執行中時，您的帳戶可能會產生費用。您在任何其他執行狀態期間不會產生費用。例如，當執行為開始或停止時，不會收取資源費用。

具有執行中狀態的執行具有下列計費影響：

- 當執行狀態為執行時，您的帳戶會產生執行儲存檔案系統用量的費用。如需執行儲存體類型的相關資訊，請參閱 在 HealthOmics 工作流程中執行儲存類型。
- 您的帳戶會根據您在工作流程定義中為每個任務指定的運算和記憶體資源，以及任務持續時間，支付執行任務的費用。如需詳細資訊，請參閱HealthOmics 任務的運算和記憶體需求。
- 每個任務的帳單閾值下限為一分鐘。如果您執行任務的時間少於一分鐘，則至少需要支付一分鐘用量的費用。如果可能，請將小型任務分組在一起以最佳化成本。分組任務也會避免啟動多個循序任務，以縮短執行時間。

如需 HealthOmics 定價的其他資訊，請參閱 HealthOmics 定價。

HealthOmics 執行輸出

當 WDL 或 CWL 執行完成時，輸出會包含輸出摘要檔案 (JSON 格式），其中會列出執行所產生的所有輸出。您可以將輸出摘要檔案用於下列目的：

- 以程式設計方式判斷執行產生的輸出檔案。
- 驗證執行是否產生所有預期的輸出。

主題

- 執行 WDL 的輸出摘要

- 執行 CWL 的輸出摘要

執行 WDL 的輸出摘要

當 WDL 執行完成時，HealthOmics 會建立名為 的輸出摘要檔案output.json。

對於工作流程的每個輸出，檔案中都有對應的索引鍵/值對。金鑰包含工作流程名稱和輸出名稱，格式如下：`WorkflowName.output_name`。對於檔案輸出，值是指向儲存檔案之 S3 中輸出位置的 S3 URI。對於 Array【File】 輸出，值是 S3 URIs的陣列。

下列範例顯示名為 之工作流程output.json的檔案BWAMappingWorkflow。

```
{
  "BWAMappingWorkflow.bam_indexes": [
    "s3://omics-outputs/8886192/out/bam_indexes/0/
pbmc8k_S1_L007_R1_001.sorted.bam.bai",
    "s3://omics-outputs/8886192/out/bam_indexes/1/pbmc8k_S1_L008_R1_001.sorted.bam.bai"
  ],
  "BWAMappingWorkflow.mapping_stats": "s3://omics-outputs/8886192/out/mapping_stats/
genome_mapping_final_stats.txt",
  "BWAMappingWorkflow.merged_bam": "s3://omics-outputs/8886192/out/merged_bam/
genome_mapping.merged.bam",
  "BWAMappingWorkflow.merged_bam_index": "s3://omics-outputs/8886192/out/
merged_bam_index/genome_mapping.merged.bam.bai",
  "BWAMappingWorkflow.reference_index_tar": "s3://omics-outputs/8886192/out/
reference_index_tar/reference_index.tar",
  "BWAMappingWorkflow.sorted_bams": [
    "s3://omics-outputs/8886192/out/sorted_bams/0/pbmc8k_S1_L007_R1_001.sorted.bam",
    "s3://omics-outputs/8886192/out/sorted_bams/1/pbmc8k_S1_L008_R1_001.sorted.bam"
  ],
  "BWAMappingWorkflow.unmapped_bams": [
    "s3://omics-outputs/8886192/out/unmapped_bams/0/
pbmc8k_S1_L007_R1_001.unmapped.bam",
    "s3://omics-outputs/8886192/out/unmapped_bams/1/pbmc8k_S1_L008_R1_001.unmapped.bam"
  ]
}
```

如果工作流程使用非檔案類型 （ 例如 String、Int、Float 或 Bool) 產生輸出，則欄位值為 JSON 基本值。例如：

```
{
```

```
  "MyWorkflow.my_int_ouput": 1,
  "MyWorkflow.my_bool_output": false,
    ...
}
```

執行 CWL 的輸出摘要

當 CWL 執行完成時，HealthOmics outputs.json會在下列位置建立名為 的輸出摘要檔案：

```
{my-S3outputpath}/{runId}/{run-uuid}/logs/outputs.json
```

輸出摘要檔案包含輸出清單。每個輸出都是金鑰/值對，其中金鑰是輸出的名稱。值是包含下列屬性的物件：

- location – 輸出檔案的完整路徑
- basename – 路徑的檔案名稱部分
- class – 輸出的類型，通常是檔案
- size – 檔案大小，以位元組為單位

在下列範例中， output.json 檔案具有兩個輸出檔案的清單。

```
{
  "example_output": {
    "location": "{my-S3outputpath}/{runId}/{run-uuid}/out/output.txt",
    "basename": "output.txt",
    "class": "File",
    "size": 13
  },
  "another_output": {
    "location": "{my-S3outputpath}/{runId}/{run-uuid}/out/metrics.json",
    "basename": "metrics.json",
    "class": "File",
    "size": 256
  }
}
```

執行失敗原因

如果執行失敗，請使用 GetRun API 操作來擷取失敗原因。

檢閱失敗原因，以協助您疑難排解執行失敗的原因。下表列出每個失敗原因以及錯誤的說明。

故障原因	錯誤說明
ASSUME_ROLE_FAILED	HealthOmics 沒有擔任該角色的許可。在角色的信任關係中指定 HealthOmics 委託人。
CANNOT_START_CONTAINER_ERROR	無法啟動工作流程任務：*##*、ID：使用映像的 *ID* 容器：*####*。請確定映像有效，然後再試一次。
CANNOT_START_CONTAINER_SIZE_ERROR	無法啟動工作流程任務：*##*、ID：使用映像的 *ID* 容器：*####*。請確定影像大小小於 25 GB，然後再試一次。
ECR_PERMISSION_ERROR	HealthOmics 沒有存取映像 URI 的許可。確認 Amazon ECR 私有儲存庫存在，並已授予 HealthOmics 服務主體的存取權。
EXPORT_FAILED	匯出失敗。檢查輸出儲存貯體是否存在，且執行角色具有對儲存貯體的寫入許可。
FILE_SYSTEM_OUT_OF_SPACE	檔案系統沒有足夠的空間。增加檔案系統大小並再次執行。
IMAGE_VERIFICATION_FAILURE	無法驗證*####*。若要修正此問題，請嘗試提取映像，然後再次將其推送到您的 ECR 儲存庫。
IMPORT_FAILED	匯入失敗。檢查輸入檔案是否存在，且執行角色可以存取輸入。
INACTIVE_OMICS_STORAGE_RESOURCE	HealthOmics 儲存 URI 未處於 ACTIVE 狀態。啟用讀取集，然後再試一次。若要進一步了解如何啟用讀取集，請參閱 在 HealthOmics 中啟用讀取集。
INPUT_URI_NOT_FOUND	提供的 URI 不存在：*uri*。檢查 URI 路徑是否存在，並確認角色可以存取物件。
INSTANCE_RESERVATION_FAILED	沒有足夠的執行個體容量來完成工作流程執行。等待並再次嘗試工作流程執行。
INVALID_ECR_IMAGE_URI	Amazon ECR 映像 URI 結構無效。請提供有效的 URI，然後再試一次。

故障原因	錯誤說明
INVALID_TASK_RESOURCE_VALUE	請求的 GPU、CPU 或記憶體對於可用的運算容量太高，或小於任務 *ID* 的最小值 1。
INVALID_URI_INPUT	URI 結構不是有效的 *uri*。請檢查 URI 結構，然後再試一次。
MODIFIED_INPUT_RESOURCE	提供的 URI *uri* 已在執行開始後修改。重試執行。
OUT_OF_MEMORY_ERROR	工作流程任務 *ID* 記憶體不足。增加工作流程定義中的記憶體值，然後再次嘗試執行。
RUN_TASK_FAILED	執行失敗，因為任務失敗。若要偵錯任務失敗，請使用 GetRunTask API 操作和 Amazon CloudWatch Logs 串流。
RUN_TIMED_OUT	在*###*後執行逾時。
SERVICE_ERROR	服務中發生暫時性錯誤。再次嘗試工作流程執行。
UNSUPPORTED_INPUT_SIZE	總輸入大小太高。減少輸入大小，然後再試一次。
WORKFLOW_RUN_FAILED	工作流程執行失敗。檢閱 CloudWatch Logs 引擎日誌串流：用於偵錯失敗的 *ID*。
WORKFLOW_VER_VALIDATION_FAILED	HealthOmics 不支援請求的 Nextflow 版本：*##* --。最新的支援版本是 *##*。將您的 Nextflow 版本修改為支援的版本，然後再試一次。
UNSUPPORTED_GPU_INSTANCE_TYPE	*##*不支援請求的執行個體類型。使用此區域支援的 GPU 執行個體類型重試執行。可用的執行個體類型為 *GPU ######*。

無回應執行的指引

開發新的工作流程時，如果程式碼發生問題，執行或特定任務可能會變成「停滯」或「停止」，而且任務無法正常結束程序。這對於故障診斷和捕捉可能具有挑戰性，因為任務長時間執行是正常的。若要防止和識別無回應的執行，請遵循以下章節中建議的最佳實務。

防止無回應執行的最佳實務

- 請確定您正在關閉任務程式碼中開啟的所有檔案。開啟太多檔案偶爾會導致工作流程引擎中的執行緒問題。
- 工作流程任務建立的背景程序應在任務結束時結束。不過，如果背景程序無法順利結束，您必須在任務程式碼中明確關閉該程序。
- 確保您的程序不會在沒有結束的情況下循環。這可能會導致無回應的執行，並且需要變更工作流程定義程式碼才能解決。
- 為您的任務提供適當的記憶體和 CPU 配置。分析 CloudWatch 日誌或在成功完成的工作流程執行執行分析器時使用 ，以確認您擁有最佳的運算配置。使用 Run Analyzer `headroom` 參數來包含額外的空間，確保程序有足夠的資源可完成。在配置的記憶體和 CPU 中包含至少 5% 的預留空間，以考慮背景作業系統程序。
  - 此外，如果執行個體需要更高的輸送量，請增加執行個體頻寬大小。少於 16 個 vCPUs Amazon EC2 執行個體可能會發生輸送量暴增。如需 Amazon EC2 執行個體輸送量的詳細資訊，請參閱 Amazon EC2 可用的執行個體頻寬。
- 請確定您在執行時使用正確的檔案系統大小。對於使用靜態執行儲存的無回應執行，請考慮增加靜態執行儲存配置，以在檔案系統上啟用更高的 IO 輸送量和儲存容量。分析執行資訊清單以查看最大檔案系統儲存，使用 Run Analyzer 判斷是否需要增加檔案系統配置。

擷取無回應執行的最佳實務

- 開發新的工作流程時，請使用已設定執行時間上限的執行群組來擷取失控程式碼。例如，如果執行需要 1 小時才能完成，請將其放在 2 或 3 小時 （或根據您的使用案例的不同期間） 後逾時的執行群組中，以擷取失控任務。此外，套用緩衝以考量處理時間的差異。
- 設定一系列具有不同最大執行時間限制的執行群組。例如，您可以將短執行指派給在幾小時後終止執行的執行群組，以及根據預期的工作流程持續時間，在幾天後終止執行的長執行群組。
- HealthOmics 的預設執行持續時間服務限制上限為 604，800 秒或 7 天，可透過配額工具中的請求進行調整。只有在您有一週內接近的執行時，才請求提高此配額的服務限制。如果您有短執行和長執行的混合，且未使用執行群組，請考慮將長執行執行放在具有較高執行持續時間服務限制上限的個別帳戶中。
- 檢查 CloudWatch 日誌是否有您懷疑可能沒有回應的任務。如果任務通常會輸出一般日誌陳述式，並且長時間沒有這樣做，則任務可能會卡住或凍結。

如果您遇到無回應的執行該怎麼辦

- 取消執行以避免產生額外費用。
- 檢查任務日誌，以檢查是否有任何程序無法正確結束。
- 檢查引擎日誌以識別任何異常的引擎行為。
- 比較無回應執行的任務和引擎日誌與相同、成功完成的執行。這有助於識別可能導致無回應行為的任何差異。
- 如果您無法判斷根本原因，請提出支援案例並包含下列項目：
  - 停滯執行的 ARN 和成功完成的相同執行的 ARN。
  - 引擎日誌（一旦執行已取消或失敗，即可使用）
  - 無回應任務的任務日誌。我們不需要工作流程中所有任務的任務日誌，即可進行故障診斷。

HealthOmics 執行中的任務生命週期

任務是執行中的單一程序。HealthOmics 會將工作流程中的每個任務映射至最適合任務所需資源的模擬運算執行個體類型。您可以在工作流程定義中指定所需的資源。如需詳細資訊，請參閱 HealthOmics 任務的運算和記憶體需求。

HealthOmics 為要使用的任務提供暫時執行儲存體。HealthOmics 會將任務輸入檔案複製到臨時執行儲存體，做為唯讀檔案。HealthOmics 提供符號連結，讓任務可以從工作目錄中存取輸入檔案。任務只能存取您在工作流程定義檔案中宣告的檔案。

任務狀態值

您可以透過監控任務狀態來追蹤任務的進度。當您開始執行時，HealthOmics Pending 會將執行中每個任務的任務狀態設定為 。當任務開始並經過生命週期時，HealthOmics 會更新狀態值，以反映其目前的進度。

您可以使用下列任一方法擷取任務狀態：

- HealthOmics 主控台會在 Run details 頁面上顯示執行中每個任務的狀態。
- GetRunTask API 操作會傳回任務狀態。
- 您可以使用 EventBridge 事件來監控任務狀態。如需詳細資訊，請參閱搭配 使用 EventBridge AWS HealthOmics。

您可以使用 GetRunTask API 操作擷取任務的目前狀態。HealthOmics 主控台會在 Run details 頁面上顯示執行中每個任務的狀態。

HealthOmics 支援下列任務狀態值：

待定

您的任務正在佇列中，正在等待啟動。任務在啟動前會短暫處於待定狀態。

- 在您的帳戶達到並行任務的數量上限之後，任務會保留在待定狀態。
- 如果執行是達到其任何資源最大值的執行群組的一部分，則任務會保持在待定狀態。
- 您可以調整執行優先順序，以便在其他排入佇列的執行之前，特定排入佇列的執行及其任務開始。如需執行優先順序的詳細資訊，請參閱 執行優先順序

啟動

HealthOmics 正在建立任務並佈建任務所需的資源，例如工作流程任務節點。

執行中

HealthOmics 正在處理任務時，任務狀態正在執行。

正在停止

完成任務處理並匯出輸出資料後，任務會轉換為停止。

- HealthOmics 會取消佈建工作流程任務節點。

已完成

HealthOmics 已完成處理任務，並將輸出資料傳輸到執行儲存檔案系統。

失敗

HealthOmics 在處理任務時發生錯誤，而且尚未完成。

- 任務會轉換為停止狀態 (HealthOmics 取消佈建資源），然後轉換為失敗狀態。
- 如果錯誤是服務錯誤 (5XX HTTP 狀態碼），且工作流程支援此任務的重試，則 HealthOmics 會嘗試再次處理任務。HealthOmics 會指派新的任務 ID 給重試。

已取消

HealthOmics 會在使用者啟動取消執行的請求後停止任務。

- 任務會轉換為停止狀態 (HealthOmics 取消佈建資源），然後轉換為已取消狀態。

對工作流程任務進行故障診斷

以下是疑難排解任務的最佳實務和考量事項。

- 任務日誌依賴STDOUT任務並STDERR產生任務。如果任務中使用的應用程式未產生其中任何一個，則不會有任務日誌。若要協助偵錯，請在 verbose 模式下使用應用程式。
- 若要檢視任務中執行的命令及其插補值，請使用 set -x Bash 命令。這有助於判斷任務是否使用正確的輸入，並識別錯誤可能讓任務無法如預期般執行的位置。
- 使用 echo命令將變數的值輸出到 STDOUT或 STDERR。這可協助您確認它們已如預期般設定。
- 使用 之類的命令來ls -l <name_of_input_file>確認輸入是否存在，且具有預期的大小。如果不是，這可能會顯示先前任務因為錯誤而產生空輸出的問題。
- 在任務指令碼df -Ph . | awk 'NR==2 {print $4}'中使用 命令來判斷任務目前可用的空間，並協助識別您可能需要使用其他儲存配置執行工作流程的情況。

在任務指令碼中包含上述任何命令，會假設任務容器也包含這些命令，而且它們位於容器環境path的上。

私有 HealthOmics 工作流程的執行最佳化

您可以針對總成本、總執行時間或兩者的組合來最佳化執行。HealthOmics 提供資料和工具，協助您做出執行最佳化決策。執行最佳化不適用於 Ready2Run 工作流程，因為您無法控制服務如何管理這些工作流程的資源佈建。

第一步是了解執行中任務目前的任務資源用量和成本，然後套用方法來最佳化執行成本和效能。

主題

執行分析器

HealthOmics 提供名為 Run Analyzer 的開放原始碼工具。此工具會擷取執行的任務層級資源用量資訊，並建議成本和執行效能的最佳化機會。

> **Note**
>
> 執行分析器會根據執行工具時的 AWS 清單價格，預估任務成本和潛在的成本節省。評估最佳化建議，並實作對您的使用案例有意義的建議。測試您採用的最佳化，以確保它們適用於您的執行。

Run Analyzer 會執行下列任務：

- 評估記憶體和運算瓶頸。
- 識別記憶體或 CPU 過度佈建的任務，並建議可降低成本的新執行個體大小。
- 計算個別任務的成本預估，並計算套用建議的潛在成本節省。
- 為您提供任務的時間軸檢視，以便您可以驗證任務相依性和處理序列。時間軸也可協助您識別長時間執行的任務。
- 提供有關執行儲存體之檔案系統大小的建議。
- 顯示任務佈建時間，讓您可以識別大型容器負載可能減慢佈建時間的區域。
- 此工具包含輸入參數（標題），可用來控制最佳化建議的積極性。

下列各節包含使用 Run Analyzer 最佳化執行的特定建議。

判斷執行成本

您可以使用下列方法和準則來判斷執行成本：

- 若要檢視計費期間的總執行成本，請遵循下列步驟：
  1. 開啟帳單和成本管理主控台，然後選擇帳單。
  2. 在依服務收費中，展開 Omic。
  3. 展開區域，然後檢視依 omics 執行個體類型、執行儲存類型和 Ready2Run 工作流程逐項列出的所有執行成本。
- 若要產生包含每次執行資訊的成本報告，請遵循下列步驟：
  1. 開啟 Billing and Cost Management 主控台，然後選擇資料匯出。
  2. 選擇建立以建立新的資料匯出。
  3. 輸入資料匯出的匯出名稱。將其他欄位保留在其預設值，以建立 CUR（成本和用量）報告。
  4. 針對時間精細程度，選取每小時或每天。

5. 在資料匯出儲存設定下，執行下列組態步驟：

   a. 設定資料匯出的 Amazon S3 儲存貯體。

   b. 針對檔案版本控制，選取是否覆寫現有的匯出檔案，或每次建立新檔案。

   系統會在接下來 24 小時內產生第一份報告，每天產生一次後續報告。

6. 如需如何建立資料匯出的詳細資訊，請參閱《資料匯出使用者指南》中的建立AWS 資料匯出。

- 您可以標記您的執行，以依類別監控和最佳化成本，例如依團隊或依專案。如果您使用標籤，請依照下列步驟，依標籤類別檢視執行成本：

  1. 開啟 Billing and Cost Management 主控台，然後選擇 Cost Explorer。
  2. 在報告參數 > 分組依據中，選擇標籤做為維度。 然後選取所需的標籤名稱。

- 若要查看任務的資源用量，請在 CloudWatch 中檢視執行資訊清單日誌。如需詳細資訊，請參閱使用 CloudWatch Logs 監控 HealthOmics 。
- 使用 執行分析器工具擷取執行的任務資源用量資訊。

判斷執行時間用量

您可以使用下列方法來協助您調查執行時間用量：

- 從 主控台的執行頁面，您可以檢視執行的總執行時間。
- 從執行詳細資訊頁面，您可以檢視下列項目：
  - 檢視執行的總執行時間。
  - 檢視執行中每個任務的執行時間。
  - 選擇其中一個連結來檢視 Amazon S3 中的日誌，或在 CloudWatch 中檢視執行日誌或執行資訊清單日誌。
- 從執行任務清單中，選擇任務的檢視日誌連結，以在 CloudWatch 中檢視任務日誌。
- `listRuns` API 操作的回應包含執行開始時間和停止時間，因此您可以計算總執行時間。
- 執行分析器 工具會在時間軸檢視上顯示任務持續時間。此工具提供任務處理序列的視覺化呈現，您可以符合預期的順序。

最佳化執行的方法

HealthOmics 會自動佈建、管理和最佳化執行資料預備的資源（例如資料匯入和資料匯出）。HealthOmics 也會啟動並執行工作流程的工作流程引擎。不過，您可以透過設定各種執行組態來影響執行開始時間、任務開始時間和整體任務執行時間。您對工作流程定義和設計的整體方法也會影響任務執行時間。下列清單說明可能影響執行和任務效能的因素：

執行儲存體類型

執行儲存類型會影響執行效能和執行佈建時間。動態執行儲存佈建更快速且永遠不會耗盡記憶體，因為它會根據您的執行儲存需求動態擴展。動態執行儲存體也非常適合開發中的工作流程，您通常可以啟動和停止工作流程來疑難排解問題。

靜態執行儲存需要較長的檔案系統佈建時間，但通常在執行具有高任務並行或需要大於 9.6 TiB 的檔案系統容量時，可以更快地完成某些執行。靜態執行儲存非常適合具有高 I/O 需求的長時間執行工作流程。

為了協助您評估特定執行之每個執行儲存類型的成本與效能，您可以嘗試 A/B 測試，以查看哪些執行儲存類型可提供更好的效能。此外，請考慮在開發週期使用動態執行儲存體，然後使用靜態執行儲存體進行大規模的生產執行。

如需執行儲存體類型的詳細資訊 在 HealthOmics 工作流程中執行儲存類型

過度佈建的執行靜態儲存

如果您的工作流程任務運算受到 I/O 限制，請考慮過度佈建靜態執行儲存。儲存成本會隨著其大小而增加，但檔案系統的最大輸送量也會增加。如果昂貴的運算任務遇到 I/O 瓶頸，增加檔案系統大小以減少任務執行時間可能會降低整體成本。

減少容器映像大小

當每個任務開始時，HealthOmics 會載入您為任務指定的容器。較大的容器需要更長的時間才能載入。將容器最佳化為盡可能小，以改善啟動新任務的效率。如果您將大型資料集新增至容器，請考慮將資料集儲存在 S3 中，並讓工作流程從 S3 匯入資料。如需 HealthOmics 支援的最大容器大小，請參閱 HealthOmics 工作流程固定大小配額。

任務大小

您可以將小型的循序任務合併為單一任務，以節省任務佈建時間。此外，HealthOmics 有一分鐘的最低任務持續時間費用，因此合併任務可能會降低成本。在合併任務中，您可以使用 Unix 管道來避免序列化和還原序列化檔案的 I/O 成本。

檔案壓縮

避免過度壓縮工作流程中繼檔案。大多數基因體格式使用「gzip」或「區塊 gzip」壓縮。解壓縮任務輸入檔案並重新壓縮任務輸出檔案，可能會耗用大量整體任務 CPU 用量。有些基因體應用程式可讓您在序列化輸出時設定壓縮層級。透過減少壓縮層級，您可以縮短 CPU 時間，但較大的檔案會增加寫入磁碟所花費的時間。根據任務和應用程式，您可以找到導致最短執行時間的中繼檔案的最佳壓縮層級。我們建議您先以具有最大輸出檔案的任務為目標。壓縮層級 2 適用於多種案例。您可以針對您的使用案例從此層級開始，並透過嘗試其他壓縮層級來比較結果。

執行緒計數

如果您在任務定義中指定執行緒，請將執行緒數目設定為與請求vCPUs 數目相同的值。

指定運算和記憶體

如果您未在任務中指定記憶體或運算資源，HealthOmics 會將最小執行個體類型 (`omics.c.large`) 指派為預設 。如果您希望 HealthOmics 指派更大的執行個體類型，請明確宣告您的記憶體和運算需求。

HealthOmics 會配置您請求vCPUs、記憶體和 GPU 資源數量。例如，如果您要求 15vCPUs和 33GiB，HealthOmics 會為您的任務配置 omics.m.4xl 執行個體 (16vCPUs64GB)，但您的任務只能使用 15 個 vCPUs和 33GiB。因此，我們建議您請求 vCPUs和符合 omics 執行個體的記憶體資源。

將多個範例批次處理為一次執行

由於檔案系統佈建在執行開始時需要時間，因此您可以將多個範例批次處理到相同的執行中，以節省佈建時間。在決定此方法之前，請考慮下列因素：

- 單一錯誤範例可能會導致工作流程失敗，因此批次處理範例可能會增加失敗的工作流程數目。如果您不確定您的工作流程在大部分時間都會成功，則每個範例執行一次可能是更好的方法。
- HealthOmics 會為整個工作流程配置一個執行儲存檔案系統。對於一批樣本，請務必指定足夠大的執行儲存量來處理所有樣本。
- 每個工作流程有最大數量的執行儲存，因此 可能會限制您可以新增至批次的樣本數量。
- 最小執行儲存體大小為 1.2 TiB，因此如果工作流程使用的儲存體遠低於每個範例的最低儲存體，批次處理可能會降低成本。
- 執行儲存可以處理多個同時連線，因此使用相同執行儲存體的多個任務不應造成 I/O 瓶頸。
- 每個執行都有自己的一組標籤。如果您使用預算或追蹤的資訊標記工作流程，最好使用不同的執行。
- IAM 角色適用於整個執行。每個使用者都可以存取批次範例的所有資料。分離工作流程可讓您使用更精細的許可。

- HealthOmics 會設定工作流程中並行工作流程數量上限和並行任務數量上限的帳戶層級配額。如需如何請求提高這些配額的詳細資訊，請參閱 HealthOmics 服務配額。

使用容器映像的參數

參數化您的容器映像，而不是在工作流程中內嵌其 URIs。它們是執行參數，HealthOmics 會在執行開始之前驗證執行是否可存取您的容器。否則，當您針對任何已完成的任務產生費用時，任務會在執行期間失敗。此外，由於這些是參數化輸入，HealthOmics 會在執行資訊清單中產生檢查總和，以改善執行來源。

使用 linter

在執行新的工作流程之前，使用 linter 尋找常見的工作流程錯誤。如需詳細資訊，請參閱HealthOmics 中的工作流程文字。

使用 EventBridge 標記問題

使用 EventBridge 自訂警示來擷取業務邏輯特有的異常。

使用序列存放區

考慮為您的來源資料使用序列存放區，以節省儲存成本。如需詳細資訊，請參閱使用 HealthOmics 部落格文章，以經濟實惠的方式在任何規模存放 omics 資料。

執行之間檔案大小差異的影響

使用者通常會使用一小組測試資料來設計和測試執行，然後在生產執行中遇到具有顯著檔案大小差異的各種資料。當您最佳化執行時，請務必考慮此差異。

下列清單說明檔案大小有顯著差異的最佳化建議：

測試資料中的不同檔案大小

嘗試在開發期間使用具有代表性變異量的測試資料。

使用 Run Analyzer

跨各種範例使用 Run Analyzer 工具，以考慮資料大小的差異。

您可以使用執行分析器來了解生產資料範例中執行之間的差異。在 Run Analyzer 中使用 `--batch` 模式來產生批次執行的統計資料，並分析處理資料集極端值所需的運算資源上限。

例如，您可以為執行分析器提供批次模式中的資料完整流程儲存格，以了解完整流程儲存格的尖峰 vCPU 和記憶體使用率。

減少輸入資料集的大小差異

如果您看到樣本大小有很大的差異，您可以對 HealthOmics 上游的樣本進行分叉，並為每個批次選取不同的檔案系統大小，以節省執行儲存成本。

在 WDL 中，使用 `size`函數為大型和小型範例的個別任務分叉資源配置。將此策略套用至最昂貴的任務，以產生最大的影響。

在 Nextflow 中，根據檔案大小或檔案名稱，使用條件式資源來分層資源配置。如需詳細資訊，請參閱 Nextflow GitHub 網站上的條件式程序資源。

不要太快最佳化

在投資大量的效能調校工作之前，請先完成您的工作流程程式碼和邏輯。變更程式碼可能會對所需資源產生重大影響。如果您在開發過程中過早最佳化執行，您可能會過度最佳化，或者如果工作流程定義稍後變更，您可能需要再次最佳化。

定期重新執行 Run Analyzer 工具

如果您隨著時間對工作流程定義進行變更，或如果您的範例變異，請定期執行 Run Analyzer 工具，以協助您進行其他最佳化。

最佳化資源並行的方法

HealthOmics 提供下列功能，協助您控制和管理大規模處理執行時的成本：

- 使用執行群組來控制您的成本和資源用量。您可以在執行群組中針對並行執行次數、vCPUs、GPUs 和每個任務的總執行時間設定最大值。如果不同的團隊或群組使用相同的帳戶，您可以為每個團隊建立個別的執行群組。您可以設定執行群組最大值，以控制每個團隊的資源用量和成本。如需詳細資訊，請參閱建立 HealthOmics 執行群組。
- 在開發期間，您可以設定具有較低最大值的個別執行群組，以擷取失控任務。
- Service Quotas 也有助於保護您的帳戶免受過多資源請求的影響。如需 Service Quotas 的相關資訊，包括如何請求提高配額值，請參閱 HealthOmics 服務配額

在 HealthOmics 中刪除執行和執行群組

當您不再需要執行或執行群組時，您可以使用 、 AWS CLI API 或 主控台將其刪除。

除了刪除執行之外，您也可以取消執行。若要取消執行，其狀態必須為 `PENDING`、`RUNNING`、`STARTING`或 `STOPPING`。

> **Note**
>
> 當您取消執行時，HealthOmics 不會儲存任何執行輸出。

下列 AWS CLI 命令顯示如何取消執行。若要執行範例，請將 取代*run id*為您要取消的執行 ID。如果成功，則沒有回應。

```
aws omics cancel-run --id run id
```

下列 AWS CLI 命令會刪除執行。只有在執行完成或取消時，才能將其刪除。若要執行範例，*run id*請以您要刪除的執行 ID 取代 。如果成功刪除執行，則沒有回應。

```
aws omics delete-run --id run id
```

您也可以刪除執行群組。只有在沒有與狀態為 PENDING、RUNNING、 或 的執行群組相關聯的執行時STARTING，才能刪除執行群組STOPPING。

下列範例示範如何使用 AWS CLI 刪除執行群組。您將不會收到回應。若要執行範例，*run group id*請以您要刪除之執行群組的 ID 取代 。

```
aws omics delete-run-group --id run group id
```

建立 HealthOmics 執行群組

您可以選擇性地建立執行群組，以限制您新增至群組之執行的運算資源。執行群組可協助您：

- 將執行排入佇列，以免超過服務限制。
- 透過設定執行持續時間上限來擷取執行中任務。
- 管理每個執行的優先順序，讓最重要的執行先完成。

如果您設定並行 vCPU、GPU 或執行次數上限，執行任務會在達到上限時排入佇列。如果您設定最長執行持續時間，如果超過最長執行持續時間，則執行會失敗。

使用執行優先順序設定在執行群組中建立優先順序。

服務限制優先於執行群組限制。例如，如果您將執行群組最大值設定為高於區域中的服務最大值，HealthOmics 會套用服務最大值。

主題

執行優先順序

您可以使用執行優先順序來建立執行群組中執行的優先順序。

如果多個執行具有相同的優先順序，則首先啟動的執行具有較高的優先順序。

您也可以為不在執行群組中的執行設定優先順序。優先順序會與不在執行群組中的所有其他執行的優先順序進行比較

您可以在開始執行時設定執行優先順序。如需詳細資訊，請參閱在 HealthOmics 中開始執行。

使用主控台建立執行群組

建立執行群組

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇執行群組。
3. 在執行群組頁面上，選擇建立執行群組。
4. 在建立執行群組詳細資訊頁面上，提供下列資訊
   - 執行群組名稱 - 此執行群組的唯一名稱。
   - 並行執行的 vCPU 上限 - 可在執行群組中的所有作用中執行同時執行的 vCPUs 數目上限。
   - 最大 GPUs：可在執行群組中的所有作用中執行中同時執行的 GPUs 數量上限。
   - 每次執行的最大執行時間 （分鐘） - 每次執行的最大時間 （分鐘）。如果執行超過執行時間上限，則執行會自動失敗。
   - 並行執行上限 - 可同時執行的執行數目上限。
5. （選用） 您最多可以將 50 個標籤新增至執行群組。
6. 選擇建立執行群組。

使用 CLI 建立執行群組

若要建立執行群組，請使用 create-run-group API 操作來建立名為 的執行群組TestRunGroup。下列範例會設定最多 20 CPUs、10 個 GPUs、5 個執行，以及最多 600 分鐘的執行持續時間。

```
aws omics create-run-group --name TestRunGroup \
--max-cpus 20 \
--max-gpus 10 \
--max-duration 600 \
--max-runs 5
```

此 API 操作的回應包含新建立 的 IDRunGroup。

```
{
    "arn": "arn:aws:omics:us-west-2:12345678901:runGroup/2839621",
    "id": "2839621",
    "tags": {}
}
```

若要取得有關執行群組的其他資訊，請將此 ID 與 get-run-group API 操作搭配使用，如下列範例所示。

```
aws omics get-run-group --id run group id
```

回應包含執行群組的限制設定和指派的標籤。

```
{
    "arn": "arn:aws:omics:us-west-2:776893852117:runGroup/2839621",
    "id": "2839621",
    "name": "TestRunGroup",
    "maxCpus": 20,
    "maxRuns": 5,
    "maxDuration": 600,
    "creationTime": "2024-06-12T15:35:39.191730+00:00",
    "tags": {},
    "maxGpus": 10
}
```

您也可以使用 list-run-group API 操作來檢視所有建立的執行群組。

```
aws omics list-run-groups
```

HealthOmics 執行的呼叫快取

AWS HealthOmics 支援私有工作流程的呼叫快取，也稱為繼續。呼叫快取會在執行完成後儲存已完成工作流程任務的輸出。後續執行可以使用快取中的任務輸出，而不是再次計算任務輸出。呼叫快取可減少運算資源用量，進而縮短執行持續時間並節省運算成本。

您可以在執行完成後存取快取的任務輸出檔案。若要執行進階任務偵錯和疑難排解，您可以在工作流程定義中將這些檔案指定為任務輸出，以快取中繼任務檔案。

您可以使用呼叫快取來儲存失敗執行的已完成任務結果。下一次執行會從上次成功完成的任務開始，而不是再次計算已完成的任務。

如果 HealthOmics 找不到任務的相符快取項目，則執行不會失敗。HealthOmics 會重新計算任務及其相依任務。

如需疑難排解呼叫快取問題的資訊，請參閱 對呼叫快取問題進行故障診斷。

主題

呼叫快取的運作方式

若要使用呼叫快取，您可以建立執行快取，並將其設定為具有快取資料的關聯 Amazon S3 位置。當您開始執行時，您可以指定執行快取。執行快取並非專用於一個工作流程。從多個工作流程執行的 可以使用相同的快取。

在執行的匯出階段，系統會將完成的任務輸出匯出至 Amazon S3 位置。若要匯出中繼任務檔案，請在工作流程定義中將這些檔案宣告為任務輸出。呼叫快取也會在內部儲存中繼資料，並為每個快取項目建立唯一的雜湊。

對於執行中的每個任務，工作流程引擎會偵測此任務是否有相符的快取項目。如果沒有相符的快取項目，HealthOmics 會計算任務。如果有相符的快取項目，引擎會擷取快取的結果。

為了符合快取項目，HealthOmics 使用原生工作流程引擎中包含的雜湊機制。HealthOmics 延伸這些現有的雜湊實作，以考量 HealthOmics 變數，例如 S3 eTags 和 ECR 容器摘要。

HealthOmics 支援這些工作流程語言版本的呼叫快取：

- WDL 1.0、1.1 版和開發版本
- Nextflow 23.10 和 24.10 版
- 所有 CWL 版本

> **Note**
>
> HealthOmics 不支援 Ready2Run 工作流程的呼叫快取。

主題

共同責任模式

使用者和 之間的共同責任 AWS 是判斷任務和執行是否為呼叫快取的良好候選者。當所有任務都是等冪時，呼叫快取會獲得最佳結果（使用相同輸入重複執行任務會產生相同的結果）。

不過，如果任務包含非確定性元素（例如隨機數產生或系統時間），則使用相同輸入重複執行任務可能會導致不同的輸出。這可能會以下列方式影響呼叫快取的有效性：

- 如果 HealthOmics 使用的快取項目（由先前的執行建立）與任務執行為目前執行產生的輸出不同，則執行可能會產生與相同執行不同的結果，而沒有快取。
- HealthOmics 可能會因為非確定性任務輸出，而找不到應相符任務的相符快取項目。如果找不到有效的快取項目，則執行會重新計算任務，進而降低使用呼叫快取的成本節省效益。

以下是已知的任務行為，可能導致影響呼叫快取結果的非確定性結果：

- 使用亂數產生器。
- 取決於系統時間。
- 使用並行 (race-conditions 可能會導致輸出差異）。
- 擷取超出任務輸入參數中指定範圍的本機或遠端檔案。

如需可能導致非確定性行為的其他案例，請參閱 Nextflow 文件網站上的非確定性程序輸入。

如果您懷疑任務產生非確定性的輸出，請考慮使用工作流程引擎功能，例如 Nextflow 中的快取選擇退出，以避免快取非確定性的特定任務。

我們建議您先徹底檢閱特定工作流程和任務需求，再在任何無效呼叫快取或輸出超出預期可能帶來風險的環境中啟用呼叫快取。例如，在判斷呼叫快取是否適合臨床使用案例時，應仔細考慮呼叫快取的潛在限制。

任務的快取需求

HealthOmics 會為符合下列要求的任務快取任務輸出：

- 任務必須定義容器。HealthOmics 不會快取沒有容器之任務的輸出。
- 任務必須產生一或多個輸出。您可以在工作流程定義中指定任務輸出。
- 工作流程定義不得使用動態值。例如，如果您將參數傳遞給具有隨每次執行遞增值的任務，HealthOmics 不會快取任務輸出。

> **Note**
>
> 如果執行中的多個任務使用相同的容器映像，HealthOmics 會為所有這些任務提供相同的映像版本。HealthOmics 提取映像後，會在執行期間忽略容器映像的任何更新。此方法提供可預測且一致的體驗，並防止因中期部署的容器映像更新而可能發生的潛在問題。

執行快取效能

當您開啟執行的呼叫快取時，您可能會注意到下列對執行效能的影響：

- 在第一次執行期間，HealthOmics 會儲存執行中任務的快取資料。此執行的匯出時間可能較長，因為呼叫快取會增加匯出資料的數量。
- 在後續執行中，從快取繼續執行時，可能會縮短處理步驟的數量並縮短執行時間。

- 如果您也選擇將中繼檔案宣告為輸出，則匯出時間可能會更長，因為這些資料可能更為詳細。

快取資料保留和失效事件

執行快取的主要目的是最佳化執行中任務的運算。如果任務有有效的相符快取項目，HealthOmics 會使用快取項目，而不是重新計算任務。否則，HealthOmics 會還原為預設服務行為，也就是重新計算任務及其相依任務。透過使用此方法，快取遺漏不會導致執行失敗。

我們建議您管理執行快取大小。隨著時間的推移，由於工作流程引擎或 HealthOmics 服務更新，或是您在執行或執行任務中所做的變更，快取項目可能不再有效。下列各節提供其他詳細資訊。

主題

資訊清單版本更新和資料更新

HealthOmics 服務可能會定期引進新功能或工作流程引擎更新，使部分或全部執行快取項目失效。在這種情況下，您的執行可能會遇到一次性快取遺漏。

HealthOmics 會為每個快取項目建立 JSON 資訊清單檔案。對於 2025 年 2 月 12 日之後開始的執行，資訊清單檔案包含版本參數。如果服務更新使任何快取項目失效，HealthOmics 會遞增版本編號，讓您可以識別要移除的舊版快取項目。

下列範例顯示 版本設定為 2 的資訊清單檔案：

```
{
     "arn": "arn:aws:omics:us-west-2:12345678901:runCache/0123456/
cacheEntry/1234567-195f-3921-a1fa-ffffcef0a6a4",
     "s3uri": "s3://example/1234567-d0d1-e230-
d599-10f1539f4a32/1348677/4795326/7e8c69b1-145f-3991-a1fa-ffffcef0a6a4",
     "taskArn": "arn:aws:omics:us-west-2:12345678901:task/4567891",
     "workDir": "/mnt/workflow/1234567-d0d1-e230-d599-10f1539f4a32/workdir/call-
TxtFileCopyTask/5w6tn5feyga7noasjuecdeoqpkltrfo3/wxz2fuddlo6hc4uh5s2lreaayczduxdm",
     "files": [
         {
             "name": "output_txt_file",
             "path": "out/output_txt_file/outfile.txt",
             "etag": "ajdhyg9736b9654673b9fbb486753bc8"
```

```
        }
    ],
    "nextflowContext": {},
    "otherOutputs": {},
    "version": 2,
}
```

