使用者指南

AWS Batch

AWS Batch: 使用者指南

Table of Contents

什麼是 AWS Batch？ 1
您是第一次 AWS Batch 使用嗎？ 2
相關服務 3
存取 AWS Batch 3
的元件 AWS Batch 3
運算環境 4
任務佇列 4
任務定義 4
任務 5
排程政策 5
消耗性資源 5
服務環境 5
服務任務 5
設定 AWS Batch 6
建立 IAM 帳戶和管理使用者 6
註冊 AWS 帳戶 6
建立具有管理存取權的使用者 7
建立 IAM 角色 8
建立金鑰對 8
建立 VPC 10
建立安全群組 11
安裝 AWS CLI 13
入門教學課程 14
使用精靈開始使用 Amazon EC2 14
概觀 14
先決條件 15
步驟 1：建立運算環境 15
步驟 2：建立任務佇列 16
步驟 3：建立任務定義 16
步驟 4：建立任務 17
步驟 4：檢閱和建立 17
步驟 6：檢視任務的輸出 17
步驟 7：清除您的教學課程資源 17
其他資源 18

使用精靈開始使用 Fargate 協同運作 18
概觀 18
先決條件 19
步驟 1：建立運算環境 19
步驟 2：建立任務佇列 20
步驟 3：建立任務定義 20
步驟 4：建立任務 21
步驟 4：檢閱和建立 21
步驟 6：檢視任務的輸出 21
步驟 7：清除您的教學課程資源 22
其他資源 22
使用 開始使用 AWS Batch 和 Fargate AWS CLI 22
先決條件 23
建立 IAM 執行角色 24
建立運算環境 25
建立任務佇列 26
建立任務定義 26
提交和監控任務 27
檢視任務輸出 28
清除資源 29
生產環境部署須知 30
後續步驟 31
Amazon EKS 入門 31
概觀 32
先決條件 33
步驟 1：建立 的 Amazon EKS 叢集 AWS Batch 34
步驟 2：準備您的 Amazon EKS 叢集 AWS Batch 34
步驟 3：建立 Amazon EKS 運算環境 38
步驟 4：建立任務佇列並連接運算環境 40
步驟 5：建立任務定義 41
步驟 6：提交任務 42
步驟 7：檢視任務的輸出 42
步驟 8：（選用） 提交具有覆寫的任務 42
步驟 9：清除您的教學課程資源 44
其他資源 44
在 Amazon EKS 私有叢集 AWS Batch 上開始使用 44

概觀 32
先決條件 47
步驟 1：建立 的 EKS 叢集 AWS Batch 48
步驟 2：準備您的 EKS 叢集 AWS Batch 49
步驟 3：建立 Amazon EKS 運算環境 53
步驟 4：建立任務佇列並連接運算環境 54
步驟 5：使用提取快取建立 Amazon ECR 55
步驟 6：註冊任務定義 56
步驟 7：提交要執行的任務 57
步驟 8：檢視任務的輸出 57
步驟 9：（選用） 提交具有覆寫的任務 57
步驟 10：清除您的教學課程資源 58
其他資源 44
故障診斷 59
SageMaker AI AWS Batch 入門 60
概觀 60
先決條件 61
步驟 1：建立 SageMaker AI 執行角色 61
步驟 2：建立您的服務環境 63
步驟 3：建立 SageMaker 任務佇列 64
步驟 4：建立並提交訓練任務 66
步驟 5：監控任務狀態 67
步驟 6：檢視任務輸出 69
步驟 7：清除您的教學課程資源 72
其他資源 72
AWS Batch Widget 儀表板 74
新增單一任務佇列小工具 74
如何新增 CloudWatch Container Insights 小工具 75
新增任務日誌小工具 75
的運算環境 AWS Batch 76
受管運算環境 76
建立多節點平行任務時的考量 78
未受管的運算環境 78
建立運算環境 80
教學課程：使用 Fargate 資源建立受管運算環境 80
教學課程：使用 Amazon EC2 資源建立受管運算環境 82