對於具有不再有效的快取項目的執行，請重建快取以建立新的有效項目。每次執行時，請執行下列步驟：

1. 在快取保留設定為 CACHE ALWAYS 的情況下啟動一次執行。此執行會建立新的快取項目。
2. 對於後續執行，請將快取保留設定為其先前的設定（一律為 CACHE 或 CACHE ON FAILURE)。

若要清除不再有效的快取項目，您可以從快取 Amazon S3 儲存貯體中刪除這些快取項目。HealthOmics 永遠不會重複使用這些快取項目。如果您選擇保留無效的項目，則不會影響執行。

> **Note**
>
> 呼叫快取會將任務輸出資料儲存在為快取指定的 Amazon S3 位置，這會對您的 產生費用 AWS 帳戶。

執行快取行為

您可以設定執行快取行為，為失敗的執行（失敗時的快取）或所有執行（一律快取）儲存任務輸出。當您建立執行快取時，您可以為使用此快取的所有執行設定預設快取行為。您可以在開始執行時覆寫預設行為。

Cache on failure 如果您要偵錯在數個任務成功完成後失敗的工作流程，則 很有用。如果雜湊考慮的所有唯一變數與先前的執行相同，則後續執行會從上次成功完成的任務繼續。

Cache always 如果您在成功完成的工作流程中更新任務， 會很有用。我們建議您遵循下列步驟：

1. 建立新的執行。將快取行為設定為一律快取，然後開始執行。
2. 執行完成後，請更新工作流程中的任務，並一律以行為集快取啟動新的執行。此執行會處理更新的任務，以及任何與更新任務相依的後續任務。所有其他任務都會使用快取的結果。
3. 視需要重複步驟 2，直到更新任務的開發完成為止。
4. 視需要在未來執行時使用更新的任務。如果您計劃在這些執行中使用新的或不同的輸入，請記得在失敗時切換後續執行至快取。

> **Note**
>
> 我們建議在使用相同測試資料集時一律快取模式，但不適用於一批執行。如果您針對大量執行設定此模式，系統可以將大量資料匯出至 Amazon S3，進而增加匯出時間和儲存成本。

控制執行快取大小

HealthOmics 不會刪除或自動封存任何執行快取資料，也不會套用 Amazon S3 清除規則來管理快取資料。我們建議您執行定期快取清除，以節省 Amazon S3 儲存成本，並保持您的執行快取大小可管理。您可以直接刪除檔案，或在執行快取儲存貯體上設定資料保留/複寫政策。

例如，您可以將 Amazon S3 生命週期政策設定為在 90 天後使物件過期，也可以在每個開發專案結束時手動清除快取資料。

下列資訊可協助您管理快取資料大小：

- 您可以檢查 Amazon S3 來檢視快取中有多少資料。HealthOmics 不會監控或報告快取大小。
- 如果您刪除有效的快取項目，後續執行不會失敗。HealthOmics 會重新計算任務及其相依任務。
- 如果您修改快取名稱或目錄結構，讓 HealthOmics 找不到任務的相符項目，HealthOmics 會重新計算任務。

如果您需要檢查快取項目是否仍然有效，請檢查快取資訊清單版本號碼。如需詳細資訊，請參閱資訊清單版本更新和資料更新。

建立執行快取

當您建立執行快取時，您可以指定快取資料的 Amazon S3 位置。此資料必須可立即存取。呼叫快取不會擷取 Glacier 中封存的物件 （ 例如 GFR 和 GDA 儲存類別 ）。

如果快取資料的 Amazon S3 儲存貯體由另一個儲存貯體擁有 AWS 帳戶，請在建立執行快取時提供該帳戶 ID。

使用主控台建立執行快取

從 主控台，依照下列步驟建立執行快取。

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇執行快取。

3. 在執行快取頁面中，選擇建立執行快取。
4. 在建立執行快取頁面的執行快取詳細資訊面板中，設定下列欄位：
   a. 輸入執行快取的名稱。
   b. (選用) 輸入描述。
   c. 輸入快取輸出的 S3 位置。選擇與工作流程位於相同區域的儲存貯體。
   d. （選用）輸入儲存貯體擁有者 AWS 帳戶 的 ，以驗證儲存貯體擁有權。如果您未輸入值，預設值為您的帳戶 ID。
   e. 在快取行為下，設定預設行為（是針對失敗的執行或所有執行快取輸出）。當您開始執行時，您可以選擇覆寫預設行為。
5. （選用）將一或多個標籤與執行快取建立關聯。
6. 選擇建立執行快取。主控台會在執行快取資料表中顯示新的執行快取。

使用 CLI 建立執行快取

使用 create-run-cache CLI 命令來建立執行快取。預設快取行為為 CACHE_ON_FAILURE。

```
aws omics create-run-cache \
      --name "workflow 123 run cache" \
      --description "my run cache" \
      --cache-s3-location "s3://amzn-s3-demo-bucket" \
      --cache-behavior "CACHE_ALWAYS"                \
      --cache-bucket-owner-id  "111122223333"
```

如果建立成功，您會收到包含下列欄位的回應。

```
{
  "arn": "string",
  "id": "string",
  "status": "ACTIVE"
  "tags": {}
  }
```

更新執行快取

您可以變更快取名稱、描述、標籤或快取行為，但不能變更快取的 S3 位置。

使用主控台更新執行快取

從 主控台，依照下列步驟更新執行快取。

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇執行快取。
3. 從執行快取資料表中，選擇要更新的執行快取，然後選擇編輯。
4. 在執行快取詳細資訊面板中，您可以更新執行快取名稱、描述和快取行為欄位。
5. （選用）將一或多個新標籤與執行快取建立關聯，或移除現有的標籤。
6. 選擇儲存執行快取。

使用 CLI 更新執行快取

使用 update-run-cache CLI 命令來更新執行快取。

```
aws omics update-run-cache \
      --name "workflow 123 run cache" \
      --id "workflow id" \
      --description "my run cache" \
      --cache-behavior "CACHE_ALWAYS"
```

如果更新成功，您會收到沒有資料欄位的回應。

刪除執行快取

如果沒有作用中的執行正在使用，您可以刪除執行快取。如果有任何執行正在使用執行快取，請等待執行完成，否則您可以取消執行。

刪除執行快取會移除資源及其中繼資料，但不會刪除 Amazon S3 中的資料。刪除快取後，您無法重新連接快取，也無法將其用於後續執行。

快取的資料會保留在 Amazon S3 中供您檢查。您可以使用標準 S3 Delete操作移除舊快取資料。或者，建立 Amazon S3 生命週期政策，讓不再使用的快取資料過期。

使用主控台刪除執行快取

從 主控台，依照下列步驟刪除執行快取。

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇執行快取。
3. 從執行快取資料表中，選擇要刪除的執行快取。
4. 從執行快取資料表功能表中，選擇刪除。
5. 從模態對話方塊中，儲存 Amazon S3 快取資料連結以供日後參考，然後確認要刪除執行快取。

   您可以使用 Amazon S3 連結來檢查快取的資料，但無法將資料重新連結到另一個執行快取。當您完成檢查時，請刪除快取資料。

使用 CLI 刪除執行快取

使用 delete-run-cache CLI 命令來刪除執行快取。

```
aws omics delete-run-cache \
      --id "my cache id"
```

如果刪除成功，您會收到沒有資料欄位的回應。

執行快取的內容

HealthOmics 會使用 S3 儲存貯體中的下列結構來組織您的執行快取：

```
s3://{cache.S3location}/{cache.uuid}/runID/taskID/{cacheentry.uuid}/
```

cache.uuid 是快取的全域唯一 ID。cacheentry.uuid 是快取任務的全域唯一 uuid。HealthOmics 會將 uuids 指派給快取和任務。

對於所有工作流程引擎，快取包含下列檔案：

- {cacheentryuuid}.json 檔案 – HealthOmics 會建立此資訊清單檔案，其中包含快取的相關資訊，包括快取中的所有項目清單，以及快取版本。
- 任務輸出檔案 – 每個任務輸出包含一或多個檔案，如任務所定義。

對於使用 Nextflow 的工作流程，Nextflow 引擎會在快取中建立這些額外的檔案：

- command.out 檔案 – 此檔案包含任務執行 stdout 內容。
- .exitcode 檔案 – 此檔案包含任務結束碼 （整數）。

> **Note**
>
> 如果您想要存取執行快取中的中繼任務檔案以進行進階故障診斷，請在工作流程定義中將這些檔案宣告為任務輸出。

引擎特定的快取功能

HealthOmics 會嘗試跨工作流程引擎提供一致的呼叫快取實作。根據每個工作流程引擎如何處理特定案例，有一些差異：

- 下一個流程
  - 無法保證能夠跨不同的 Nextflow 版本進行快取。例如，如果您在 v23.10.0 中執行任務，並在 v24.10.8 中執行相同的任務，HealthOmics 可能會將第二個執行視為快取遺漏。
  - 您可以使用快取false指令關閉個別任務的快取。如需此指令的相關資訊，請參閱 Nextflow 規格中的處理程序。
  - HealthOmics 使用 Nextflow 寬鬆模式，但不支援深度快取模式。
  - 如果您在任務的輸入 S3 路徑中使用 glob 模式，快取會評估每個個別 S3 物件。如果您新增物件，HealthOmics 只會重新計算使用新物件的任務。
  - HealthOmics 不會快取任務重試。此行為與 Nextflow 的預設行為一致。
- WDL
  - 當您使用 WDL 工作流程的開發版本時，HealthOmics 支援新的輸入「目錄」類型。對於呼叫快取，如果目錄中的任何物件變更，HealthOmics 會重新計算輸入目錄的所有任務。
  - HealthOmics 支援任務層級快取，但不支援工作流程層級快取。
- CWL
  - 來自任務的常數輸出不會從資訊清單中明確顯示。HealthOmics 會將常數輸出快取為中繼檔案。

使用執行快取

根據預設，執行不會使用執行快取。若要使用執行的快取，請在開始執行時指定執行快取和執行快取行為。

執行完成後，您可以使用 主控台、CloudWatch Logs 或 API 操作來追蹤快取命中或疑難排解快取問題。如需詳細資訊，請參閱 追蹤呼叫快取資訊 和 對呼叫快取問題進行故障診斷。

如果執行中的一或多個任務產生非確定性輸出，強烈建議您不要使用呼叫快取來執行，或選擇不快取這些特定任務。如需詳細資訊，請參閱共同責任模式。

> **Note**
>
> 您在開始執行時提供 IAM 服務角色。若要使用呼叫快取，服務角色需要存取執行快取 Amazon S3 位置的許可。如需詳細資訊，請參閱的服務角色 AWS HealthOmics。

主題

使用主控台設定具有執行快取的執行

從 主控台，您可以在開始執行時設定執行的執行快取。

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇執行。
3. 在執行頁面上，選擇要開始的執行。
4. 選擇開始執行並完成開始執行的步驟 1 和 2，如中所述使用主控台啟動執行。
5. 在開始執行的步驟 3 中，選擇選取現有的執行快取。
6. 從執行快取 ID 下拉式清單中選取快取。
7. 若要覆寫預設的執行快取行為，請選擇執行的快取行為。如需詳細資訊，請參閱執行快取行為。
8. 繼續執行開始執行的步驟 4。

使用 CLI 設定具有執行快取的執行

若要啟動使用執行快取的執行，請將 cache-id 參數新增至 start-run CLI 命令。或者，使用 `cache-behavior` 參數覆寫您為執行快取設定的預設行為。下列範例僅顯示 命令的快取欄位：

```
aws omics start-run \
        ...
```

```
    --cache-id "xxxxxx"    \
    --cache-behavior  CACHE_ALWAYS
```

如果操作成功，您會收到沒有資料欄位的回應。

執行快取的錯誤案例

針對下列案例，HealthOmics 可能不會快取任務輸出，即使執行快取行為設為一律快取。

- 如果執行在第一個任務成功完成之前遇到錯誤，則沒有要匯出的快取輸出。
- 如果匯出程序失敗，HealthOmics 不會將任務輸出儲存到 Amazon S3 快取位置。
- 如果執行因filesystem out of space錯誤而失敗，呼叫快取不會儲存任何任務輸出。
- 如果您取消執行，呼叫快取不會儲存任何任務輸出。
- 如果執行遇到執行逾時，呼叫快取不會儲存任何任務輸出，即使您將執行設定為在失敗時使用快取。

追蹤呼叫快取資訊

您可以使用主控台、CLI 或 CloudWatch Logs 追蹤呼叫快取事件 （ 例如執行快取命中 ）。

主題

- 使用主控台追蹤快取命中
- 使用 CLI 追蹤呼叫快取
- 使用 CloudWatch Logs 追蹤呼叫快取

使用主控台追蹤快取命中

在執行的執行詳細資訊頁面中，執行任務資料表會顯示每個任務的快取命中資訊。資料表也包含關聯快取項目的連結。使用下列程序來檢視執行的快取命中資訊。

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇執行。
3. 在執行頁面上，選擇要檢查的執行。
4. 在執行詳細資訊頁面上，選擇執行任務索引標籤以顯示任務資料表。
5. 如果任務發生快取命中，快取命中資料欄會包含 Amazon S3 中執行快取項目位置的連結。
6. 選擇連結以檢查執行快取項目。

使用 CLI 追蹤呼叫快取

使用 get-run CLI 命令確認執行是否使用呼叫快取。

```
aws omics get-run --id 1234567
```

在回應中，如果已設定 `cacheId` 欄位，則執行會使用該快取。

使用 list-run-tasks CLI 命令，擷取執行中每個快取任務的快取資料位置。

```
aws omics list-run-tasks --id 1234567
```

在回應中，如果任務的 cacheHit 欄位為 true，cacheS3Uri 欄位會提供該任務的快取資料位置。

您也可以使用 get-run-task CLI 命令來擷取特定任務的快取資料位置：

```
aws omics get-run-task --id 1234567 --task-id <task_id>
```

使用 CloudWatch Logs 追蹤呼叫快取

HealthOmics 會在 `/aws/omics/WorkflowLog` CloudWatch 日誌群組中建立快取活動日誌。每個執行快取都有一個日誌串流：runCache/<cache_id>/<cache_uuid>。

對於使用呼叫快取的執行，HealthOmics 會為這些事件產生 CloudWatch Logs 項目：

- 建立快取項目 (CACHE_ENTRY_CREATED)
- 符合快取項目 (CACHE_HIT)
- 不符合快取項目 (CACHE_MISS)

如需這些日誌的詳細資訊，請參閱 CloudWatch 中的日誌 。

在`/aws/omics/WorkflowLog`日誌群組上使用下列 CloudWatch Insights 查詢，傳回此快取每次執行的快取命中次數：

```
filter @logStream like 'runCache/<CACHE_ID>/'
 fields @timestamp, @message
 filter logMessage like 'CACHE_HIT'
 parse "run: *," as run
 stats count(*) as cacheHits by run
```

使用下列查詢傳回每次執行所建立的快取項目數量：

```
filter @logStream like 'runCache/<CACHE_ID>/'
 fields @timestamp, @message
 filter logMessage like 'CACHE_ENTRY_CREATED'
 parse "run: *," as run
 stats count(*) as cacheEntries by run
```

共用 HealthOmics 工作流程

身為私有工作流程的擁有者，您可以與相同 AWS 帳戶 區域中的 共用工作流程。若要與多個工作流程共用 AWS 帳戶，您可以建立相同工作流程的多個共用。

身為擁有者，您可以透過刪除共用來撤銷對共用工作流程的存取。

> **Note**
>
> HealthOmics 會自動允許共用工作流程在訂閱者帳戶中執行工作流程時存取 Amazon ECR 儲存庫。您不需要為共用工作流程授予其他儲存庫存取權。

當您共用工作流程時，訂閱者可以使用任何工作流程版本。如果您需要共用工作流程的版本層級存取控制，建議您建立單獨的工作流程，而不是使用工作流程版本。

主題

訂閱共用工作流程

若要訂閱共用工作流程，請遵循這些整體步驟來接受和使用工作流程：

1. 使用 主控台或 API 來接受共享。將您目前的區域設定為與共享請求相同的區域。

- 若要在主控台中尋找共用請求，請導覽至所有資源共用頁面，然後選擇與我共用索引標籤。

2. 使用 主控台或 API 為共用工作流程建立執行。
   - 若要在主控台中尋找工作流程詳細資訊頁面，請導覽至與我共用 （請參閱步驟 1），然後選擇共用工作流程的資源連結。
3. 您可以為工作流程提供自己的輸入資料。
4. 共用工作流程會在您的 中執行 AWS 帳戶。

身為共用工作流程的訂閱者，系統會阻止您執行下列工作流程動作：

- 匯出共用工作流程
- 重新執行共用工作流程
  - 您可以為共用工作流程建立新的執行。
- 重新共用工作流程。
- 將標籤指派給工作流程。
- 刪除工作流程。
  - 當您不再需要工作流程時，您會刪除工作流程共享。

如需資源共用的其他資訊中的跨帳戶資源共用 AWS HealthOmics，請參閱 。

監控工作流程共享的狀態

HealthOmics 會針對工作流程共享的每個狀態變更，將事件傳送至 EventBridge。如果您想要接收特定狀態變更的通知，請設定 EventBridge 規則來監控工作流程共用狀態變更事件。例如：

- 您希望每次收到工作流程共享請求，以及每次使用者撤銷工作流程共享時收到通知。
- 啟動工作流程共享請求後，您希望在使用者接受或拒絕請求時收到通知。

如需使用事件的詳細資訊，請參閱 搭配 使用 EventBridge AWS HealthOmics。

使用主控台共用私有工作流程

從 主控台，您可以與 AWS 帳戶 與工作流程位於相同區域中的 共用私有工作流程。

共用私有工作流程

1. 開啟 HealthOmics 主控台。

2. 在左側導覽窗格中，選擇私有工作流程。
3. 從私有工作流程頁面上的工作流程表格中，選取要共用的工作流程，然後選擇共用。
4. 在共享工作流程頁面的共享詳細資訊面板中，輸入共享 AWS 帳戶 的描述性名稱，然後輸入訂閱者的 。
5. 選擇共用資源。主控台會在所有資源共用頁面中顯示資源共用。

共享的初始狀態為待定。訂閱者接受共享後，狀態會變更為作用中。

使用 CLI 共用私有工作流程

使用 create-share API 操作來建立工作流程共用。主要訂閱者是將存取工作流程的使用者 AWS 帳戶的 。

```
aws omics create-share \
  --resource-arn "arn:aws:omics:us-west-2:555555555555:workflow/123456" \
  --principal-subscriber "123456789012" \
  --name "my_Share-123"
```

如果建立成功，您會收到共用 ID 和狀態的回應。

```
{
"shareId": "495c21bedc889d07d0ab69d710a6841e-dd75ab7a1a9c384fa848b5bd8e5a7e0a",
"name": "my_Share-123",
"status": "PENDING"
}
```

共享會保持待定狀態，直到訂閱者使用 `accept-share` API 操作接受為止。

如需其他 API 使用範例中的跨帳戶資源共用 AWS HealthOmics，請參閱 。

使用主控台接受共用工作流程

您可以使用 主控台來接受提供的工作流程共享。請務必將主控台設定為與工作流程相同的區域。

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇所有資源共用，然後選擇與我共用索引標籤。
3. 從與我共用的資源資料表 中，選取工作流程共用，然後選擇接受。

接受工作流程後，請選擇共用工作流程的資源連結以檢視其詳細資訊。

使用主控台執行共用工作流程

接受工作流程共享後，您可以在工作流程上開始執行。

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇所有資源共用，然後選擇與我共用索引標籤。
3. 從與我共用的資源表格中，選擇共用工作流程的資源連結。
4. 在工作流程詳細資訊頁面中，選擇建立執行。

   主控台會開啟建立執行頁面，並預先填入工作流程類型（共用）和工作流程 ID。
5. 在建立執行表單中設定其餘欄位。如需其他資訊，請參閱 使用主控台啟動執行。

使用 API 執行共用工作流程

使用 get-workflow 擷取共用工作流程的 ARN。

```
aws omics get-workflow --id 1234567 \
--workflow-owner-id 55555555555
```

當您執行工作流程時，請提供工作流程擁有者的 AWS 帳戶 ID 和共用工作流程的 ARN。

```
aws omics start-run --id 1234567 --workflow-owner-id 55555555555 \
--role-arn arn:aws:iam::1234567892012:role/service-role/OmicsWorkflow-20221004T164236 \
--name ArchiveTest --retention-mode REMOVE
```

HealthOmics 中的 Ready2Run 工作流程

Ready2Run 工作流程是由第三方發佈者發佈的預先設定工作流程。有些發佈者，例如 Sentieon Inc，提供以訂閱為基礎的工作流程。其他 Ready2Run 工作流程不需要訂閱，而且某些工作流程是開放原始碼，例如 NF-Core 工作流程。

Ready2Run 工作流程非常適合下列案例：

- 您想要專注於管道輸出的分析和產生結果，而不需要設定基礎基礎設施。
- 您想要使用已建立的工作流程複寫結果。
- 身為軟體開發人員，您想要直接整合應用程式與 HealthOmics SDK。

HealthOmics 支援 Ready2Run 工作流程的版本控制。對於提供 版本的 Ready2Run 工作流程，您可以在開始執行時指定版本名稱。

所有 Ready2Run 工作流程都提供日誌，包括 CloudWatch 日誌，可用於故障診斷。

> **Note**
>
> Sentieon Ready2Run 工作流程以訂閱為基礎。當您第一次在帳戶中執行 Sentieon Ready2Run 工作流程時，Sentieon 會自動為您的 建立兩週的評估授權 AWS 帳戶。授權適用於所有 Sentieon Ready2Run 工作流程。在評估期間結束後，您可以請求永久授權或請求延長評估授權。如需詳細資訊，請參閱 Subscribing to Sentieon Ready2Run workflows。

主題

HealthOmics 中可用的 Ready2Run 工作流程

下表列出 HealthOmics 中可用的 Ready2Run 工作流程。

您可以登入 HealthOmics 主控台來檢視這些工作流程的詳細資訊，包括輸入參數和工作流程圖表。如需 Ready2Run 工作流程的定價資訊，請參閱 HealthOmics 定價。

> **Note**
>
> 每個 Ready2Run 工作流程都有最大輸入檔案大小。這些檔案大小上限無法調整。

工作流程名稱	發布者	需要訂閱？	輸入檔案大小上限 (GiB)	預估執行時間 (HH：MM)
AlphaFold 用於 601-1200 殘差	Google DeepMind	否	1	11：15
AlphaFold 最多可用於 600 個殘差	Google DeepMind	否	1	7：30
Bases2Fastq for 2x150	Element Biosciences	否	1000	1：45
Bases2Fastq for 2x300	Element Biosciences	否	1000	1：30
Bases2Fastq for 2x75	Element Biosciences	否	500	0：45
ESMFold 最多可用於 800 個殘差	Meta Research	否	1	0：15
GATK-BP fq2bam	廣泛研究所	否	64	10：10
適用於 30 倍基因體的 GATK-BP Germline bam2vcf	廣泛研究所	否	39	2：45

工作流程名稱	發布者	需要訂閱？	輸入檔案大小上限 (GiB)	預估執行時間 (HH：MM)
適用於 30 倍基因體的 GATK-BP Germline fq2vcf	廣泛研究所	否	64	12：30
GATK-BP 體力 WES bam2vcf	廣泛研究所	否	86	1：30
NVIDIA Parabricks BAM2FQ2BAM WGS，最多可達 30X	NVIDIA Corporation	否	80	1：39
NVIDIA Parabricks BAM2FQ2BAM WGS，最多可達 50X	NVIDIA Corporation	否	120	2：45
NVIDIA Parabricks BAM2FQ2BAM WGS，最多可達 5X	NVIDIA Corporation	否	20	0：18
NVIDIA Parabricks FQ2BAM WGS 最多可達 30X	NVIDIA Corporation	否	71	1:00
高達 50X 的 NVIDIA Parabricks FQ2BAM WGS	NVIDIA Corporation	否	137	1：45

工作流程名稱	發布者	需要訂閱？	輸入檔案大小上限 (GiB)	預估執行時間 (HH：MM)
NVIDIA Parabricks FQ2BAM WGS 最多可達 5X	NVIDIA Corporation	否	13	0：15
高達 30X 的 NVIDIA Parabricks Germline DeepVariant WGS	NVIDIA Corporation	否	71	2:00
NVIDIA Parabricks Germline DeepVariant WGS，最多可達 50X	NVIDIA Corporation	否	137	3：30
NVIDIA Parabricks Germline DeepVariant WGS 最多可達 5X	NVIDIA Corporation	否	12	0：30
NVIDIA Parabricks Germline HaplotypeCaller WGS 最多可達 30X	NVIDIA Corporation	否	71	1：15

工作流程名稱	發布者	需要訂閱？	輸入檔案大小上限 (GiB)	預估執行時間 (HH：MM)
NVIDIA Parabricks Germline HaplotypeCaller WGS 最多可達 50X	NVIDIA Corporation	否	137	2:00
NVIDIA Parabricks Germline HaplotypeCaller WGS 最多可達 5X	NVIDIA Corporation	否	13	0：15
NVIDIA Parabricks 體力 Mutect2 WGS，最多可達 50X	NVIDIA Corporation	否	196	0：45
scRNAseq 與 KallistoBUStools	NF-Core	否	119	1：30
scRNAseq 搭配 Salmon Alevin-fry	NF-Core	否	119	2：30
scRNAseq 搭配 STARsolo	NF-Core	否	119	2：30
Sentieon Germline BAM WES 高達 300 倍	Sentieon， Inc.	是	9	1:00

工作流程名稱	發布者	需要訂閱？	輸入檔案大小上限 (GiB)	預估執行時間 (HH：MM)
Sentieon Germline BAM WGS 高達 32 倍	Sentieon， Inc.	是	18	1：30
Sentieon Germline FASTQ WES 高達 100 倍	Sentieon， Inc.	是	5	0：45
Sentieon Germline FASTQ WES 高達 300 倍	Sentieon， Inc.	是	26	2:00
Sentieon Germline FASTQ WGS 最多可達 32 倍	Sentieon， Inc.	是	51	3：30
適用於 ONT 的 Sentieon LongRead	Sentieon， Inc.	是	25	1：30
適用於 PacBio HiFi 的 Sentieon LongRead	Sentieon， Inc.	是	58	4:00
Sentieon 體操 WES	Sentieon， Inc.	是	50	2：30
Sentieon 體操 WGS	Sentieon， Inc.	是	113	4：30
Ultima Genomics DeepVariant 最多可達 40 倍	Ultima Genomics	否	91	1：55

當您使用 Ready2Run 工作流程時，您的工作流程已預先設定且無法編輯。與私有工作流程相反，Ready2Run 工作流程不支援下列項目：

- 增加最大輸入檔案大小
- 變更運算資源或執行儲存
- 變更工作流程定義或容器
- 將執行新增至執行群組
- 共用工作流程

如果發佈者已在 GitHub 上共用 Ready2Run 工作流程，您可以根據 Ready2Run 工作流程建立自己的私有工作流程。下表提供每個發佈者的 GitHub 工作流程連結。

發布者	GitHub 上的工作流程
Google DeepMind，中繼研究	蛋白質摺疊工作流程
Element Biosciences	如需詳細資訊，請聯絡 Element Biosciences
廣泛研究所	GATK 工作流程
NVIDIA Corporation	Parabricks 工作流程
nf 核心	NF-Core 工作流程
Sentieon	Sentieon 工作流程
Ultima Genomics	Ultima Genomics 工作流程

訂閱 Sentieon Ready2Run 工作流程

Sentieon Ready2Run 工作流程以訂閱為基礎。當您第一次在帳戶中執行 Sentieon Ready2Run 工作流程時，Sentieon 會自動為您的 建立兩週的評估授權 AWS 帳戶。授權適用於所有 Sentieon Ready2Run 工作流程。在評估期間結束後，您可以請求永久授權或請求延長評估授權。

請依照下列步驟訂閱 Sentieon Ready2Run 工作流程：

- 請依照這些指示尋找您的 AWS 正式使用者 ID。

- 傳送電子郵件至 Sentieon 支援群組 (support@sentieon.com) 以請求軟體授權。在電子郵件中提供 AWS 正式使用者 ID。

使用主控台啟動 HealthOmics Ready2Run 工作流程

在主控台中使用 Ready2Run 工作流程類似於使用私有工作流程。其中一個關鍵區別是工作流程發佈者提供範例資料，因此您可以嘗試工作流程，而無需建立自己的資料。

在主控台中使用 Ready2Run 工作流程

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇 Ready2Run 工作流程。
3. 在 Ready2Run 工作流程頁面上，選擇您要使用的工作流程。主控台會開啟該工作流程的詳細資訊頁面。
4. 詳細資訊索引標籤會列出資訊，例如名稱、每次執行的清單價格、描述、工作流程語言類型、執行儲存容量、狀態、建立日期，以及具有描述的參數。詳細資訊索引標籤也會告訴您工作流程是否需要訂閱。
5. 若要使用工作流程，請選擇建立執行
6. 在指定執行詳細資訊頁面中，輸入執行名稱。或者，您可以指定工作流程版本。您也可以將執行優先順序新增至執行。
7. 輸入或選取執行輸出的 Amazon S3 位置。
8. 針對執行中繼資料保留模式，選擇是否保留或移除執行中繼資料。
9. 在服務角色面板中，選擇是否使用現有的服務角色或建立新的服務角色。
10. （選用）新增標籤以協助識別和管理執行。
11. 選擇下一步。
12. 在新增參數頁面中，選擇其中一個選項來新增執行參數值：
    - 從 Amazon S3 位置選取參數檔案 (JSON 格式）。
    - 從本機磁碟機選取參數檔案 (JSON 格式）。
    - 手動輸入參數值。
    - 使用工作流程發佈者提供的 Ready2Run 範例資料來執行工作流程。
13. 如果您上傳 JSON 檔案，主控台會剖析檔案並執行內嵌驗證。然後，您可以視需要手動更新參數的值。
14. 選擇下一步。

15. 檢閱您的輸入，然後選擇開始執行。

使用 API 啟動 HealthOmics Ready2Run 工作流程

大多數 API 操作的行為方式與 Ready2Run 工作流程和私有工作流程類似。

若要傳回可用的 Ready2Run 工作流程清單，請使用將 `type` 參數設為 READY2RUN 的 list-workflows。

```
aws omics list-workflows --type READY2RUN
```

在您識別要從 list-workflows 回應執行的工作流程之後，您可以使用 get-workflow 搭配 `--id` 參數來取得更多詳細資訊。

```
aws omics get-workflow --type READY2RUN --id workflow id
```

若要執行 Ready2Run 工作流程，您可以使用 start-run API 操作，並將工作流程類型參數設為 READY2RUN，如下列範例所示

```
aws-omics start-run \
  --workflow-type READY2RUN \
  --workflow-id workflow id \
  --output-uri &example-s3-bucket; \
  --role-arn arn:aws:iam::1234567892012:role/service-role/OmicsWorkflow-20221004T164236
 \
  --parameters file:///path/to/parameters.json
```

若要指定工作流程版本，請使用 workflow-version 參數，如本範例所示。

```
aws-omics start-run \
  --workflow-type READY2RUN \
  ...
  --version-name '3.0.0'
```

若要監控您的執行，您可以使用 get-run API 操作，如下所示。

```
aws-omics get-run \
  --id run id
```

HealthOmics 儲存體

使用 HealthOmics 儲存以低成本有效率地存放、擷取、組織和共用基因體資料。HealthOmics 儲存體了解不同資料物件之間的關係，因此您可以定義哪些讀取集源自相同的來源資料。這為您提供了資料來源。

儲存為 ACTIVE 狀態的資料可立即擷取。30 天以上未存取的資料會儲存為 ARCHIVE 狀態。若要存取封存的資料，您可以透過 API 操作或主控台重新啟用。

HealthOmics 序列存放區旨在保留檔案的內容完整性。不過，因為在作用中和封存分層期間壓縮，所以不會保留匯入資料檔案和匯出檔案的位元相等性。

在擷取期間，HealthOmics 會產生實體標籤或 HealthOmics ETag，以驗證資料檔案的內容完整性。定序部分會在讀取集的來源層級識別並擷取為 ETag。ETag 計算不會改變實際的檔案或基因體資料。建立讀取集之後，ETag 不應在讀取集來源的整個生命週期中變更。這表示重新匯入相同的檔案會導致計算相同的 ETag 值。

主題

HealthOmics ETags和資料來源

HealthOmics ETag （實體標籤） 是序列存放區中擷取內容的雜湊。這可簡化資料擷取和處理，同時維持擷取資料檔案的內容完整性。ETag 會反映物件語意內容的變更，而非其中繼資料。指定的讀取集類型和演算法決定如何計算 ETag。ETag 計算不會改變實際的檔案或基因體資料。當讀取集的檔案類型結構描述允許時，序列存放區會更新連結至資料來源的欄位。

檔案具有位元身分和語意身分。位元身分表示檔案的位元相同，語意身分表示檔案的內容相同。語意身分在擷取檔案的內容完整性時，對中繼資料變更和壓縮變更具有彈性。

HealthOmics 序列存放區中的讀取集會在整個物件的生命週期進行壓縮/解壓縮週期和資料來源追蹤。在此處理期間，擷取檔案的位元身分可能會變更，並預期每次啟動檔案時都會變更；不過，會維護檔案的語意身分。語意身分會擷取為 HealthOmics 實體標籤，或在序列存放區擷取期間計算的 ETag，並以讀取集中繼資料的形式提供。

當讀取集的檔案類型結構描述允許時，序列存放區更新欄位會連結至資料來源。對於 uBAM、BAM 和 CRAM 檔案，新的 `@CO`或 `Comment`標籤會新增至 標頭。註解包含序列存放區 ID 和擷取時間戳記。

Amazon S3 ETags

使用 Amazon S3 URI 存取檔案時，Amazon S3 API 操作也可能傳回 Amazon S3 ETag 和檢查總和值。Amazon S3 ETag 和檢查總和值與 HealthOmics ETags 不同，因為它們代表檔案的位元身分。若要進一步了解描述性中繼資料和物件，請參閱 Amazon S3 物件 API 文件。Amazon S3 ETag 值可能會隨著讀取集的每個啟用週期而變更，您可以使用它們來驗證檔案的讀取。不過，請勿快取用於檔案生命週期內檔案身分驗證的 Amazon S3 ETag 值，因為這些值無法保持一致。相反地，HealthOmics ETag 在整個讀取集的生命週期中保持一致。

HealthOmics 如何計算 ETags

ETag 是從擷取檔案內容的雜湊產生。ETag 演算法系列預設為 MD5up，但可以在序列存放區建立期間以不同方式設定。計算 ETag 時，演算法和計算的雜湊會新增至讀取集。檔案類型支援的 MD5 演算法如下所示。

- FASTQ_MD5up – 計算未壓縮、完整 FASTQ 讀取集來源的 MD5 雜湊。
- BAM_MD5up – 計算未壓縮 BAM 或 uBAM 讀取集來源的對齊區段的 MD5 雜湊，如 SAM 所示，如果有連結的參考可用的話。
- CRAM_MD5up – 根據連結的參考，計算未壓縮 CRAM 讀取集來源之對齊區段的 MD5 雜湊，如 SAM 中所表示。

> **Note**
>
> 已知 MD5 雜湊容易遭受碰撞。因此，如果兩個不同的檔案為了利用已知的碰撞而製造，則可能具有相同的 ETag。

SHA256 系列支援下列演算法。演算法的計算方式如下：

- FASTQ_SHA256up – 計算未壓縮、完整 FASTQ 讀取集來源的 SHA-256 雜湊。
- BAM_SHA256up – 計算未壓縮 BAM 或 uBAM 讀取集來源的對齊區段的 SHA-256 雜湊，如 SAM 所示，如果可用，則根據連結的參考。
- CRAM_SHA256up – 根據連結的參考，計算未壓縮 CRAM 讀取集來源之對齊區段的 SHA-256 雜湊，如 SAM 所示。

SHA512 系列支援下列演算法。演算法的計算方式如下：

- FASTQ_SHA512up – 計算未壓縮、完整 FASTQ 讀取集來源的 SHA-512 雜湊。
- BAM_SHA512up – 計算未壓縮 BAM 或 uBAM 讀取集來源的對齊區段的 SHA-512 雜湊，如 SAM 所示，如果可用，則根據連結的參考。
- CRAM_SHA512up – 根據連結的參考，計算未壓縮 CRAM 讀取集來源之對齊區段的 SHA-512 雜湊，如 SAM 所示。

建立 HealthOmics 參考存放區

HealthOmics 中的參考存放區是用於儲存參考基因體的資料存放區。您可以在每個 AWS 帳戶 和區域中擁有單一參考存放區。您可以使用 主控台或 CLI 建立參考存放區。

主題

使用主控台建立參考存放區

建立參考存放區

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇開始使用 HealthOmics。
3. 從 Genomics 資料儲存選項中選擇參考基因體。
4. 您可以選擇先前匯入的參考基因體，或匯入新的參考基因體。如果您尚未匯入參考基因體，請選擇右上角的匯入參考基因體。

5. 在建立參考基因體匯入任務頁面上，選擇快速建立或手動建立選項來建立參考存放區，然後提供下列資訊。

   - 參考基因體名稱 - 此存放區的唯一名稱。
   - 描述 （選用） - 此參考存放區的描述。
   - IAM 角色 - 選取可存取參考基因體的角色。
   - 來自 Amazon S3 的參考 - 選取 Amazon S3 儲存貯體中的參考序列檔案。
   - 標籤 （選用） - 為此參考存放區提供最多 50 個標籤。

使用 CLI 建立參考存放區

下列範例示範如何使用 建立參考存放區 AWS CLI。每個 AWS 區域可以有一個參考存放區。

參考存放區支援儲存副檔名為 `.fasta`、`.fa`、`.fas`、`.fsa`、`.faa`、`.fna`、`.ffn`、`.frn.mpfa`、、、 `.seq`的 FASTA 檔案`.txt`。也支援這些擴充功能的`bgzip`版本。

在下列範例中，將 *`reference store name`*取代為您為參考存放區選擇的名稱。

```
aws omics create-reference-store --name "reference store name"
```

您會收到 JSON 回應，其中包含參考存放區 ID 和名稱、ARN，以及建立參考存放區時的時間戳記。

```
{
    "id": "3242349265",
    "arn": "arn:aws:omics:us-west-2:555555555555:referenceStore/3242349265",
    "name": "MyReferenceStore",
    "creationTime": "2022-07-01T20:58:42.878Z"
}
```

您可以在其他 AWS CLI 命令中使用參考存放區 ID。您可以使用 list-reference-stores 命令來擷取連結至您帳戶的參考存放區 IDs 清單，如下列範例所示。

```
aws omics list-reference-stores
```

為了回應，您收到新建立的參考存放區名稱。

```
{
    "referenceStores": [
```

```
        {
              "id": "3242349265",
              "arn": "arn:aws:omics:us-west-2:555555555555:referenceStore/3242349265",
              "name": "MyReferenceStore",
             "creationTime": "2022-07-01T20:58:42.878Z"
         }
     ]
}
```

建立參考存放區之後，您可以建立匯入任務，將基因體參考檔案載入其中。若要這樣做，您必須使用或建立 IAM 角色來存取資料。政策範例如下。

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject",
                "s3:GetBucketLocation"

            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket1",
                "arn:aws:s3:::amzn-s3-demo-bucket1/*"
            ]
        }
    ]
}
```

您還必須擁有類似下列範例的信任政策。

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
```

```
            "Effect": "Allow",
            "Principal": {
                "Service": [
                   "omics.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

您現在可以匯入參考基因體。此範例使用 Genome Reference Consortium Human Build 38 (hg38)，這是開放存取，可從 上的開放資料登錄 AWS檔取得。託管此資料的儲存貯體位於美國東部 （俄亥俄）。若要在其他 AWS 區域中使用儲存貯體，您可以將資料複製到您區域中託管的 Amazon S3 儲存貯體。使用下列 AWS CLI 命令將基因體複製到 Amazon S3 儲存貯體。

```
aws s3 cp s3://broad-references/hg38/v0/Homo_sapiens_assembly38.fasta s3://amzn-s3-
demo-bucket
```

然後，您可以開始匯入任務。*source file path* 使用您自己的輸入取代 *reference store IDrole ARN*、 和 。

```
aws omics start-reference-import-job --reference-store-id reference store ID --role-
arn role ARN --sources source file path
```

匯入資料後，您會以 JSON 收到下列回應。

```
{
        "id": "7252016478",
        "referenceStoreId": "3242349265",
        "roleArn": "arn:aws:iam::111122223333:role/OmicsReferenceImport",
        "status": "CREATED",
        "creationTime": "2022-07-01T21:15:13.727Z"
}
```

您可以使用下列命令來監控任務的狀態。在下列範例中，將 *reference store ID*和 取代*job ID*為您的參考存放區 ID，以及您想要進一步了解的任務 ID。

```
aws omics get-reference-import-job --reference-store-id reference store ID --id job ID
```

為了回應，您會收到回應，其中包含該參考存放區的詳細資訊及其狀態。