教學課程：使用 Amazon EC2 資源建立未受管的運算環境 89
教學課程：使用 Amazon EKS 資源建立受管運算環境 90
教學課程：使用 Amazon EKS 資源建立未受管的運算環境 95
資源：運算環境範本 96
執行個體類型運算資料表 97
更新運算環境 100
運算環境更新策略 100
選擇正確的更新策略 101
執行擴展更新 103
執行基礎設施更新 105
執行藍/綠更新 109
運算資源 AMI 114
運算資源 AMI 規格 115
教學課程：建立運算資源 AMI 117
使用 GPU 工作負載 AMI 119
Amazon Linux 棄用 125
Amazon EKS Amazon Linux 2 AMI 棄用 125
Amazon ECS Amazon Linux 2 AMI 棄用 126
使用 Amazon EC2 啟動範本 126
預設和覆寫啟動範本 128
啟動範本中的 Amazon EC2 使用者資料 129
參考：啟動範本範例 130
執行個體中繼資料服務 (IMDS) 組態 132
組態案例 133
EC2 組態 134
如何從 ECS AL2 遷移至 ECS AL2023 135
執行個體類型配置策略 136
記憶體管理 138
預留系統記憶體 139
教學課程：檢視運算資源記憶體 139
Amazon EKS AWS Batch 上的 記憶體和 vCPU 考量事項 139
Fargate 運算環境 144
何時使用 Fargate 145
Fargate 上的任務定義 145
Fargate 上的任務佇列 147
Fargate 上的運算環境 148

Amazon EKS 運算環境 148
Amazon EKS 151
Amazon EKS 預設 AMI 151
混合 AMI 環境 152
支援的 Kubernetes 版本 153
更新運算環境的Kubernetes版本 154
Kubernetes 節點的共同責任 154
在 AWS Batch 受管節點DaemonSet上執行 155
自訂 Amazon EKS 啟動範本 156
如何從 EKS AL2 升級到 EKS AL2023 160
服務環境 162
什麼是服務環境 162
服務環境如何與其他 AWS Batch 元件搭配使用 163
服務環境的最佳實務 163
服務環境狀態和生命週期 164
服務環境狀態定義 164
建立服務環境 165
先決條件 165
更新服務環境 167
刪除服務環境 168
刪除先決條件 169
任務佇列 171
建立任務佇列 171
建立 Amazon EC2 任務佇列 171
建立 Fargate 任務佇列 172
建立 Amazon EKS 任務佇列 173
建立 SageMaker 任務佇列 174
任務佇列範本 177
檢視任務佇列 178
檢視任務佇列資訊 178
刪除任務佇列 180
公平共用排程政策 180
使用共用識別符 181
使用排程政策 182
使用公平共用排程 182
教學課程：建立排程政策 183

參考：排程政策範本 184
資源感知排程 185
建立消耗性資源 186
指定任務的資源 186
檢查資源用量 187
更新使用中資源數量 189
尋找需要消耗性資源的任務 190
刪除消耗性資源 190
任務定義 192
建立單一節點任務定義 192
教學課程：在 Amazon EC2 資源上建立單一節點任務定義 193
在 Fargate 資源上建立單一節點任務定義 198
在 Amazon EKS 資源上建立單一節點任務定義 203
在 Amazon EC2 資源上建立具有多個容器的單一節點任務定義 207
建立多節點平行任務定義 213
教學課程：在 Amazon EC2 資源上建立多節點平行任務定義 213
參考：使用 ContainerProperties 的任務定義範本 219
ContainerProperties 的任務定義參數 227
使用 EcsProperties 建立任務定義 266
ContainerProperties 與EcsProperties任務定義 266
AWS Batch APIs的一般變更 267
Amazon ECS 的多容器任務定義 268
Amazon EKS 的多容器任務定義 269
使用 EcsProperties 的 Reference： AWS Batch job 案例 269
使用 awslogs 日誌驅動程式 276
JobDefiniton 資料類型中的 AWS Batch awslogs 日誌驅動程式選項 276
在任務定義中指定日誌組態 279
指定敏感資料 280
使用 Secrets Manager 281
使用 Systems Manager 參數存放區 288
任務的私有登錄檔身分驗證 291
私有登錄檔身分驗證所需的 IAM 許可 292
教學課程：建立私有登錄檔身分驗證的秘密 293
Amazon EFS 磁碟區 294
Amazon EFS 磁碟區考量事項 295
使用 Amazon EFS 存取點 296