```
{
    "id": "7252016478",
    "referenceStoreId": "3242349265",
    "roleArn": "arn:aws:iam::555555555555:role/OmicsReferenceImport",
    "status": "RUNNING",
    "creationTime": "2022-07-01T21:15:13.727Z",
    "sources": [
        {
            "sourceFile": "s3://amzn-s3-demo-bucket/Homo_sapiens_assembly38.fasta",
            "status": "IN_PROGRESS",
            "name": "MyReference"
        }
    ]
}
```

您也可以列出您的參考，並根據參考名稱篩選它們，以找到匯入的參考。*reference store ID* 將取代為您的參考存放區 ID，並新增選用篩選條件以縮小清單範圍。

```
aws omics list-references --reference-store-id reference store ID --filter
 name=MyReference
```

為了回應，您會收到下列資訊。

```
{
    "references": [
        {
            "id": "1234567890",
            "arn": "arn:aws:omics:us-west-2:555555555555:referenceStore/1234567890/
reference/1234567890",
            "referenceStoreId": "12345678",
            "md5": "7ff134953dcca8c8997453bbb80b6b5e",
            "status": "ACTIVE",
            "name": "MyReference",
            "creationTime": "2022-07-02T00:15:19.787Z",
            "updateTime": "2022-07-02T00:15:19.787Z"
        }
```

```
    ]
}
```

若要進一步了解參考中繼資料，請使用 get-reference-metadata API 操作。在下列範例中，將 *reference store ID* 取代為您的參考存放區 ID，並將 取代*reference ID*為您想要進一步了解的參考 ID。

```
aws omics get-reference-metadata --reference-store-id reference store ID --id reference
 ID
```

您會收到以下資訊以回應。

```
{
    "id": "1234567890",
    "arn": "arn:aws:omics:us-west-2:555555555555:referenceStore/referencestoreID/
reference/referenceID",
    "referenceStoreId": "1234567890",
    "md5": "7ff134953dcca8c8997453bbb80b6b5e",
    "status": "ACTIVE",
    "name": "MyReference",
    "creationTime": "2022-07-02T00:15:19.787Z",
    "updateTime": "2022-07-02T00:15:19.787Z",
    "files": {
        "source": {
            "totalParts": 31,
            "partSize": 104857600,
            "contentLength": 3249912778
        },
        "index": {
            "totalParts": 1,
            "partSize": 104857600,
            "contentLength": 160928
        }
    }
}
```

您也可以使用 get-reference 下載部分參考檔案。在下列範例中，將 取代*reference store ID*為您的參考存放區 ID，並將 *reference ID* 取代為您要從中下載的參考 ID。

```
aws omics get-reference --reference-store-id reference store ID --id reference ID --
part-number 1 outfile.fa
```

建立 HealthOmics 序列存放區

HealthOmics 序列存放區支援以未對齊格式 FASTQ (僅限 Gzip) 和 儲存基因體檔案uBAM。它也支援 BAM和 的對齊格式CRAM。

匯入的檔案會儲存為讀取集。您可以新增標籤至讀取集，並使用 IAM 政策來控制讀取集的存取。對齊的讀取集需要參考基因體才能對齊基因體序列，但未對齊的讀取集是選用的。

若要存放讀取集，請先建立序列存放區。建立序列存放區時，您可以將選用的 Amazon S3 儲存貯體指定為備用位置，以及存放 S3 存取日誌的位置。備用位置用於儲存無法在直接上傳期間建立讀取集的任何檔案。備用位置適用於 2023 年 5 月 15 日之後建立的序列存放區。您可以在建立序列存放區時指定備用位置。

您最多可以指定五個讀取集標籤金鑰。當您使用符合其中一個金鑰的標籤金鑰建立或更新讀取集時，讀取集標籤會傳播到對應的 Amazon S3 物件。HealthOmics 建立的系統標籤預設會傳播。

主題

使用主控台建立序列存放區

建立序列存放區

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇序列存放區。
3. 在建立序列存放區頁面上，提供下列資訊
   - 序列存放區名稱 - 此存放區的唯一名稱。
   - 描述 （選用） - 此序列存放區的描述。
4. 針對 S3 中的備用位置，指定 Amazon S3 位置。HealthOmics 使用備用位置來存放無法在直接上傳期間建立讀取集的任何檔案。您需要授予 HealthOmics 服務對 Amazon S3 備用位置的寫入存取權。如需政策範例，請參閱 設定備用位置。

備用位置不適用於 2023 年 5 月 16 日之前建立的序列存放區。

5. （選用）對於 S3 傳播的讀取集標籤金鑰，您可以輸入最多五個讀取集金鑰，以從讀取集傳播到基礎 S3 物件。透過將標籤從讀取集傳播到 S3 物件，您可以根據標籤和/或最終使用者授予 S3 存取許可，以透過 Amazon S3 getObjectTagging API 操作查看傳播的標籤。

   a. 在文字方塊中輸入一個索引鍵值。主控台會建立新的文字方塊，以新增下一個金鑰。

   b. （選用）選擇移除以移除所有金鑰。

6. 在資料加密下，選取您希望資料加密由 擁有和管理， AWS 還是使用客戶受管 CMK。

7. （選用）在 S3 資料存取下，選取是否要建立新的角色和政策，以透過 Amazon S3 存取序列存放區。

8. （選用）對於 S3 存取記錄，Enabled如果您希望 Amazon S3 收集存取日誌記錄，請選取 。

   對於 S3 中的存取記錄位置，指定要存放日誌的 Amazon S3 位置。只有在您啟用 S3 存取記錄時，才會顯示此欄位。

9. 標籤（選用）- 為此序列存放區提供最多 50 個標籤。這些標籤與讀取集匯入/標籤更新期間設定的讀取集標籤不同

建立 存放區之後，即可使用 匯入基因體檔案。

使用 CLI 建立序列存放區

在下列範例中，*sequence store name*將 取代為您為序列存放區選擇的名稱。

```
aws omics create-sequence-store --name sequence store name --fallback-location "s3://
amzn-s3-demo-bucket"
```

您會以 JSON 收到下列回應，其中包含新建立序列存放區的 ID 號碼。

```
{
    "id": "3936421177",
    "arn": "arn:aws:omics:us-west-2:111122223333:sequenceStore/3936421177",
    "name": "sequence_store_example_name",
    "creationTime": "2022-07-13T20:09:26.038Z"
    "fallbackLocation" : "s3://amzn-s3-demo-bucket"
}
```

您也可以使用 list-sequence-stores 命令來檢視與您的帳戶相關聯的所有序列存放區，如下所示。

```
aws omics list-sequence-stores
```

您會收到下列回應。

```
{
    "sequenceStores": [
        {
            "arn": "arn:aws:omics:us-west-2:111122223333:sequenceStore/3936421177",
            "id": "3936421177",
            "name": "MySequenceStore",
            "creationTime": "2022-07-13T20:09:26.038Z",
            "updatedTime": "2024-09-13T04:11:31.242Z",
            "fallbackLocation" : "s3://amzn-s3-demo-bucket",
            "status": "Active"
        }
    ]
}
```

您可以使用 get-sequence-store 來進一步了解序列存放區，方法是使用其 ID，如下列範例所示：

```
aws omics get-sequence-store --id sequence store ID
```

您會收到下列回應：

```
{
  "arn": "arn:aws:omics:us-west-2:123456789012:sequenceStore/sequencestoreID",
  "creationTime": "2024-01-12T04:45:29.857Z",
  "updatedTime": "2024-09-13T04:11:31.242Z",
  "description": null,
  "fallbackLocation": null,
  "id": "2015356892",
  "name": "MySequenceStore",
  "s3Access": {
      "s3AccessPointArn": "arn:aws:s3:us-
west-2:123456789012:accesspoint/592761533288-2015356892",
      "s3Uri": "s3://592761533288-2015356892-ajdpi90jdas90a79fh9a8ja98jdfa9jf98-
s3alias/592761533288/sequenceStore/2015356892/",
      "accessLogLocation": "s3://IAD-seq-store-log/2015356892/"
  },
  "sseConfig": {
      "keyArn": "arn:aws:kms:us-west-2:123456789012:key/eb2b30f5-635d-4b6d-b0f9-
d3889fe0e648",
```

```
      "type": "KMS"
  },
  "status": "Active",
  "statusMessage": null,
  "setTagsToSync": ["withdrawn","protocol"],
}
```

建立之後，也可以更新數個存放區參數。這可以透過 主控台或 API `updateSequenceStore`操作來完成。

更新序列存放區

若要更新序列存放區，請遵循下列步驟：

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇序列存放區。
3. 選擇要更新的序列存放區。
4. 在詳細資訊面板中，選擇編輯。
5. 在編輯詳細資訊頁面上，您可以更新下列欄位：

   - 序列存放區名稱 - 此存放區的唯一名稱。
   - 描述 - 此序列存放區的描述。
   - S3 中的備用位置，指定 Amazon S3 位置。HealthOmics 使用備用位置來存放無法在直接上傳期間建立讀取集的任何檔案。
   - S3 傳播的讀取集標籤金鑰 您可以輸入最多五個讀取集金鑰以傳播到 Amazon S3。
   - （選用）對於 S3 存取記錄，`Enabled`如果您希望 Amazon S3 收集存取日誌記錄，請選取 。

     對於 S3 中的存取記錄位置，指定要存放日誌的 Amazon S3 位置。只有在您啟用 S3 存取記錄時，才會顯示此欄位。
   - 標籤（選用）- 為此序列存放區提供最多 50 個標籤。

更新序列存放區的讀取集標籤

若要更新序列存放區的讀取集標籤或其他欄位，請遵循下列步驟：

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇序列存放區。

3. 選擇您要更新的序列存放區。
4. 選擇詳細資訊索引標籤。
5. 選擇編輯。
6. 視需要新增新的讀取集標籤或刪除現有的標籤。
7. 視需要更新名稱、描述、備用位置或 S3 資料存取。
8. 選擇儲存變更。

匯入基因體檔案

若要將基因體檔案匯入序列存放區，請遵循下列步驟：

匯入基因體檔案

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇序列存放區。
3. 在序列存放區頁面上，選擇要匯入檔案的序列存放區。
4. 在個別序列存放區頁面上，選擇匯入基因體檔案。
5. 在指定匯入詳細資訊頁面上，提供下列資訊

   - IAM 角色 - 可存取 Amazon S3 上基因體檔案的 IAM 角色。
   - 參考基因體 - 此基因體資料的參考基因體。
6. 在指定匯入資訊清單頁面上，指定下列資訊資訊清單檔案。資訊清單檔案是 JSON 或 YAML 檔案，描述基因體資料的重要資訊。如需資訊清單檔案的資訊，請參閱 將讀取集匯入 HealthOmics 序列存放區。
7. 按一下建立匯入任務。

刪除 HealthOmics 參考和序列存放區

您可以刪除參考和序列存放區。序列存放區只能在不包含讀取集時刪除，而且只有在不包含參考時，才能刪除參考存放區。刪除序列或參考存放區也會刪除與該存放區相關聯的任何標籤。

下列範例示範如何使用 刪除參考存放區 AWS CLI。如果動作成功，您將不會收到回應。在下列範例中，將 取代*reference store ID*為您的參考存放區 ID。

```
aws omics delete-reference-store --id reference store ID
```

下列範例示範如何刪除序列存放區。如果動作成功，您不會收到回應。在下列範例中，將 取代*sequence store ID*為您的序列存放區 ID。

```
aws omics delete-sequence-store --id sequence store ID
```

您也可以刪除參考存放區中的參考，如下列範例所示。只有在未用於讀取集、變體存放區或註釋存放區時，才能刪除參考。在下列範例中，將 取代*reference store ID*為參考存放區 ID，並將 取代*reference ID*為要刪除的參考 ID。

```
aws omics delete-reference  --id reference ID --reference-store-id reference store ID
```

將讀取集匯入 HealthOmics 序列存放區

建立序列存放區之後，請建立匯入任務，將讀取集上傳至資料存放區。您可以從 Amazon S3 儲存貯體上傳檔案，也可以使用同步 API 操作直接上傳。Amazon S3 儲存貯體必須與序列存放區位於相同的區域。

您可以將任何已對齊和未對齊的讀取集組合上傳至序列存放區，不過，如果匯入中的任何讀取集已對齊，您必須包含參考基因體。

您可以重複使用用來建立參考存放區的 IAM 存取政策。

下列主題說明將讀取集匯入序列存放區，然後取得匯入資料的相關資訊時所遵循的主要步驟。

主題

- 將檔案上傳至 Amazon S3
- 建立清單檔案
- 啟動匯入任務
- 監控匯入任務
- 尋找匯入的序列檔案
- 取得讀取集的詳細資訊
- 下載讀取集資料檔案

將檔案上傳至 Amazon S3

下列範例示範如何將檔案移至 Amazon S3 儲存貯體。

```
aws s3 cp s3://1000genomes/phase1/data/HG00100/alignment/
HG00100.chrom20.ILLUMINA.bwa.GBR.low_coverage.20101123.bam s3://your-bucket
aws s3 cp s3://1000genomes/phase3/data/HG00146/sequence_read/SRR233106_1.filt.fastq.gz
 s3://your-bucket
aws s3 cp s3://1000genomes/phase3/data/HG00146/sequence_read/SRR233106_2.filt.fastq.gz
 s3://your-bucket
aws s3 cp s3://1000genomes/data/HG00096/alignment/
HG00096.alt_bwamem_GRCh38DH.20150718.GBR.low_coverage.cram s3://your-bucket
aws s3 cp s3://gatk-test-data/wgs_ubam/NA12878_20k/NA12878_A.bam s3://your-bucket
```

此範例中CRAM使用的範例 BAM 和 需要不同的基因體參考 Hg19和 Hg38。若要進一步了解或存取這些參考，請參閱 上的開放資料登錄檔中的廣泛 Genome 參考 AWS。

建立清單檔案

您還必須在 JSON 中建立資訊清單檔案，以在 中建立匯入任務的模型 `import.json`（請參閱下列範例）。如果您在主控台中建立序列存放區，則不需要指定 `sequenceStoreId`或 `roleARN`，因此資訊清單檔案會從 `sources` 輸入開始。

API manifest

下列範例使用 API 匯入三個讀取集：一個 BAM、FASTQ一個 和一個 CRAM。

```
{
  "sequenceStoreId": "3936421177",
  "roleArn": "arn:aws:iam::555555555555:role/OmicsImport",
  "sources":
  [
      {
          "sourceFiles":
          {
              "source1": "s3://amzn-s3-demo-bucket/
HG00100.chrom20.ILLUMINA.bwa.GBR.low_coverage.20101123.bam"
          },
          "sourceFileType": "BAM",
          "subjectId": "mySubject",
          "sampleId": "mySample",
          "referenceArn": "arn:aws:omics:us-
west-2:555555555555:referenceStore/0123456789/reference/0000000001",
          "name": "HG00100",
          "description": "BAM for HG00100",
```

```
            "generatedFrom": "1000 Genomes"
        },
        {
            "sourceFiles":
            {
                "source1": "s3://amzn-s3-demo-bucket/SRR233106_1.filt.fastq.gz",
                "source2": "s3://amzn-s3-demo-bucket/SRR233106_2.filt.fastq.gz"
            },
            "sourceFileType": "FASTQ",
            "subjectId": "mySubject",
            "sampleId": "mySample",
            // NOTE: there is no reference arn required here
            "name": "HG00146",
            "description": "FASTQ for HG00146",
            "generatedFrom": "1000 Genomes"
        },
        {
            "sourceFiles":
            {
                "source1": "s3://amzn-s3-demo-bucket/
HG00096.alt_bwamem_GRCh38DH.20150718.GBR.low_coverage.cram"
            },
            "sourceFileType": "CRAM",
            "subjectId": "mySubject",
            "sampleId": "mySample",
            "referenceArn": "arn:aws:omics:us-
west-2:555555555555:referenceStore/0123456789/reference/0000000001",
            "name": "HG00096",
            "description": "CRAM for HG00096",
            "generatedFrom": "1000 Genomes"
        },
        {
            "sourceFiles":
            {
                "source1": "s3://amzn-s3-demo-bucket/NA12878_A.bam"
            },
            "sourceFileType": "UBAM",
            "subjectId": "mySubject",
            "sampleId": "mySample",
            // NOTE: there is no reference arn required here
            "name": "NA12878_A",
            "description": "uBAM for NA12878",
            "generatedFrom": "GATK Test Data"
        }
```

```
  ]
}
```

Console manifest

此範例程式碼用於使用 主控台匯入單一讀取集。

```
[
  {
      "sourceFiles":
      {
          "source1": "s3://amzn-s3-demo-bucket/
HG00100.chrom20.ILLUMINA.bwa.GBR.low_coverage.20101123.bam"
      },
      "sourceFileType": "BAM",
      "subjectId": "mySubject",
      "sampleId": "mySample",
      "name": "HG00100",
      "description": "BAM for HG00100",
      "generatedFrom": "1000 Genomes"
  },
  {
      "sourceFiles":
      {
          "source1": "s3://amzn-s3-demo-bucket/SRR233106_1.filt.fastq.gz",
          "source2": "s3://amzn-s3-demo-bucket/SRR233106_2.filt.fastq.gz"
      },
      "sourceFileType": "FASTQ",
      "subjectId": "mySubject",
      "sampleId": "mySample",
      "name": "HG00146",
      "description": "FASTQ for HG00146",
      "generatedFrom": "1000 Genomes"
  },
  {
      "sourceFiles":
      {
          "source1": "s3://your-bucket/
HG00096.alt_bwamem_GRCh38DH.20150718.GBR.low_coverage.cram"
      },
      "sourceFileType": "CRAM",
      "subjectId": "mySubject",
      "sampleId": "mySample",
      "name": "HG00096",
```

```
        "description": "CRAM for HG00096",
        "generatedFrom": "1000 Genomes"
    },
    {
        "sourceFiles":
        {
            "source1": "s3://amzn-s3-demo-bucket/NA12878_A.bam"
        },
        "sourceFileType": "UBAM",
        "subjectId": "mySubject",
        "sampleId": "mySample",
        "name": "NA12878_A",
        "description": "uBAM for NA12878",
        "generatedFrom": "GATK Test Data"
    }
]
```

或者，您可以上傳 YAML 格式的資訊清單檔案。

啟動匯入任務

若要啟動匯入任務，請使用下列 AWS CLI 命令。

```
aws omics start-read-set-import-job --cli-input-json file://import.json
```

您會收到以下回應，表示任務建立成功。

```
{
  "id": "3660451514",
  "sequenceStoreId": "3936421177",
  "roleArn": "arn:aws:iam::111122223333:role/OmicsImport",
  "status": "CREATED",
  "creationTime": "2022-07-13T22:14:59.309Z"
}
```

監控匯入任務

匯入任務開始後，您可以使用下列命令來監控其進度。在下列範例中，將 取代*sequence store id*為您的序列存放區 ID，並將 取代*job import ID*為匯入 ID。

```
aws omics get-read-set-import-job --sequence-store-id sequence store id --id job import
 ID
```

以下顯示與指定序列存放區 ID 相關聯的所有匯入任務的狀態。

```
{
  "id": "1234567890",
  "sequenceStoreId": "1234567890",
  "roleArn": "arn:aws:iam::111122223333:role/OmicsImport",
  "status": "RUNNING",
  "statusMessage": "The job is currently in progress.",
  "creationTime": "2022-07-13T22:14:59.309Z",
  "sources": [
      {
          "sourceFiles":
          {
              "source1": "s3://amzn-s3-demo-bucket/
HG00100.chrom20.ILLUMINA.bwa.GBR.low_coverage.20101123.bam"
          },
          "sourceFileType": "BAM",
          "status": "IN_PROGRESS",
          "statusMessage": "The job is currently in progress."
          "subjectId": "mySubject",
          "sampleId": "mySample",
          "referenceArn": "arn:aws:omics:us-
west-2:111122223333:referenceStore/3242349265/reference/8625408453",
          "name": "HG00100",
          "description": "BAM for HG00100",
          "generatedFrom": "1000 Genomes",
          "readSetID": "1234567890"
      },
      {
          "sourceFiles":
          {
              "source1": "s3://amzn-s3-demo-bucket/SRR233106_1.filt.fastq.gz",
              "source2": "s3://amzn-s3-demo-bucket/SRR233106_2.filt.fastq.gz"
          },
          "sourceFileType": "FASTQ",
          "status": "IN_PROGRESS",
          "statusMessage": "The job is currently in progress."
          "subjectId": "mySubject",
          "sampleId": "mySample",
          "name": "HG00146",
```

```
            "description": "FASTQ for HG00146",
            "generatedFrom": "1000 Genomes",
            "readSetID": "1234567890"
        },
        {
            "sourceFiles":
            {
                "source1": "s3://amzn-s3-demo-bucket/
HG00096.alt_bwamem_GRCh38DH.20150718.GBR.low_coverage.cram"
            },
            "sourceFileType": "CRAM",
            "status": "IN_PROGRESS",
            "statusMessage": "The job is currently in progress."
            "subjectId": "mySubject",
            "sampleId": "mySample",
            "referenceArn": "arn:aws:omics:us-
west-2:111122223333:referenceStore/3242349265/reference/1234568870",
            "name": "HG00096",
            "description": "CRAM for HG00096",
            "generatedFrom": "1000 Genomes",
            "readSetID": "1234567890"
        },
        {
            "sourceFiles":
            {
                "source1": "s3://amzn-s3-demo-bucket/NA12878_A.bam"
            },
            "sourceFileType": "UBAM",
            "status": "IN_PROGRESS",
            "statusMessage": "The job is currently in progress."
            "subjectId": "mySubject",
            "sampleId": "mySample",
            "name": "NA12878_A",
            "description": "uBAM for NA12878",
            "generatedFrom": "GATK Test Data",
            "readSetID": "1234567890"
        }
    ]
}
```

尋找匯入的序列檔案

任務完成後，您可以使用 list-read-sets API 操作來尋找匯入的序列檔案。在下列範例中，將 取代*sequence store id*為您的序列存放區 ID。

```
aws omics list-read-sets --sequence-store-id sequence store id
```

您會收到下列回應。

```
{
  "readSets": [
      {
          "id": "0000000001",
          "arn": "arn:aws:omics:us-west-2:111122223333:sequenceStore/01234567890/
readSet/0000000001",
          "sequenceStoreId": "1234567890",
          "subjectId": "mySubject",
          "sampleId": "mySample",
          "status": "ACTIVE",
          "name": "HG00100",
          "description": "BAM for HG00100",
          "referenceArn": "arn:aws:omics:us-
west-2:111122223333:referenceStore/01234567890/reference/0000000001",
          "fileType": "BAM",
          "sequenceInformation": {
              "totalReadCount": 9194,
              "totalBaseCount": 928594,
              "generatedFrom": "1000 Genomes",
              "alignment": "ALIGNED"
          },
          "creationTime": "2022-07-13T23:25:20Z"
          "creationType": "IMPORT",
          "etag": {
              "algorithm": "BAM_MD5up",
              "source1": "d1d65429212d61d115bb19f510d4bd02"
          }
      },
      {
          "id": "0000000002",
          "arn": "arn:aws:omics:us-west-2:111122223333:sequenceStore/0123456789/
readSet/0000000002",
          "sequenceStoreId": "0123456789",
          "subjectId": "mySubject",
```

```
            "sampleId": "mySample",
            "status": "ACTIVE",
            "name": "HG00146",
            "description": "FASTQ for HG00146",
            "fileType": "FASTQ",
            "sequenceInformation": {
                "totalReadCount": 8000000,
                "totalBaseCount": 1184000000,
                "generatedFrom": "1000 Genomes",
                "alignment": "UNALIGNED"
            },
            "creationTime": "2022-07-13T23:26:43Z"
            "creationType": "IMPORT",
            "etag": {
                "algorithm": "FASTQ_MD5up",
                "source1": "ca78f685c26e7cc2bf3e28e3ec4d49cd"
            }
        },
        {
            "id": "0000000003",
            "arn": "arn:aws:omics:us-west-2:111122223333:sequenceStore/0123456789/
readSet/0000000003",
            "sequenceStoreId": "0123456789",
            "subjectId": "mySubject",
            "sampleId": "mySample",
            "status": "ACTIVE",
            "name": "HG00096",
            "description": "CRAM for HG00096",
            "referenceArn": "arn:aws:omics:us-
west-2:111122223333:referenceStore/0123456789/reference/0000000001",
            "fileType": "CRAM",
            "sequenceInformation": {
                "totalReadCount": 85466534,
                "totalBaseCount": 24000004881,
                "generatedFrom": "1000 Genomes",
                "alignment": "ALIGNED"
            },
            "creationTime": "2022-07-13T23:30:41Z"
            "creationType": "IMPORT",
            "etag": {
                "algorithm": "CRAM_MD5up",
                "source1": "66817940f3025a760e6da4652f3e927e"
            }
        },
```

```
      {
          "id": "0000000004",
          "arn": "arn:aws:omics:us-west-2:111122223333:sequenceStore/0123456789/
readSet/0000000004",
          "sequenceStoreId": "0123456789",
          "subjectId": "mySubject",
          "sampleId": "mySample",
          "status": "ACTIVE",
          "name": "NA12878_A",
          "description": "uBAM for NA12878",
          "fileType": "UBAM",
          "sequenceInformation": {
              "totalReadCount": 20000,
              "totalBaseCount": 5000000,
              "generatedFrom": "GATK Test Data",
              "alignment": "ALIGNED"
          },
          "creationTime": "2022-07-13T23:30:41Z"
          "creationType": "IMPORT",
          "etag": {
              "algorithm": "BAM_MD5up",
              "source1": "640eb686263e9f63bcda12c35b84f5c7"
          }
      }
  ]
}
```

取得讀取集的詳細資訊

若要檢視讀取集的詳細資訊，請使用 GetReadSetMetadata API 操作。在下列範例中，將 取代*sequence store id*為您的序列存放區 ID，並將 取代*read set id*為您的讀取集 ID。

```
aws omics get-read-set-metadata --sequence-store-id sequence store id --id read set id
```

您會收到下列回應。

```
{
"arn": "arn:aws:omics:us-west-2:123456789012:sequenceStore/2015356892/
readSet/9515444019",
"creationTime": "2024-01-12T04:50:33.548Z",
"creationType": "IMPORT",
```

```
"creationJobId": "33222111",
"description": null,
"etag": {
  "algorithm": "FASTQ_MD5up",
  "source1": "00d0885ba3eeb211c8c84520d3fa26ec",
  "source2": "00d0885ba3eeb211c8c84520d3fa26ec"
},
"fileType": "FASTQ",
"files": {
  "index": null,
  "source1": {
    "contentLength": 10818,
    "partSize": 104857600,
    "s3Access": {
      "s3Uri": "s3://accountID-sequence store ID-ajdpi90jdas90a79fh9a8ja98jdfa9jf98-
s3alias/592761533288/sequenceStore/2015356892/readSet/9515444019/
import_source1.fastq.gz"
},
    "totalParts": 1
  },
  "source2": {
    "contentLength": 10818,
    "partSize": 104857600,
    "s3Access": {
      "s3Uri": "s3://accountID-sequence store ID-ajdpi90jdas90a79fh9a8ja98jdfa9jf98-
s3alias/592761533288/sequenceStore/2015356892/readSet/9515444019/
import_source1.fastq.gz"
    },
    "totalParts": 1
  }
},
"id": "9515444019",
"name": "paired-fastq-import",
"sampleId": "sampleId-paired-fastq-import",
"sequenceInformation": {
  "alignment": "UNALIGNED",
  "generatedFrom": null,
  "totalBaseCount": 30000,
  "totalReadCount": 200
},
"sequenceStoreId": "2015356892",
"status": "ACTIVE",
"statusMessage": null,
"subjectId": "subjectId-paired-fastq-import"
```

```
}
```

下載讀取集資料檔案

您可以使用 Amazon S3 GetObject API 操作存取作用中讀取集的物件。物件的 URI 會在 GetReadSetMetadata API 回應中傳回。如需詳細資訊，請參閱使用 Amazon S3 URIs 存取 HealthOmics 讀取集。

或者，使用 HealthOmics GetReadSet API 操作。您可以使用 下載個別組件GetReadSet，以平行下載 。這些部分類似於 Amazon S3 部分。以下是如何從讀取集下載第 1 部分的範例。在下列範例中，將 取代*sequence store id*為您的序列存放區 ID，並將 取代*read set id*為您的讀取集 ID。

```
aws omics get-read-set --sequence-store-id sequence store id --id read set id  --part-
number 1 outfile.bam
```

您也可以使用 HealthOmics Transfer Manager 下載 HealthOmics 參考或讀取集的檔案。您可以在這裡下載 HealthOmics Transfer Manager。如需使用和設定 Transfer Manager 的詳細資訊，請參閱此 GitHub 儲存庫。

直接上傳至 HealthOmics 序列存放區

我們建議您使用 HealthOmics Transfer Manager 將檔案新增至序列存放區。如需使用 Transfer Manager 的詳細資訊，請參閱此 GitHub 儲存庫。您也可以透過直接上傳 API 操作，將讀取集直接上傳至序列存放區。

直接上傳讀取集會先處於 PROCESSING_UPLOAD 狀態。這表示檔案部分目前正在上傳，而且您可以存取讀取集中繼資料。上傳組件並驗證檢查總和後，讀取集會變成ACTIVE和行為與匯入的讀取集相同。

如果直接上傳失敗，讀取集狀態會顯示為 UPLOAD_FAILED。您可以將 Amazon S3 儲存貯體設定為無法上傳檔案的備用位置。備用位置適用於 2023 年 5 月 15 日之後建立的序列存放區。

主題

- 使用 直接上傳至序列存放區 AWS CLI
- 設定備用位置

使用 直接上傳至序列存放區 AWS CLI

若要開始，請開始分段上傳。您可以使用 來執行此操作 AWS CLI，如下列範例所示。

使用 AWS CLI 命令直接上傳

1. 透過分隔資料來建立組件，如下列範例所示。

```
split -b 100MiB SRR233106_1.filt.fastq.gz source1_part_
```

2. 將來源檔案分段後，請建立分段讀取集上傳，如下列範例所示。*sequence store ID* 將 和其他參數取代為您的序列存放區 ID 和其他值。

```
aws omics create-multipart-read-set-upload \
--sequence-store-id sequence store ID \
--name upload name \
--source-file-type FASTQ \
--subject-id subject ID \
--sample-id sample ID \
--description "FASTQ for HG00146" "description of upload" \
--generated-from "1000 Genomes""source of imported files"
```

您可以在回應中取得 uploadID和其他中繼資料。使用 uploadID進行上傳程序的下一個步驟。

```
{
"sequenceStoreId": "1504776472",
"uploadId": "7640892890",
"sourceFileType": "FASTQ",
"subjectId": "mySubject",
"sampleId": "mySample",
"generatedFrom": "1000 Genomes",
"name": "HG00146",
"description": "FASTQ for HG00146",
"creationTime": "2023-11-20T23:40:47.437522+00:00"
}
```

3. 將讀取集新增至上傳。如果您的檔案夠小，您只需要執行此步驟一次。對於較大的檔案，您可以對檔案的每個部分執行此步驟。如果您使用先前使用的組件編號上傳新組件，則會覆寫先前上傳的組件。

在下列範例中，將 *sequence store ID*、 *upload ID*和其他參數取代為您的值。

```
aws omics upload-read-set-part \
--sequence-store-id sequence store ID \
--upload-id upload ID \
--part-source SOURCE1 \
```

```
--part-number part number \
--payload  source1/source1_part_aa.fastq.gz
```

回應是您可以用來驗證上傳檔案是否符合您預期檔案的 ID。

```
{
"checksum": "984979b9928ae8d8622286c4a9cd8e99d964a22d59ed0f5722e1733eb280e635"
}
```

4. 如有必要，請繼續上傳檔案的部分。若要確認您的讀取集已上傳，請使用 list-read-set-upload-parts API 操作，如下所示。在下列範例中，將 *sequence store ID* 、 *upload ID*和 取代*part source*為您自己的輸入。

```
aws omics list-read-set-upload-parts \
 --sequence-store-id sequence store ID \
 --upload-id upload ID \
 --part-source SOURCE1
```

回應會傳回讀取集的數量、大小，以及最近更新的時間戳記。

```
{
"parts": [
    {
        "partNumber": 1,
        "partSize": 104857600,
        "partSource": "SOURCE1",
        "checksum": "MVMQk+vB9C3Ge8ADHkbKq752n3BCUzyl41qEkqlOD5M=",
        "creationTime": "2023-11-20T23:58:03.500823+00:00",
        "lastUpdatedTime": "2023-11-20T23:58:03.500831+00:00"
    },
    {
        "partNumber": 2,
        "partSize": 104857600,
        "partSource": "SOURCE1",
        "checksum": "keZzVzJNChAqgOdZMvOmjBwrOPM0enPj1UAfs0nvRto=",
        "creationTime": "2023-11-21T00:02:03.813013+00:00",
        "lastUpdatedTime": "2023-11-21T00:02:03.813025+00:00"
    },
    {
        "partNumber": 3,
        "partSize": 100339539,
        "partSource": "SOURCE1",
```

```
        "checksum": "TBkNfMsaeDpXzEf3ldlbi0ipFDPaohKHyZ+LF1J4CHk=",
        "creationTime": "2023-11-21T00:09:11.705198+00:00",
        "lastUpdatedTime": "2023-11-21T00:09:11.705208+00:00"
    }
]
}
```

5. 若要檢視所有作用中的分段讀取集上傳，請使用 list-multipart-read-set-uploads，如下所示。*sequence store ID* 將 取代為您自己的序列存放區的 ID。

```
aws omics list-multipart-read-set-uploads --sequence-store-id
            sequence store ID
```

此 API 只會傳回進行中的分段讀取集上傳。擷取的讀取集為 之後ACTIVE，或者如果上傳失敗，將不會在 list-multipart-read-set-uploads API 的回應中傳回上傳。若要檢視作用中的讀取集，請使用 list-read-sets API。listlist-multipart-read-set-uploads 的範例回應如下所示。

```
{
"uploads": [
    {
        "sequenceStoreId": "1234567890",
        "uploadId": "8749584421",
        "sourceFileType": "FASTQ",
        "subjectId": "mySubject",
        "sampleId": "mySample",
        "generatedFrom": "1000 Genomes",
        "name": "HG00146",
        "description": "FASTQ for HG00146",
        "creationTime": "2023-11-29T19:22:51.349298+00:00"
    },
    {
        "sequenceStoreId": "1234567890",
        "uploadId": "5290538638",
        "sourceFileType": "BAM",
        "subjectId": "mySubject",
        "sampleId": "mySample",
        "generatedFrom": "1000 Genomes",
        "referenceArn": "arn:aws:omics:us-
west-2:123456789012:referenceStore/8168613728/reference/2190697383",
        "name": "HG00146",
        "description": "BAM for HG00146",
```

```
        "creationTime": "2023-11-29T19:23:33.116516+00:00"
    },
    {
        "sequenceStoreId": "1234567890",
        "uploadId": "4174220862",
        "sourceFileType": "BAM",
        "subjectId": "mySubject",
        "sampleId": "mySample",
        "generatedFrom": "1000 Genomes",
        "referenceArn": "arn:aws:omics:us-
west-2:123456789012:referenceStore/8168613728/reference/2190697383",
        "name": "HG00147",
        "description": "BAM for HG00147",
        "creationTime": "2023-11-29T19:23:47.007866+00:00"
    }
]
}
```

6. 上傳檔案的所有部分之後，請使用 complete-multipart-read-set-upload 結束上傳程序，如下列範例所示。使用您自己的值取代組件的 *sequence store IDupload ID*、 和 參數。

```
aws omics complete-multipart-read-set-upload \
--sequence-store-id sequence store ID \
--upload-id upload ID \
--parts '[{"checksum":"gaCBQMe+rpCFZxLpoP6gydBoXaKKDA/
Vobh5zBDb4W4=","partNumber":1,"partSource":"SOURCE1"}]'
```

complete-multipart-read-set-upload 的回應是匯入讀取集IDs。

```
{
"readSetId": "0000000001"
}
```

7. 若要停止上傳，請使用 abort-multipart-read-set-upload 搭配上傳 ID 來結束上傳程序。*upload ID* 使用您自己的參數值取代 *sequence store ID*和 。

```
aws omics abort-multipart-read-set-upload \
--sequence-store-id sequence store ID \
--upload-id upload ID
```

8. 上傳完成後，請使用 get-read-set 從讀取集擷取您的資料，如下所示。如果上傳仍在處理中，get-read-set 會傳回有限的中繼資料，而且產生的索引檔案無法使用。使用您自己的輸入取代 *sequence store ID* 和其他參數。

```
aws omics get-read-set
 --sequence-store-id sequence store ID \
 --id read set ID \
 --file SOURCE1 \
 --part-number 1 myfile.fastq.gz
```

9. 若要檢查中繼資料，包括上傳的狀態，請使用 get-read-set-metadata API 操作。

```
aws omics get-read-set-metadata --sequence-store-id sequence store ID --id read set
 ID
```

回應包含中繼資料詳細資訊，例如檔案類型、參考 ARN、檔案數量，以及序列的長度。它還包含狀態。可能的狀態為 PROCESSING_UPLOAD、 ACTIVE和 UPLOAD_FAILED。

```
{
"id": "0000000001",
"arn": "arn:aws:omics:us-west-2:555555555555:sequenceStore/0123456789/
readSet/0000000001",
"sequenceStoreId": "0123456789",
"subjectId": "mySubject",
"sampleId": "mySample",
"status": "PROCESSING_UPLOAD",
"name": "HG00146",
"description": "FASTQ for HG00146",
"fileType": "FASTQ",
"creationTime": "2022-07-13T23:25:20Z",
"files": {
    "source1": {
        "totalParts": 5,
        "partSize": 123456789012,
        "contentLength": 6836725,

    },
    "source2": {
        "totalParts": 5,
        "partSize": 123456789056,
        "contentLength": 6836726
    }
```

```
},
'creationType": "UPLOAD"
}
```

設定備用位置

當您建立或更新序列存放區時，您可以將 Amazon S3 儲存貯體設定為無法上傳檔案的備用位置。這些讀取集的檔案部分會傳輸至後援位置。備用位置適用於 2023 年 5 月 15 日之後建立的序列存放區。

建立 Amazon S3 儲存貯體政策，以授予 HealthOmics 對 Amazon S3 備用位置的寫入存取權，如下列範例所示：

```
{
    "Effect": "Allow",
    "Principal": {
        "Service": "omics.amazonaws.com"
    },
    "Action": "s3:PutObject",
    "Resource": "arn:aws:s3:::amzn-s3-demo-bucket/*"
}
```

如果備用或存取日誌的 Amazon S3 儲存貯體使用客戶受管金鑰，請將下列許可新增至金鑰政策：

```
 {
    "Sid": "Allow use of key",
    "Effect": "Allow",
    "Principal": {
        "Service": "omics.amazonaws.com"
    },
    "Action": [
        "kms:Decrypt",
        "kms:GenerateDataKey*"
    ],
    "Resource": "*"
}
```

將 HealthOmics 讀取集匯出至 Amazon S3 儲存貯體

您可以將讀取集匯出為批次匯出任務到 Amazon S3 儲存貯體。若要這樣做，請先建立具有 儲存貯體寫入存取權的 IAM 政策，類似於下列 IAM 政策範例。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:PutObject",
        "s3:GetBucketLocation"
      ],
      "Resource": [
        "arn:aws:s3:::amzn-s3-demo-bucket1",
        "arn:aws:s3:::amzn-s3-demo-bucket1/*"
      ]
    }
  ]
}
```

JSON

```
{
"Version": "2012-10-17",
"Statement": [
  {
      "Effect": "Allow",
      "Principal": {
          "Service": [
              "omics.amazonaws.com"
          ]
      },
      "Action": "sts:AssumeRole"
  }
]
}
```

IAM 政策就緒後，開始您的讀取集匯出任務。下列範例示範如何使用 start-read-set-export-job API 操作來執行此操作。在下列範例中，以您的輸入取代所有參數，例如 *sources*、、 *sequence store ID* *destination role ARN*和 。

```
aws omics start-read-set-export-job
--sequence-store-id sequence store id \
--destination valid s3 uri \
--role-arn role ARN \
--sources readSetId=read set id_1 readSetId=read set id_2
```

您會收到下列回應，其中包含原始伺服器序列存放區和目的地 Amazon S3 儲存貯體的資訊。

```
{
"id": <job-id>,
"sequenceStoreId": <sequence-store-id>,
"destination": <destination-s3-uri>,
"status": "SUBMITTED",
"creationTime": "2022-10-22T01:33:38.079000+00:00"
}
```

任務開始後，您可以使用 get-read-set-export-job API 操作來判斷其狀態，如下所示。將 *sequence store ID*和 分別取代*job ID*為您的序列存放區 ID 和任務 ID。

```
aws omics get-read-set-export-job --id job-id --sequence-store-id sequence store ID
```

您可以使用 list-read-set-export-jobs API 操作來檢視序列存放區初始化的所有匯出任務，如下所示。將 取代*sequence store ID*為您的序列存放區 ID。

```
aws omics list-read-set-export-jobs --sequence-store-id sequence store ID.
```

```
{
"exportJobs": [
  {
      "id": <job-id>,
      "sequenceStoreId": <sequence-store-id>,
      "destination": <destination-s3-uri>,
      "status": "COMPLETED",
      "creationTime": "2022-10-22T01:33:38.079000+00:00",
      "completionTime": "2022-10-22T01:34:28.941000+00:00"
  }
]
}
```

除了匯出您的讀取集之外，您也可以使用 Amazon S3 存取 URIs 來共用它們。如需進一步了解，請參閱使用 Amazon S3 URIs 存取 HealthOmics 讀取集。

使用 Amazon S3 URIs 存取 HealthOmics 讀取集

您可以使用 Amazon S3 URI 路徑來存取作用中的序列存放區讀取集。

透過 Amazon S3 URI 路徑，您可以使用 Amazon S3 操作來列出、共用和下載您的讀取集。透過 S3 APIs存取 可加速協作和工具整合，因為許多產業工具已建置為可從 S3 讀取。此外，您可以與其他 帳戶共用對 S3 APIs存取權，並提供資料的跨區域讀取存取權。

HealthOmics 不支援 Amazon S3 URI 存取封存的讀取集。當您啟用讀取集時，每次都會還原至相同的 URI 路徑。

將資料載入 HealthOmics 存放區時，由於 Amazon S3 URI 是以 Amazon S3 存取點為基礎，因此您可以直接與讀取 Amazon S3 URIs 的業界標準工具整合，如下所示：

- 視覺化分析應用程式，例如 Integrative Genomics Viewer (IGV) 或 UCSC Genome 瀏覽器。
- 具有 Amazon S3 延伸模組的常見工作流程，例如 CWL、WDL 和 Nextflow。
- 任何可以驗證和讀取存取點 Amazon S3 URIs 或讀取預先簽章 Amazon S3 URIs 的工具。
- Amazon S3 公用程式，例如掛載點或 CloudFront。

Amazon S3 掛載點可讓您使用 Amazon S3 儲存貯體做為本機檔案系統。若要進一步了解掛載點並安裝以供使用，請參閱適用於 Amazon S3 的掛載點。

Amazon CloudFront 是一種內容交付網路 (CDN) 服務，專為高效能、安全性和開發人員便利性而打造。若要進一步了解如何使用 Amazon CloudFront，請參閱 Amazon CloudFront 文件。若要使用序列存放區設定 CloudFront，請聯絡 AWS HealthOmics 團隊。

序列存放區字首上的動作 S3：GetObject、S3：GetObjectTagging 和 S3：List 儲存貯體已啟用資料擁有者根帳戶。若要讓帳戶中的使用者存取資料，您可以建立 IAM 政策，並將其連接到使用者或角色。如需政策範例，請參閱 使用 Amazon S3 URIs 存取資料的許可。

您可以在作用中讀取集上使用下列 Amazon S3 API 操作來列出和擷取您的資料。啟用封存的讀取集後，您可以透過 Amazon S3 URIs 存取它們。

- GetObject - 從 Amazon S3 擷取物件。
- HeadObject – HEAD 操作會從物件擷取中繼資料，而不會傳回物件本身。如果您只想要物件的中繼資料，此操作很有用。

- ListObjects 和 ListObject v2 — 傳回儲存貯體中部分或全部（最多 1，000 個）的物件。
- CopyObject — 建立已存放在 Amazon S3 中的物件複本。HealthOmics 支援複製到 Amazon S3 存取點，但不支援寫入存取點。

HealthOmics 序列存放區會透過 ETags 維護檔案的語意身分。在檔案的整個生命週期中，以位元身分為基礎的 Amazon S3 ETag 可能會變更，不過，HealthOmics ETag 會保持不變。如需詳細資訊，請參閱 HealthOmics ETags和資料來源。

主題

- HealthOmics 儲存體中的 Amazon S3 URI 結構
- 使用託管或本機 IGV 存取讀取集
- 在 HealthOmics 中使用 Samtools 或 HTSlib
- 使用掛載點 HealthOmics
- 搭配 HealthOmics 使用 CloudFront

HealthOmics 儲存體中的 Amazon S3 URI 結構

具有 Amazon S3 URIs 的所有檔案都有 `omics:sampleId` `omics:subjectId`和資源標籤。您可以使用這些標籤，透過 等模式使用 IAM 政策來共用存取權`"s3:ExistingObjectTag/omics:subjectId": "pattern desired"`。

檔案結構如下：

`.../`*`account_id`*`/sequenceStore/`*`seq_store_id`*`/readSet/`*`read_set_id`*`/`*`files`*`.`

對於從 Amazon S3 匯入序列存放區的檔案，序列存放區會嘗試維護原始來源名稱。當名稱發生衝突時，系統會附加讀取集資訊，以確保檔案名稱是唯一的。例如，對於 fastq 讀取集，如果兩個檔案名稱相同，為了使名稱唯一，`sourceX`會在 .fastq.gz 或 .fq.gz 之前插入。對於直接上傳，檔案名稱會遵循下列模式：

- 對於 FASTQ — *`read_set_name_sourcex`*.fastq.gz
- 對於 uBAM/BAM/CRAM— *`read_set_name.file ##`*名，副檔名為 `.bam`或 `.cram`。例如，`NA193948.bam`。

對於 BAM 或 CRAM 的讀取集，索引檔案會在擷取過程中自動產生。對於產生的索引檔案，檔案名稱結尾會套用適當的索引副檔名。它具有模式 *`##########>.<########`*索引延伸為 `.bai`或 `.crai`。

使用託管或本機 IGV 存取讀取集

IGV 是一種用於分析 BAM 和 CRAM 檔案的基因體瀏覽器。它同時需要 檔案和 索引，因為它一次只會顯示一部分的基因體。IGV 可在本機下載和使用，並有建立 AWS 託管 IGV 的指南。不支援公有 Web 版本，因為它需要 CORS。

本機 IGV 依賴本機 AWS 組態來存取檔案。確定該組態中使用的角色已連接政策，以啟用對所存取讀取集的 s3 URI 的 kms:Decrypt 和 s3：GetObject 許可。之後，在 IGV 中，您可以使用「從 URL 檔案 > 載入」並在 URI 中貼上來源和索引。或者，可以用相同的方式產生和使用預先簽章URLs，這會略過 AWS 組態。請注意，Amazon S3 URI 存取不支援 CORS，因此不支援依賴 CORS 的請求。

AWS 託管 IGV 範例依賴 AWS Cognito 在環境中建立正確的組態和許可。確保已建立政策以啟用 Amazon S3 URI 的 Amazon S3 URI：Decrypt 和 s3：GetObject 許可，並將此政策新增至指派給 Cognito 使用者集區的角色。之後，在 IGV 中，您可以使用「從 URL 檔案 > 載入」，然後在 URI 中輸入來源和索引。或者，可以用相同的方式產生和使用預先簽章URLs，這會略過 AWS 組態。

請注意，序列存放區不會出現在「Amazon」索引標籤下，因為 只會顯示您在設定 AWS 設定檔的區域中擁有的儲存貯體。

在 HealthOmics 中使用 Samtools 或 HTSlib

HTSlib 是由多種工具共用的核心程式庫，例如 Samtools、rSamtools、PySam 等。使用 HTSlib 1.20 版或更新版本，取得 Amazon S3 存取點的無縫支援。對於較舊版本的 HTSlib 程式庫，您可以使用下列解決方法：

- 使用 設定 HTS Amazon S3 主機的環境變數：`export HTS_S3_HOST="s3.region.amazonaws.com"`。
- 為您要使用的檔案產生預先簽章的 URL。如果使用 BAM 或 CRAM，請確保為檔案和索引產生預先簽章的 URL。之後，這兩個檔案都可以與程式庫搭配使用。
- 使用掛載點，在您使用 HTSlib 程式庫的相同環境中掛載序列存放區或讀取集字首。從這裡，可以使用本機檔案路徑存取檔案。

使用掛載點 HealthOmics

Amazon S3 掛載點是簡單的高輸送量檔案用戶端，可將 Amazon S3 儲存貯體掛載為本機檔案系統。透過適用於 Amazon S3 的掛載點，您的應用程式可以透過開啟和讀取等檔案操作存取存放在 Amazon S3 中的物件。Amazon S3 掛載點會自動將這些操作轉譯為 Amazon S3 物件 API 呼叫，讓您的應用程式透過檔案界面存取 Amazon S3 的彈性儲存體和輸送量。

您可以使用掛載點安裝指示來安裝掛載點。掛載點使用安裝的本機 AWS 設定檔，可在 Amazon S3 字首層級運作。確保使用的設定檔具有啟用 s3：GetObject、s3：ListBucket 和 kms:Decrypt 許可的政策，以存取讀取集（或序列存放區）的 Amazon S3 URI 字首。之後，可以使用下列路徑掛載儲存貯體：

```
mount-s3 access point arn local path to mount --prefix prefix to sequence store or read
 set --region region
```

搭配 HealthOmics 使用 CloudFront

Amazon CloudFront 是一種內容交付網路 (CDN) 服務，專為高效能、安全性和開發人員便利性而打造。想要使用 CloudFront 的客戶必須與 Service 團隊合作，以開啟 CloudFront 分佈。與您的客戶團隊合作，讓 HealthOmics 服務團隊參與。

在 HealthOmics 中啟用讀取集

您可以使用 start-read-set-activation-job API 操作或透過 啟用封存的讀取集 AWS CLI，如下列範例所示。將 *sequence store ID*和 取代*read set id*為您的序列存放區 ID 和讀取集 IDs。

```
aws omics start-read-set-activation-job
     --sequence-store-id sequence store ID \
     --sources readSetId=read set ID readSetId=read set id_1 read set id_2
```

您會收到包含啟用任務資訊的回應，如下所示。

```
{
    "id": "12345678",
    "sequenceStoreId": "1234567890",
    "status": "SUBMITTED",
    "creationTime": "2022-10-22T00:50:54.670000+00:00"
}
```

啟動任務後，您可以使用 get-read-set-activation-job API 操作來監控其進度。以下是如何使用 AWS CLI 來檢查啟用任務狀態的範例。將 *job ID*和 分別取代*sequence store ID*為您的序列存放區 ID 和任務 IDs。

```
aws omics get-read-set-activation-job --id job ID --sequence-store-id sequence store ID
```

回應摘要說明啟用任務，如下所示。

```
{
    "id": 123567890,
    "sequenceStoreId": 123467890,
    "status": "SUBMITTED",
    "statusUpdateReason": "The job is submitted and will start soon.",
    "creationTime": "2022-10-22T00:50:54.670000+00:00",
    "sources": [
        {
            "readSetId": <reads set id_1>,
            "status": "NOT_STARTED",
            "statusUpdateReason": "The source is queued for the job."
        },
        {
            "readSetId": <read set id_2>,
            "status": "NOT_STARTED",
            "statusUpdateReason": "The source is queued for the job."
        }
    ]
}
```

您可以使用 get-read-set-metadata API 操作檢查啟用任務的狀態。可能的狀態為 ACTIVE、ACTIVATING和 ARCHIVED。在下列範例中，將 取代*sequence store ID*為您的序列存放區 ID，並將 取代*read set ID*為您的讀取集 ID。

```
aws omics get-read-set-metadata --sequence-store-id sequence store ID --id read set ID
```

下列回應顯示讀取集處於作用中狀態。

```
{
    "id": "12345678",
    "arn": "arn:aws:omics:us-west-2:555555555555:sequenceStore/1234567890/
readSet/12345678",
    "sequenceStoreId": "0123456789",
    "subjectId": "mySubject",
    "sampleId": "mySample",
    "status": "ACTIVE",
    "name": "HG00100",
    "description": "HG00100 aligned to HG38 BAM",
    "fileType": "BAM",
    "creationTime": "2022-07-13T23:25:20Z",
```

```
    "sequenceInformation": {
        "totalReadCount": 1513467,
        "totalBaseCount": 163454436,
        "generatedFrom": "Pulled from SRA",
        "alignment": "ALIGNED"
    },
    "referenceArn": "arn:aws:omics:us-west-2:555555555555:referenceStore/0123456789/
reference/0000000001",
    "files": {
        "source1": {
            "totalParts": 2,
            "partSize":  10485760,
            "contentLength": 17112283,
            "s3Access": {
        "s3Uri": "s3://accountID-sequence store ID-ajdpi90jdas90a79fh9a8ja98jdfa9jf98-
s3alias/592761533288/sequenceStore/2015356892/readSet/9515444019/
import_source1.fastq.gz"
},
         },
        "index": {
            "totalParts": 1,
            "partSize": 53216,
            "contentLength": 10485760
            "s3Access": {
        "s3Uri": "s3://accountID-sequence store ID-ajdpi90jdas90a79fh9a8ja98jdfa9jf98-
s3alias/592761533288/sequenceStore/2015356892/readSet/9515444019/
import_source1.fastq.gz"
},
        }
    },
    "creationType": "IMPORT",
    "etag": {
        "algorithm": "BAM_MD5up",
        "source1": "d1d65429212d61d115bb19f510d4bd02"
    }
}
```

您可以使用 list-read-set-activation-jobs 檢視所有讀取集啟用任務，如下列範例所示。在下列範例中，將 取代*sequence store ID*為您的序列存放區 ID。

```
aws omics list-read-set-activation-jobs --sequence-store-id sequence store ID
```

您會收到下列回應。

```
{
    "activationJobs": [
        {
            "id": 1234657890,
            "sequenceStoreId": "1234567890",
            "status": "COMPLETED",
            "creationTime": "2022-10-22T01:33:38.079000+00:00",
            "completionTime": "2022-10-22T01:34:28.941000+00:00"
        }
    ]
}
```

HealthOmics 分析

HealthOmics 分析支援儲存和分析基因體變體和註釋。Analytics 提供兩種類型的儲存資源：變體存放區和註釋存放區。您可以使用這些資源來存放、轉換和查詢基因體變體資料和註釋資料。將資料匯入資料存放區後，您可以使用 Athena 對資料進行進階分析。

您可以使用 HealthOmics 主控台或 API 來建立和管理存放區、匯入資料，以及與協作者共用分析存放區資料。

變體會以 VCF 格式存放支援資料，而註釋存放區支援 TSV/CSV 和 GFF3 格式。基因體座標表示為以零為基礎的半封閉半開間隔。當您的資料位於 HealthOmics 分析資料存放區時，將透過 管理對 VCF 檔案的存取 AWS Lake Formation。然後，您可以使用 Amazon Athena 查詢 VCF 檔案。查詢必須使用 Athena 查詢引擎版本 3。若要進一步了解 Athena 查詢引擎版本，請參閱 Amazon Athena 文件。

主題

建立 HealthOmics 變體存放區

下列主題說明如何使用 主控台和 API 建立 HealthOmics 變體存放區。

主題

使用主控台建立變體存放區

您可以使用 HealthOmics 主控台建立變體存放區。

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇變體存放區。
3. 在建立變體存放區頁面上，提供下列資訊
   - 變體存放區名稱 - 此存放區的唯一名稱。
   - 描述 （選用） - 此變體存放區的描述。
   - 參考基因體 - 此變體存放區的參考基因體。
   - 資料加密 - 選擇您希望資料加密是由 AWS 或您自己擁有和管理。
   - 標籤 （選用） - 為此變體存放區提供最多 50 個標籤。
4. 選擇建立變體存放區。

使用 API 建立變體存放區

使用 HealthOmics `CreateVariantStore` API 操作來建立變體存放區。您也可以使用 執行此操作 AWS CLI。

若要建立變體存放區，請提供存放區的名稱和參考存放區的 ARN。當變體存放區的狀態變更為 READY 時，已準備好擷取資料。

下列範例使用 AWS CLI 來建立變體存放區。

```
aws omics create-variant-store --name myvariantstore \
    --reference referenceArn="arn:aws:omics:us-
west-2:555555555555:referenceStore/123456789/reference/5987565360"
```

若要確認建立變體存放區，您會收到下列回應。

```
{
    "creationTime": "2022-11-03T18:19:52.296368+00:00",
    "id": "45aeb91d5678",
    "name": "myvariantstore",
    "reference": {
        "referenceArn": "arn:aws:omics:us-west-2:555555555555:referenceStore/123456789/
reference/5987565360"
```

```
    },
    "status": "CREATING"
}
```

若要進一步了解變體存放區，請使用 get-variant-store API。

```
aws omics get-variant-store --name myvariantstore
```

您會收到下列回應。

```
{
    "id": "45aeb91d5678",
    "reference": {
        "referenceArn": "arn:aws:omics:us-west-2:555555555555:referenceStore/123456789/
reference/5987565360"
    },
    "status": "ACTIVE",
    "storeArn": "arn:aws:omics:us-west-2:555555555555:variantStore/myvariantstore",
    "name": "myvariantstore",
    "creationTime": "2022-11-03T18:19:52.296368+00:00",
    "updateTime": "2022-11-03T18:30:56.272792+00:00",
    "tags": {},
    "storeSizeBytes": 0
}
```

若要檢視與帳戶相關聯的所有變體存放區，請使用 list-variant-stores API。

```
aws omics list-variant-stores
```

您會收到一個回應，列出所有變體存放區及其 IDs、狀態和其他詳細資訊，如下列範例回應所示。

```
{
    "variantStores": [
        {
            "id": "45aeb91d5678",
            "reference": {
                "referenceArn": "arn:aws:omics:us-
west-2:55555555555:referenceStore/5506874698"
            },
            "status": "ACTIVE",
            "storeArn": "arn:aws:omics:us-west-2:55555555555:variantStore/
new_variant_store",
```

```
            "name": "variantstore",
            "creationTime": "2022-11-03T18:19:52.296368+00:00",
            "updateTime": "2022-11-03T18:30:56.272792+00:00",
            "statusMessage": "",
            "storeSizeBytes": 141526
        }
    ]
}
```

您也可以根據狀態或其他條件，篩選 list-variant-stores API 的回應。

匯入 2023 年 5 月 15 日當天或之後建立的分析存放區的 VCF 檔案已定義變體效果預測器 (VEP) 註釋的結構描述。這可讓您更輕鬆地查詢和剖析匯入的 VCF 資料。變更不會影響 2023 年 5 月 15 日之前建立的存放區，除非annotation fields參數包含在 API 或 CLI 呼叫中。對於這些存放區，使用 annotation fields 參數會導致請求失敗。

建立 HealthOmics 變體存放區匯入任務

下列範例示範如何使用 AWS CLI 為變體存放區建立匯入任務。

```
aws omics start-variant-import-job \
       --destination-name myvariantstore \
       --runLeftNormalization false \
       --role-arn  arn:aws:iam::55555555555:role/roleName \
       --items source=s3://my-omics-bucket/sample.vcf.gz source=s3://my-omics-bucket/
sample2.vcf.gz
```

```
{
    "destinationName": "store_a",
    "roleArn": "....",
    "runLeftNormalization": false,
    "items": [
        {"source": "s3://my-omics-bucket/sample.vcf.gz"},
        {"source": "s3://my-omics-bucket/sample2.vcf.gz"}
    ]
}
```

對於 2023 年 5 月 15 日之後建立的存放區，下列範例顯示如何新增 --annotation-fields 參數。註釋欄位是使用匯入來定義。

```
aws omics start-variant-import-job \
```

```
   --destination-name annotationparsingvariantstore \
   --role-arn arn:aws:iam::123456789012:role/<role_name> \
   --items source=s3://pathToS3/sample.vcf
   --annotation-fields '{"VEP": "CSQ"}'
```

```
{
    "jobId": "981e2286-e954-4391-8a97-09aefc343861"
}
```

使用 get-variant-import-job 檢查狀態。

```
aws omics get-variant-import-job --job-id 08279950-a9e3-4cc3-9a3c-a574f9c9e229
```

您將會收到 JSON 回應，顯示匯入任務的狀態。VCF 中的 VEP 註釋會剖析為 ID/值對儲存在 INFO 欄中的資訊。Ensembl 變體效果預測器註釋 INFO 欄的預設 ID 為 CSQ，但您可以使用 `--annotation-fields` 參數來指示 INFO 欄中使用的自訂值。VEP 註釋目前支援剖析。

對於在 2023 年 5 月 15 日之前建立的存放區，或未包含 VEP 註釋的 VCF 檔案，回應不包含任何註釋欄位。

```
{
    "creationTime": "2023-04-11T17:52:37.241958+00:00",
    "destinationName": "annotationparsingvariantstore",
    "id": "7a1c67e3-b7f9-434d-817b-9c571fd63bea",
    "items": [

    {
       "jobStatus": "COMPLETED",
       "source": "s3://amzn-s3-demo-bucket/NA12878.2k.garvan.vcf"
    }
 ],
    "roleArn": "arn:aws:iam::555555555555:role/<role_name>",

    "runLeftNormalization": false,
    "status": "COMPLETED",
    "updateTime": "2023-04-11T17:58:22.676043+00:00",
}
```

屬於 VCF 檔案一部分的 VEP 註釋會儲存為具有下列結構的預先定義結構描述。額外欄位可用來存放預設結構描述中未包含的任何其他 VEP 欄位。

```
annotations struct<
   vep: array<struct<
      allele:string,
      consequence: array<string>,
      impact:string,
      symbol:string,
      gene:string,
      `feature_type`: string,
      feature: string,
      biotype: string,
      exon: struct<rank:string, total:string>,
      intron: struct<rank:string, total:string>,
      hgvsc: string,
      hgvsp: string,
      `cdna_position`: string,
      `cds_position`: string,
      `protein_position`: string,
      `amino_acids`: struct<reference:string, variant: string>,
      codons: struct<reference:string, variant: string>,
      `existing_variation`: array<string>,
      distance: string,
      strand: string,
      flags: array<string>,
      symbol_source: string,
      hgnc_id: string,
      `extras`: map<string, string>
    >>
>
```

剖析是以盡最大努力的方法執行。如果 VEP 項目未遵循 VEP 標準規格，則不會剖析，且陣列中的資料列將為空白。

對於新的變體存放區，get-variant-import-job 的回應將包含註釋欄位，如下所示。

```
aws omics get-variant-import-job --job-id 08279950-a9e3-4cc3-9a3c-a574f9c9e229
```

您會收到 JSON 回應，顯示匯入任務的狀態。

```
{
    "creationTime": "2023-04-11T17:52:37.241958+00:00",
    "destinationName": "annotationparsingvariantstore",
```

```
    "id": "7a1c67e3-b7f9-434d-817b-9c571fd63bea",
    "items": [

    {
    "jobStatus": "COMPLETED",
    "source": "s3://amzn-s3-demo-bucket/NA12878.2k.garvan.vcf"
    }
 ],
    "roleArn": "arn:aws:iam::123456789012:role/<role_name>",
    "runLeftNormalization": false,
    "status": "COMPLETED",
    "updateTime": "2023-04-11T17:58:22.676043+00:00",
    "annotationFields" : {"VEP": "CSQ"}
  }
}
```

您可以使用 list-variant-import-jobs 來查看所有匯入任務及其狀態。

```
aws omics list-variant-import-jobs --ids 7a1c67e3-b7f9-434d-817b-9c571fd63bea
```

回應包含的資訊如下所示。

```
{
    "variantImportJobs": [
    {
        "creationTime": "2023-04-11T17:52:37.241958+00:00",
        "destinationName": "annotationparsingvariantstore",
        "id": "7a1c67e3-b7f9-434d-817b-9c571fd63bea",
        "roleArn": "arn:aws:iam::55555555555:role/roleName",
        "runLeftNormalization": false,
        "status": "COMPLETED",
        "updateTime": "2023-04-11T17:58:22.676043+00:00",
        "annotationFields" : {"VEP": "CSQ"}
        }
    ]
  }
}
```

如有必要，您可以使用下列命令取消匯入任務。

```
aws omics cancel-variant-import-job
     --job-id edd7b8ce-xmpl-47e2-bc99-258cac95a508
```

建立 HealthOmics 註釋存放區

註釋存放區是代表註釋資料庫的資料存放區，例如來自 TSV、VPF 或 GFF 檔案的資料存放區。如果指定相同的參考基因體，註釋存放區會在匯入期間映射到與變體存放區相同的座標系統。下列主題說明如何使用 HealthOmics 主控台和 AWS CLI 來建立和管理註釋存放區。

主題

使用主控台建立註釋存放區

使用下列程序，透過 HealthOmics 主控台建立註釋存放區。

建立註釋存放區

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇註釋存放區。
3. 在註釋存放區頁面上，選擇建立註釋存放區。
4. 在建立註釋存放區頁面上，提供下列資訊
   - 註釋存放區名稱 - 此存放區的唯一名稱。
   - 描述 （選用） - 此參考基因體的描述。
   - 資料格式和結構描述詳細資訊 - 選取資料檔案格式並上傳此存放區的結構描述定義。
   - 參考基因體 - 此註釋的參考基因體。
   - 資料加密 - 選擇您希望 AWS 或您自己擁有和管理資料加密。
   - 標籤 （選用） - 為此註釋存放區提供最多 50 個標籤。
5. 選擇建立註釋存放區。

使用 API 建立註釋存放區

下列範例示範如何使用 建立註釋存放區 AWS CLI。對於所有 AWS CLI 和 API 操作，您必須指定資料的格式。

```
aws omics create-annotation-store --name my_annotation_store \
```

```
           --store-format GFF \
           --reference referenceArn="arn:aws:omics:us-
west-2:555555555555:referenceStore/6505293348/reference/5987565360"
           --version-name new_version
```

您會收到以下回應，以確認建立註釋存放區。

```
{
           "creationTime": "2022-08-24T20:34:19.229500Z",
           "id": "3b93cdef69d2",
           "name": "my_annotation_store",
           "reference": {
               "referenceArn": "arn:aws:omics:us-
west-2:555555555555:referenceStore/6505293348/reference/5987565360"
           },
           "status": "CREATING"
           "versionName": "my_version"
       }
```

若要進一步了解註釋存放區，請使用 get-annotation-store API。

```
aws omics get-annotation-store --name my_annotation_store
```

您會收到下列回應。

```
{
          "id": "eeb019ac79c2",
          "reference": {
              "referenceArn": "arn:aws:omics:us-
west-2:555555555555:referenceStore/5638433913/reference/5871590330“
          },
          "status": "ACTIVE",
          "storeArn": "arn:aws:omics:us-west-2:555555555555:annotationStore/gffstore",
          "name": "my_annotation_store",
          "creationTime": "2022-11-05T00:05:19.136131+00:00",
          "updateTime": "2022-11-05T00:10:36.944839+00:00",
          "tags": {},
          "storeFormat": "GFF",
          "statusMessage": "",
          "storeSizeBytes": 0,
          "numVersions": 1
      }
```

若要檢視與帳戶相關聯的所有註釋存放區，請使用 list-annotation-stores API 操作。

```
aws omics list-annotation-stores
```

您會收到一個回應，列出所有註釋存放區及其 IDs、狀態和其他詳細資訊，如下列範例回應所示。

```
{
           "annotationStores": [
               {
                  "id": "4d8f3eada259",
                   "reference":
                       "referenceArn": "arn:aws:omics:us-
west-2:555555555555:referenceStore/5638433913/reference/5871590330"
                   },
                   "status": "CREATING",
                   "name": "gffstore",
                   "creationTime": "2022-09-27T17:30:52.182990+00:00",
                   "updateTime": "2022-09-27T17:30:53.025362+00:00"
               }
           ]
       }
```

您也可以根據狀態或其他條件來篩選回應。

為 HealthOmics 註釋存放區建立匯入任務

主題

- 使用 API 建立註釋匯入任務
- TSV 和 VCF 格式的其他參數
- 建立 TSV 格式的註釋存放區
- 啟動 VCF 格式的匯入任務

使用 API 建立註釋匯入任務

下列範例示範如何使用 AWS CLI 來啟動註釋匯入任務。

```
aws omics start-annotation-import-job \
           --destination-name myannostore \
```

```
            --version-name myannostore \
            --role-arn arn:aws:iam::123456789012:role/roleName \
            --items source=s3://my-omics-bucket/sample.vcf.gz
            --annotation-fields '{"VEP": "CSQ"}'
```

如果包含註釋欄位，則在 2023 年 5 月 15 日之前建立的註釋存放區會傳回錯誤訊息。它們不會針對涉及註釋存放區匯入任務的任何 API 操作傳回輸出。

然後，您可以使用 get-annotation-import-job API 操作和 `job ID` 參數來進一步了解註釋匯入任務的詳細資訊。

```
aws omics get-annotation-import-job --job-id 9e4198fb-fa85-446c-9301-9b823a1a8ba8
```

您會收到下列回應，包括註釋欄位。

```
{
          "creationTime": "2023-04-11T19:09:25.049767+00:00",
          "destinationName": "parsingannotationstore",
          "versionName": "parsingannotationstore",
          "id": "9e4198fb-fa85-446c-9301-9b823a1a8ba8",
          "items": [
              {
                  "jobStatus": "COMPLETED",
                  "source": "s3://my-omics-bucket/sample.vep.vcf"
              }
          ],
          "roleArn": "arn:aws:iam::55555555555:role/roleName",
          "runLeftNormalization": false,
          "status": "COMPLETED",
          "updateTime": "2023-04-11T19:13:09.110130+00:00",
          "annotationFields" : {"VEP": "CSQ"}
       }
```

若要檢視所有註釋存放區匯入任務，請使用 list-annotation-import-jobs 。

```
aws omics list-annotation-import-jobs --ids 9e4198fb-fa85-446c-9301-9b823a1a8ba8
```

回應包含註釋存放區匯入任務的詳細資訊和狀態。

```
{
          "annotationImportJobs": [
          {
              "creationTime": "2023-04-11T19:09:25.049767+00:00",
              "destinationName": "parsingannotationstore",
              "versionName": "parsingannotationstore",
              "id": "9e4198fb-fa85-446c-9301-9b823a1a8ba8",
              "roleArn": "arn:aws:iam::55555555555:role/roleName",
              "runLeftNormalization": false,
              "status": "COMPLETED",
              "updateTime": "2023-04-11T19:13:09.110130+00:00",
              "annotationFields" : {"VEP": "CSQ"}
          }
          ]
      }
```

TSV 和 VCF 格式的其他參數

對於 TSV 和 VCF 格式，還有其他參數可通知 API 如何剖析您的輸入。

> **Important**
>
> 使用查詢引擎匯出的 CSV 註釋資料會直接從資料集匯入傳回資訊。如果匯入的資料包含公式或命令，則檔案可能需要 CSV 插入。因此，使用查詢引擎匯出的檔案可能會提示安全性警告。若要避免惡意活動，請在讀取匯出檔案時關閉連結和巨集。

TSV 剖析器也會執行基本的生物資訊學操作，例如基因體座標的左側標準化和標準化，如下表所列。

格式類型	描述
一般	一般文字檔案。沒有基因體資訊。
CHR_POS	開始位置 - 1，新增結束位置，這與 相同POS。
CHR_POS_REF_ALT	包含 contig、1 基位置、ref 和 alt 等位基因資訊。
CHR_START_END_REF_ALT_ONE_BASE	包含 contig、start、end、ref 和 alt 等位基因資訊。座標以 1 為基礎。

格式類型	描述
CHR_START_END_ZERO_BASE	包含連續、開始和結束位置。座標以 0 為基礎。
CHR_START_END_ONE_BASE	包含連續、開始和結束位置。座標以 1 為基礎。
CHR_START_END_REF_ALT_ZERO_BASE	包含 contig、start、end、ref 和 alt 等位基因資訊。座標以 0 為基礎。

TSV 匯入註釋存放區請求如下所示。

```
aws omics start-annotation-import-job \
--destination-name tsv_anno_example \
--role-arn arn:aws:iam::555555555555:role/demoRole \
--items source=s3://demodata/genomic_data.bed.gz \
--format-options '{ "tsvOptions": {
        "readOptions": {
            "header": false,
            "sep": "\t"
        }
    }
}'
```

建立 TSV 格式的註釋存放區

下列範例使用包含標頭、資料列和註解的索引標籤限制檔案來建立註釋存放區。座標為 CHR_START_END_ONE_BASED，其中包含來自 OMIM 人類基因貼圖摘要的 HG19 基因貼圖。 https://www.omim.org/downloads

```
aws omics create-annotation-store --name mimgenemap \
  --store-format TSV \
  --reference=referenceArn=arn:aws:omics:us-
west-2:555555555555:referenceStore/6505293348/reference/2310864158 \
  --store-options=tsvStoreOptions='{
    annotationType=CHR_START_END_ONE_BASE,
    formatToHeader={CHR=chromosome, START=genomic_position_start,
 END=genomic_position_end},
    schema=[
      {chromosome=STRING},
```

```
      {genomic_position_start=LONG},
      {genomic_position_end=LONG},
      {cyto_location=STRING},
      {computed_cyto_location=STRING},
      {mim_number=STRING},
      {gene_symbols=STRING},
      {gene_name=STRING},
      {approved_gene_name=STRING},
      {entrez_gene_id=STRING},
      {ensembl_gene_id=STRING},
      {comments=STRING},
      {phenotypes=STRING},
      {mouse_gene_symbol=STRING}]}'
```

您可以使用或不使用標頭匯入檔案。若要在 CLI 請求中指出這一點，請使用 header=false，如下列匯入任務範例所示。

```
aws omics start-annotation-import-job \
   --role-arn arn:aws:iam::555555555555:role/demoRole \
   --items=source=s3://amzn-s3-demo-bucket/annotation-examples/hg38_genemap2.txt \
   --destination-name output-bucket \
   --format-options=tsvOptions='{readOptions={sep="\t",header=false,comment="#"}}'
```

下列範例會為床鋪檔案建立註釋存放區。床位檔案是簡單的標籤分隔檔案。在這個範例中，資料欄是「」、「開始」、「結束」和「區域名稱」。座標為零，且資料沒有標頭。

```
aws omics create-annotation-store \
   --name cexbed --store-format TSV \
   --reference=referenceArn=arn:aws:omics:us-
west-2:555555555555:referenceStore/6505293348/reference/2310864158 \
   --store-options=tsvStoreOptions='{
   annotationType=CHR_START_END_ZERO_BASE,
   formatToHeader={CHR=chromosome, START=start, END=end},
   schema=[{chromosome=STRING}, {start=LONG}, {end=LONG}, {name=STRING}]}'
```

然後，您可以使用下列 CLI 命令，將床位檔案匯入註釋存放區。

```
aws omics start-annotation-import-job \
   --role-arn arn:aws:iam::555555555555:role/demoRole \
   --items=source=s3://amzn-s3-demo-bucket/TruSeq_Exome_TargetedRegions_v1.2.bed \
   --destination-name cexbed \
```

```
   --format-options=tsvOptions='{readOptions={sep="\t",header=false,comment="#"}}'
```

下列範例會為以標籤分隔的檔案建立註釋存放區，其中包含 VCF 檔案的前幾個資料欄，後面接著包含註釋資訊的資料欄。它包含具有有關 、啟動、參考和替代等位基因資訊的基因組位置，並且包含 標頭。

```
aws omics create-annotation-store --name gnomadchrx --store-format TSV \
--reference=referenceArn=arn:aws:omics:us-
west-2:555555555555:referenceStore/6505293348/reference/2310864158 \
--store-options=tsvStoreOptions='{
    annotationType=CHR_POS_REF_ALT,
    formatToHeader={CHR=chromosome, POS=start, REF=ref, ALT=alt},
    schema=[
        {chromosome=STRING},
        {start=LONG},
        {ref=STRING},
        {alt=STRING},
        {filters=STRING},
        {ac_hom=STRING},
        {ac_het=STRING},
        {af_hom=STRING},
        {af_het=STRING},
        {an=STRING},
        {max_observed_heteroplasmy=STRING}]}'
```

然後，您可以使用下列 CLI 命令將檔案匯入註釋存放區。

```
aws omics start-annotation-import-job \
  --role-arn arn:aws:iam::555555555555:role/demoRole \
   --items=source=s3://amzn-s3-demo-bucket/
gnomad.genomes.v3.1.sites.chrM.reduced_annotations.tsv \
   --destination-name gnomadchrx \
   --format-options=tsvOptions='{readOptions={sep="\t",header=true,comment="#"}}'
```

下列範例顯示客戶如何為 mim2gene 檔案建立註釋存放區。mim2gene 檔案提供 OMIM 中的基因與另一個基因識別符之間的連結。它以標籤分隔，並包含註解。

```
aws omics create-annotation-store \
  --name mim2gene \
```

```
  --store-format TSV \
  --reference=referenceArn=arn:aws:omics:us-
west-2:555555555555:referenceStore/6505293348/reference/2310864158 \
  --store-options=tsvStoreOptions='
    {annotationType=GENERIC,
    formatToHeader={},
    schema=[
        {mim_gene_id=STRING},
        {mim_type=STRING},
        {entrez_id=STRING},
        {hgnc=STRING},
        {ensembl=STRING}]}'
```

然後，您可以將資料匯入您的 存放區，如下所示。

```
aws omics start-annotation-import-job \
   --role-arn arn:aws:iam::555555555555:role/demoRole \
   --items=source=s3://xquek-dev-aws/annotation-examples/mim2gene.txt \
   --destination-name mim2gene \
   --format-options=tsvOptions='{readOptions={sep="\t",header=false,comment="#"}}'
```

啟動 VCF 格式的匯入任務

對於 VCF 檔案，有兩個額外的輸入 `ignoreQualField`和 `ignoreFilterField`，可忽略或包含這些參數，如下所示。

```
aws omics start-annotation-import-job --destination-name annotation_example\
  --role-arn arn:aws:iam::555555555555:role/demoRole \
  --items source=s3://demodata/example.garvan.vcf \
  --format-options '{ "vcfOptions": {
    "ignoreQualField": false,
    "ignoreFilterField": false
    }
   }'
```

您也可以取消註釋存放區匯入，如下所示。如果取消成功，您不會收到此 AWS CLI 呼叫的回應。不過，如果找不到匯入任務 ID 或匯入任務已完成，您會收到錯誤訊息。

```
aws omics cancel-annotation-import-job --job-id edd7b8ce-xmpl-47e2-bc99-258cac95a508
```

> **Note**
>
> 您的 get-annotation-import-job、get-variant-import-job、list-annotation-import-jobs 和 list-variant-import-jobs 中繼資料匯入任務歷史記錄會在兩年後自動刪除。匯入的變體和註釋資料不會自動刪除，並保留在您的資料存放區中。

建立新的 HealthOmics 註釋存放區版本

您可以建立新的註釋存放區版本，以收集不同版本的註釋資料庫。這可協助您整理註釋資料，這會定期更新。

若要建立新的現有註釋存放區版本，請使用 create-annotation-store-version API，如下列範例所示。

```
aws omics create-annotation-store-version \
     --name my_annotation_store \
     --version-name my_version
```

您將會收到包含註釋存放區版本 ID 的下列回應，確認已建立新版本的註釋。

```
{
     "creationTime": "2023-07-21T17:15:49.251040+00:00",
     "id": "3b93cdef69d2",
     "name": "my_annotation_store",
     "reference": {
         "referenceArn": "arn:aws:omics:us-
west-2:555555555555:referenceStore/6505293348/reference/5987565360"
     },
     "status": "CREATING",
     "versionName": "my_version"
}
```

若要更新註釋存放區版本的描述，您可以使用 update-annotation-store-version 將更新新增至註釋存放區版本。

```
aws omics update-annotation-store-version \
    --name my_annotation_store \
    --version-name my_version \
    --description "New Description"
```

您將會收到下列回應，確認註釋存放區版本已更新。

```
{
     "storeId": "4934045d1c6d",
     "id": "2a3f4a44aa7b",
     "description":"New Description",
     "status": "ACTIVE",
     "name": "my_annotation_store",
     "versionName": "my_version",
     "creation Time": "2023-07-21T17:20:59.380043+00:00",
     "updateTime": "2023-07-21T17:26:17.892034+00:00"
}
```

若要檢視註釋存放區版本的詳細資訊，請使用 get-annotation-store-version。

```
aws omics get-annotation-store-version --name my_annotation_store --version-name
 my_version
```

您將會收到包含版本名稱、狀態和其他詳細資訊的回應。

```
{
     "storeId": "4934045d1c6d",
     "id": "2a3f4a44aa7b",
     "status": "ACTIVE",
     "versionArn": "arn:aws:omics:us-west-2:555555555555:annotationStore/
my_annotation_store/version/my_version",
     "name": "my_annotation_store",
     "versionName": "my_version",
     "creationTime": "2023-07-21T17:15:49.251040+00:00",
     "updateTime": "2023-07-21T17:15:56.434223+00:00",
     "statusMessage": "",
     "versionSizeBytes": 0
    }
```

若要檢視註釋存放區的所有版本，您可以使用 list-annotation-store-versions，如下列範例所示。

```
aws omics list-annotation-store-versions --name my_annotation_store
```

您將會收到包含下列資訊的回應

```
{
  "annotationStoreVersions": [
    {
     "storeId": "4934045d1c6d",
     "id": "2a3f4a44aa7b",
     "status": "CREATING",
     "versionArn": "arn:aws:omics:us-west-2:555555555555:annotationStore/
my_annotation_store/version/my_version_2",
     "name": "my_annotation_store",
     "versionName": "my_version_2",
     "creation Time": "2023-07-21T17:20:59.380043+00:00",
     "versionSizeBytes": 0
    },
    {
     "storeId": "4934045d1c6d",
     "id": "4934045d1c6d",
     "status": "ACTIVE",
     "versionArn": "arn:aws:omics:us-west-2:555555555555:annotationStore/
my_annotation_store/version/my_version_1",
     "name": "my_annotation_store",
     "versionName": "my_version_1",
     "creationTime": "2023-07-21T17:15:49.251040+00:00",
     "updateTime": "2023-07-21T17:15:56.434223+00:00",
     "statusMessage": "",
     "versionSizeBytes": 0
    }
}
```

如果您不再需要註釋存放區版本，您可以使用 delete-annotation-store-versions 刪除註釋存放區版本，如下列範例所示。

```
aws omics delete-annotation-store-versions --name my_annotation_store --versions
 my_version
```