在任務定義中指定 Amazon EFS 檔案系統 296
任務定義範例 299
環境變數 299
參數替換 300
測試 GPU 功能 301
多節點平行任務 302
任務 304
教學課程：提交任務 304
服務任務 307
服務任務承載 308
提交服務任務 309
服務任務狀態 310
服務任務重試策略 311
監控佇列中的服務任務 314
終止服務任務 316
任務狀態 316
任務環境變數 318
自動化任務重試 320
任務相依性 321
任務逾時 321
Amazon EKS 任務 322
教學課程：將執行中的任務映射至 Pod 和節點 323
教學課程：將執行中的 Pod 映射回其任務 324
多節點平行任務 326
環境變數 326
節點群組 327
任務生命週期 328
運算環境考量 328
Amazon EKS 上的多節點平行任務 329
執行 MNP 任務 329
建立 Amazon EKS MNP 任務定義 331
提交 Amazon EKS MNP 任務 333
覆寫 Amazon EKS MNP 任務定義 333
陣列任務 334
陣列任務工作流程的範例 336
使用陣列任務索引 339

執行 GPU 任務 345
在 Amazon EKS 上建立 GPU 型Kubernetes叢集 349
建立 Amazon EKS GPU 任務定義 351
在 Amazon EKS 叢集中執行 GPU 任務 352
檢視任務佇列中的任務 353
搜尋任務佇列中的任務 353
搜尋 AWS Batch 任務AWS （主控台） 354
搜尋和篩選 AWS Batch 任務 (AWS CLI) 355
AWS Batch 任務的網路模式 356
在 CloudWatch Logs 中檢視任務日誌 357
檢閱 AWS Batch 任務資訊 358
中的安全性 AWS Batch 360
身分和存取權管理 360
目標對象 361
使用身分驗證 361
使用政策管理存取權 362
AWS Batch 如何使用 IAM 363
身分型政策範例 367
AWS 受管政策 370
IAM 政策、角色和許可 374
IAM 政策結構 374
資源：範例政策 377
Resource： AWS Batch managed 政策 387
AWS Batch IAM 執行角色 387
支援的資源層級許可 388
教學課程：建立 IAM 執行角色 388
教學課程：檢查 IAM 執行角色 389
使用服務連結角色 390
Amazon ECS 執行個體角色 401
Amazon EC2 Spot 機群角色 404
EventBridge IAM 角色 407
建立 Virtual Private Cloud 408
建立 VPC 409
後續步驟 409
VPC 端點 410
考量事項 410

建立介面端點 411
建立端點政策 412
法規遵循驗證 413
基礎設施安全性 413
預防跨服務混淆代理人 413
範例：僅存取一個運算環境的角色 414
範例：存取多個運算環境的角色 415
CloudTrail 416
AWS Batch CloudTrail 中的資訊 416
參考：了解 AWS Batch 日誌檔案項目 417
IAM AWS Batch 故障診斷 419
我未獲授權，不得在 AWS Batch中執行動作 419
我未獲得執行 iam:PassRole 的授權 419
我想要允許 AWS 帳戶外的人員存取我的 AWS Batch 資源 420
AWS 步驟函數 421
教學課程：檢視狀態機器詳細資訊 421
教學課程：編輯狀態機器 422
教學課程：執行狀態機器 422
Amazon EventBridge 423
AWS Batch 事件 423
任務狀態變更事件 424
任務佇列封鎖事件 425
服務任務狀態變更事件 427
服務任務佇列封鎖事件 429
教學課程：搭配 使用 AWS 使用者通知 AWS Batch 430
AWS Batch 任務做為 EventBridge 目標 430
教學課程：建立排程任務 431
教學課程：建立具有事件模式的規則 433
教學課程：通過輸入轉換器 435
教學課程：聆聽 AWS Batch 任務事件 438
先決條件 438
教學課程：建立 Lambda 函數 438
教學課程：註冊事件規則 439
教學課程：測試您的組態 441
教學課程：針對失敗的任務事件傳送 Amazon Simple Notification Service 提醒 441
先決條件 441

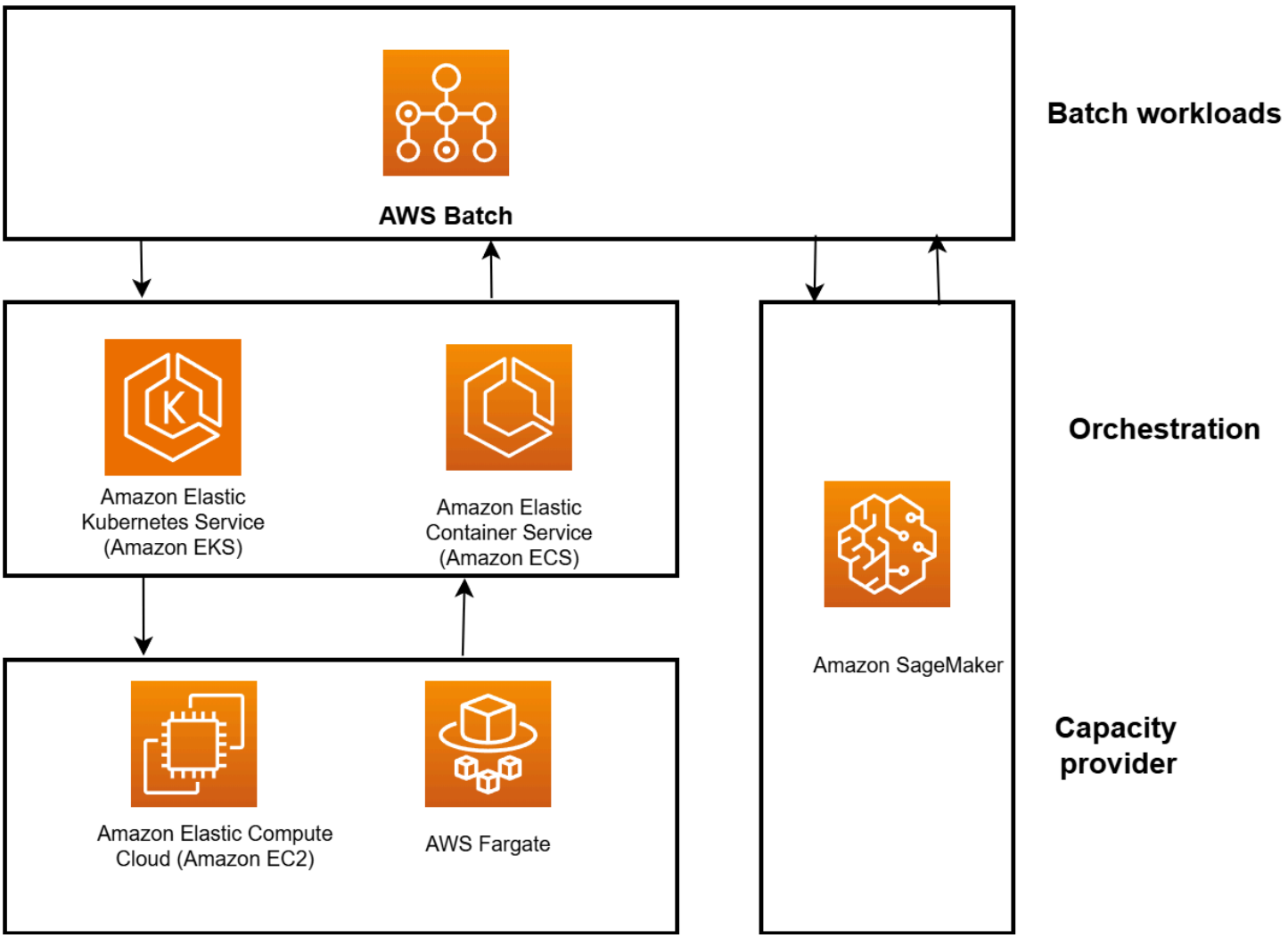
教學課程：建立和訂閱 Amazon SNS 主題 442
教學課程：註冊事件規則 442
教學課程：測試您的規則 444
替代規則：批次任務佇列已封鎖 444
Elastic Fabric Adapter 446
監控 AWS Batch 448
CloudWatch Logs 448
教學課程：新增 CloudWatch Logs IAM 政策 449
安裝及設定 CloudWatch 代理程式 451
教學課程：檢視 CloudWatch Logs 451
CloudWatch 容器洞見 453
開啟容器洞見 453
使用 CloudWatch Logs 監控 AWS Batch Amazon EKS 任務 454
先決條件 455
安裝 附加元件 455
標記您的 資源 456
標籤基本概念 456
標記您的 資源 456
標籤限制 457
教學課程：使用主控台管理標籤 458
建立時在個別資源上新增標籤 458
新增和刪除個別資源上的標籤 458
使用 CLI 或 API 管理標籤 459
最佳實務 461
使用時機 AWS Batch 461
大規模執行的檢查清單 461
最佳化容器和 AMIs 462
選擇正確的運算環境資源 463
Amazon EC2 隨需或 Amazon EC2 Spot 464
使用適用於 的 Amazon EC2 Spot 最佳實務 AWS Batch 465
常見錯誤和故障診斷 466
疑難排解 469
AWS Batch 470
接收自動執行個體系列更新的最佳執行個體類型組態 470
INVALID 運算環境 471
任務停滯在 RUNNABLE 狀態 473