如果刪除的儲存體版本沒有錯誤，您將會收到下列回應。

```
{
  "errors": []
}
```

如果有錯誤，您將收到包含錯誤詳細資訊的回應，如下所示。

```
{
  "errors": [
    {
      "versionName": "my_version",
      "message": "Version with versionName: my_version was not found."
    }
  ]
}
```

如果您嘗試刪除具有作用中匯入任務的註釋存放區版本，您將收到包含錯誤的回應，如下所示。

```
{
  "errors": [
    {
      "versionName": "my_version",
      "message": "version has an inflight import running"
    }
  ]
}
```

在此情況下，您可以強制刪除註釋存放區版本，如下列範例所示。

```
aws omics delete-annotation-store-versions --name my_annotation_store --versions
 my_version --force
```

刪除 HealthOmics 分析存放區

當您刪除變體或註釋存放區時，系統也會刪除該存放區中的所有匯入資料，以及任何相關聯的標籤。

下列範例示範如何使用 刪除變體存放區 AWS CLI。如果動作成功，變體存放區狀態會轉換為 DELETING。

```
aws omics delete-variant-store --id <variant-store-id>
```

下列範例示範如何刪除註釋存放區。如果動作成功，註釋存放區狀態會轉換為 DELETING。如果存在多個版本，則無法刪除註釋存放區。

```
aws omics delete-annotation-store --id <annotation-store-id>
```

查詢 HealthOmics 分析資料

您可以使用 AWS Lake Formation 和 Amazon Athena 或 Amazon EMR 在變體存放區上執行查詢。執行任何查詢之前，請先完成 Lake Formation 和 Amazon Athena 的設定程序 （如以下章節所述）。

如需 Amazon EMR 的相關資訊，請參閱教學課程：Amazon EMR 入門

對於 2024 年 9 月 26 日之後建立的變體存放區，HealthOmics 會依範例 ID 分割存放區。此分割表示 HealthOmics 使用範例 ID 來最佳化變體資訊的儲存。使用範例資訊做為篩選條件的查詢會更快傳回結果，因為查詢掃描的資料較少。

HealthOmics 使用範例 IDs做為分割區檔案名稱。在您擷取資料之前，請檢查範例 ID 是否包含任何 PHI 資料。如果是這樣，請在擷取資料之前變更範例 ID。如需在範例 IDs中包含和不包含哪些內容的詳細資訊，請參閱 AWS HIPAA 合規網頁上的指導。

主題

設定 Lake Formation 以使用 HealthOmics

在您使用 Lake Formation 管理 HealthOmics 資料存放區之前，請執行下列 Lake Formation 組態程序。

主題

建立或驗證 Lake Formation 管理員

您必須先定義一或多個管理員，才能在 Lake Formation 中建立資料湖。

管理員是具有建立資源連結許可的使用者和角色。您為每個區域的每個帳戶設定資料湖管理員。

在 Lake Formation 主控台中建立管理員使用者

1. 開啟 AWS Lake Formation 主控台：Lake Formation 主控台
2. 如果主控台顯示歡迎使用 Lake Formation 面板，請選擇開始使用。

   Lake Formation 會將您新增至資料湖管理員資料表。
3. 否則，從左側功能表中，選擇管理角色和任務。
4. 視需要新增任何其他管理員。

使用 Lake Formation 主控台建立資源連結

若要建立使用者可以查詢的共用資源，必須停用預設存取控制。若要進一步了解停用預設存取控制，請參閱 Lake Formation 文件中的變更資料湖的預設安全設定。您可以個別或群組建立資源連結，以便存取 Amazon Athena 或其他 AWS 服務（例如 Amazon EMR) 中的資料。

在 AWS Lake Formation 主控台中建立資源連結，並與 HealthOmics Analytics 使用者共用

1. 開啟 AWS Lake Formation 主控台：Lake Formation 主控台
2. 在主要導覽列中，選擇資料庫。
3. 在資料庫表格中，選取所需的資料庫。
4. 從建立功能表中，選擇資源連結。
5. 輸入資源連結名稱。如果您計劃從 Athena 存取資料庫，請使用小寫字母（最多 256 個字元）輸入名稱。
6. 選擇建立。
7. 新的資源連結現在會列在資料庫下。

使用 Lake Formation 主控台授予共用資源的存取權

Lake Formation 資料庫管理員可以使用下列程序授予共用資源的存取權。

1. 開啟 AWS Lake Formation 主控台：https：//https://console.aws.amazon.com/lakeformation/
2. 在主要導覽列中，選擇資料庫。
3. 在資料庫頁面上，選取您先前建立的資源連結。
4. 從動作功能表中，選擇對目標授予。
5. 在主體下的授予資料許可頁面上，選擇 IAM 使用者或角色。

6. 從 IAM 使用者或角色下拉式功能表中，尋找您要授予存取權的使用者。
7. 接著，在 LF 標籤或目錄資源卡下，選取具名資料目錄資源選項。
8. 從資料表選用下拉式功能表中，選取所有資料表或您先前建立的資料表。
9. 在資料表許可卡的資料表許可下，選擇描述和選取。
10. 接著，選擇授予。

若要檢視 Lake Formation 許可，請從主要導覽窗格中選擇 Data lake 許可。資料表顯示可用的資料庫和資源連結。

設定 AWS RAM 資源共用的許可

在 AWS Lake Formation 主控台中，選擇主導覽列中的資料湖許可來檢視許可。在資料許可頁面上，您可以檢視顯示資源類型、資料庫，以及與 RAM Resource Share 下共用資源**ARN**相關的資料表。如果您需要接受 AWS Resource Access Manager (AWS RAM) 資源共享， 會在 主控台中 AWS Lake Formation 通知您。

HealthOmics 可以在商店建立期間隱含接受 AWS RAM 資源共用。若要接受 AWS RAM 資源共用，呼叫 或 `CreateAnnotationStore` API 操作的 IAM 使用者`CreateVariantStore`或角色必須允許下列動作：

- `ram:GetResourceShareInvitations` - 此動作可讓 HealthOmics 尋找邀請。
- `ram:AcceptResourceShareInvitation` - 此動作允許 HealthOmics 使用 FAS 字符接受邀請。

如果沒有這些許可，您會在建立存放區期間看到授權錯誤。

以下是包含這些動作的範例政策。將此政策新增至接受 AWS RAM 資源共用的 IAM 使用者或角色。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "omics:*",
        "ram:AcceptResourceShareInvitation",
```

```
        "ram:GetResourceShareInvitations"
      ],
      "Resource": "*"
    }
  ]
}
```

為查詢設定 Athena

您可以使用 Athena 查詢變體和註釋。執行任何查詢之前，請執行下列設定任務：

主題

- 使用 Athena 主控台設定查詢結果位置
- 使用 Athena 引擎 v3 設定工作群組

使用 Athena 主控台設定查詢結果位置

若要設定查詢結果位置，請遵循下列步驟。

1. 開啟 Athena 主控台：Athena 主控台
2. 在主要導覽列中，選擇查詢編輯器。
3. 在查詢編輯器中，選擇設定索引標籤，然後選擇管理。
4. 輸入位置的 S3 字首以儲存查詢結果。

使用 Athena 引擎 v3 設定工作群組

若要設定工作群組，請依照下列步驟進行。

1. 開啟 Athena 主控台：Athena 主控台
2. 在主要導覽列中，選擇工作群組，然後選擇建立工作群組。
3. 輸入工作群組的名稱。
4. 選取 Athena SQL 做為引擎類型。
5. 在升級查詢引擎下，選取手動。
6. 在查詢版本引擎下，選取 Athena 第 3 版。
7. 選擇建立工作群組。

在 HealthOmics 變體存放區上執行查詢

您可以使用 Amazon Athena 在變體存放區上執行查詢。請注意，變體和註釋存放區中的基因體座標表示為以零為基礎、半封閉的半開間隔。

使用 Athena 主控台執行簡單的查詢

下列範例示範如何執行簡單的查詢。

1. 開啟 Athena 查詢編輯器：Athena 查詢編輯器
2. 在工作群組下，選取您在設定期間建立的工作群組。
3. 確認資料來源是 AwsDataCatalog。
4. 針對資料庫，選取您在 Lake Formation 設定期間建立的資料庫資源連結。
5. 將下列查詢複製到查詢 1 索引標籤下的查詢編輯器：

```
SELECT * from omicsvariants limit 10
```

6. 選擇執行以執行查詢。主控台會將資料表的前 10 列填入結果omicsvariants資料表。

使用 Athena 主控台執行複雜的查詢

下列範例示範如何執行複雜的查詢。若要執行此查詢，`ClinVar`請將 匯入註釋存放區。

執行複雜的查詢

1. 開啟 Athena 查詢編輯器：Athena 查詢編輯器
2. 在工作群組下，選取您在設定期間建立的工作群組。
3. 確認資料來源是 AwsDataCatalog。
4. 針對資料庫，選取您在 Lake Formation 設定期間建立的資料庫資源連結。
5. 選擇右上角+的 ，以建立新的查詢索引標籤，名為查詢 2。
6. 將下列查詢複製到查詢 2 索引標籤下的查詢編輯器：

```
SELECT variants.sampleid,
  variants.contigname,
  variants.start,
  variants."end",
  variants.referenceallele,
  variants.alternatealleles,
```

```
  variants.attributes AS variant_attributes,
  clinvar.attributes AS clinvar_attributes
FROM omicsvariants as variants
INNER JOIN omicsannotations as clinvar ON
  variants.contigname=CONCAT('chr',clinvar.contigname)
  AND variants.start=clinvar.start
  AND variants."end"=clinvar."end"
  AND variants.referenceallele=clinvar.referenceallele
  AND variants.alternatealleles=clinvar.alternatealleles
WHERE clinvar.attributes['CLNSIG']='Likely_pathogenic'
```

7. 選擇執行以開始執行查詢。

共用 HealthOmics 分析存放區

身為變體存放區或註釋存放區的擁有者，您可以與其他 AWS 帳戶共用存放區。擁有者可以透過刪除共用來撤銷對共用資源的存取。

身為共用存放區的訂閱者，您必須先接受共用。然後，您可以定義使用共用存放區的工作流程。資料會在 AWS Glue 和 Lake Formation 中顯示為資料表。

當您不再需要存取存放區時，您會刪除共用。

如需資源共用的其他資訊中的跨帳戶資源共用 AWS HealthOmics，請參閱 。

建立存放區共用

若要建立存放區共享，請使用 create-share API 操作。委託人訂閱者是將訂閱共享 AWS 帳戶 之使用者的 。下列範例會為變體存放區建立共享。若要與多個帳戶共用商店，您可以建立相同商店的多個共用。

```
aws omics create-share  \
        --resource-arn "arn:aws:omics:us-west-2:555555555555:variantStore/
omics_dev_var_store" \
        --principal-subscriber "123456789012" \
        --name "my_Share-123"
```

如果建立成功，您會收到共用 ID 和狀態的回應。

```
{
```

```
    "shareId": "495c21bedc889d07d0ab69d710a6841e-dd75ab7a1a9c384fa848b5bd8e5a7e0a",
    "name": "my_Share-123",
    "status": "PENDING"
}
```

在訂閱者使用接受共用 API 操作接受共用之前，共用會保持待定狀態。

中的跨帳戶資源共用 AWS HealthOmics

使用跨帳戶共用與協作者共用資源，而無需建立複本或修改 IAM 資源政策。下列資源支援跨帳戶共用：

- HealthOmics 變體存放區
- HealthOmics 註釋存放區
- 私有工作流程

共用資源包含下列步驟：

1. 資源擁有者會建立共享，並指定資源的 ARN 和 AWS 帳戶 預期訂閱者的 。資源共享會保持待定狀態，直到訂閱者接受共享為止。
2. 訂閱者接受資源共享來存取資源。資源共用會轉換為啟用狀態。
3. HealthOmics 服務可讓訂閱者帳戶存取 資源。
4. 資源擁有者可以刪除共用，或訂閱者可以撤銷對共用的存取權。訂閱者無法刪除共用或相關聯的資源。

主題

建立共享

您可以使用 create-share API 操作來建立共享。主要訂閱者是將訂閱共用資源 AWS 帳戶 之使用者的 。下列範例會為變體存放區建立共享。

```
aws omics create-share \
    --resource-arn "arn:aws:omics:us-west-2:555555555555:variantStore/
omics_dev_var_store" \
    --principal-subscriber "123456789012"  \
```

```
    --name "my_Share-123"
```

如果建立成功，您會收到共用 ID 和狀態的回應。

```
{
"shareId": "495c21bedc889d07d0ab69d710a6841e-dd75ab7a1a9c384fa848b5bd8e5a7e0a",
     "name": "my_Share-123",
     "status": "PENDING"
}
```

在訂閱者使用 accept-share API 操作接受共用之前，共用會保持待定狀態。

```
 aws omics accept-share \
    --share-id "495c21bedc889d07d0ab69d710a6841e-dd75ab7a1a9c384fa848b5bd8e5a7e0a"
```

訂閱者接受共用後，共用會轉換為作用中狀態。

```
{
"status": "ACTIVATING"
}
```

擷取共享的相關資訊

使用 get-share API 操作來擷取共用的相關資訊。

```
aws omics get-share --share-id "495c21bedc889d07d0ab69d710a6841e-
dd75ab7a1a9c384fa848b5bd8e5a7e0a"
```

API 回應包含共享的中繼資料資訊。

```
{
  "share":
    {
      "shareId": "495c21bedc889d07d0ab69d710a6841e-dd75ab7a1a9c384fa848b5bd8e5a7e0a",
      "name": "my_Share-123",
      "resourceArn": "arn:aws:omics:us-west-2:555555555555:variantStore/
omics_dev_var_store",
```

```
      "principalSubscriber": "123456789012",
      "ownerId": "555555555555",
      "status": "PENDING"
    }
}
```

檢視您擁有的共享

使用 list-shares API 擷取您擁有的每個共享的相關資訊。

```
aws omics list-shares  --resource-owner SELF
```

API 回應包含您擁有的每個共享的中繼資料。

從其他帳戶檢視接受的共享

使用 list-shares API 來檢視您從其他帳戶接受的所有共用。

```
aws omics list-shares  --resource-owner OTHER
```

API 回應包含您接受的每個共用的中繼資料。

刪除共享

在您不再需要共享之後，請使用 delete-share API 來刪除共享。

```
aws omics delete-share \
    --share-id "495c21bedc889d07d0ab69d710a6841e-dd75ab7a1a9c384fa848b5bd8e5a7e0a"
```

在 HealthOmics 中標記資源

主題

重要通知

HealthOmics 根據 AWS 共同責任模型政策保護客戶資料。這表示所有客戶資料都會在轉換中和靜態時加密。不過，並非所有客戶輸入的資源名稱都會加密，例如資料存放區或任務型操作。不應包含個人身分識別資訊或受保護的健康資訊。如需詳細資訊，請參閱AWS HealthOmics 的安全性。

標記 HealthOmics 資源

您可以使用標籤將中繼資料指派給 AWS 資源。每個標籤都是由使用者定義的金鑰和值組成的標籤。標籤可協助您管理、識別、組織、搜尋及篩選資源。

本主題描述常用的標記類別和策略，協助您實作一致且有效的標記策略。下列各節假設對 AWS 資源、標記、詳細帳單和 有基本知識 AWS Identity and Access Management。

每個標籤有兩個部分：

- 標籤金鑰 (例如，CostCenter、Environment 或 Project)。標籤鍵會區分大小寫。
- 標籤值 （ 例如 111122223333 或 Production)。與標籤鍵相同，標籤值會區分大小寫。

您可使用標籤來依照用途、擁有者、環境或其他條件分類資源。如需更多資訊，請參閱 AWS 標記策略。

您可以從資源的服務主控台、服務 API 或 新增、變更或移除資源的標籤 AWS CLI。

若要啟用標記，請確定已授權 TagResources。您可以連接 IAM 政策來授權 TagResources，如下列範例所示。

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "omics:Create*",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "omics:Start*",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "omics:Tag*",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "omics:Untag*",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "omics:List*",
            "Resource": "*"
        }
    ]
}
```

最佳實務

當您為 AWS 資源建立標記策略時，請遵循最佳實務：

- 請勿在標籤中存放個人身分識別資訊 (PII)、受保護醫療資訊 (PHI) 或其他敏感資訊。
- 使用標準化、區分大小寫的標籤格式，並統一套用在所有資源類型上。
- 考慮支援多種用途的標籤準則，例如資源存取控制管理、成本追蹤、自動化和組織。

- 使用自動化工具來協助管理資源標籤。AWS Resource Groups 和 Resource Groups Tagging API 可讓您以程式設計方式控制標籤，進而自動管理、搜尋和篩選標籤和資源。
- 當您使用更多標籤時，標記會更有效率。
- 當使用者需要變更時，可以編輯或修改標籤。不過，若要更新存取控制標籤，您還必須更新參考這些標籤的政策，以控制對 資源的存取。

標記需求

標籤均擁有以下要求：

- 金鑰不能以 aws： 為字首。
- 索引鍵在標籤集內必須是唯一的。
- 索引鍵必須介於 1 到 128 個允許的字元之間。
- 值必須介於 0 到 256 個允許的字元之間。
- 每個標籤集的值不需要是唯一的。
- 索引鍵和值的允許字元為 Unicode 字母、數字、空格和下列任何符號：_ . : / = + - @。
- 金鑰和值會區分大小寫。

序列存放區讀取集標籤

對於序列存放區，在讀取集上建立的標籤位於讀取集資源層級。讀取集下也包含可使用 S3 APIs 存取、搜尋和限制的物件。根據預設，範例 ID (omics：sampleId) 和主體 ID (omics：subjectId) 會新增至物件。

此外，讀取集和其下的物件之間最多可同步五個標籤。要同步的標籤組態是在使用 `propogatedSetLevelTags` 參數建立或更新存放區期間設定的存放區層級組態。

如果存放區中已有資料，更新金鑰可能需要一些時間。在此更新期間，HealthOmics 會將儲存狀態變更為 Updating。完成後，HealthOmics 會將存放區狀態設定為 Active。當標籤正在傳播時，可能無法強制執行依賴標籤的許可。標籤傳播完成後，就會強制執行許可。

在讀取集上設定或更新標籤時，系統會根據存放區組態決定是否更新該讀取集的物件。

將標籤新增至 HealthOmics 資源

將標籤新增至資源可協助您識別和組織 AWS 資源，並管理對這些資源的存取。首先，您將一或多個標籤（鍵/值對）新增至資源。每個資源最多可以使用 50 個標籤。對於您可以在索引鍵和值欄位中使用的字元也有限制。

新增標籤後，您可以建立 IAM 政策，以根據這些標籤管理對 資源的 AWS 存取。您可以使用 HealthOmics 主控台或 AWS CLI 將標籤新增至資源。新增標籤到儲存庫可能會影響存取該儲存庫。將標籤新增至資料存放區之前，請檢閱任何可能使用標籤來控制存取資料存放區等資源的 IAM 政策。

服務標籤會針對主旨和序列存放區的範例 ID 自動產生。

請依照下列步驟，使用 AWS CLI 將標籤新增至 HealthOmics 資源。例如，若要在建立序列存放區時將標籤新增至序列存放區，您可以在 中使用下列命令 AWS CLI。序列存放區的名稱是 MySequenceStore，而兩個新增的索引鍵標籤是 key1 和 key2，值分別是 value1 和 value2：

```
aws omics create-sequence-store --name "MySequenceStore" --tags key1=value1,key2=value2
```

輸出不會列出標籤。它會傳回下列回應。

```
{
    "id": "6860403586",
    "referenceStoreId": "4889894479",
    "roleArn": "arn:aws:iam::555555555555:role/ImportTest",
    "status": "CREATED",
    "creationTime": "2022-07-21T01:19:07.194Z"
}
```

若要將標籤新增至現有資源，請執行下列範例命令。

```
aws omics tag-resource --resource-arn arn:aws:omics:us-
west-2:555555555555:sequenceStore/2275234794 --tags key1=value1,key2=value2
```

如果成功，此命令不會傳回任何回應。

列出資源的標籤

請依照下列步驟使用 AWS CLI 來檢視 HealthOmics 資源的 AWS 標籤清單。若未新增標籤，傳回的清單空白。

在終端機或命令列，執行 list-tags-for-resource 命令，如下列範例所示。

```
aws omics list-tags-for-resource --resource-arn arn:aws:omics:us-
west-2:555555555555:sequenceStore/2275234794
```

您將會收到 JSON 格式的標籤回應清單。

```
 {
    "tags": {
        "key1": "value1",
        "key2": "value2"
    }
}
```

從資料存放區移除標籤

您可以移除與資源相關聯的一或多個標籤。移除標籤不會從與該標籤相關聯的其他 AWS 資源中刪除標籤。

在終端機或命令列，執行 untag-resource 命令，指定您要移除標籤之資源的 Amazon Resource Name (ARN)，以及您要移除之標籤的標籤索引鍵。

```
aws omics untag-resource --resource-arn arn:aws:omics:us-
west-2:555555555555:sequenceStore/2275234794 --tag-keys key1,key2
```

如果成功，此命令不會傳回回應。若要驗證與資源相關聯的標籤，請執行 list-tags-for-resource 命令。

HealthOmics 的 IAM 許可

您可以使用 AWS Identity and Access Management (IAM) 來管理 HealthOmics API 和資源的存取，例如存放區和工作流程。對於您帳戶中使用 HealthOmics 的使用者和應用程式，您可以在可套用至 IAM 使用者、群組或角色的許可政策中管理許可。

若要管理帳戶中使用者和應用程式的許可，請使用 HealthOmics 提供的政策，或自行撰寫。HealthOmics 主控台使用多個 服務來取得函數組態和觸發條件的相關資訊。您可以使用提供的 政策，或做為更嚴格政策的起點。

HealthOmics 使用 IAM 服務角色代表您存取其他 服務。例如，當您執行從 Amazon S3 讀取資料的工作流程時，您會建立或選擇服務角色。對於某些功能，您也需要設定其他 服務中資源的許可。在您開始使用 HealthOmics 之前，請先檢閱這些要求

如需 IAM 的詳細資訊，請參閱 IAM 使用者指南中的什麼是 IAM？。

主題

- HealthOmics 的身分型 IAM 政策
- 的服務角色 AWS HealthOmics
- 資源許可
- 使用 Amazon S3 URIs 存取資料的許可

HealthOmics 的身分型 IAM 政策

若要授予您帳戶中的使用者 HealthOmics 的存取權，請在 AWS Identity and Access Management (IAM) 中使用身分型政策。身分型政策可以直接套用至 IAM 使用者，或與使用者相關聯的 IAM 群組和角色。您也可以授予其他帳戶中的使用者在帳戶中擔任角色並存取 HealthOmics 資源的許可。

若要授予使用者對工作流程版本執行動作的許可，您必須將工作流程和特定工作流程版本新增至資源清單。

下列 IAM 政策允許使用者存取所有 HealthOmics API 動作，並將服務角色傳遞給 HealthOmics。

Example 使用者政策

JSON

```
{
```

```
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "omics:*"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "iam:PassRole"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "iam:PassedToService": "omics.amazonaws.com"
        }
      }
    }
  ]
}
```

當您使用 HealthOmics 時，也會與其他 AWS 服務互動。若要存取這些服務，請使用每個服務提供的受管政策。若要限制對資源子集的存取，您可以使用 受管政策做為起點，以建立更嚴格的政策。

- AmazonS3FullAccess – 存取任務使用的 Amazon S3 儲存貯體和物件。
- AmazonEC2ContainerRegistryFullAccess – 存取工作流程容器映像的 Amazon ECR 登錄檔和儲存庫。
- AWSLakeFormationDataAdmin – 存取分析存放區建立的 Lake Formation 資料庫和資料表。
- ResourceGroupsandTagEditorFullAccess – 使用 HealthOmics 標記 API 操作來標記 HealthOmics 資源。

上述政策不允許使用者建立 IAM 角色。對於具有這些執行任務許可的使用者，管理員必須建立授予 HealthOmics 存取資料來源許可的服務角色。如需詳細資訊，請參閱的服務角色 AWS HealthOmics。

定義執行的自訂 IAM 許可

您可以在授權StartRun請求中包含請求參考的任何工作流程、執行或執行群組。若要這樣做，請在 IAM 政策中列出所需的工作流程、執行或執行群組組合。例如，您可以將工作流程的使用限制在特定執行或執行群組。您也可以指定工作流程只能與執行群組搭配使用。

以下是允許具有單一執行群組的單一工作流程的 IAM 政策範例。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
      {
          "Effect": "Allow",
          "Action": [
              "omics:StartRun"
          ],
          "Resource": [
              "arn:aws:omics:us-west-2:123456789012:workflow/1234567",
              "arn:aws:omics:us-west-2:123456789012:runGroup/2345678"
          ]
      },
      {
          "Effect": "Allow",
          "Action": [
              "omics:StartRun"
          ],
          "Resource": [
              "arn:aws:omics:us-west-2:123456789012:run/*",
              "arn:aws:omics:us-west-2:123456789012:runGroup/2345678"
          ]
      },
      {
          "Effect": "Allow",
          "Action": [
              "omics:GetRun",
              "omics:ListRunTasks",
              "omics:GetRunTask",
              "omics:CancelRun",
              "omics:DeleteRun"
          ],
```

```
            "Resource": [
                "arn:aws:omics:us-west-2:123456789012:run/*"
            ]
        }
    ]
}
```

的服務角色 AWS HealthOmics

服務角色是 AWS Identity and Access Management (IAM) 角色，授予 AWS 服務存取您帳戶中資源的許可。當您啟動匯入任務或開始執行 AWS HealthOmics 時，您會向 提供服務角色。

HealthOmics 主控台可以為您建立所需的角色。如果您使用 HealthOmics API 來管理資源，請使用 IAM 主控台建立服務角色。如需詳細資訊，請參閱建立角色以將許可委派給 AWS 服務。

服務角色必須具有下列信任政策。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "omics.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

信任政策允許 HealthOmics 服務擔任該角色。

主題

- IAM 服務政策範例
- 範本範例 AWS CloudFormation

IAM 服務政策範例

在這些範例中，資源名稱和帳戶 IDs是您以實際值取代 的預留位置。

下列範例顯示可用於啟動執行之服務角色的政策。政策會授予許可，以存取執行的 Amazon S3 輸出位置、工作流程日誌群組和 Amazon ECR 容器。

> **Note**
>
> 如果您使用 呼叫快取執行，請在 s3 許可中將執行快取 Amazon S3 位置新增為資源。

Example 啟動執行的服務角色政策

JSON

```
{
"Version": "2012-10-17",
"Statement": [
      {
          "Effect": "Allow",
          "Action": [
              "s3:GetObject",
              "s3:PutObject"
          ],
          "Resource": [
              "arn:aws:s3:::amzn-s3-demo-bucket1/*"
          ]
      },
      {
          "Effect": "Allow",
          "Action": [
              "s3:ListBucket"
          ],
          "Resource": [
              "arn:aws:s3:::amzn-s3-demo-bucket1"
          ]
      },
      {
          "Effect": "Allow",
          "Action": [
              "logs:DescribeLogStreams",
```

```
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": [
                "arn:aws:logs:us-east-1:123456789012:log-group:/aws/omics/
WorkflowLog:log-stream:*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup"
            ],
            "Resource": [
                "arn:aws:logs:us-east-1:123456789012:log-group:/aws/omics/
WorkflowLog:*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ecr:BatchGetImage",
                "ecr:GetDownloadUrlForLayer",
                "ecr:BatchCheckLayerAvailability"
            ],
            "Resource": [
                "arn:aws:ecr:us-east-1:123456789012:repository/*"
            ]
        }
    ]
}
```

下列範例顯示您可以用於存放區匯入任務的服務角色政策。政策授予存取 Amazon S3 輸入位置 的許可。

Example 參考存放區任務的服務角色

JSON

```
{
"Version": "2012-10-17",
"Statement": [
```

```
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket/*"
            ]
        },

        {
            "Effect": "Allow",
            "Action": [
                "s3:GetBucketLocation"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket"
            ]
        }

    ]
}
```

範本範例 AWS CloudFormation

下列範例 AWS CloudFormation 範本會建立服務角色，讓 HealthOmics 有權存取名稱字首為 的 Amazon S3 儲存貯體omics-，以及上傳工作流程日誌。

Example 參考存放區、Amazon S3 和 CloudWatch Logs 許可

```
Parameters:
  bucketName:
    Description: Bucket name
    Type: String

Resources:
  serviceRole:
    Type: AWS::IAM::Role
    Properties:
      Policies:
        - PolicyName: read-reference
          PolicyDocument:
```

```
            Version: 2012-10-17
            Statement:
            - Effect: Allow
              Action:
                - omics:*
              Resource: !Sub arn:${AWS::Partition}:omics:${AWS::Region}:
${AWS::AccountId}:referenceStore/*
        - PolicyName: read-s3
          PolicyDocument:
            Version: 2012-10-17
            Statement:
            - Effect: Allow
              Action:
                - s3:ListBucket
              Resource: !Sub arn:${AWS::Partition}:s3:::${bucketName}
            - Effect: Allow
              Action:
                - s3:GetObject
                - s3:PutObject
              Resource: !Sub arn:${AWS::Partition}:s3:::${bucketName}/*
        - PolicyName: upload-logs
          PolicyDocument:
            Version: 2012-10-17
            Statement:
            - Effect: Allow
              Action:
                - logs:DescribeLogStreams
                - logs:CreateLogStream
                - logs:PutLogEvents
              Resource: !Sub arn:${AWS::Partition}:logs:${AWS::Region}:
${AWS::AccountId}:loggroup:/aws/omics/WorkflowLog:log-stream:*
            - Effect: Allow
              Action:
                - logs:CreateLogGroup
              Resource: !Sub arn:${AWS::Partition}:logs:${AWS::Region}:
${AWS::AccountId}:loggroup:/aws/omics/WorkflowLog:*
      AssumeRolePolicyDocument: |
        {
          "Version": "2012-10-17",
          "Statement": [
            {
              "Action": [
                "sts:AssumeRole"
              ],
```

```
            "Effect": "Allow",
            "Principal": {
              "Service": [
                "omics.amazonaws.com"
              ]
            }
          }
        ]
      }
```

資源許可

AWS HealthOmics 當您執行任務或建立存放區時， 會代表您建立和存取其他 服務中的資源。在某些情況下，您需要在其他 服務中設定許可，以存取資源或允許 HealthOmics 存取資源。

章節

- Amazon ECR 許可
- Lake Formation 許可

Amazon ECR 許可

在 HealthOmics 服務從您的私有 Amazon ECR 儲存庫在容器中執行工作流程之前，您可以為容器建立資源政策。政策會授予 HealthOmics 服務使用容器的許可。您可以將此資源政策新增至工作流程參考的每個私有儲存庫。

> **Note**
>
> 私有儲存庫和工作流程必須位於相同區域。

下列各節說明必要的政策組態。

主題

- 建立 Amazon ECR 儲存庫的資源政策
- 使用跨帳戶容器執行工作流程
- 共用工作流程的 Amazon ECR 儲存庫政策

建立 Amazon ECR 儲存庫的資源政策

建立資源政策，以允許 HealthOmics 服務使用儲存庫中的容器執行工作流程。此政策會授予 HealthOmics 服務主體存取所需 Amazon ECR 動作的許可。

請依照下列步驟建立政策：

1. 在 Amazon ECR 主控台中開啟私有儲存庫頁面，然後選取您要授予存取權的儲存庫。
2. 從側邊列導覽中，選取許可。
3. 選擇編輯 JSON。
4. 選擇新增陳述式。
5. 新增下列政策陳述式，然後選取儲存政策。

   JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "omics workflow access",
            "Effect": "Allow",
            "Principal": {
                "Service": "omics.amazonaws.com"
            },
            "Action": [
                "ecr:GetDownloadUrlForLayer",
                "ecr:BatchGetImage",
                "ecr:BatchCheckLayerAvailability"
            ],
            "Resource": "*"
        }
    ]
}
```

使用跨帳戶容器執行工作流程

如果不同的 AWS 帳戶擁有工作流程和容器，您需要設定下列跨帳戶許可：

1. 更新儲存庫的 Amazon ECR 政策，將許可明確授予擁有工作流程的帳戶。

2. 更新擁有工作流程的帳戶的服務角色，以授予容器映像的存取權。

下列範例示範 Amazon ECR 資源政策，授予擁有工作流程之帳戶的存取權。

在此範例中：

- 工作流程帳戶 ID：111122223333
- 容器儲存庫帳戶 ID：444455556666
- 容器名稱：samtools

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "omics.amazonaws.com"
            },
            "Action": [
                "ecr:BatchCheckLayerAvailability",
                "ecr:BatchGetImage",
                "ecr:GetDownloadUrlForLayer"
            ]
        },
        {
            "Sid": "allow access to the service role of the account that owns the
 workflow",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::111122223333:role/DemoCustomer"
            },
            "Action": [
                "ecr:BatchCheckLayerAvailability",
                "ecr:BatchGetImage",
                "ecr:GetDownloadUrlForLayer"
            ]
        }
    ]
```

```
}
```

若要完成設定，請將下列政策陳述式新增至擁有工作流程之帳戶的服務角色。此政策授予服務角色存取「samtools」容器映像的許可。請務必使用您自己的值取代帳戶號碼、容器名稱和區域。

```
{
    "Sid": "CrossAccountEcrRepoPolicy",
    "Effect": "Allow",
    "Action": ["ecr:BatchCheckLayerAvailability", "ecr:BatchGetImage",
 "ecr:GetDownloadUrlForLayer"],
    "Resource": "arn:aws:ecr:us-west-2:444455556666:repository/samtools"
}
```

共用工作流程的 Amazon ECR 儲存庫政策

> **Note**
>
> HealthOmics 會自動允許共用工作流程存取工作流程擁有者帳戶中的 Amazon ECR 儲存庫，而工作流程是在訂閱者的帳戶中執行。您不需要為共用工作流程授予其他儲存庫存取權。如需詳細資訊，請參閱共用 HealthOmics 工作流程。

根據預設，訂閱者無法存取 Amazon ECR 儲存庫來使用基礎容器。或者，您可以將條件索引鍵新增至儲存庫的資源政策，以自訂對 Amazon ECR 儲存庫的存取。以下各節提供範例政策。

限制對特定工作流程的存取

您可以在條件陳述式中列出個別工作流程，因此只有這些工作流程可以使用儲存庫中的容器。SourceArn 條件金鑰指定共用工作流程的 ARN。下列範例會授予指定工作流程使用此儲存庫的許可。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
     {
       "Sid": "OmicsAccessPrincipal",
       "Effect": "Allow",
```

```
        "Principal": {
          "Service": "omics.amazonaws.com"
        },
        "Action": [
          "ecr:GetDownloadUrlForLayer",
          "ecr:BatchGetImage",
          "ecr:BatchCheckLayerAvailability"
        ],
        "Resource": "*",
        "Condition": {
           "StringEquals": {
              "aws:SourceArn": "arn:aws:omics:us-
east-1:111122223333:workflow/1234567"
           }
        }
      }
  ]
}
```

限制對特定帳戶的存取

您可以在條件陳述式中列出訂閱者帳戶，以便只有這些帳戶具有在儲存庫中使用容器的許可。SourceAccount 條件索引鍵會指定訂閱 AWS 帳戶 者的 。下列範例會授予指定帳戶使用此儲存庫的許可。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "OmicsAccessPrincipal",
      "Effect": "Allow",
      "Principal": {
        "Service": "omics.amazonaws.com"
      },
      "Action": [
        "ecr:GetDownloadUrlForLayer",
        "ecr:BatchGetImage",
        "ecr:BatchCheckLayerAvailability"
      ],
```

```
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "111122223333"
        }
      }
    }
  ]
}
```

您也可以拒絕特定訂閱者的 Amazon ECR 許可，如下列範例政策所示。

JSON

```
{
      "Version": "2012-10-17",
      "Statement": [
          {
              "Sid": "OmicsAccessPrincipal",
              "Effect": "Allow",
              "Principal": {
                  "Service": "omics.amazonaws.com"
              },
              "Action": [
                  "ecr:GetDownloadUrlForLayer",
                  "ecr:BatchGetImage",
                  "ecr:BatchCheckLayerAvailability"
              ],
              "Resource": "*",
              "Condition": {
                "StringNotEquals": {
                  "aws:SourceAccount": "111122223333"
                }
            }
         }
      ]
  }
```

Lake Formation 許可

在 HealthOmics 中使用分析功能之前，請在 Lake Formation 中設定預設資料庫設定。

在 Lake Formation 中設定資源許可

1. 在 Lake Formation 主控台中開啟資料目錄設定頁面。
2. 在新建立的資料庫和資料表的預設許可下，取消勾選資料庫和資料表的 IAM 存取控制需求。
3. 選擇儲存。

如果您的服務政策具有正確的 RAM 許可，HealthOmics Analytics 會自動接受資料，例如下列範例。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "omics:*"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "ram:AcceptResourceShareInvitation",
        "ram:GetResourceShareInvitations"
      ],
      "Resource": "*"
    }
  ]
}
```

使用 Amazon S3 URIs 存取資料的許可

您可以使用 HealthOmics API 操作或 Amazon S3 API 操作來存取序列存放區資料。

對於 HealthOmics API 存取，HealthOmics 許可是透過 IAM 政策管理。不過，S3 存取需要兩個層級的組態：在存放區的 S3 存取政策和 IAM 政策中明確允許。若要進一步了解如何搭配 HealthOmics 使用 IAM 政策，請參閱 HealthOmics 的服務角色。

有三種方式可以共用使用 Amazon S3 APIs 讀取物件的功能：

1. 以政策為基礎的共用 – 此共用需要在 S3 存取政策中啟用 IAM 主體，並撰寫 IAM 政策並將其連接到 IAM 主體。如需詳細資訊，請參閱下一個主題。
2. 預先簽章URLs – 您也可以為序列存放區中的檔案產生可共用的預先簽章 URL。若要進一步了解如何使用 Amazon S3 建立預先簽章URLs，請參閱 Amazon S3 文件中的使用預先簽章URLs。序列存放區 S3 存取政策支援用於限制預先簽章 URL 功能的陳述式。
3. 擔任的角色 – 在資料擁有者的 帳戶中建立角色，其具有允許使用者擔任該角色的存取政策。

主題

以政策為基礎的共用

如果您使用直接 S3 URI 存取序列存放區資料，HealthOmics 會為相關聯的 S3 儲存貯體存取政策提供增強的安全措施。

下列規則適用於新的 S3 存取政策。對於現有政策，規則會在您下次更新政策時套用：

- S3 存取政策支援下列政策元素
  - 版本、ID、陳述式、Sid、Effect、委託人、動作、資源、條件
- S3 存取政策支援下列條件金鑰：
  - s3：ExistingObjectTag/<key>、s3：prefix、s3：signatureversion、s3：TlsVersion
  - 政策也支援使用下列條件運算子的 aws：PrincipalArn：ArnEquals 和 ArnLike

如果您嘗試新增或更新政策以包含不支援的元素或條件，系統會拒絕請求。

主題

- IAM 政策
- 標籤式存取控制

預設 S3 存取政策

當您建立序列存放區時，HealthOmics 會建立預設 S3 存取政策，授予資料存放區擁有者的根帳戶序列存放區中所有可存取物件的下列許可：S3：GetObject、S3GetObjectTagging 和 S3：ListBucket。預設建立的政策為：

JSON

```
{
    "Version": "2012-10-17",
    "Statement":
    [
        {
            "Effect": "Allow",
            "Principal":
            {
                "AWS": "arn:aws:iam::111111111111:root"
            },
            "Action":
            [
                "s3:GetObject",
                "s3:GetObjectTagging"
            ],
            "Resource": "arn:aws:s3:us-
west-2:222222222222:accesspoint/111111111111-1234567890/object/111111111111/
sequenceStore/1234567890/*"
        },
        {
            "Effect": "Allow",
            "Principal":
            {
                "AWS": "arn:aws:iam::111111111111:root"
            },
            "Action": "s3:ListBucket",
            "Resource": "arn:aws:s3:us-
west-2:222222222222:accesspoint/111111111111-1234567890/111111111111/
sequenceStore/1234567890/*"
        }
```

```
    ]
}
```

自訂存取政策

如果 S3 存取政策為空白，則不允許 S3 存取。如果有現有政策且您需要移除 s3 存取權，請使用 `deleteS3AccessPolicy`移除所有存取權。

若要新增共用的限制或授予其他帳戶的存取權，您可以使用 `PutS3AccessPolicy` API 更新政策。政策的更新不能超過序列存放區的字首或指定的動作。

IAM 政策

若要允許使用者或 IAM 主體使用 Amazon S3 APIs 存取，除了 S3 存取政策中的許可之外，還需要建立 IAM 政策並連接到主體以授予存取權。允許 Amazon S3 API 存取的政策可以在序列存放區層級或讀取集層級套用。在讀取集層級，許可可以透過字首或使用資源標籤篩選條件來限制範例或主體 ID 模式。

如果序列存放區使用客戶受管金鑰 (CMK)，委託人也必須具有使用 KMS 金鑰進行解密的權限。如需詳細資訊，請參閱《 AWS Key Management Service 開發人員指南》中的跨帳戶 KMS 存取。