建立時未標記 Spot 執行個體 477
Spot 執行個體未縮減規模 478
無法擷取 Secrets Manager 秘密 479
無法覆寫任務定義資源需求 479
更新desiredvCpus設定時的錯誤訊息 481
AWS Batch 在 Amazon EKS 上 481
INVALID 運算環境 481
AWS Batch Amazon EKS 任務上的 卡在 RUNNABLE 狀態 484
AWS Batch Amazon EKS 任務上的 卡在 STARTING 狀態 485
確認 aws-auth ConfigMap 已正確設定 486
RBAC 許可或繫結未正確設定 487
資源：服務配額 489
文件歷史紀錄 491
..... cdxcvi

什麼是 AWS Batch？

AWS Batch 可協助您在 上執行批次運算工作負載 AWS 雲端。批次運算是開發人員、科學家和工程師存取大量運算資源的常見方式。 AWS Batch 免除了設定和管理所需基礎設施的繁重工作，類似於傳統的批次運算軟體。這項服務可以快速佈建資源，回應提交的任務，以便消除容量限制，降低運算成本，進而加快結果產生。

作為全受管服務， AWS Batch 可協助您執行任何規模的批次運算工作負載。 AWS Batch 會自動佈建運算資源，並根據工作負載的數量和規模來最佳化工作負載分佈。透過 AWS Batch，您不需要安裝或管理批次運算軟體，因此您可以專注於分析結果和解決問題。

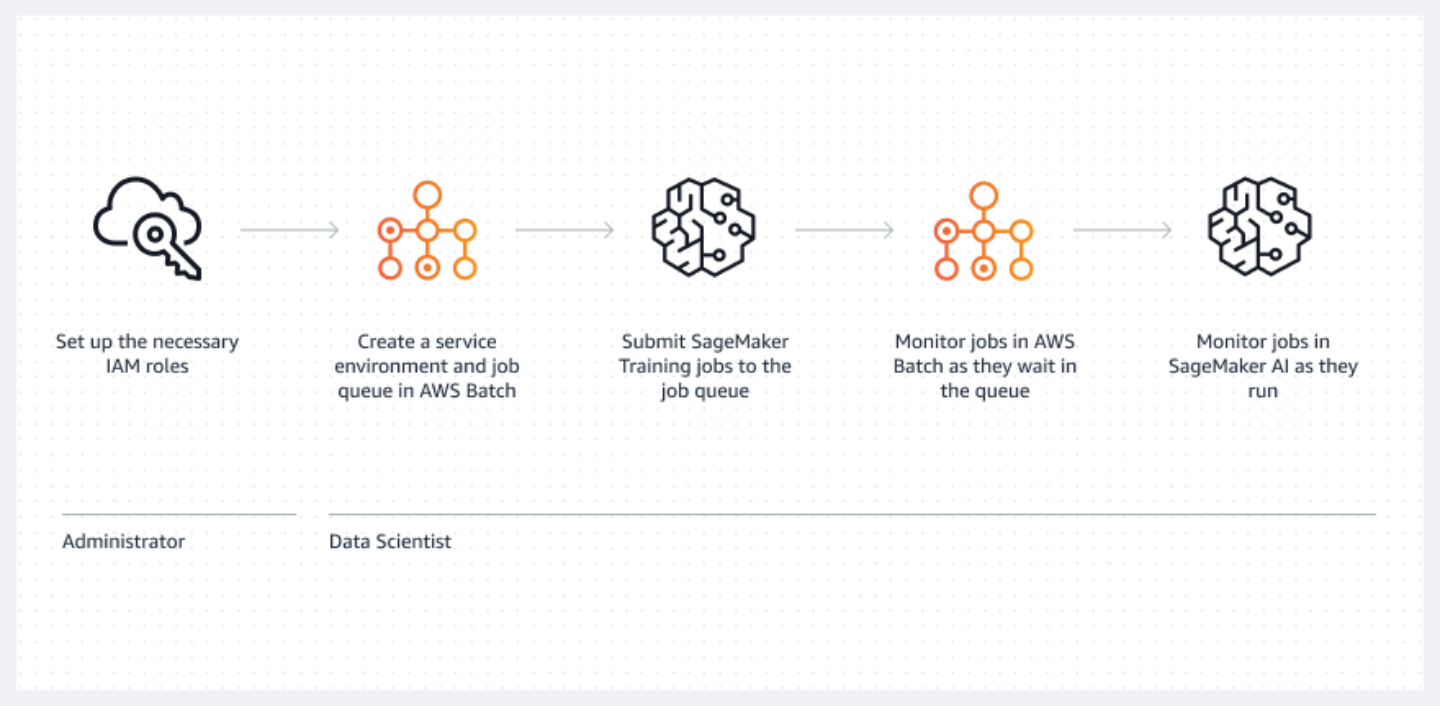


AWS Batch 提供所有必要功能，以在 AWS 受管容器協同運作服務上執行大規模的運算密集型工作負載。Amazon ECS 和 Amazon EKS AWS Batch 能夠在 Amazon EC2 執行個體和 Fargate 資源上擴展運算容量。

AWS Batch 為批次工作負載提供全受管服務，並提供操作功能，以最佳化這些工作負載類型的輸送量、速度、資源效率和成本。

AWS Batch 也啟用 SageMaker Training 任務佇列，允許資料科學家和 ML 工程師將具有優先順序的訓練任務提交至可設定的佇列。此功能可確保 ML 工作負載在資源可用時立即自動執行，無需手動協調並改善資源使用率。

對於機器學習工作負載， 為 SageMaker Training 任務 AWS Batch 提供佇列功能。您可以使用特定政策來設定佇列，以最佳化 ML Training 工作負載的成本、效能和資源配置。



這提供了一個共同的責任模型，其中管理員設定基礎設施和許可，而資料科學家可以專注於提交和監控其 ML 訓練工作負載。任務會根據設定的優先順序和資源可用性自動排入佇列並執行。

您是第一次 AWS Batch 使用嗎？

如果您是第一次使用 AWS Batch，建議您先閱讀以下章節：

- 的元件 AWS Batch
- 建立 IAM 帳戶和管理使用者
- 設定 AWS Batch
- AWS Batch 教學課程入門
- SageMaker AI AWS Batch 入門

相關服務

AWS Batch 是一項全受管批次運算服務，可在 AWS Amazon ECS、Amazon EKS 和 Spot 或隨需執行個體等各種運算產品中，規劃、排程及執行容器化批次 ML AWS Fargate、模擬和分析工作負載。如需每個受管運算服務的詳細資訊，請參閱：

- Amazon EC2 使用者指南
- 《AWS Fargate 開發人員指南》
- Amazon EKS 使用者指南
- 《Amazon SageMaker AI 開發人員指南》