下列範例可讓使用者存取序列存放區。您可以使用其他條件或資源型篩選條件來微調存取權。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::111111111111:root"
      },
      "Action":
      [
         "s3:GetObject",
         "s3:GetObjectTagging"
      ],
      "Resource": "arn:aws:s3:us-
west-2:222222222222:accesspoint/111111111111-1234567890/object/111111111111/
sequenceStore/1234567890/*",
```

```
      "Condition": {
        "StringEquals": {
          "s3:ExistingObjectTag/omics:readSetStatus": "ACTIVE"
        }
      }
    },
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::111111111111:root"
      },
      "Action": "s3:ListBucket",
      "Resource": "arn:aws:s3:us-
west-2:222222222222:accesspoint/111111111111-1234567890",
      "Condition": {
        "StringLike": {
          "s3:prefix": "111111111111/sequenceStore/1234567890/*"
        }
      }
    }
  ]
}
```

標籤式存取控制

若要使用標籤型存取控制，必須先更新序列存放區，以傳播將使用的標籤金鑰。此組態是在序列存放區建立或更新期間設定。標籤傳播後，即可使用標籤條件來進一步新增限制。這些限制可以放置在 S3 存取政策或 IAM 政策中。以下是將會設定的以標籤為基礎的 S3 存取政策範例：

```
{
    "Sid": "tagRestrictedGets",
    "Effect": "Allow",
    "Principal":
    {
        "AWS": "arn:aws:iam::<target_restricted_account_id>:root"
    },
    "Action":
    [
        "s3:GetObject",
        "s3:GetObjectTagging"
    ],
```

```
    "Resource": "arn:aws:s3:us-west-2:222222222222:accesspoint/111111111111-1234567890/
object/111111111111/sequenceStore/1234567890/*",
    "Condition":
    {
        "StringEquals":
        {
            "s3:ExistingObjectTag/tagKey1": "tagValue1",
            "s3:ExistingObjectTag/tagKey2": "tagValue2"
        }
    }
}
```

限制範例

案例：建立共用，其中資料擁有者可以限制使用者下載「撤銷」資料的能力。

在此案例中，資料擁有者 （帳戶 #111111111111) 管理資料存放區。此資料擁有者會與廣泛的第三方使用者共用資料，包括研究人員 （帳戶編號 999999999999)。在管理資料的過程中，資料擁有者會定期收到撤銷參與者資料的請求。若要管理此撤回，資料擁有者會先限制接收請求時的直接下載存取權，最後根據其需求刪除資料。

為了滿足此需求，資料擁有者會設定序列存放區，而且每個讀取集都會收到「狀態」的標籤，如果提取請求通過，則會將其設定為「撤銷」。對於標籤設定為此值的資料，他們想要確保沒有任何使用者可以在此檔案上執行「getObject」。若要執行此設定，資料擁有者將需要確保採取兩個步驟。

步驟 1. 對於序列存放區，請確定狀態標籤已更新為要傳播。方法是在呼叫 `createSequenceStore`或 `propogatedSetLevelTags`時，將「狀態」金鑰新增至 。`updateSequenceStore.`

步驟 2. 更新存放區的 s3 存取政策，以限制狀態標籤設為撤銷之物件的 getObject。方法是使用 `PutS3AccesPolicy` API 更新存放區存取政策。下列政策可讓客戶在列出物件時仍能看到撤銷的檔案，但無法加以存取：

- 陳述式 1 (restrictedGetWithdrawal：帳戶 999999999999 無法擷取已撤銷的物件。
- 陳述式 2 (ownerGetAll)：資料擁有者帳戶 111111111111 可以擷取所有物件，包括已撤銷的物件。
- 陳述式 3 (everyoneListAll)：所有共用帳戶 111111111111 和 999999999999 都可以在整個字首上執行 ListBucket操作。

JSON

```
{
    "Version": "2012-10-17",
    "Statement":
    [
        {
            "Sid": "restrictedGetWithdrawal",
            "Effect": "Allow",
            "Principal":
            {
                "AWS": "arn:aws:iam::999999999999:root"
            },
            "Action":
            [
                "s3:GetObject",
                "s3:GetObjectTagging"
            ],
            "Resource": "arn:aws:s3:us-
west-2:222222222222:accesspoint/111111111111-1234567890/object/111111111111/
sequenceStore/1234567890/*",
            "Condition":
            {
                "StringNotEquals":
                {
                    "s3:ExistingObjectTag/status": "withdrawn"
                }
            }
        },
        {
            "Sid": "ownerGetAll",
            "Effect": "Allow",
            "Principal":
            {
                "AWS": "arn:aws:iam::111111111111:root"
            },
            "Action":
            [
                "s3:GetObject",
                "s3:GetObjectTagging"
            ],
```

```
            "Resource": "arn:aws:s3:us-
west-2:222222222222:accesspoint/111111111111-1234567890/object/111111111111/
sequenceStore/1234567890/*",
            "Condition":
            {
                "StringEquals":
                {
                    "s3:ExistingObjectTag/omics:readSetStatus": "ACTIVE"
                }
            }
        },
        {
            "Sid": "everyoneListAll",
            "Effect": "Allow",
            "Principal":
            {
                "AWS": [
                    "arn:aws:iam::111111111111:root",
                    "arn:aws:iam::999999999999:root"
                ]
            },
            "Action": "s3:ListBucket",
            "Resource": "arn:aws:s3:us-
west-2:222222222222:accesspoint/111111111111-1234567890",
            "Condition":
            {
                "StringLike":
                {
                    "s3:prefix": "111111111111/sequenceStore/1234567890/*"
                }
            }
        }
    ]
}
```

AWS HealthOmics 的安全性

的雲端安全性 AWS 是最高優先順序。身為 AWS 客戶，您可以受益於資料中心和網路架構，這些架構專為符合最安全敏感組織的需求而建置。

安全性是 AWS 與您之間的共同責任。共同責任模型將其描述為雲端的安全性和雲端中的安全性：

- 雲端的安全性 – AWS 負責保護在 中執行 AWS 服務的基礎設施 AWS 雲端。 AWS 也為您提供可安全使用的服務。作為AWS 合規計畫的一部分，第三方稽核人員會定期測試和驗證我們安全的有效性。若要了解適用於 AWS HealthOmics 的合規計劃，請參閱AWS 合規計劃的 服務範圍。
- 雲端的安全性 – 您的責任取決於您使用 AWS 的服務。您也必須對其他因素負責，包括資料的機密性、您公司的要求和適用法律和法規。

本文件可協助您了解如何在使用 AWS HealthOmics 時套用共同責任模型。下列主題說明如何設定 AWS HealthOmics 以符合您的安全與合規目標。您也會了解如何使用其他 AWS 服務來協助您監控和保護 AWS HealthOmics 資源。

主題

中的資料保護 AWS HealthOmics

AWS 共同責任模型適用於 AWS HealthOmics 中的資料保護。如此模型所述， AWS 負責保護執行所有 的 全球基礎設施 AWS 雲端。您負責維護在此基礎設施上託管內容的控制權。您也同時負責所使用 AWS 服務 的安全組態和管理任務。如需資料隱私權的詳細資訊，請參閱資料隱私權常見問答集。如需有關歐洲資料保護的相關資訊，請參閱 AWS 安全性部落格上的 AWS 共同的責任模型和 GDPR 部落格文章。

基於資料保護目的，我們建議您保護 AWS 帳戶 登入資料，並使用 AWS IAM Identity Center 或 AWS Identity and Access Management (IAM) 設定個別使用者。如此一來，每個使用者都只會獲得授與完成其任務所必須的許可。我們也建議您採用下列方式保護資料：

- 每個帳戶均要使用多重要素驗證 (MFA)。
- 使用 SSL/TLS 與 AWS 資源通訊。我們需要 TLS 1.2 並建議使用 TLS 1.3。
- 使用 設定 API 和使用者活動記錄 AWS CloudTrail。如需有關使用 CloudTrail 追蹤擷取 AWS 活動的資訊，請參閱AWS CloudTrail 《 使用者指南》中的使用 CloudTrail 追蹤。
- 使用 AWS 加密解決方案，以及其中的所有預設安全控制 AWS 服務。
- 使用進階的受管安全服務 (例如 Amazon Macie)，協助探索和保護儲存在 Amazon S3 的敏感資料。
- 如果您在 AWS 透過命令列界面或 API 存取 時需要 FIPS 140-3 驗證的密碼編譯模組，請使用 FIPS 端點。如需有關 FIPS 和 FIPS 端點的更多相關資訊，請參閱聯邦資訊處理標準 (FIPS) 140-3。

我們強烈建議您絕對不要將客戶的電子郵件地址等機密或敏感資訊，放在標籤或自由格式的文字欄位中，例如名稱欄位。這包括當您使用 AWS HealthOmics 或使用 AWS 服務 主控台、API AWS CLI或其他 AWS SDKs 時。您在標籤或自由格式文字欄位中輸入的任何資料都可能用於計費或診斷日誌。如果您提供外部伺服器的 URL，我們強烈建議請勿在驗證您對該伺服器請求的 URL 中包含憑證資訊。

靜態加密

主題

- AWS 擁有的金鑰
- 客戶受管金鑰
- 建立客戶受管金鑰
- 使用客戶受管金鑰所需的 IAM 許可
- 進一步了解

為了保護靜態敏感客戶資料， 預設會使用服務擁有的 AWS Key Management Service (AWS KMS) 金鑰 AWS HealthOmics 提供加密。也支援客戶受管金鑰。若要進一步了解客戶受管金鑰，請參閱 Amazon Key Management Service。

所有 HealthOmics 資料存放區 （ 儲存和分析 ） 都支援使用客戶受管金鑰。建立資料存放區之後，就無法變更加密組態。如果資料存放區使用 AWS 擁有的金鑰，則會將其表示為 AWS_OWNED_KMS_KEY，而且您不會看到用於靜態加密的特定金鑰。

對於 HealthOmics 工作流程，暫存檔案系統不支援客戶受管金鑰；不過，所有資料都會使用 XTS-AES-256 區塊加密演算法自動加密靜態資料，以加密檔案系統。用於啟動工作流程執行的 IAM 使用者

和角色也必須能夠存取用於工作流程輸入和輸出儲存貯體的 AWS KMS 金鑰。工作流程不使用授予，且 AWS KMS 加密僅限於輸入和輸出 Amazon S3 儲存貯體。用於工作流程 APIs 的 IAM 角色也必須能夠存取使用的 AWS KMS 金鑰，以及輸入和輸出 Amazon S3 儲存貯體。您可以使用 IAM 角色和許可來控制存取或 AWS KMS 政策。若要進一步了解，請參閱 的身分驗證和存取控制 AWS KMS。

當您 AWS Lake Formation 搭配 HealthOmics Analytics 使用 時，與 Lake Formation 相關聯的任何解密許可也會授予輸入和輸出 Amazon S3 儲存貯體。如需有關 如何 AWS Lake Formation 管理許可的詳細資訊，請參閱 AWS Lake Formation 文件。

HealthOmics Analytics 授予 Lake Formation kms:Decrypt 讀取 Amazon S3 儲存貯體中加密資料的許可。只要您有透過 Lake Formation 查詢資料的許可，您就可以讀取加密的資料。對資料的存取是透過 Lake Formation 中的資料存取控制來控制，而不是透過 KMS 金鑰政策。若要進一步了解，請參閱 Lake Formation 文件中的AWS 整合式 AWS 服務請求。

AWS 擁有的金鑰

根據預設，HealthOmics 會使用 AWS 擁有的金鑰 自動加密靜態資料，因為此資料可能包含敏感資訊，例如個人身分識別資訊 (PII) 或受保護醫療資訊 (PHI)。 AWS 擁有的金鑰 不會存放在您的帳戶中。它們是 AWS 擁有和管理用於多個 AWS 帳戶的 KMS 金鑰集合的一部分。

AWS 服務可以使用 AWS 擁有的金鑰 來保護您的資料。您無法檢視、管理或存取 AWS 擁有的金鑰，或稽核其使用方式。不過，您不需要執行任何工作或變更任何程式來保護加密資料的金鑰。

您不需要支付每月費用或使用費 AWS 擁有的金鑰，也不會計入您帳戶的 AWS KMS 配額。如需詳細資訊，請參閱AWS 受管金鑰。

客戶受管金鑰

HealthOmics 支援使用您建立、擁有和管理的對稱客戶受管金鑰，以透過現有 AWS 擁有的加密新增第二層加密。您可以完全控制此層加密，因此能執行以下任務：

- 建立和維護金鑰政策、IAM 政策和授予
- 輪換金鑰密碼編譯資料
- 啟用和停用金鑰政策
- 新增標籤
- 建立金鑰別名
- 安排金鑰供刪除

您也可以使用 CloudTrail 來追蹤 HealthOmics AWS KMS 代表您傳送到 的請求。需支付額外費用 AWS KMS 。如需詳細資訊，請參閱客戶受管金鑰。

建立客戶受管金鑰

您可以使用 AWS 管理主控台或 AWS KMS APIs 來建立對稱客戶受管金鑰。

請遵循 AWS Key Management Service 開發人員指南中建立對稱客戶受管金鑰的步驟。

金鑰政策會控制客戶受管金鑰的存取權限。每個客戶受管金鑰都必須只有一個金鑰政策，其中包含決定誰可以使用金鑰及其使用方式的陳述式。當您建立客戶受管金鑰時，您可以指定金鑰政策。如需詳細資訊，請參閱《AWS Key Management Service 開發人員指南》中的管理對客戶受管金鑰的存取。

若要搭配 HealthOmics Analytics 資源使用客戶受管金鑰，呼叫委託人需要在金鑰政策中執行 kms:CreateGrant 操作。這可讓系統使用 FAS 字符來建立對客戶受管金鑰的授予，以控制對指定 KMS 金鑰的存取。此金鑰可讓使用者存取 HealthOmics 所需的 kms:grant 操作。如需詳細資訊，請參閱使用授予。

對於 HealthOmics 分析，呼叫委託人必須允許下列 API 操作：

- kms:CreateGrant 將授予新增至特定客戶受管金鑰，允許存取 HealthOmics Analytics 中的授予操作。
- kms:DescribeKey 提供驗證金鑰所需的客戶受管金鑰詳細資訊。這是所有 操作的必要項目。
- kms:GenerateDataKey 提供存取權，以加密所有寫入操作的靜態資源。此外，此動作提供客戶受管金鑰詳細資訊，服務可用來驗證發起人是否具有使用該金鑰的存取權。
- kms:Decrypt 可讓您存取加密資源的讀取或搜尋操作。

若要搭配 HealthOmics 儲存資源使用客戶受管金鑰，必須在金鑰政策中允許 HealthOmics 服務主體和呼叫主體。這可讓服務驗證發起人是否有權存取 金鑰，並使用 服務主體來使用客戶受管金鑰執行儲存體管理。對於 HealthOmics 儲存，服務主體的金鑰政策必須允許下列 API 操作：

- kms:DescribeKey 提供驗證金鑰所需的客戶受管金鑰詳細資訊。這是所有 操作的必要項目。
- kms:GenerateDataKey 提供存取權，以加密所有寫入操作的靜態資源。此外，此動作提供客戶受管金鑰詳細資訊，服務可用來驗證發起人是否具有使用該金鑰的存取權。
- kms:Decrypt 可讓您存取加密資源的讀取或搜尋操作。

下列範例顯示政策陳述式，允許服務主體建立並使用客戶受管金鑰加密的 HealthOmics 序列或參考存放區，並與之互動：

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
{
    "Effect": "Allow",
    "Principal": {
        "Service": "omics.amazonaws.com"
    },
    "Action": [
        "kms:Decrypt",
        "kms:DescribeKey",
        "kms:Encrypt",
        "kms:GenerateDataKey*"
    ],
    "Resource": "*"
        }
    ]
}
```

下列範例顯示為資料存放區建立許可以從 Amazon S3 儲存貯體解密資料的政策。

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "omics:GetReference",
                "omics:GetReferenceMetadata"
            ],
            "Resource": [
                "arn:AWS:omics:{{region}}:{{accountId}}:referenceStore/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
```

```
                "s3:GetObject"
            ],
            "Resource": [
                "arn:AWS:s3:::[[s3path]]/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:AWS:kms:{{region}}:{111122223333}:key/{{key_id}}"
            ],
            "Condition": {
              "StringEquals": {
                  "kms:ViaService": [
                    "s3.{{region}}.amazonAWS.com"
                  ]
              }
           }
        }
    ]
}
```

使用客戶受管金鑰所需的 IAM 許可

建立資源時，例如使用客戶受管金鑰進行 AWS KMS 加密的資料存放區，金鑰政策和 IAM 使用者或角色的 IAM 政策都有必要的許可。

您可以使用 kms:ViaService 條件金鑰，將 KMS 金鑰的使用限制為僅來自 HealthOmics 的請求。

如需金鑰政策的詳細資訊，請參閱《AWS Key Management Service 開發人員指南》中的啟用 IAM 政策。

主題

- Analytics API 許可
- 儲存 API 許可
- HealthOmics 如何在 AWS KMS 中使用授予
- 監控 AWS HealthOmics 的加密金鑰

Analytics API 許可

對於 HealthOmics 分析，建立存放區的 IAM 使用者或角色必須具有 kms:CreateGrant、kms:GenerateDataKey、 kms:Decrypt 和 kms:DescribeKey 許可，以及必要的 HealthOmics 許可。

儲存 API 許可

對於 HealthOmics 儲存 APIs，呼叫下列 API 操作的 IAM 使用者或角色需要列出的許可：

CreateReferenceStore、CreateSequenceStore

若要建立存放區，IAM 呼叫者必須具有`kms:DescribeKey`許可加上必要的 HealthOmics 許可。HealthOmics 服務主體會呼叫 `kms:GenerateDataKeyWithoutPlaintext` 來執行資料載入和存取的存取驗證檢查。

StartReadSetImportJob、StartReferenceImportJob

若要開始資料匯入任務，IAM 發起人必須擁有儲存體上 KMS 金鑰的 `kms:Decrypt` 和 `kms:GenerateDataKey` 許可，才能匯入，以及包含要匯入物件的 Amazon S3 儲存貯體上的 `kms:Decrypt` 許可。此外，傳遞至呼叫的角色必須具有 Amazon S3 儲存貯體的`kms:Decrypt`許可，其中包含要匯入的物件。IAM 發起人也必須具有將角色傳遞至任務的許可。

CreateMultipartReadSetUpload、UploadReadSetPart、 CompleteMultipartReadSetUpload

若要完成分段上傳，IAM 發起人必須擁有 `kms:Decrypt`和 ，`kms:GenerateDataKey`才能建立、上傳和完成分段上傳。

StartReadSetExportJob

若要啟動資料匯出任務，IAM 發起人必須具有儲存體上 KMS 金鑰的`kms:Decrypt`許可，才能從接收物件的 Amazon S3 儲存貯體匯出 和 `kms:GenerateDataKey`和 `kms:Decrypt`許可。此外，傳入呼叫的角色必須具有接收物件之 Amazon S3 儲存貯體的`kms:Decrypt`許可。IAM 發起人也必須具有將角色傳遞至任務的許可。

StartReadsetActivationJob

若要啟動讀取集啟用任務，IAM 發起人必須具有 物件的 `kms:Decrypt`和 `kms:GenerateDataKey`許可。

GetReference、GetReadSet

若要從存放區讀取物件，IAM 呼叫者必須具有物件的`kms:Decrypt`許可。

讀取集合 S3 GetObject

若要使用 Amazon S3 `GetObject` API 從商店讀取物件，IAM 發起人必須具有物件的`kms:Decrypt`許可。為客戶受管金鑰和 AWS 擁有的金鑰 組態設定此許可。

HealthOmics 如何在 AWS KMS 中使用授予

HealthOmics Analytics 需要授予才能使用客戶受管 KMS 金鑰。HealthOmics 工作流程不需要或使用授予。HealthOmics Storage 會直接從服務委託人使用客戶受管金鑰，因此請勿使用授予。當您建立使用客戶受管金鑰加密的分析存放區時，HealthOmics 分析會透過傳送 CreateGrant 請求至 AWS KMS 來代表您建立授予。AWS KMS 中的授予用於讓 HealthOmics 存取客戶帳戶中的 KMS 金鑰。

不建議撤銷或淘汰 HealthOmics 分析代表您建立的授予。如果您撤銷或淘汰授予 HealthOmics 許可以使用您帳戶中的 AWS KMS 金鑰的授予，HealthOmics 無法存取此資料、加密推送到資料存放區的新資源，或在提取時解密這些資源。

當您撤銷或淘汰 HealthOmics 的授予時，變更會立即發生。若要撤銷存取權，建議您刪除資料存放區，而不是撤銷授予。當您刪除資料存放區時，HealthOmics 會代表您淘汰授予。

監控 AWS HealthOmics 的加密金鑰

您可以使用 CloudTrail 來追蹤在使用客戶受管金鑰時代表您 AWS HealthOmics 傳送給 AWS KMS 的請求。CloudTrail 日誌中的日誌項目會在 userAgent 欄位中顯示 HealthOmics.amazonAWS.com，以清楚區分 HealthOmics 提出的請求。

下列範例是 CreateGrant、GenerateDataKey、Decrypt 和 DescribeKey 的 CloudTrail 事件，用於監控 HealthOmics 呼叫 AWS KMS 的操作，以存取客戶受管金鑰加密的資料。

以下也說明如何使用 CreateGrant 來允許 HealthOmics 分析存取客戶提供的 KMS 金鑰，讓 HealthOmics 能夠使用該 KMS 金鑰來加密所有靜態客戶資料。

您不需要建立自己的授予。HealthOmics 透過傳送 CreateGrant 請求至 AWS KMS 來代表您建立授予。中的授予 AWS KMS 用於授予 HealthOmics 存取客戶帳戶中 AWS KMS 金鑰的權限。

```
{
    "eventVersion": "1.08",
    "userIdentity": {
        "type": "AssumedRole",
        "principalId": "xx:test",
        "arn": "arn:AWS:sts::555555555555:assumed-role/user-admin/test",
        "accountId": "xx",
        "accessKeyId": "xxx",
```

```
        "sessionContext": {
            "sessionIssuer": {
                "type": "Role",
                "principalId": "xxxx",
                "arn": "arn:AWS:iam::555555555555:role/user-admin",
                "accountId": "555555555555",
                "userName": "user-admin"
            },
            "webIdFederationData": {},
            "attributes": {
                "creationDate": "2022-11-11T01:36:17Z",
                "mfaAuthenticated": "false"
            }
        },
        "invokedBy": "apigateway.amazonAWS.com"
    },
    "eventTime": "2022-11-11T02:34:41Z",
    "eventSource": "kms.amazonAWS.com",
    "eventName": "CreateGrant",
    "AWSRegion": "us-west-2",
    "sourceIPAddress": "apigateway.amazonAWS.com",
    "userAgent": "apigateway.amazonAWS.com",
    "requestParameters": {
        "granteePrincipal": "AWS Internal",
        "keyId": "arn:AWS:kms:us-west-2:555555555555:key/a6e87d77-cc3e-4a98-a354-
e4c275d775ef",
        "operations": [
            "CreateGrant",
            "RetireGrant",
            "Decrypt",
            "GenerateDataKey"
        ]
    },
    "responseElements": {
        "grantId": "4869b81e0e1db234342842af9f5531d692a76edaff03e94f4645d493f4620ed7",
        "keyId": "arn:AWS:kms:us-west-2:245126421963:key/xx-cc3e-4a98-a354-
e4c275d775ef"
    },
    "requestID": "d31d23d6-b6ce-41b3-bbca-6e0757f7c59a",
    "eventID": "3a746636-20ef-426b-861f-e77efc56e23c",
    "readOnly": false,
    "resources": [
        {
            "accountId": "245126421963",
```

```
            "type": "AWS::KMS::Key",
            "ARN": "arn:AWS:kms:us-west-2:245126421963:key/xx-cc3e-4a98-a354-
e4c275d775ef"
        }
    ],
    "eventType": "AWSApiCall",
    "managementEvent": true,
    "recipientAccountId": "245126421963",
    "eventCategory": "Management"
}
```

下列範例示範如何使用 GenerateDataKey，確保使用者擁有在儲存資料之前加密資料的必要許可。

```
{
    "eventVersion": "1.08",
    "userIdentity": {
        "type": "AssumedRole",
        "principalId": "EXAMPLEUSER",
        "arn": "arn:AWS:sts::111122223333:assumed-role/Sampleuser01",
        "accountId": "111122223333",
        "accessKeyId": "EXAMPLEKEYID",
        "sessionContext": {
            "sessionIssuer": {
                "type": "Role",
                "principalId": "EXAMPLEROLE",
                "arn": "arn:AWS:iam::111122223333:role/Sampleuser01",
                "accountId": "111122223333",
                "userName": "Sampleuser01"
            },
            "webIdFederationData": {},
            "attributes": {
                "creationDate": "2021-06-30T21:17:06Z",
                "mfaAuthenticated": "false"
            }
        },
        "invokedBy": "omics.amazonAWS.com"
    },
    "eventTime": "2021-06-30T21:17:37Z",
    "eventSource": "kms.amazonAWS.com",
    "eventName": "GenerateDataKey",
    "AWSRegion": "us-east-1",
    "sourceIPAddress": "omics.amazonAWS.com",
```

```
    "userAgent": "omics.amazonAWS.com",
    "requestParameters": {
        "keySpec": "AES_256",
        "keyId": "arn:AWS:kms:us-east-1:111122223333:key/EXAMPLE_KEY_ARN"
    },
    "responseElements": null,
    "requestID": "EXAMPLE_ID_01",
    "eventID": "EXAMPLE_ID_02",
    "readOnly": true,
    "resources": [
        {
            "accountId": "111122223333",
            "type": "AWS::KMS::Key",
            "ARN": "arn:AWS:kms:us-east-1:111122223333:key/EXAMPLE_KEY_ARN"
        }
    ],
    "eventType": "AWSApiCall",
    "managementEvent": true,
    "recipientAccountId": "111122223333",
    "eventCategory": "Management"
}
```


進一步了解

下列資源提供有關靜態資料加密的詳細資訊。

如需 AWS Key Management Service 基本概念的詳細資訊，請參閱 AWS KMS 文件。

如需 AWS KMS 文件中安全最佳實務的詳細資訊。

傳輸中加密

AWS HealthOmics 使用 TLS 1.2+ 透過公有端點和後端服務加密傳輸中的資料。

HealthOmics 中的身分和存取管理

AWS Identity and Access Management (IAM) 是 AWS 服務 ，可協助管理員安全地控制對 AWS 資源的存取。IAM 管理員可控制誰可以進行身分驗證 （登入） 和授權 （具有許可），以使用 AWS HealthOmics 資源。IAM 是 AWS 服務 您可以免費使用的 。

主題

目標對象

您使用 AWS Identity and Access Management (IAM) 的方式會有所不同，取決於您在 AWS HealthOmics 中執行的工作。

服務使用者 – 如果您使用 AWS HealthOmics 服務來執行任務，您的管理員會為您提供所需的登入資料和許可。當您使用更多 AWS HealthOmics 功能來執行工作時，您可能需要額外的許可。了解存取許可的管理方式可協助您向管理員請求正確的許可。如果您無法存取 AWS HealthOmics 中的功能，請參閱對 AWS HealthOmics 身分和存取進行故障診斷。

服務管理員 – 如果您在公司負責 AWS HealthOmics 資源，您可能擁有 AWS HealthOmics 的完整存取權。您的任務是判斷服務使用者應存取的 AWS HealthOmics 功能和資源。接著，您必須將請求提交給您的 IAM 管理員，來變更您服務使用者的許可。檢閱此頁面上的資訊，了解 IAM 的基本概念。若要進一步了解貴公司如何搭配 AWS HealthOmics 使用 IAM，請參閱 AWS HealthOmics 如何使用 IAM。

IAM 管理員 – 如果您是 IAM 管理員，建議您進一步了解如何撰寫政策以管理對 AWS HealthOmics 的存取。若要檢視您可以在 IAM 中使用的 AWS HealthOmics 身分型政策範例，請參閱 的身分型政策範例 AWS HealthOmics。

使用身分驗證

身分驗證是您 AWS 使用身分憑證登入 的方式。您必須以 AWS 帳戶根使用者身分、IAM 使用者身分或擔任 IAM 角色來驗證 （ 登入 AWS)。

您可以使用透過身分來源提供的憑證，以聯合身分 AWS 身分身分登入 。 AWS IAM Identity Center (IAM Identity Center) 使用者、您公司的單一登入身分驗證，以及您的 Google 或 Facebook 登入資料，都是聯合身分的範例。您以聯合身分登入時，您的管理員先前已設定使用 IAM 角色的聯合身分。當您使用聯合 AWS 身分存取 時，您會間接擔任角色。

根據您的使用者類型，您可以登入 AWS Management Console 或 AWS 存取入口網站。如需登入的詳細資訊 AWS，請參閱AWS 登入 《 使用者指南》中的如何登入您的 AWS 帳戶 。

如果您以 AWS 程式設計方式存取 ， AWS 會提供軟體開發套件 (SDK) 和命令列界面 (CLI)，以使用您的登入資料以密碼編譯方式簽署您的請求。如果您不使用 AWS 工具，則必須自行簽署請求。如需使用建議的方法自行簽署請求的詳細資訊，請參閱《IAM 使用者指南》中的適用於 API 請求的AWS Signature 第 4 版。

無論您使用何種身分驗證方法，您可能都需要提供額外的安全性資訊。例如， AWS 建議您使用多重驗證 (MFA) 來提高帳戶的安全性。如需更多資訊，請參閱《AWS IAM Identity Center 使用者指南》中的多重要素驗證和《IAM 使用者指南》中的 IAM 中的AWS 多重要素驗證。

AWS 帳戶 根使用者

當您建立 時 AWS 帳戶，您會從一個登入身分開始，該身分可完整存取帳戶中的所有 AWS 服務 和資源。此身分稱為 AWS 帳戶 Theroot 使用者，可透過使用您用來建立帳戶的電子郵件地址和密碼登入來存取。強烈建議您不要以根使用者處理日常任務。保護您的根使用者憑證，並將其用來執行只能由根使用者執行的任務。如需這些任務的完整清單，了解需以根使用者登入的任務，請參閱 IAM 使用者指南中的需要根使用者憑證的任務。

聯合身分

最佳實務是， 要求人類使用者，包括需要管理員存取權的使用者，使用臨時 AWS 服務 憑證與身分提供者聯合來存取 。

聯合身分是您企業使用者目錄、Web 身分提供者、 AWS Directory Service、Identity Center 目錄，或 AWS 服務 是透過身分來源提供的登入資料存取的任何使用者。當聯合身分存取時 AWS 帳戶，它們會擔任 角色，而角色會提供臨時登入資料。

對於集中式存取權管理，我們建議您使用 AWS IAM Identity Center。您可以在 IAM Identity Center 中建立使用者和群組，也可以連接並同步到您自己的身分來源中的一組使用者 AWS 帳戶 和群組，以便在所有 和應用程式中使用。如需 IAM Identity Center 的詳細資訊，請參閱 AWS IAM Identity Center 使用者指南中的什麼是 IAM Identity Center？。

IAM 使用者和群組

IAM 使用者是 中的身分 AWS 帳戶 ，具有單一人員或應用程式的特定許可。建議您盡可能依賴臨時憑證，而不是擁有建立長期憑證 (例如密碼和存取金鑰) 的 IAM 使用者。但是如果特定使用案例需要擁有長期憑證的 IAM 使用者，建議您輪換存取金鑰。如需更多資訊，請參閱 IAM 使用者指南中的為需要長期憑證的使用案例定期輪換存取金鑰。

IAM 群組是一種指定 IAM 使用者集合的身分。您無法以群組身分簽署。您可以使用群組來一次為多名使用者指定許可。群組可讓管理大量使用者許可的程序變得更為容易。例如，您可以擁有一個名為 IAMAdmins 的群組，並給予該群組管理 IAM 資源的許可。

使用者與角色不同。使用者只會與單一人員或應用程式建立關聯，但角色的目的是在由任何需要它的人員取得。使用者擁有永久的長期憑證，但角色僅提供臨時憑證。如需更多資訊，請參閱《IAM 使用者指南》中的 IAM 使用者的使用案例。

IAM 角色

IAM 角色是 中具有特定許可 AWS 帳戶 的身分。它類似 IAM 使用者，但不與特定的人員相關聯。若要暫時在 中擔任 IAM 角色 AWS Management Console，您可以從使用者切換至 IAM 角色（主控台）。您可以透過呼叫 AWS CLI 或 AWS API 操作或使用自訂 URL 來擔任角色。如需使用角色的方法詳細資訊，請參閱《IAM 使用者指南》中的擔任角色的方法。

使用臨時憑證的 IAM 角色在下列情況中非常有用：

- 聯合身分使用者存取 — 如需向聯合身分指派許可，請建立角色，並為角色定義許可。當聯合身分進行身分驗證時，該身分會與角色建立關聯，並獲授予由角色定義的許可。如需有關聯合角色的相關資訊，請參閱《IAM 使用者指南》中的為第三方身分提供者 (聯合) 建立角色。如果您使用 IAM Identity Center，則需要設定許可集。為控制身分驗證後可以存取的內容，IAM Identity Center 將許可集與 IAM 中的角色相關聯。如需有關許可集的資訊，請參閱 AWS IAM Identity Center 使用者指南中的許可集。
- 暫時 IAM 使用者許可 – IAM 使用者或角色可以擔任 IAM 角色來暫時針對特定任務採用不同的許可。
- 跨帳戶存取權：您可以使用 IAM 角色，允許不同帳戶中的某人 (信任的主體) 存取您帳戶的資源。角色是授予跨帳戶存取權的主要方式。不過，對於某些 AWS 服務，您可以將政策直接連接到資源（而不是使用角色做為代理）。如需了解使用角色和資源型政策進行跨帳戶存取之間的差異，請參閱《IAM 使用者指南》中的 IAM 中的跨帳戶資源存取。
- 跨服務存取 – 有些 AWS 服務 使用其他 中的功能 AWS 服務。例如，當您在服務中進行呼叫時，該服務通常會在 Amazon EC2 中執行應用程式或將物件儲存在 Amazon Simple Storage Service (Amazon S3) 中。服務可能會使用呼叫主體的許可、使用服務角色或使用服務連結角色來執行此作業。
  - 轉送存取工作階段 (FAS) – 當您使用 IAM 使用者或角色在其中執行動作時 AWS，您會被視為委託人。使用某些服務時，您可能會執行某個動作，進而在不同服務中啟動另一個動作。FAS 使用呼叫 的委託人許可 AWS 服務，結合 AWS 服務 請求向下游服務提出請求。只有當服務收到需要與其他 AWS 服務 或 資源互動才能完成的請求時，才會提出 FAS 請求。在此情況下，您必須具有執行這兩個動作的許可。如需提出 FAS 請求時的政策詳細資訊，請參閱《轉發存取工作階段》。

- 服務角色 – 服務角色是服務擔任的 IAM 角色，可代表您執行動作。IAM 管理員可以從 IAM 內建立、修改和刪除服務角色。如需詳細資訊，請參閱《IAM 使用者指南》中的建立角色以委派許可權給 AWS 服務。
- 服務連結角色 – 服務連結角色是一種連結至 的服務角色。 AWS 服務服務可以擔任代表您執行動作的角色。服務連結角色會出現在您的 中 AWS 帳戶 ，並由服務擁有。IAM 管理員可以檢視，但不能編輯服務連結角色的許可。

- 在 Amazon EC2 上執行的應用程式 – 您可以使用 IAM 角色來管理在 EC2 執行個體上執行之應用程式的臨時登入資料，以及提出 AWS CLI 或 AWS API 請求。這是在 EC2 執行個體內儲存存取金鑰的較好方式。若要將 AWS 角色指派給 EC2 執行個體並將其提供給其所有應用程式，您可以建立連接至執行個體的執行個體描述檔。執行個體設定檔包含該角色，並且可讓 EC2 執行個體上執行的程式取得臨時憑證。如需詳細資訊，請參閱《IAM 使用者指南》中的使用 IAM 角色來授予許可權給 Amazon EC2 執行個體上執行的應用程式。

使用政策管理存取權

您可以透過 AWS 建立政策並將其連接到身分或資源來控制 AWS 中的存取。政策是 中的物件， AWS 當與身分或資源相關聯時， 會定義其許可。當委託人（使用者、根使用者或角色工作階段）發出請求時， 會 AWS 評估這些政策。政策中的許可決定是否允許或拒絕請求。大多數政策會以 JSON 文件 AWS 的形式存放在 中。如需 JSON 政策文件結構和內容的詳細資訊，請參閱 IAM 使用者指南中的 JSON 政策概觀。

管理員可以使用 AWS JSON 政策來指定誰可以存取內容。也就是說，哪個主體在什麼條件下可以對什麼資源執行哪些動作。

預設情況下，使用者和角色沒有許可。若要授予使用者對其所需資源執行動作的許可，IAM 管理員可以建立 IAM 政策。然後，管理員可以將 IAM 政策新增至角色，使用者便能擔任這些角色。

IAM 政策定義該動作的許可，無論您使用何種方法來執行操作。例如，假設您有一個允許 `iam:GetRole` 動作的政策。具有該政策的使用者可以從 AWS Management Console AWS CLI、 或 API AWS 取得角色資訊。

身分型政策

身分型政策是可以附加到身分 (例如 IAM 使用者、使用者群組或角色) 的 JSON 許可政策文件。這些政策可控制身分在何種條件下能對哪些資源執行哪些動作。如需了解如何建立身分型政策，請參閱《IAM 使用者指南》中的透過客戶管理政策定義自訂 IAM 許可。

身分型政策可進一步分類成內嵌政策或受管政策。內嵌政策會直接內嵌到單一使用者、群組或角色。受管政策是獨立的政策，您可以連接到 中的多個使用者、群組和角色 AWS 帳戶。受管政策包括 AWS 受管政策和客戶受管政策。如需了解如何在受管政策及內嵌政策之間選擇，請參閱《IAM 使用者指南》中的在受管政策和內嵌政策間選擇。

資源型政策

資源型政策是連接到資源的 JSON 政策文件。資源型政策的最常見範例是 IAM 角色信任政策和 Amazon S3 儲存貯體政策。在支援資源型政策的服務中，服務管理員可以使用它們來控制對特定資源的存取權限。對於附加政策的資源，政策會定義指定的主體可以對該資源執行的動作以及在何種條件下執行的動作。您必須在資源型政策中指定主體。委託人可以包含帳戶、使用者、角色、聯合身分使用者或 AWS 服務。

資源型政策是位於該服務中的內嵌政策。您無法在資源型政策中使用來自 IAM 的 AWS 受管政策。

存取控制清單 (ACL)

存取控制清單 (ACL) 可控制哪些主體 (帳戶成員、使用者或角色) 擁有存取某資源的許可。ACL 類似於資源型政策，但它們不使用 JSON 政策文件格式。

Amazon S3 AWS WAF和 Amazon VPC 是支援 ACLs的服務範例。如需進一步了解 ACL，請參閱 Amazon Simple Storage Service 開發人員指南中的存取控制清單 (ACL) 概觀。

其他政策類型

AWS 支援其他較不常見的政策類型。這些政策類型可設定較常見政策類型授予您的最大許可。

- 許可界限 – 許可範圍是一種進階功能，可供您設定身分型政策能授予 IAM 實體 (IAM 使用者或角色) 的最大許可。您可以為實體設定許可界限。所產生的許可會是實體的身分型政策和其許可界限的交集。會在 `Principal` 欄位中指定使用者或角色的資源型政策則不會受到許可界限限制。所有這類政策中的明確拒絕都會覆寫該允許。如需許可界限的詳細資訊，請參閱 IAM 使用者指南中的 IAM 實體許可界限。
- 服務控制政策 SCPs) – SCPs是 JSON 政策，可指定 中組織或組織單位 (OU) 的最大許可 AWS Organizations。 AWS Organizations 是一種用於分組和集中管理您企業擁有 AWS 帳戶 的多個 的服務。若您啟用組織中的所有功能，您可以將服務控制政策 (SCP) 套用到任何或所有帳戶。SCP 會限制成員帳戶中實體的許可，包括每個實體 AWS 帳戶根使用者。如需 Organizations 和 SCP 的詳細資訊，請參閱《AWS Organizations 使用者指南》中的服務控制政策。
- 資源控制政策 (RCP) - RCP 是 JSON 政策，可用來設定您帳戶中資源的可用許可上限，採取這種方式就不需要更新附加至您所擁有的每個資源的 IAM 政策。RCP 會限制成員帳戶中資源的許

可，並可能影響身分的有效許可，包括 AWS 帳戶根使用者，無論它們是否屬於您的組織。如需 Organizations 和 RCPs的詳細資訊，包括 AWS 服務 支援 RCPs 清單，請參閱AWS Organizations 《 使用者指南》中的資源控制政策 (RCPs)。

- 工作階段政策 – 工作階段政策是一種進階政策，您可以在透過撰寫程式的方式建立角色或聯合使用者的暫時工作階段時，做為參數傳遞。所產生工作階段的許可會是使用者或角色的身分型政策和工作階段政策的交集。許可也可以來自資源型政策。所有這類政策中的明確拒絕都會覆寫該允許。如需詳細資訊，請參閱 IAM 使用者指南中的工作階段政策。

多種政策類型

將多種政策類型套用到請求時，其結果形成的許可會更為複雜、更加難以理解。若要了解如何 AWS 在涉及多種政策類型時決定是否允許請求，請參閱《IAM 使用者指南》中的政策評估邏輯。

AWS HealthOmics 如何使用 IAM

在您使用 IAM 管理對 AWS HealthOmics 的存取之前，請先了解哪些 IAM 功能可與 AWS HealthOmics 搭配使用。

您可以搭配 使用的 IAM 功能 AWS HealthOmics

IAM 功能	HealthOmics 支援
身分型政策	是
資源型政策	否
政策動作	是
政策資源	是
政策條件索引鍵	否
ACL	否
ABAC （政策中的標籤）	是
臨時憑證	是
主體許可	是

IAM 功能	HealthOmics 支援
服務角色	是
服務連結角色	否

若要全面了解 HealthOmics 和其他 AWS 服務如何與大多數 IAM 功能搭配使用，請參閱《AWS IAM 使用者指南》中的與 IAM 搭配使用的 服務。

預防跨服務混淆代理人

混淆代理人問題屬於安全性議題，其中沒有執行動作許可的實體可以強制具有更多許可的實體執行該動作。在 中 AWS，跨服務模擬可能會導致混淆代理人問題。在某個服務 (呼叫服務) 呼叫另一個服務 (被呼叫服務) 時，可能會發生跨服務模擬。可以操縱呼叫服務來使用其許可，以其不應有存取許可的方式對其他客戶的資源採取動作。為了預防這種情況， AWS 提供的工具可協助您保護所有服務的資料，而這些服務主體已獲得您帳戶中資源的存取權。

我們建議在資源政策中使用 `aws:SourceArn`和 `aws:SourceAccount`全域條件內容金鑰，以限制 AWS HealthOmics 為資源提供其他服務的許可。

若要避免 HealthOmics 擔任之角色中的混淆代理人問題，請在角色的信任政策`aws:SourceArn`中將 `arn:aws:omics:region:accountNumber:*` 的值設為 。萬用字元 (*) 會套用所有 HealthOmics 資源的條件。

下列信任關係政策會授予 HealthOmics 存取 資源的權限，並使用 `aws:SourceArn`和 `aws:SourceAccount`全域條件內容金鑰來防止混淆代理人問題。當您為 HealthOmics 建立角色時，請使用此政策。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": [
```

```
            "omics.amazonaws.com"
          ]
        },
        "Action": "sts:AssumeRole",
        "Condition": {
          "StringEquals": {
            "aws:SourceAccount": "accountNumber"
          },
          "ArnLike": {
            "aws:SourceArn": "arn:aws:omics:region:accountNumber:*"
          }
        }
      }
    ]
}
```

HealthOmics 的身分型政策

支援身分型政策：是

身分型政策是可以附加到身分 (例如 IAM 使用者、使用者群組或角色) 的 JSON 許可政策文件。這些政策可控制身分在何種條件下能對哪些資源執行哪些動作。如需了解如何建立身分型政策，請參閱《IAM 使用者指南》中的透過客戶管理政策定義自訂 IAM 許可。

使用 IAM 身分型政策，您可以指定允許或拒絕的動作和資源，以及在何種條件下允許或拒絕動作。您無法在身分型政策中指定主體，因為這會套用至連接的使用者或角色。如要了解您在 JSON 政策中使用的所有元素，請參閱《IAM 使用者指南》中的 IAM JSON 政策元素參考。

HealthOmics 的身分型政策範例

若要檢視 AWS HealthOmics 身分型政策的範例，請參閱 的身分型政策範例 AWS HealthOmics。

HealthOmics 中的資源型政策

支援資源型政策：否

資源型政策是附加到資源的 JSON 政策文件。資源型政策的最常見範例是 IAM 角色信任政策和 Amazon S3 儲存貯體政策。在支援資源型政策的服務中，服務管理員可以使用它們來控制對特定資源的存取權限。對於附加政策的資源，政策會定義指定的主體可以對該資源執行的動作以及在何種條件下執行的動作。您必須在資源型政策中指定主體。委託人可以包含帳戶、使用者、角色、聯合身分使用者或 AWS 服務。

如需啟用跨帳戶存取權，您可以指定在其他帳戶內的所有帳戶或 IAM 實體，做為資源型政策的主體。新增跨帳戶主體至資源型政策，只是建立信任關係的一半。當主體和資源位於不同位置時 AWS 帳戶，信任帳戶中的 IAM 管理員也必須授予主體實體（使用者或角色）存取資源的許可。其透過將身分型政策連接到實體來授與許可。不過，如果資源型政策會為相同帳戶中的主體授予存取，這時就不需要額外的身分型政策。如需詳細資訊，請參閱《IAM 使用者指南》中的 IAM 中的快帳戶資源存取。

HealthOmics 的政策動作

支援政策動作：是

管理員可以使用 AWS JSON 政策來指定誰可以存取內容。也就是說，哪個主體在什麼條件下可以對什麼資源執行哪些動作。

JSON 政策的 `Action` 元素描述您可以用來允許或拒絕政策中存取的動作。政策動作通常具有與相關聯 AWS API 操作相同的名稱。有一些例外狀況，例如沒有相符的 API 操作的僅限許可動作。也有一些作業需要政策中的多個動作。這些額外的動作稱為相依動作。

政策會使用動作來授予執行相關聯動作的許可。

若要查看 HealthOmics 動作的清單，請參閱《服務授權參考》中的 AWS HealthOmics 定義的動作。

HealthOmics 中的政策動作在動作之前使用下列字首：

```
omics
```

若要在單一陳述式中指定多個動作，請用逗號分隔。

```
"Action": [
      "omics:action1",
      "omics:action2"
         ]
```

若要檢視 AWS HealthOmics 身分型政策的範例，請參閱 的身分型政策範例 AWS HealthOmics。

HealthOmics 的政策資源

支援政策資源：是

管理員可以使用 AWS JSON 政策來指定誰可以存取內容。也就是說，哪個主體在什麼條件下可以對什麼資源執行哪些動作。

Resource JSON 政策元素可指定要套用動作的物件。陳述式必須包含 Resource 或 NotResource 元素。最佳實務是使用其 Amazon Resource Name (ARN) 來指定資源。您可以針對支援特定資源類型的動作 (稱為資源層級許可) 來這麼做。

對於不支援資源層級許可的動作 (例如列出操作)，請使用萬用字元 (*) 來表示陳述式適用於所有資源。

```
"Resource": "*"
```

若要查看 HealthOmics 資源類型及其 ARNs的清單，請參閱服務授權參考中的 AWS HealthOmics 定義的資源。若要了解您可以使用哪些動作指定每個資源的 ARN，請參閱 AWS HealthOmics 定義的動作。

若要檢視 AWS HealthOmics 身分型政策的範例，請參閱 的身分型政策範例 AWS HealthOmics。

HealthOmics 的政策條件索引鍵

HealthOmics 不支援政策條件索引鍵。

HealthOmics 中的存取控制清單 (ACLs)

支援 ACL：否

存取控制清單 (ACL) 可控制哪些主體 (帳戶成員、使用者或角色) 擁有存取某資源的許可。ACL 類似於資源型政策，但它們不使用 JSON 政策文件格式。

使用 HealthOmics 的屬性型存取控制 (ABAC)

支援 ABAC (政策中的標籤)：是

屬性型存取控制 (ABAC) 是一種授權策略，可根據屬性來定義許可。在 中 AWS，這些屬性稱為標籤。您可以將標籤連接到 IAM 實體 （使用者或角色） 和許多 AWS 資源。為實體和資源加上標籤是 ABAC 的第一步。您接著要設計 ABAC 政策，允許在主體的標籤與其嘗試存取的資源標籤相符時操作。

ABAC 在成長快速的環境中相當有幫助，並能在政策管理變得繁瑣時提供協助。

如需根據標籤控制存取，請使用 aws:ResourceTag/*key-name*、aws:RequestTag/*key-name* 或 aws:TagKeys 條件索引鍵，在政策的條件元素中，提供標籤資訊。

如果服務支援每個資源類型的全部三個條件金鑰，則對該服務而言，值為 Yes。如果服務僅支援某些資源類型的全部三個條件金鑰，則值為 Partial。

如需 ABAC 的詳細資訊，請參閱《IAM 使用者指南》中的使用 ABAC 授權定義許可。如要查看含有設定 ABAC 步驟的教學課程，請參閱 IAM 使用者指南中的使用屬性型存取控制 (ABAC)。

如需標記 HealthOmics 資源的詳細資訊，請參閱 在 HealthOmics 中標記資源。

下列範例示範如何撰寫 IAM 政策，拒絕在沒有特定標籤的情況下存取資源。

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Deny",
            "Action": [
                "omics:*"
            ],
            "Resource": [
                "*"
            ],
            "Condition": {
                "Null": {
                  "aws:RequestTag/MyCustomTag": "true"
                }
            }
        }
    ]
}
```

搭配 HealthOmics 使用臨時登入資料

支援臨時憑證：是

當您使用臨時登入資料登入時，有些 AWS 服務 無法運作。如需詳細資訊，包括哪些 AWS 服務 使用臨時登入資料，請參閱《AWS 服務 IAM 使用者指南》中的使用 IAM 的 。

如果您 AWS Management Console 使用使用者名稱和密碼以外的任何方法登入 ，則會使用暫時登入資料。例如，當您 AWS 使用公司的單一登入 (SSO) 連結存取 時，該程序會自動建立臨時登入資料。當您以使用者身分登入主控台，然後切換角色時，也會自動建立臨時憑證。如需切換角色的詳細資訊，請參閱《IAM 使用者指南》中的從使用者切換至 IAM 角色 (主控台)。

您可以使用 AWS CLI 或 AWS API 手動建立臨時登入資料。然後，您可以使用這些臨時登入資料來存取 AWS。 AWS 建議您動態產生臨時登入資料，而不是使用長期存取金鑰。如需詳細資訊，請參閱 IAM 中的暫時性安全憑證。

HealthOmics 的跨服務主體許可

支援轉寄存取工作階段 (FAS)：是

當您使用 IAM 使用者或角色在 中執行動作時 AWS，您會被視為委託人。使用某些服務時，您可能會執行某個動作，進而在不同服務中啟動另一個動作。FAS 使用呼叫 的委託人許可 AWS 服務，結合 AWS 服務 請求向下游服務提出請求。只有在服務收到需要與其他 AWS 服務 或 資源互動才能完成的請求時，才會提出 FAS 請求。在此情況下，您必須具有執行這兩個動作的許可。如需提出 FAS 請求時的政策詳細資訊，請參閱轉發存取工作階段。

HealthOmics 的服務角色

支援服務角色：是

服務角色是服務擔任的 IAM 角色，可代您執行動作。IAM 管理員可以從 IAM 內建立、修改和刪除服務角色。如需詳細資訊，請參閱《IAM 使用者指南》中的建立角色以委派許可權給 AWS 服務。

> **Warning**
>
> 變更服務角色的許可可能會中斷 HealthOmics 功能。只有在 HealthOmics 提供指引時，才能編輯服務角色。

HealthOmics 的服務連結角色

支援服務連結角色：否

服務連結角色是連結至 的一種服務角色 AWS 服務。服務可以擔任代表您執行動作的角色。服務連結角色會出現在您的 中 AWS 帳戶 ，並由服務擁有。IAM 管理員可以檢視，但不能編輯服務連結角色的許可。

如需建立或管理服務連結角色的詳細資訊，請參閱可搭配 IAM 運作的AWS 服務。在表格中尋找服務，其中包含服務連結角色欄中的 Yes。選擇是連結，以檢視該服務的服務連結角色文件。

的身分型政策範例 AWS HealthOmics

根據預設，使用者和角色沒有建立或修改 AWS HealthOmics 資源的許可。他們也無法使用 AWS Management Console、 AWS Command Line Interface (AWS CLI) 或 AWS API 來執行任務。若要授予使用者對其所需資源執行動作的許可，IAM 管理員可以建立 IAM 政策。然後，管理員可以將 IAM 政策新增至角色，使用者便能擔任這些角色。

如需了解如何使用這些範例 JSON 政策文件建立 IAM 身分型政策，請參閱《IAM 使用者指南》中的建立 IAM 政策 (主控台)。

如需 AWS HealthOmics 定義的動作和資源類型的詳細資訊，包括每種資源類型的 ARNs 格式，請參閱《服務授權參考》中的 AWS HealthOmics 的動作、資源和條件金鑰。

主題

- 政策最佳實務
- 使用 HealthOmics 主控台
- 允許使用者檢視他們自己的許可

政策最佳實務

身分型政策會判斷您帳戶中的某個人員是否可以建立、存取或刪除 AWS HealthOmics 資源。這些動作可能會讓您的 AWS 帳戶產生費用。當您建立或編輯身分型政策時，請遵循下列準則及建議事項：

- 開始使用 AWS 受管政策並轉向最低權限許可 – 若要開始將許可授予您的使用者和工作負載，請使用將許可授予許多常見使用案例的 AWS 受管政策。它們可在您的 中使用 AWS 帳戶。我們建議您定義特定於使用案例 AWS 的客戶受管政策，以進一步減少許可。如需更多資訊，請參閱 IAM 使用者指南中的 AWS 受管政策或任務職能的AWS 受管政策。
- 套用最低權限許可 – 設定 IAM 政策的許可時，請僅授予執行任務所需的許可。為實現此目的，您可以定義在特定條件下可以對特定資源採取的動作，這也稱為最低權限許可。如需使用 IAM 套用許可的更多相關資訊，請參閱 IAM 使用者指南中的 IAM 中的政策和許可。
- 使用 IAM 政策中的條件進一步限制存取權 – 您可以將條件新增至政策，以限制動作和資源的存取。例如，您可以撰寫政策條件，指定必須使用 SSL 傳送所有請求。如果透過特定 等使用服務動作 AWS 服務，您也可以使用條件來授予其存取權 AWS CloudFormation。如需詳細資訊，請參閱 IAM 使用者指南中的 IAM JSON 政策元素：條件。

- 使用 IAM Access Analyzer 驗證 IAM 政策，確保許可安全且可正常運作 – IAM Access Analyzer 驗證新政策和現有政策，確保這些政策遵從 IAM 政策語言 (JSON) 和 IAM 最佳實務。IAM Access Analyzer 提供 100 多項政策檢查及切實可行的建議，可協助您撰寫安全且實用的政策。如需詳細資訊，請參閱《IAM 使用者指南》中的使用 IAM Access Analyzer 驗證政策。
- 需要多重要素驗證 (MFA) – 如果您的案例需要 IAM 使用者或 中的根使用者 AWS 帳戶，請開啟 MFA 以提高安全性。如需在呼叫 API 操作時請求 MFA，請將 MFA 條件新增至您的政策。如需詳細資訊，請參閱《IAM 使用者指南》https://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_mfa_configure-api-require.html中的透過 MFA 的安全 API 存取。

如需 IAM 中最佳實務的相關資訊，請參閱 IAM 使用者指南中的 IAM 安全最佳實務。

使用 HealthOmics 主控台

若要存取 AWS HealthOmics 主控台，您必須擁有一組最低許可。這些許可必須允許您列出和檢視 中 AWS HealthOmics 資源的詳細資訊 AWS 帳戶。如果您建立比最基本必要許可更嚴格的身分型政策，則對於具有該政策的實體 (使用者或角色) 而言，主控台就無法如預期運作。

對於僅呼叫 AWS CLI 或 AWS API 的使用者，您不需要允許最低主控台許可。反之，只需允許存取符合他們嘗試執行之 API 操作的動作就可以了。

允許使用者檢視他們自己的許可

此範例會示範如何建立政策，允許 IAM 使用者檢視附加到他們使用者身分的內嵌及受管政策。此政策包含在主控台或使用 或 AWS CLI AWS API 以程式設計方式完成此動作的許可。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ViewOwnUserInfo",
            "Effect": "Allow",
            "Action": [
                "iam:GetUserPolicy",
                "iam:ListGroupsForUser",
                "iam:ListAttachedUserPolicies",
                "iam:ListUserPolicies",
                "iam:GetUser"
            ],
            "Resource": ["arn:aws:iam::*:user/${aws:username}"]
        },
```

```
        {
            "Sid": "NavigateInConsole",
            "Effect": "Allow",
            "Action": [
                "iam:GetGroupPolicy",
                "iam:GetPolicyVersion",
                "iam:GetPolicy",
                "iam:ListAttachedGroupPolicies",
                "iam:ListGroupPolicies",
                "iam:ListPolicyVersions",
                "iam:ListPolicies",
                "iam:ListUsers"
            ],
            "Resource": "*"
        }
    ]
}
```

AWS 的 受管政策 AWS HealthOmics

AWS 受管政策是由 AWS AWS 受管政策建立和管理的獨立政策旨在為許多常用案例提供許可，以便您可以開始將許可指派給使用者、群組和角色。

請記住， AWS 受管政策可能不會授予特定使用案例的最低權限許可，因為這些許可可供所有 AWS 客戶使用。我們建議您定義使用案例專屬的客戶管理政策，以便進一步減少許可。

您無法變更 AWS 受管政策中定義的許可。如果 AWS 更新受 AWS 管政策中定義的許可，則更新會影響政策連接的所有主體身分（使用者、群組和角色）。 AWS 服務 當新的 啟動或新的 API 操作可供現有服務使用時， AWS 最有可能更新 AWS 受管政策。

如需詳細資訊，請參閱《IAM 使用者指南》中的 AWS 受管政策。

AWS 受管政策：AmazonOmicsFullAccess

您可以將`AmazonOmicsFullAccess`政策連接至 IAM 身分，讓其完整存取 HealthOmics。

此政策會授予所有 HealthOmics 動作的完整存取許可。當您建立註釋或變體存放區時，Omics 也會讓您透過 Resource Access Manager (RAM) 主控台中的資源共享邀請存取該存放區。如需透過 Lake Formation 進行資源共用邀請的詳細資訊，請參閱 Lake Formation 中的跨帳戶資料共用。對於 Omics 管理員政策，您也需要下列許可才能存取 Amazon S3 儲存貯體。

- PutObject
- GetObject
- ListBucket
- AbortMultipartUpload
- ListMultipartUploadParts

JSON

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Effect": "Allow",
   "Action": [
    "omics:*"
   ],
   "Resource": "*"
  },
  {
   "Effect": "Allow",
   "Action": [
    "ram:AcceptResourceShareInvitation",
    "ram:GetResourceShareInvitations"
   ],
   "Resource": "*",
   "Condition": {
    "StringEquals": {
     "aws:CalledViaLast": "omics.amazonaws.com"
    }
   }
  },
  {
   "Effect": "Allow",
   "Action": "iam:PassRole",
```

```
   "Resource": "*",
   "Condition": {
    "StringEquals": {
     "iam:PassedToService": "omics.amazonaws.com"
    }
   }
  }
 ]
}
```

AWS 受管政策：AmazonOmicsReadOnlyAccess

當您想要將該身分的許可限制為唯讀存取時，您可以將AWSOmicsReadOnlyAccess政策連接到 IAM 身分。

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "omics:Get*",
                "omics:List*"
            ],
            "Resource": "*"
        }
    ]
}
```

HealthOmics AWS 受管政策的更新

檢視自此服務開始追蹤這些變更以來 HealthOmics AWS 受管政策更新的詳細資訊。如需此頁面變更的自動提醒，請訂閱 HealthOmics 文件歷史記錄頁面上的 RSS 摘要。

變更	描述	日期
AmazonOmicsFullAccess - 新增政策	HealthOmics 新增了新的政策，以授予使用者對所有動作和資源的完整存取權。若要進一步了解，請參閱 AmazonOmicsFullAccess。	2023 年 2 月 23 日
HealthOmics 已開始追蹤變更	HealthOmics 開始追蹤其 AWS 受管政策的變更。	2022 年 11 月 29 日
AmazonOmicsReadOnlyAccess - 新增政策	HealthOmics 新增了限制唯讀存取的新政策。若要進一步了解，AmazonOmicsReadOnlyAccess。	2022 年 11 月 29 日

對 AWS HealthOmics 身分和存取進行故障診斷

使用以下資訊來協助您診斷和修正使用 AWS HealthOmics 和 IAM 時可能遇到的常見問題。

主題

- 我無權在 HealthOmics 中執行動作
- 我未獲得執行 iam:PassRole 的授權
- 我想要允許 以外的人員 AWS 帳戶 存取我的 HealthOmics 資源

我無權在 HealthOmics 中執行動作

如果您收到錯誤，告知您未獲授權執行動作，您的政策必須更新，允許您執行動作。

下列範例錯誤會在mateojackson IAM 使用者嘗試使用主控台檢視一個虛構 *my-example-widget* 資源的詳細資訊，但卻無虛構 omics:*GetWidget* 許可時發生。

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
 omics:GetWidget on resource: my-example-widget
```

在此情況下，必須更新 mateojackson 使用者的政策，允許使用 omics:*GetWidget* 動作存取 *my-example-widget* 資源。

如果您需要協助，請聯絡您的 AWS 管理員。您的管理員提供您的簽署憑證。

我未獲得執行 iam:PassRole 的授權

如果您收到錯誤，告知您無權執行 iam:PassRole動作，您的政策必須更新，以允許您將角色傳遞給 AWS HealthOmics。

有些 AWS 服務 可讓您將現有角色傳遞給該服務，而不是建立新的服務角色或服務連結角色。如需執行此作業，您必須擁有將角色傳遞至該服務的許可。

當名為 的 IAM marymajor 使用者嘗試使用主控台在 AWS HealthOmics 中執行動作時，會發生下列範例錯誤。但是，動作請求服務具備服務角色授予的許可。Mary 沒有將角色傳遞至該服務的許可。

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
 iam:PassRole
```

在這種情況下，Mary 的政策必須更新，允許她執行 iam:PassRole 動作。

如果您需要協助，請聯絡您的 AWS 管理員。您的管理員提供您的簽署憑證。

我想要允許 以外的人員 AWS 帳戶 存取我的 HealthOmics 資源

您可以建立一個角色，讓其他帳戶中的使用者或您組織外部的人員存取您的資源。您可以指定要允許哪些信任物件取得該角色。針對支援基於資源的政策或存取控制清單 (ACL) 的服務，您可以使用那些政策來授予人員存取您的資源的許可。

如需進一步了解，請參閱以下內容：

- 若要了解 AWS HealthOmics 是否支援這些功能，請參閱 AWS HealthOmics 如何使用 IAM。
- 若要了解如何在您擁有 AWS 帳戶 的 資源之間提供存取權，請參閱《IAM 使用者指南》中的在您擁有 AWS 帳戶 的另一個 IAM 使用者中提供存取權。
- 若要了解如何將資源的存取權提供給第三方 AWS 帳戶，請參閱《IAM 使用者指南》中的將存取權提供給第三方 AWS 帳戶 擁有。
- 如需了解如何透過聯合身分提供存取權，請參閱 IAM 使用者指南中的將存取權提供給在外部進行身分驗證的使用者 (聯合身分)。

- 如需了解使用角色和資源型政策進行跨帳戶存取之間的差異，請參閱《IAM 使用者指南》中的 IAM 中的跨帳戶資源存取。

的合規驗證 AWS HealthOmics

在多個合規計畫 AWS HealthOmics 中，第三方稽核人員會評估 的安全與 AWS 合規。這包括 HIPAA、FedRAMP 和其他。下表顯示 HealthOmics 服務的合規認證。

認證	連結
HIPAA	HIPAA 合格服務參考
HiTrust-CSF	健康資訊信任聯盟通用安全架構
FedRAMP Moderate (East/West)	聯邦風險與授權管理計劃
ISO/CSA STAR	ISO 和 CSA STAR 認證
C5	雲端運算合規控制目錄
DoD CC SRG IL2	國防部雲端運算安全需求指南
ENS 高級	Esquema Nacional de Seguridad
FINMA	瑞士金融市場監管局
ISMAP	資訊系統安全管理和評估計劃
OSPAR	委外服務供應商的稽核報告
PCI	支付卡產業資料安全標準
拼法	銀行關聯 CCI - 第三方資格
PiTuKri	評估雲端服務資訊安全的條件
SOC 1、2、3	系統和組織控制

如需特定合規計劃範圍內所有 AWS 服務的清單，請參閱合規計劃範圍內的 AWS 服務。如需一般資訊，請參閱 AWS 合規計劃 。

您可以使用 下載第三方稽核報告 AWS Artifact。如需詳細資訊，請參閱在 中下載報告 AWS Artifact。

HealthOmics 資料存放區使用範例 ID 進行內部檔案命名和標記資源。在您擷取資料之前，請檢查範例 ID 是否包含任何 PHI 資料。如果是這樣，請在擷取資料之前變更範例 ID。如需詳細資訊，請參閱 AWS HIPAA 合規網頁上的指引。

您使用 時的合規責任 AWS HealthOmics 取決於資料的機密性、您公司的合規目標，以及適用的法律和法規。 AWS 提供下列資源來協助合規：

- 安全與合規快速入門指南：這些部署指南討論架構考量，並提供在 AWS上部署以安全及合規為重心之基準環境的步驟。
- HIPAA 安全與合規架構白皮書 – 此白皮書說明公司如何使用 AWS 來建立符合 HIPAA 規範的應用程式。
- AWS 合規資源 – 此工作手冊和指南集合可能適用於您的產業和據點。
- 《 AWS Config 開發人員指南》中的使用規則評估資源 AWS Config：評估資源組態符合內部實務、產業準則和法規的程度。
- AWS Security Hub – AWS 此服務提供 內安全狀態的完整檢視 AWS ，協助您檢查是否符合安全產業標準和最佳實務。

HealthOmics 中的彈性

AWS 全球基礎設施是以 AWS 區域 和 可用區域為基礎建置。 AWS 區域 提供多個實體隔離和隔離的可用區域，這些可用區域與低延遲、高輸送量和高度備援聯網連接。透過可用區域，您可以設計與操作的應用程式和資料庫，在可用區域之間自動容錯移轉而不會發生中斷。可用區域的可用性、容錯能力和擴展能力，均較單一或多個資料中心的傳統基礎設施還高。

如需 AWS 區域 和可用區域的詳細資訊，請參閱 AWS 全球基礎設施。

除了 AWS 全球基礎設施之外，AWS HealthOmics 還提供數種功能，以協助支援您的資料彈性和備份需求。

AWS HealthOmics 和介面 VPC 端點 (AWS PrivateLink)

您可以在 VPC 和 之間建立私有連線， AWS HealthOmics 方法是建立介面 VPC 端點。介面端點採用 AWS PrivateLink技術，可讓您在沒有網際網路閘道、NAT 裝置、VPN 連線或 AWS Direct Connect 連線的情況下，私下存取 HealthOmics API 操作。VPC 中的執行個體不需要公有 IP 地址，即可與 HealthOmics API 操作通訊。VPC 和 HealthOmics 之間的流量不會超出 Amazon 網路。

每個介面端點都是由您子網路中的一或多個彈性網路介面表示。

如需詳細資訊，請參閱《Amazon VPC 使用者指南》中的界面 VPC 端點 (AWS PrivateLink)。

除了以色列（特拉維夫）之外，所有區域的 HealthOmics 都支援 VPC 端點政策。根據預設，允許透過端點完整存取 HealthOmics。

HealthOmics VPC 端點的考量事項

為 HealthOmics 設定介面 VPC 端點之前，請務必檢閱《Amazon VPC 使用者指南》中的介面端點屬性和限制。

HealthOmics 支援從您的 VPC 呼叫所有 HealthOmics Storage API 動作。

根據預設，HealthOmics 不支援 VPC 端點政策，但您可以為 HealthOmics Storage 操作建立完整 HealthOmics 存取的 VPC 端點。如需詳細資訊，請參閱《Amazon VPC 使用者指南》中的使用 VPC 端點控制對服務的存取。

建立 HealthOmics 的介面 VPC 端點

您可以使用 Amazon VPC 主控台或 AWS Command Line Interface () 為 HealthOmics 服務建立 VPC 端點AWS CLI。如需詳細資訊，請參閱《Amazon VPC 使用者指南》中的建立介面端點。

使用下列服務名稱為 HealthOmics 建立 VPC 端點：

- com.amazonaws.*region*.storage-omics
- com.amazonaws.*region*.control-storage-omics
- com.amazonaws.*region*.analytics-omics
- com.amazonaws.*region*.workflows-omics
- com.amazonaws.*region*.tags-omics

美國東部（維吉尼亞北部）和美國西部（奧勒岡）區域支援 AWS PrivateLink FIPS 端點。對於這些區域，您也可以使用下列服務名稱：

- com.amazonaws.*region*.storage-omics-fips
- com.amazonaws.*region*.control-storage-omics-fips
- com.amazonaws.*region*.analytics-omics-fips
- com.amazonaws.*region*.workflows-omics-fips

- com.amazonaws.*region*.tags-omics-fips

如果您開啟端點的私有 DNS，您可以使用區域的預設 DNS 名稱向 HealthOmics 提出 API 請求，例如 `omics.us-east-1.amazonaws.com`。

如需詳細資訊，請參閱《Amazon VPC 使用者指南》中的透過介面端點存取服務。

為 HealthOmics 建立 VPC 端點政策

您可以將端點政策連接至控制 HealthOmics 存取的 VPC 端點。此政策會指定下列資訊：

- 可執行動作的委託人
- 可執行的動作
- 可在其中執行動作的資源

如需詳細資訊，請參閱《Amazon VPC 使用者指南》中的使用 VPC 端點控制對服務的存取。

範例：HealthOmics 動作的 VPC 端點政策。

以下是 HealthOmics 端點政策的範例。連接到端點時，此政策會授予所有資源上所有主體對 HealthOmics 動作的存取權。

API

```
{
   "Statement":[
      {
         "Principal":"*",
         "Effect":"Allow",
         "Action":[
            "omics:List*"
         ],
         "Resource":"*"
      }
   ]
}
```

AWS CLI

```
aws ec2 modify-vpc-endpoint \
```

```
    --vpc-endpoint-id vpce-id \
    --region us-west-2 \
    --policy-document \
    "{\"Statement\":[{\"Principal\":\"*\",\"Effect\":\"Allow\",\"Action\":
[\"omics:List*\"],\"Resource\":\"*\"}]}"
```

使用 Amazon S3 URIs 存取讀取集的特殊考量

若要在使用私有連線時透過 Amazon S3 URIs 存取讀取集，請在序列存放區上設定 PrivateLink 介面端點。在您設定它們之後，端點具有下列格式：

```
com.amazonaws.region.storage-omics
com.amazonaws.region.control-storage-omics
```

若要使用閘道端點，請遵循 Amazon S3 的閘道端點指南來設定閘道端點。HealthOmics 擁有 Amazon S3 儲存貯體，因此您不需要建立或調整儲存貯體政策。閘道端點依賴連接到存取資料的使用者或角色的政策，但您也可以使用更嚴格的政策來設定端點。這些政策可以包含根據 Amazon S3 存取點 ARN 和 Amazon S3 動作的存取限制。

監控 AWS HealthOmics

監控是維護 AWS HealthOmics 和其他 AWS 解決方案可靠性、可用性和效能的重要部分。 AWS 提供下列監控工具來監看 AWS HealthOmics、在發生錯誤時回報，並適時採取自動動作：

- Amazon CloudWatch AWS 會即時監控您的 AWS 資源和您在 上執行的應用程式。您可以收集和追蹤指標、建立自訂儀板表，以及設定警示，在特定指標達到您指定的閾值時通知您或採取動作。例如，您可以讓 CloudWatch 追蹤 CPU 使用量或其他 Amazon EC2 執行個體指標，並在需要時自動啟動新的執行個體。如需詳細資訊，請參閱 Amazon CloudWatch 使用者指南。
- Amazon CloudWatch Logs 可讓您監控、存放和存取來自 Amazon EC2 執行個體、CloudTrail 及其他來源的日誌檔案。CloudWatch Logs 可監控日誌檔案中的資訊，並在達到特定閾值時通知您。您也可以將日誌資料存檔在高耐用性的儲存空間。如需詳細資訊，請參閱 Amazon CloudWatch Logs 使用者指南。
- AWS CloudTrail 擷取您 AWS 帳戶 發出或代表發出的 API 呼叫和相關事件，並傳送日誌檔案至您指定的 Amazon S3 儲存貯體。您可以找出哪些使用者和帳戶呼叫 AWS、發出呼叫的來源 IP 地址，以及呼叫的發生時間。如需詳細資訊，請參閱 AWS CloudTrail 使用者指南。
- Amazon EventBridge 為無伺服器事件匯流排服務，可讓您輕鬆將應用程式與來自各種來源的資料互相連線。EventBridge 可從您自己的應用程式、Software-as-a-Service(SaaS) 應用程式 AWS 和服務提供即時資料串流，並將該資料路由到 Lambda 等目標。這可讓您監控在服務中發生的事件，並建置事件導向的架構。如需詳細資訊，請參閱 Amazon EventBridge 使用者指南。

> **Note**
>
> 如需服務更新，請設定和監控您的個人運作狀態儀表板。如需如何管理儀表板的詳細資訊，請參閱 AWS Health Dashboard 入門。

主題

S3 存取記錄

您可以使用存放區建立的存取日誌，監控 Amazon S3 API 對 HealthOmics 序列存放區的存取。您可以使用 CloudWatch 從 HealthOmics API 操作 監控 S3 存取。CloudWatch 可讓您從自己的帳戶查看 Amazon S3 存取。身為資料擁有者，如果您共用對第三方帳戶的存取權，則無法在 CloudWatch 中使用存取記錄。反之，請使用存放區的 S3 存取日誌。這會記錄所設定 Amazon S3 儲存貯體 中資料的所有 S3 存取。 Amazon S3

使用 `CreateSequenceStore`或 `UpdateSequenceStore` API 操作設定 S3 存取日誌。此外，請確定 HealthOmics 服務主體 (`omics.amazonaws.com`) 具有所設定 S3 字首的`s3:PutObject`許可。

> **Note**
>
> 日誌使用目的地儲存貯體的預設加密組態。如果儲存貯體使用客戶受管金鑰，服務主體必須具有使用金鑰進行寫入的存取權。

若要關閉存取日誌，請使用 並將存取日誌組態`UpdateSequenceStore`設為空白。

使用 CloudWatch 指標監控 HealthOmics

您可以使用 CloudWatch 監控 HealthOmics，這會收集原始資料並將其處理為可讀且近乎即時的指標。這些統計資料會保留 15 個月，以便您存取歷史資訊，並更清楚 Web 應用程式或服務的執行效能。您也可以設定留意特定閾值的警示，當滿足這些閾值時傳送通知或採取動作。如需詳細資訊，請參閱 Amazon CloudWatch 使用者指南。

AWS HealthOmics 服務會在 `AWS/Omics` 命名空間中報告下列指標。

系統會針對下列 AWS HealthOmics APIs 回報 API 呼叫計數指標。只會報告 API 操作維度。

- 參考和參考存放區 APIs — CreateReferenceStore、DeleteReferenceStore、StartReferenceImportJob
- 序列存放區和讀取集 APIs — CreateSequenceStore、DeleteSequenceStore、StartReadSetImportJob、StartReadSetActivationJob、St
- 變體存放區 APIs — CreateVariantStore、DeleteVariantStore、StartVariantImportJob、CancelVariantImportJob
- 註釋存放區 APIs — CreateAnnotationStore、DeleteAnotationStore、StartAnnotationImportJob、CancelAnnotationImportJob

- 工作流程、執行和執行群組 APIs — CreateWorkflow、DeleteWorkflow、StartRun、CancelRun、DeleteRun、CreateRunGroup、DeleteRunGr

檢視 *AWS HealthOmics* 指標

的 CloudWatch AWS HealthOmics 指標可在 CloudWatch 主控台中檢視。

若要檢視指標 (CloudWatch 主控台)

1. 登入 AWS 管理主控台並開啟 CloudWatch 主控台。
2. 選擇指標，選擇所有指標，然後選擇 AWS/用量。
3. 適用於 的 Filter ServiceAWS HealthOmics。
4. 選擇維度、選擇指標名稱，再選擇 Add to graph (新增至圖形)。
5. 選擇日期範圍的值。所選日期範圍的指標計數會顯示在圖形中。

使用 CloudWatch 建立警示

CloudWatch 警示會監看指定期間內的單一指標，並執行一或多個動作：傳送通知至 Amazon Simple Notification Service (Amazon SNS) 主題或 Auto Scaling 政策。動作是根據指標在您指定的數個期間內相對於指定閾值的值。當警示變更狀態時，CloudWatch 也可以傳送 Amazon SNS 訊息給您。

CloudWatch 警示只有在狀態變更且在您指定的期間持續存在時，才會叫用動作。

若要檢視指標 (CloudWatch 主控台)

1. 登入 AWS 管理主控台並開啟 CloudWatch 主控台。
2. 選擇 Alarms (警示)，然後選擇 Create Alarm (建立警示)。
3. 選擇 AWS/Usage，然後使用服務維度選擇 AWS HealthOmics 指標。
4. 對於 Time Range (時間範圍)，選擇要監控的時間範圍，然後選擇 Next (下一步)。
5. 輸入 Name (名稱) 和 Description (描述)。
6. 針對隨時，選擇 >=，然後輸入最大值。
7. 如果您希望 CloudWatch 在達到警示狀態時傳送電子郵件，請在動作區段中，針對每當此警示，選擇狀態為 ALARM。針對傳送通知至，選擇郵寄清單，或選擇新增清單並建立新的郵寄清單。
8. 在 Alarm Preview (警示預覽) 區段中預覽警示。如果警示符合您的要求，選擇 Create Alarm (建立警示)。

使用 CloudWatch Logs 監控 HealthOmics

HealthOmics 會產生各種日誌，協助您了解和疑難排解您的執行。日誌提供兩個位置：CloudWatch 和 Amazon S3。

根據預設， 執行已開啟記錄。您可以選擇性地關閉執行的記錄，方法是在startrun請求`LogLevel = OFF`中設定 。

> **Note**
>
> 如需服務更新，請設定和監控您的個人運作狀態儀表板。如需如何管理儀表板的詳細資訊，請參閱 AWS Health Dashboard 入門。

主題

HealthOmics 工作流程的日誌類型

HealthOmics 為工作流程提供下列類型的日誌：

- 引擎日誌 – 基礎工作流程引擎 (Nextflow、WDL 和 CWL) 會產生執行的引擎日誌。這些日誌可協助您對工作流程定義問題進行疑難排解。
- 執行資訊清單日誌 – 這些日誌提供有關每個執行任務的高階資訊，例如任務狀態、開始時間、停止時間和失敗原因 （ 如果任務失敗 ）。

  執行資訊清單日誌也會報告資源使用率統計資料，有助於了解資源最佳化機會。這些統計資料包括：

  - cpusAverage
  - cpusMaximum
  - cpusReserved
  - gpusReserved
  - memoryAverageGiB

  - memoryMaximumGiB
  - memoryReservedGiB
  - runningSeconds
- 執行日誌 – 執行日誌提供整體執行狀態，以及個別任務開始、執行、停止和完成的時間。執行日誌也可讓您了解檔案匯入和匯出步驟。
- 任務日誌 – 任務日誌提供執行中個別任務的詳細記錄資訊。任務日誌中的輸出取決於任務定義，以及您在程式碼中使用日誌陳述式的位置。如果您的任務日誌未提供您所需的洞見層級，請考慮將其他日誌陳述式新增至您的任務定義，以產生更深入的任務日誌。
- 執行快取日誌 – 執行快取日誌提供執行快取的整體狀態和任務輸出的快取。執行快取日誌可讓您了解使用快取的每個執行的快取命中和未命中。
- Outputs.json – 對於 WDL 和 CWL 工作流程，HealthOmics 會在執行完成後將名為 的引擎產生檔案交付`outputs.json`至您的 Amazon S3 儲存貯體。此檔案包含執行的所有輸出的清單和映射。

CloudWatch 中的日誌

您可以在下列日誌群組中找到 HealthOmics CloudWatch 工作流程日誌：`/aws/omics/WorkflowLog`。此外，get-run API 操作的輸出會提供引擎日誌和執行日誌的 CloudWatch 日誌串流 ARNs。

根據預設， 會無限期 AWS 保留 CloudWatch Logs。您可以調整日誌群組的保留政策，將保留期間設定為 10 年到 1 天。

下表提供 HealthOmics 中的 CloudWatch Logs 摘要。 HealthOmics

日誌名稱	可在 CloudWatch Logs 中使用	何時可使用日誌	日誌串流格式
引擎日誌	是，對於失敗的執行	執行完成後	run/*runID*/engine
執行資訊清單日誌	是	執行完成後	manifest/run/*runID*/*runUUID*
執行日誌	是	即時	run/*runID*
任務日誌	是	即時	run/*runID*/task/*taskID*

日誌名稱	可在 CloudWatch Logs 中使用	何時可使用日誌	日誌串流格式
執行快取日誌	是	即時	runCache/*runCacheId* /*runCacheUUID*
Outputs.json (WDL 和 CWL)	否	N/A	無

Amazon S3 中的日誌

執行完成後，引擎日誌會交付到您的 S3 儲存貯體，並無限期提供，直到您將其刪除為止。這些日誌位於您為工作流程指定的 S3 輸出 URI 的日誌目錄中。

日誌目錄的路徑格式如下：`s3://{user_provided_path}/logs/`。

下表提供 Amazon S3 儲存貯體中可用的 HealthOmics 日誌摘要。

日誌名稱	在 Amazon S3 中提供	何時可使用日誌	日誌串流路徑
引擎日誌	是	執行完成後	s3：//*user_provided_path* /logs/engine.log
Outputs.json (WDL 和 CWL)	是	執行完成後	s3：//*user_provided_path* /*runID*/*runUUID*/logs/outputs.json
執行資訊清單日誌、執行日誌和任務日誌	否	N/A	無

CLI 中的互動式 CloudWatch Logs

您可以使用互動式模式下的 Live Tail 命令，以互動方式檢視 CloudWatch Logs。您可以即時追蹤執行進度，並定義最多 5 個關鍵字以在日誌中反白顯示：

```
aws logs start-live-tail  \
  --mode interactive  \
  --log-group-identifiers arn:aws:logs:region:account-ID:log-group:/aws/omics/
WorkflowLog
```

如需詳細資訊，請參閱《 AWS CLI 命令參考》中的啟動即時結尾。

從主控台存取 CloudWatch Logs

若要存取執行的日誌，您可以從 HealthOmics 主控台的執行詳細資訊頁面直接連結至這些日誌。

1. 開啟 HealthOmics 主控台。
2. 在左側導覽窗格中，選擇執行。
3. 從執行資料表中選取執行。
4. 在執行詳細資訊頁面中，您可以選擇下列任何動作：
   a. 從執行摘要中，選擇檢視執行日誌。主控台會在 CloudWatch 主控台中開啟執行日誌。
   b. 在執行摘要中，選擇在 Amazon S3 中檢視日誌。主控台會在 Amazon S3 主控台中開啟日誌資料夾。
   c. 在執行任務中，選擇檢視日誌、檢視執行日誌或檢視任務的執行資訊清單日誌。主控台會在 CloudWatch 主控台中開啟日誌。

您也可以從 CloudWatch 主控台導覽至日誌：

1. 開啟 CloudWatch 主控台 https：//https://console.aws.amazon.com/cloudwatch/。
2. 從左側功能表中，選擇日誌群組。
3. 選擇 `/aws/omics/WorkflowLog` 群組。

   如果日誌群組清單很長，您可以在搜尋文字方塊中輸入 omics 以縮小清單範圍。
4. 當日誌群組詳細資訊頁面開啟時，選擇您要檢視的日誌串流。主控台會顯示此日誌串流的事件。

使用 記錄 AWS HealthOmics API 呼叫 AWS CloudTrail

AWS HealthOmics 已與 服務整合 AWS CloudTrail，此服務提供使用者、角色或 HealthOmics 中 AWS 服務所採取動作的記錄。CloudTrail 會將 HealthOmics 的所有 API 呼叫擷取為事件。擷取的呼

叫包括來自 HealthOmics 主控台的呼叫，以及對 HealthOmics API 操作的程式碼呼叫。如果您建立線索，則可以將 CloudTrail 事件持續交付至 Amazon S3 儲存貯體，包括 HealthOmics 的事件。即使您未設定追蹤，依然可以透過 CloudTrail 主控台中的事件歷史記錄檢視最新事件。您可以使用 CloudTrail 所收集的資訊，判斷向 HealthOmics 提出的請求、提出請求的 IP 地址、提出請求的人員、提出請求的時間，以及其他詳細資訊。

若要進一步了解 CloudTrail，請參閱「AWS CloudTrail 使用者指南」。

CloudTrail 中的 HealthOmics 資訊

當您建立帳戶 AWS 帳戶 時，您的 上會啟用 CloudTrail。在 HealthOmics 中發生活動時，該活動會與事件歷史記錄中的其他 AWS 服務事件一起記錄在 CloudTrail 事件中。您可以在 中檢視、搜尋和下載最近的事件 AWS 帳戶。如需詳細資訊，請參閱「使用 CloudTrail 事件歷史記錄檢視事件」。

若要持續記錄 中的事件 AWS 帳戶，包括 HealthOmics 的事件，請建立追蹤。線索能讓 CloudTrail 將日誌檔案交付至 Amazon S3 儲存貯體。依預設，當您在主控台中建立追蹤時，該追蹤會套用至所有的 AWS 區域。線索會記錄 AWS 分割區中所有區域的事件，並將日誌檔案傳送到您指定的 Amazon S3 儲存貯體。此外，您可以設定其他 AWS 服務，以進一步分析和處理 CloudTrail 日誌中收集的事件資料。如需詳細資訊，請參閱下列內容：

- 建立追蹤的概觀
- CloudTrail 支援的服務和整合
- 設定 CloudTrail 的 Amazon SNS 通知
- 接收多個區域的 CloudTrail 日誌檔案和接收多個帳戶的 CloudTrail 日誌檔案

CloudTrail 會記錄所有 HealthOmics 動作，並記錄在 AWS HealthOmics API 參考中。例如，對 `CreateReferenceStore`、`StartVariantImportJob` 和 `CreateWorkflow` 動作發出的呼叫會在 CloudTrail 記錄檔案中產生專案。

每一筆事件或日誌專案都會包含產生請求者的資訊。身分資訊可協助您判斷下列事項：

- 是否使用 IAM 使用者登入資料提出請求。
- 提出該請求時，是否使用了特定角色或聯合身分使用者的暫時安全憑證。
- 請求是否由其他 AWS 服務提出。

如需詳細資訊，請參閱 CloudTrail userIdentity 元素。

了解 HealthOmics 日誌檔案項目

追蹤是一種組態，能讓事件以日誌檔案的形式交付到您指定的 Amazon S3 儲存貯體。CloudTrail 日誌檔案包含一或多個日誌專案。一個事件為任何來源提出的單一請求，並包含請求動作、請求的日期和時間、請求參數等資訊。CloudTrail 日誌檔並非依公有 API 呼叫的堆疊追蹤排序，因此不會以任何特定順序出現。

下列範例顯示示範 CreateWorkflow 動作的 CloudTrail 日誌項目。


```
{
    "eventVersion": "1.08",
    "userIdentity": {
        "type": "AssumedRole",
        "principalId": "AROAIU53LOGOMTOPXXNPG:username",
        "arn": "arn:aws:sts::account:assumed-role/admin/username",
        "accountId": "account-id",
        "accessKeyId": "accessKeyId",
        "sessionContext": {
            "sessionIssuer": {
                "type": "Role",
                "principalId": "AROAIU53LOGOMTOPXXNPG",
                "arn": "arn:aws:iam::account:role/admin",
                "accountId": "account",
                "userName": "admin"
            },
            "webIdFederationData": {},
            "attributes": {
                "creationDate": "2022-07-23T18:26:09Z",
                "mfaAuthenticated": "false"
            }
        }
    },
    "eventTime": "2022-07-23T18:46:42Z",
    "eventSource": "omics.amazonaws.com",
    "eventName": "CreateWorkflow",
    "awsRegion": "us-west-2",
    "sourceIPAddress": "205.251.233.176",
    "userAgent": "aws-cli/1.22.45 Python/3.9.13 Darwin/20.6.0 botocore/1.23.45",
    "requestParameters": {
        "name": "parameter_name",
        "definitionZip": "czM6Ly93b3JrZmxvd2RlZi1oZWxsby9kZWZpbml0aW9uLnppcA==",
        "requestId": "d788a73c-b81b-45fb-a8a6-d8bb4449ec8a"
    },
```

```
    "responseElements": {
        "id": "1002571",
        "arn": "arn:aws:omics:us-west-2:555555555555:instance/i-b188560f ",
        "status": "CREATING",
        "tags": {
            "resourceArn": "arn:aws:omics:us-west-2:083685709690:workflow/1002571"
        }
    },
    "requestID": "842d731d-f264-4b08-a2c9-2f7d45e1eaa3",
    "eventID": "76872ca2-f208-4193-807d-7dd7ea34e6b2",
    "readOnly": false,
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "recipientAccountId": "083685709690",
    "eventCategory": "Management"
}
```

搭配 使用 EventBridge AWS HealthOmics

當資源變更狀態時，HealthOmics 會將事件傳送至 Amazon EventBridge。資源包括匯入任務、匯出任務、資源共用、工作流程、任務和執行。對於每種類型的資源，都有產生事件的狀態變更清單。

事件匯流排是接收事件並將其交付至目的地的路由器。您的帳戶包含自動從 AWS 服務接收事件的預設事件匯流排。您可以建立其他自訂事件匯流排。

您可以建立 EventBridge 規則來指定事件匯流排接收事件時要採取的動作。例如，您可以建立規則，通知您資源的狀態變更。

使用事件的常見案例包括：

- 監控使用者何時與您共用資源或撤銷共用。
- 監控執行是否失敗或成功完成。

如需使用 EventBridge 的詳細資訊，請參閱什麼是 Amazon EventBridge？

主題

- 設定 HealthOmics 的 EventBridge
- HealthOmics 中的 EventBridge 事件
- 事件訊息結構

- 事件訊息範例

設定 HealthOmics 的 EventBridge

在您可以監控 EventBridge 事件之前，請先建立 EventBridge 匯流排，並為感興趣的事件建立規則。

設定 EventBridge 匯流排

您可以使用 的預設事件匯流排， AWS 帳戶 或設定自訂事件匯流排。若要設定自訂事件匯流排，請遵循下列步驟：

1. 開啟 EventBridge 主控台：https：//https://console.aws.amazon.com/events/。
2. 在左側導覽中，選擇事件匯流排。
3. 選擇 Create event bus (建立事件匯流排)。
4. 在建立事件匯流排表單中，輸入匯流排的名稱。
5. 選擇建立以建立匯流排。

建立 EventBridge 規則

下列程序說明如何建立簡單的規則。如需規則的詳細資訊，請參閱 EventBridge 中的規則。

1. 開啟 EventBridge 主控台：https：//https://console.aws.amazon.com/events/。
2. 在左側導覽中，選擇 規則。
3. 選擇建立規則。主控台會開啟建立規則表單。
4. 在定義規則詳細資訊中，提供規則的名稱。
   - 在名稱中，輸入匯流排的名稱。
   - 針對事件匯流排，選取此規則的匯流排。
   - 選擇下一步。
5. 在建置事件模式中，在事件來源下選取 AWS 事件或 EventBridge 合作夥伴事件。
6. 向下捲動至事件模式。
   a. 針對事件來源，選取 AWS 服務。
   b. 針對 AWS 服務，在文字篩選條件中輸入 omics，然後選取 AWS HealthOmics做為服務。
   c. 針對事件類型，選取感興趣的事件 （或所有事件)。

d. 選擇下一步。

7. 在選取目標 (Select target) 中，選取事件的目標。例如，選擇 AWS 服務、選擇 CloudWatch 日誌群組，然後設定日誌群組。

   對於許多目標類型，EventBridge 需要許可才能將事件傳送到目標。主控台會為您建立這些許可。

8. （選用）在設定標籤中，將標籤與規則建立關聯。
9. 在檢閱和更新中，檢閱組態，然後選擇建立規則。

HealthOmics 中的 EventBridge 事件

下表列出 HealthOmics 傳送至 EventBridge 的事件，以及事件的可能狀態值清單。

事件名稱	可能的狀態值
註釋匯入任務狀態變更	已提交、進行中、已取消、已完成、失敗或已完成但失敗
註釋存放區共用狀態變更	待處理、啟用、作用中、刪除、刪除、失敗
註釋存放區狀態變更	建立、建立、更新、更新、刪除、刪除或建立失敗
讀取設定啟用任務狀態變更	已提交、進行中、已完成、失敗或已完成但失敗
讀取設定匯出任務狀態變更	已提交、進行中、已完成、失敗或已完成但失敗
讀取集匯入任務狀態變更	已提交、進行中、已完成、失敗或已完成但失敗
讀取集狀態變更	處理上傳、上傳失敗、作用中、封存、啟用或刪除
參考匯入任務狀態變更	已提交、進行中、已完成、失敗或已完成但失敗
參考狀態變更	作用中或已刪除
參考存放區狀態變更	已建立、更新、作用中或刪除
執行狀態變更	待定、開始、執行、停止、完成、刪除、失敗或取消

事件名稱	可能的狀態值
序列存放區狀態變更	已建立、更新、作用中或刪除
任務狀態變更	待定、開始、執行、停止、完成、刪除、失敗或取消
變體匯入任務狀態變更	已提交、進行中、已取消、已完成、失敗或已完成但失敗
變體存放區共用狀態變更	待處理、啟用、作用中、刪除、刪除、失敗
變體存放區狀態變更	建立、建立、更新、更新、刪除、刪除或建立失敗
工作流程共用狀態變更	待處理、啟用、作用中、刪除、刪除、失敗
工作流程狀態變更	建立成功、建立失敗、刪除成功或刪除失敗

事件訊息結構

HealthOmics 會竭盡全力將狀態變更事件訊息傳送至 EventBridge。事件是具有 JSON 結構的物件，也包含中繼資料詳細資訊。您可以使用中繼資料做為輸入，以重新建立事件或了解詳細資訊。事件包括下列欄位：

- `version` — 目前所有事件為 0（零）。
- `id` — 為每個事件產生的第 4 版 UUID。
- `detail-type` — 正在傳送的事件類型。
- `account` — 儲存貯體擁有者的 12 位數 AWS 帳戶 ID。
- `source` — 識別產生事件的服務。
- `time` — 事件發生的時間。
- `region` — 識別儲存貯 AWS 區域 體的 。
- `resources` — 包含儲存貯體 Amazon Resource Name (ARN) 的 JSON 陣列。
- `detail` — 包含事件相關資訊的 JSON 物件。

執行事件包括下列欄位：

- uuid — 執行的通用唯一識別符。
- workflowId — 與此執行相關聯的工作流程的工作流程識別符。
- workflowName — 與此執行相關聯的工作流程名稱。
- runId — 執行識別符。
- runName — 執行名稱。
- runOutputUri — 執行將寫入其輸出資料的 URI。

事件訊息範例

下列範例是執行狀態變更的事件，顯示其他欄位。

```
{
    "version":"0",
    "id":"c0e540f4-df38-b986-86c1-3e3730f971fe",
    "detail-type":"Run Status Change",
    "source":"aws.omics",
    "account":"123456789012",
    "time":"2022-10-20T22:07:35Z",
    "region":"us-west-2",
    "resources":[
        "arn:aws:omics:us-west-2:123456789012:run/2101313"
    ],
    "detail":{
        "omicsVersion":"1.0.0",
        "arn":"arn:aws:omics:us-west-2:123456789012:run/2101313",
        "status":"COMPLETED",
        "uuid":"153893cd-097a-40ec-aec7-838a97cd2b21",
        "runId": "1234567",
        "runName": "run name",
        "runOutputUri": "s3://amzn-s3-demo-bucket/run-output/2101313",
        "workflowId": "1234567",
        "workflowName": "workflow name"
    }
}
```

下列範例是任務狀態變更的事件。

```
{
    "version": "0",
```

```
    "id": "718d6817-c868-26d3-8ef0-0dc9b2ac73f4",
    "detail-type": "Task Status Change",
    "source": "aws.omics",
    "account": "123456789012",
    "time": "2024-10-30T09:05:44Z",
    "region": "us-west-2",
    "resources": ["arn:aws:omics:us-west-2:123456789012:task/8888888"],
    "detail": {
        "omicsVersion": "1.0.0",
        "arn": "arn:aws:omics:us-west-2:123456789012:task/8888888",
        "status": "COMPLETED",
        "runArn": "arn:aws:omics:us-west-2:123456789012:run/2101313",
        "runUuid": "153893cd-097a-40ec-aec7-838a97cd2b21",
        "runId": "1234567",
        "runName": "run name",
        "workflowId": "1234567",
        "workflowName": "workflow name"
    }
}
```

以下是讀取集狀態變更的事件範例。

```
{
  "version": "0",
  "id": "64ca0eda-9751-dc55-c41a-1bd50b4fc9b7",
  "detail-type": "Read Set Status Change",
  "source": "aws.omics",
  "account": "123456789012",
  "time": "2023-04-04T17:53:06Z",
  "region": "us-west-2",
  "resources": ["arn:aws:omics:us-west-2:123456789012:sequenceStore/1234567890/
readSet/3456789012"],
  "detail": {
     "omicsVersion": "1.0.0",
     "arn": "arn:aws:omics:us-west-2:123456789012:sequenceStore/1234567890/
readSet/3456789012",
     "sequenceStoreId" : "1234567890",
     "id": "3456789012",
     "status": "PROCESSING_UPLOAD"
  }
}
```

系統會為變體存放區匯入任務建立類似的事件。

```
{
    "version": "0",
    "id": "6a7e8feb-b491-4cf7-a9f1-bf3703467718",
    "detail-type": "Variant Store Status Change",
    "source": "aws.omics",
    "account": "123456789012",
    "time": "2015-12-22T18:43:48Z",
    "region": "us-east-1",
    "resources": ["arn:aws:omics:us-east-1:123456789012:myvariantstore2"],
    "detail": {
       "omicsVersion": "1.0.0",
       "arn": "arn:aws:omics:us-east-1:123456789012:myvariantstore2",
       "status": "CREATED",
       "storeId": "6710c5f02610",
       "storeName": "myvariantstore2"
    }
}
```

以下是匯入任務狀態變更的事件。

```
{
    "version": "0",
    "id": "6a7e8feb-b491-4cf7-a9f1-bf3703467718",
    "detail-type": "Variant Import Job Status Change",
    "source": "aws.omics",
    "account": "123456789012",
    "time": "2015-12-22T18:43:48Z",
    "region": "us-east-1",
    "resources": ["arn:aws:omics:us-east-1:123456789012:my_variant_store/
b64ea9a3-459f-4b68-92c3-3ddb83209fe9"],
    "detail": {
       "omicsVersion": "1.0.0",
       "arn": "arn:aws:omics:us-east-1:123456789012:my_variant_store/
b64ea9a3-459f-4b68-92c3-3ddb83209fe9",
       "status": "COMPLETED",
       "jobId": "b64ea9a3-459f-4b68-92c3-3ddb83209fe9",
       "storeId": "a74869f91e20",
       "storeName": "my_variant_store"
    }
}
```

故障診斷

下列主題可協助您針對使用 HealthOmics 工作流程和資料存放區時遇到的問題進行疑難排解。

主題

對工作流程進行故障診斷

主題

如何對失敗的執行進行故障診斷？

使用 GetRun API 操作來擷取失敗原因。如需詳細資訊，請參閱執行失敗原因。

如何對失敗的任務進行故障診斷？

檢閱任務失敗訊息中的錯誤代碼，以了解失敗。檢閱 CloudWatch 中的任務日誌，以查看任務的詳細記錄訊息。如果您沒有收到詳細的日誌訊息，您可以修改工作流程以輸出其他日誌陳述式。如需詳細資訊，請參閱使用 CloudWatch Logs 監控 HealthOmics 。

在哪裡可以找到成功完成執行的引擎日誌？

HealthOmics 只會針對失敗的執行發佈日誌至 CloudWatch。如果執行成功完成，HealthOmics 會將引擎日誌交付到您的 Amazon S3 儲存貯體。如需詳細資訊，請參閱Amazon S3 中的日誌。

如何減少工作流程的輸入參數大小？

您可以為工作流程指定最多 50 KB 的輸入參數。您可以使用目錄匯入或範例工作表來維持在此大小限制內。如需詳細資訊，請參閱管理執行參數大小。

為什麼我的執行未完成？

如果您的程式碼發生問題，且程序未正確結束，您的執行可能會變得沒有回應或「卡住」。如需如何防止和捕捉無回應執行的詳細資訊，請參閱 無回應執行的指引。

對呼叫快取問題進行故障診斷

下列主題可協助您針對呼叫快取時遇到的問題進行疑難排解。

主題

為什麼我的執行不會儲存至快取？

1. 透過檢查 GetRun API 操作回應中的 cacheId 欄位，確認執行已設定為使用快取。使用 CLI，執行此命令：`aws omics get-run –id <run_id>`。
2. 如果執行成功，請確認 GetRun 回應中傳回的快取行為為 CACHE_ALWAYS。如果快取行為設定為 CACHE_ON_FAILURE，則執行只會在失敗時儲存至快取。

為什麼任務不使用快取項目？

在 `/aws/omics/WorkflowLog` CloudWatch 日誌群組中，開啟執行快取的日誌串流：runCache/<cache_id>/<cache_uuid>。

1. 確認先前的執行已為您預期要快取的任務建立快取項目。儲存到快取的執行會以 CACHE_ENTRY_CREATED 的日誌訊息記錄。
2. 找到任務的 CACHE_MISS 日誌，並執行已完成的日誌。如果沒有日誌項目，請檢查執行是否已設定為使用快取。
3. 如果已建立快取項目，請確認兩個任務CPUs、記憶體、GPUs 和容器摘要都相同。建立快取項目之任務的任務 ARN 位於日誌訊息中。

4. 如果兩個任務的運算需求相符，請確認任務之間的輸入未變更。若要這樣做，請開啟引擎日誌。如果執行的狀態為 FAILED，則日誌將在 Cloudwatch Log Group /aws/omics/WorkflowLog 中。否則，您可以在執行的輸出目錄中找到引擎日誌。

對資料存放區進行故障診斷

主題

為什麼我的讀取集上的 S3 GetObject 失敗？

最常見的是，失敗是因為缺少許可。序列存放區 S3 讀取許可是一種雙向組態，需要序列存放區 S3 存取政策允許存取，且 IAM 主體必須連接允許存取的政策。如需政策需求的詳細資訊，請參閱 使用 Amazon S3 URIs 存取資料的許可。檢查下列組態是否已就位：

- 序列存放區 S3 存取政策明確允許存取 IAM 主體或主體帳戶的根目錄。
- 檢查 IAM 主體是否具有明確提供存取資源許可的政策。請注意，定義許可時，IAM 主體政策必須使用存取點 ARN，而不是存取點別名型路徑，而且 ARN 處於 條件，而不是用來指定資源。
- 如果您的存放區使用客戶受管金鑰 (CMK-KMS)，請確定 IAM 主體對金鑰具有 kms:decrypt 許可。如需跨帳戶設定用量，請參閱 KMS 跨帳戶存取指南。

如果您有使用標籤型存取控制的政策，請確定下列事項：

- 確保序列存放區已完成標籤同步。因此，存放區的狀態必須為 active，而不是 updating。
- 確保讀取集和政策的標籤索引鍵或索引鍵值中沒有錯別字。

為什麼我在 Athena 中看不到註釋存放區或變體存放區？

在 Lake Formation 中，請務必根據與您共用的存放區建立資源連結。建立您有權存取的資源連結後，應該會在 Athena 中顯示該存放區。如需詳細資訊，請參閱設定 Lake Formation 以使用 HealthOmics。

為什麼我無法存取 Athena 中的資料存放區？

如果您的註釋或變體存放區可見，但您收到錯誤訊息，指出存取遭拒，請檢查您正在使用的查詢引擎版本。僅支援使用引擎版本 3 執行的查詢。若要進一步了解 Athena 查詢引擎版本，請參閱 Amazon Athena 文件。

的配額 AWS HealthOmics

AWS 會使用 HealthOmics 配額的預設值填入您的帳戶。除非另有說明，否則每個配額值都是每個區域的最大值。

> **Important**
>
> 您可以請求提高大部分的服務配額和 API 配額。請參閱以下主題以了解詳細資訊。

主題

HealthOmics 服務配額

下表列出 HealthOmics 服務配額及其預設值。若要檢視每個區域的目前配額，請開啟 Service Quotas 主控台。

> **Important**
>
> 您可以使用 Service Quotas 主控台請求增加可調整配額。

如需服務配額的詳細資訊，請參閱Service Quotas 使用者指南》中的請求提高配額。對於 Service Quotas 主控台中無法使用的配額，請使用配額增加表單。

名稱	預設	可調整	描述
分析 - 註釋存放區上限	每個受支援的區域：10	是	目前 AWS 區域中的註釋存放區數量上限

名稱	預設	可調整	描述
分析 - 並行變體或註釋存放區匯入任務上限	每個受支援的區域：5	是	目前 AWS 區域中並行匯入任務的數量上限
分析 - 每個變體存放區匯入任務的檔案上限	每個受支援的區域：1,000	是	目前 AWS 區域中每個變體匯入任務的檔案數量上限
分析 - 每個註釋存放區的最大共用數	每個受支援的區域：10	是	目前 AWS 區域中每個註釋存放區的共用數量上限
分析 - 每個變體存放區的最大共用	每個受支援的區域：10	是	目前 AWS 區域中每個變體存放區的共用數量上限
分析 - 變體匯入任務中每個檔案的大小上限	每個支援的區域：20 GB	是	目前 AWS 區域中變體匯入任務中一個檔案的大小上限
分析 - 註釋匯入任務中每個檔案的大小上限	每個支援的區域：20 GB	是	目前 AWS 區域中註釋匯入任務中一個檔案的大小上限
分析 - 變體存放區上限	每個受支援的區域：10	是	目前 AWS 區域中的變體存放區數量上限
分析 - 每個註釋存放區的版本上限	每個受支援的區域：10	是	目前 AWS 區域中每個註釋存放區的版本數量上限
儲存 - 並行讀取集啟用任務上限	每個受支援的區域：25	是	目前 AWS 區域中並行讀取集啟用任務的數量上限
儲存 - 最大並行序列和參考存放區匯出任務	每個受支援的區域：5	是	目前 AWS 區域中序列或參考存放區的並行匯出任務數目上限

名稱	預設	可調整	描述
儲存 - 最大並行序列或參考存放區匯入任務	每個受支援的區域：5	是	目前 AWS 區域中序列或參考存放區的並行匯入任務數目上限
儲存 - 每個序列存放區的讀取集上限	每個受支援的區域：1,000,000	是	目前 AWS 區域中序列存放區中的讀取集數目上限
儲存 - 每個參考存放區的最大參考數	每個受支援的區域：50	是	目前 AWS 區域中參考存放區中的參考數目上限
儲存 - 序列儲存上限	每個受支援的區域：20	是	目前 AWS 區域中的序列存放區數目上限
工作流程 - 作用中 GPUs 上限	每個支援的區域：12	是	目前 AWS 區域中並行作用中 GPUs 的數量上限。在 us-east-1 和 us-west-2 中，最多 500 個值的配額增加請求會自動核准。
工作流程 - 使用動態執行儲存體的並行作用中執行上限	每個受支援的區域：50	是	在目前 AWS 區域中使用動態執行儲存體的作用中執行數量上限。最多 200 個值的配額增加請求會自動核准。
工作流程 - 使用靜態執行儲存體的並行作用中執行上限	每個受支援的區域：10	是	在目前 AWS 區域中使用靜態執行儲存體的作用中執行數量上限。最多 50 個值的配額增加請求會自動核准。

名稱	預設	可調整	描述
工作流程 - 每次執行的並行任務上限	每個受支援的區域：25	是	目前 AWS 區域中每次執行的並行任務數目上限。在 us-east-1 和 us-west-2 中，最多 100 個值的配額增加請求會自動核准。
工作流程 - 執行持續時間上限	每個支援的區域：604，800 秒	是	目前 AWS 區域中的工作流程執行持續時間上限。
工作流程 - 執行上限（作用中或非作用中）	每個受支援的區域：5,000	是	目前 AWS 區域中的執行次數上限（作用中或非作用中）。
工作流程 - 每個工作流程的共用數量上限	每個受支援的區域：100	是	目前 AWS 區域中每個工作流程的共用數量上限
工作流程 - 每次執行的靜態執行儲存容量上限	每個支援的區域：9，600	是	目前 AWS 區域中每次執行的最大靜態執行儲存容量，以 GB (GiB) 為單位。在 us-east-1 和 us-west-2 中，配額增加請求最多 50，000 個值會自動核准。
工作流程 - 工作流程上限	每個受支援的區域：1,000	是	目前 AWS 區域中的工作流程數目上限。
工作流程 - StartRun 操作的每秒交易數 (TPS)	每個支援的區域：0.1	是	目前 AWS 區域中 StartRun 操作的每秒交易數上限 (TPS)。最多 1 個值的配額增加請求會自動核准。

HealthOmics 固定大小配額

除了 之外HealthOmics 服務配額，HealthOmics 還包含具有固定大小的配額。您無法請求增加這些值。

除非另有說明，否則每個配額都會列出每個區域的最大值。

主題

HealthOmics 分析固定大小配額

下表顯示分析配額支援的值上限。這些值無法調整。

名稱	描述	最大	可調整 是/否
分析 - 每個註釋存放區匯入任務的檔案上限	每個註釋匯入任務的檔案數量上限。	1	否

HealthOmics 儲存體固定大小配額

下表顯示儲存檔案支援的值上限。這些值無法調整。

名稱	描述	最大	可調整 是/否
儲存 - S3 存取資源政策大小上限	S3 存取資源政策的大小上限	15 KB	否
儲存 - 傳播集層級標籤上限	每個存放區傳播到 S3 物件的設定層級標籤索引鍵數目上限	5	否

名稱	描述	最大	可調整 是/否
儲存 - 每個啟用任務的讀取集上限	每個啟用任務的讀取集數量上限。	20	否
儲存 - 每個匯出任務的讀取集上限	每個匯出任務的讀取集數目上限。	100	否
儲存 - 每個匯入任務的讀取集上限	每個匯入任務的讀取集數目上限。	100	否
儲存 - 最大參考儲存	參考存放區的數目上限。	1	否
儲存 - 直接上傳的部分大小上限	直接上傳至序列存放區的組件大小上限。	100 MB	否
儲存 - 檔案中用於直接上傳的部分上限	要直接上傳至序列存放區的檔案中的組件數量上限。	10,000	否
儲存 - 參考大小上限	可匯入至參考存放區的參考檔案大小上限。	15 GB	否
儲存 - 讀取集來源大小上限	讀取集中單一來源檔案的大小上限，可匯入序列存放區。	976 GB	否

HealthOmics 工作流程固定大小配額

下表顯示工作流程配額支援的值上限。這些值無法調整。

名稱	描述	大小上限	可調整 是/否
工作流程 - 執行群組上限	執行群組的數量上限。	1000	否

名稱	描述	大小上限	可調整 是/否
工作流程 - 執行快取上限	您可以為一個帳戶建立的執行快取數目上限。 一個或多個執行可以共用相同的執行快取。HealthOmics 可以為每個帳戶快取的執行次數沒有配額。	1000	否
工作流程 - 工作流程版本上限	每個工作流程的工作流程版本數目上限。	1000	否
工作流程 - CPU 執行個體容器大小	CPU 執行個體的容器映像大小上限。	45 GiB	否
工作流程 - GPU 執行個體容器大小	GPU 執行個體的容器映像大小上限。	95 GiB	否
GPU 執行個體 /dev/shm 共用記憶體	每個 GPU 執行個體的共用記憶體數量上限。	每個 GPU 8 GB	否
工作流程 - 執行參數檔案	執行參數檔案的大小上限。	50，000 個位元組	否
工作流程 - 工作流程參數範本檔案	工作流程參數範本檔案的項目數量上限和檔案大小上限。此配額適用於您使用主控台或 API 建立的工作流程。	1，000 個項目，400 KB	否
工作流程 - 工作流程定義檔案大小 - API	當您使用 API 操作或 AWS SDK 建立工作流程時，工作流程定義檔案的大小上限。	100 MB	否

名稱	描述	大小上限	可調整 是/否
工作流程 - 工作流程定義檔案大小 - 主控台（直接上傳）	當您使用主控台建立工作流程時，您可以直接上傳的工作流程定義檔案大小上限。	4.4 MB	否
工作流程 - 工作流程定義檔案大小 - 主控台（從 Amazon S3 上傳）	當您使用主控台建立工作流程時，可以從 Amazon S3 上傳提供的工作流程定義檔案大小上限。	100 MB	否
工作流程 - 儲存庫大小	外部程式碼儲存庫的大小上限。	1 GiB	否
工作流程 - 儲存庫個別檔案大小	來自外部程式碼儲存庫的個別檔案大小上限。	100 MiB	否
工作流程 - README 檔案大小	README 檔案的大小上限。	500 KiB	否

如需如何減少執行參數檔案大小的建議，請參閱 管理執行參數大小。

HealthOmics Ready2Run 工作流程固定大小配額

每個 Ready2Run 工作流程都有最大輸入檔案大小。在下表中，檔案大小單位會以 Gibibytes (GiB) 列出。這些檔案大小上限無法調整。

Ready2Run 工作流程名稱	輸入檔案大小上限 (GiB)	可調整 （是/否）
AlphaFold 用於 601-1200 殘差	1	否
AlphaFold 最多可用於 600 個殘差	1	否
Bases2Fastq for 2x150	1000	否

Ready2Run 工作流程名稱	輸入檔案大小上限 (GiB)	可調整 （是/否）
Bases2Fastq for 2x300	1000	否
Bases2Fastq for 2x75	500	否
ESMFold 最多可用於 800 個殘差	1	否
GATK-BP fq2bam	64	否
適用於 30 倍基因體的 GATK-BP Germline bam2vcf	39	否
適用於 30 倍基因體的 GATK-BP Germline fq2vcf	64	否
GATK-BP 體力 WES bam2vcf	86	否
NVIDIA Parabricks BAM2FQ2BAM WGS，最多可達 30X	80	否
NVIDIA Parabricks BAM2FQ2BAM WGS，最多可達 50X	120	否
NVIDIA Parabricks BAM2FQ2BAM WGS 最多可達 5X	20	否
NVIDIA Parabricks FQ2BAM WGS 最多可達 30X	71	否
高達 50X的 NVIDIA Parabricks FQ2BAM WGS	137	否
NVIDIA Parabricks FQ2BAM WGS 最多可達 5X	13	否

Ready2Run 工作流程名稱	輸入檔案大小上限 (GiB)	可調整 （是/否）
高達 30X的 NVIDIA Parabricks Germline DeepVariant WGS	71	否
高達 50X的 NVIDIA Parabricks Germline DeepVariant WGS	137	否
NVIDIA Parabricks Germline DeepVariant WGS 最多可達 5X	12	否
NVIDIA Parabricks Germline HaplotypeCaller WGS 最多可達 30X	71	否
NVIDIA Parabricks Germline HaplotypeCaller WGS 最多可達 50X	137	否
NVIDIA Parabricks Germline HaplotypeCaller WGS 最多可達 5X	13	否
NVIDIA Parabricks 體力 Mutect2 WGS，最多可達 50X	196	否
scRNAseq 與 KallistoBUStools	119	否
scRNAseq 搭配 Salmon Alevin-fry	119	否
scRNAseq 搭配 STARsolo	119	否
Sentieon Germline BAM WES 高達 300 倍	9	否
Sentieon Germline BAM WGS 最多可達 32 倍	18	否

Ready2Run 工作流程名稱	輸入檔案大小上限 (GiB)	可調整 （是/否）
Sentieon Germline FASTQ WES 高達 100 倍	5	否
Sentieon Germline FASTQ WES 高達 300 倍	26	否
Sentieon Germline FASTQ WGS 最多可達 32 倍	51	否
適用於 ONT 的 Sentieon LongRead	25	否
適用於 PacBio HiFi 的 Sentieon LongRead	58	否
Sentieon 體操 WES	50	否
Sentieon 體操 WGS	113	否
Ultima Genomics DeepVariant 最多可達 40 倍	91	否

HealthOmics API 配額

HealthOmics 具有與 API 操作相關的下列配額。若有指示，配額是可調整的。若要請求增加，請使用配額增加表單。

對於列出的每個 API 操作，配額是每個區域中該 API 操作的每秒交易數上限 (TPS)。

主題

- 一般 API 配額
- 儲存 API 配額
- 工作流程 API 配額
- Analytics API 配額

一般 API 配額

下表列出套用至多個類別（儲存、工作流程和分析）的一般 API 操作。

API 操作	預設 TPS 上限	可調整（是/否）
AcceptShare、CreateShare、DeleteShare、GetShare、ListShares	1 TPS	是

儲存 API 配額

下表列出儲存 API 操作。

儲存 API 操作	預設 TPS 上限	可調整（是/否）
CreateSequenceStore、UpdateSequenceStore、DeleteSequenceStore、CreateReferenceStore、DeleteReferenceStore	1 TPS	是
BatchDeleteReadSet、DeleteReference	1 TPS	是
CreateMultipartReadSetUpload、 CompleteMultipartReadSetUpload、AbortMultipartReadSetUpload	1 TPS	否
GetS3AccessPolicy, PutS3AccessPolicy, DeleteS3AccessPolicy	1 TPS	是
GetReference	10 TPS	是
UploadReadSetPart	10 TPS	是

儲存 API 操作	預設 TPS 上限	可調整（是/否）
GetReadSet	30 TPS	是
GetSequenceStore、ListSequenceStores	5 TPS	是
GetReadSetMetadata、ListReadSets	5 TPS	是
StartReadSetImportJob、GetReadSetImportJob、ListReadSetImportJobs	5 TPS	是
StartReadSetExportJob、GetReadSetExportJob、ListReadSetExportJobs	5 TPS	是
ListReferenceStores	5 TPS	是
StartReferencetImportJob、GetReferenceImportJob、ListReferenceImportJobs	5 TPS	是
ListReferences、GetReferenceMetadata	5 TPS	是
StartReadsetActivationJob	5 TPS	是
ListReadsetActivationJobs、GetReadSetActivationJob	5 TPS	是
ListMultipartReadSetUploads、ListReadSetUploadParts	5 TPS	是
TagResource、UntagResource、ListTagsForResource	5 TPS	是

工作流程 API 配額

下表列出工作流程 API 操作。

工作流程 API 操作	預設 TPS 上限	可調整（是/否）
StartRun	0.1 TPS	是
CreateWorkflow	5 TPS	是
CancelRun、DeleteRun、GetRun、GetRunTask、ListRunTasks、ListRuns	10 TPS	是
CreateRunGroup、DeleteRunGroup、GetRunGroup、ListRunGroups、UpdateRunGroup	10 TPS	是
CreateRunCache、UpdateRunCache、DeleteRunCache、GetRunCache、ListRunCaches	10 TPS	是
DeleteWorkflow、GetWorkflow、ListWorkflows、UpdateWorkflow	10 TPS	是

Analytics API 配額

下表列出分析 API 操作。

Analytics API 操作	預設 TPS 上限	可調整（是/否）
CreateVariantStore、DeleteVariantStore、GetVariantStore、ListVariantStores、UpdateVariantStore	1 TPS	否

Analytics API 操作	預設 TPS 上限	可調整（是/否）
StartVariantImportJob、CancelVariantImportJob、GetVariantImportJob、ListVariantImportJobs	1 TPS	否
CreateAnnotationStore、DeleteAnnotationStore、GetAnnotationStore、ListAnnotationStores、UpdateAnnotationStore	1 TPS	否
StartAnnotationImportJob、ListAnnotationImportJobs、GetAnnotationImportJob、CancelAnnotationImportJob	1 TPS	否

HealthOmics 使用者指南的文件歷史記錄

下表說明 HealthOmics 的文件版本。

變更	描述	日期
新的 README 和儲存庫整合功能	新增從外部程式碼儲存庫和 README 檔案建立工作流程的支援。	2025 年 7 月 24 日
新功能	HealthOmics 新增對 Nextflow 自動參數插補的支援。若要進一步了解，請參閱 HealthOmics 工作流程的參數範本檔案。	2025 年 6 月 27 日
新功能	HealthOmics 新增工作流程版本控制的支援。若要進一步了解，請參閱 HealthOmics 中的工作流程版本控制。	2025 年 4 月 18 日
新功能	HealthOmics 為動態執行儲存增加彈性輸送量。若要進一步了解，請參閱在 HealthOmics 中執行儲存類型。	2025 年 4 月 16 日
新功能	HealthOmics 新增序列存放區 S3 位置的屬性型存取控制，以及最多將五個讀取集標籤同步至序列存放區 S3 物件的彈性。若要進一步了解，請參閱建立 HealthOmics 序列存放區。	2024 年 11 月 22 日
新功能	HealthOmics 新增了對私有工作流程呼叫快取的支援，也稱為繼續。若要進一步了解，請參閱呼叫快取。	2024 年 11 月 20 日

新功能	HealthOmics 新增了新的 API 欄位，協助您在序列存放區輸入任務和讀取集之間進行映射。	2024 年 8 月 29 日
新功能	HealthOmics 新增了管理 Nextflow 版本的支援。若要進一步了解，請參閱 Nextflow 版本。	2024 年 8 月 14 日
新功能	HealthOmics 新增了對共用工作流程和動態執行儲存的支援。	2024 年 4 月 30 日
新功能	HealthOmics 新增了對 Amazon S3 存取參考和序列存放區的支援，以及對 SHA256 ETags 的支援。	2024 年 4 月 15 日
新功能	HealthOmics 新增序列存放區的實體標籤 ETags)。	2023 年 10 月 6 日
新功能	HealthOmics 新增註釋存放區版本控制和分析存放區共用。	2023 年 8 月 15 日
新功能	HealthOmics 新增通用工作流程語言 (CWL) 做為 HealthOmics 工作流程的支援語言。	2023 年 6 月 30 日
新功能	HealthOmics 新增了新的 Ready2Run 工作流程、工作流程的 GPU 支援、註釋存放區的資料剖析、直接上傳至 HealthOmics 儲存，以及與 EventBridge 的整合。	2023 年 5 月 15 日

新的 受管政策	HealthOmics 新增了提供完整存取權的新受管政策。若要進一步了解，請參閱 AWS 受管政策。	2023 年 2 月 23 日
新的 受管政策	HealthOmics 新增了新的受管政策，將存取權限制為唯讀。若要進一步了解，請參閱 AWS 受管政策。	2022 年 11 月 29 日
初始版本	HealthOmics 使用者指南的初始版本	2022 年 11 月 29 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。