存取 AWS Batch

您可以使用下列 AWS Batch 方式存取 ：

AWS Batch 主控台

您建立和管理 資源的 Web 界面。

AWS Command Line Interface

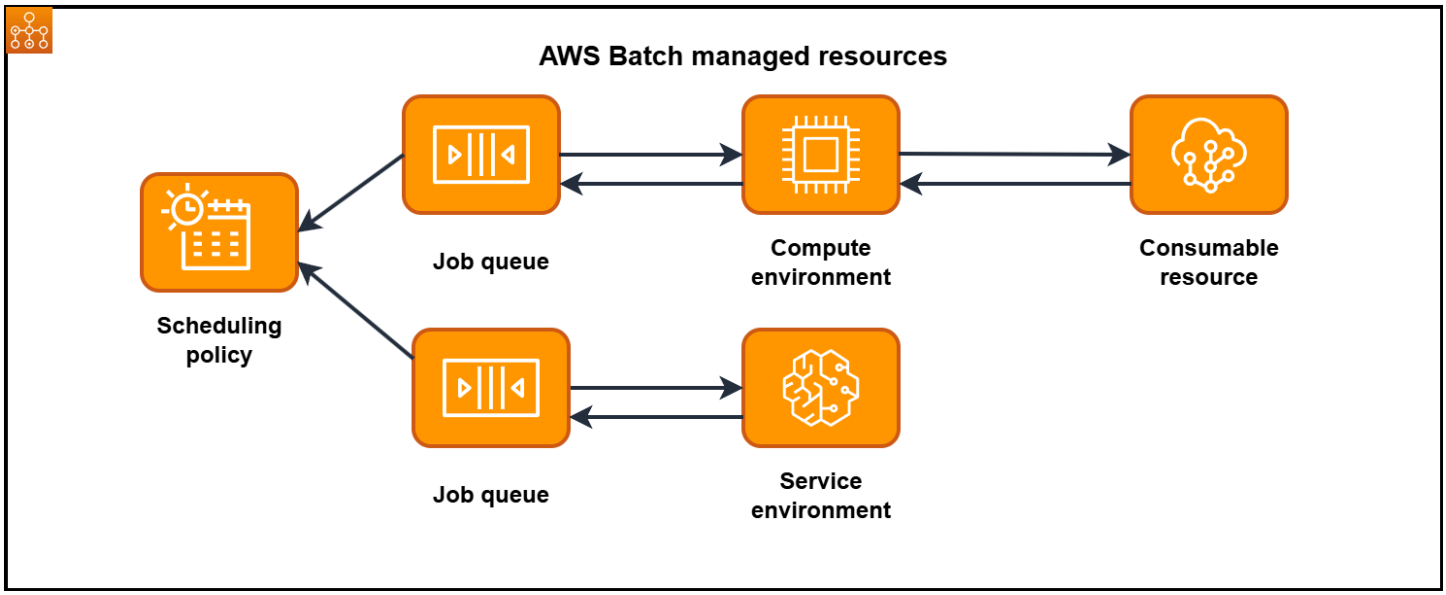
在 AWS 服務 命令列 Shell 中使用命令與 互動。Windows、macOS 和 Linux AWS Command Line Interface 支援 。如需 的詳細資訊 AWS CLI，請參閱 AWS Command Line Interface 使用者指南。您可以在 AWS Batch 命令AWS CLI 參考中找到命令。

AWS SDKs

如果您偏好使用特定語言 APIs 來建置應用程式，而不是透過 HTTP 或 HTTPS 提交請求，請使用 提供的程式庫、範例程式碼、教學課程和其他資源 AWS。這些程式庫提供基本函數來自動化任務，例如以密碼編譯方式簽署您的請求、重試請求，以及處理錯誤回應。這些函數可讓您更有效率地開始使用。如需詳細資訊，請參閱要建置的工具 AWS。

的元件 AWS Batch

AWS Batch 簡化跨區域內多個可用區域的執行中批次任務。您可以在新的或現有的 VPC 中建立 AWS Batch 運算環境。在運算環境設置完畢並與任務佇列關聯後，您可以定義任務定義，即指定由哪個 Docker 容器映像來執行您的任務。容器映像是從容器登錄檔儲存和提取，可能來自您的 AWS 基礎設施的內部或外部。



運算環境

運算環境是一組受管的或未受管的運算資源，用於執行任務。使用 受管運算環境，您可以在多個細節層級指定所需的運算類型 (Fargate 或 EC2)。您可以設定使用特定 EC2 執行個體類型的運算環境，例如 `c5.2xlarge`或 。 `m5.10xlarge`或者，您可以選擇只指定您想要使用最新的執行個體類型。您也可以指定環境的最小、所需和最大 vCPUs 數量，以及您願意為 Spot 執行個體支付的金額，以隨需執行個體價格和一組 VPC 子網路的目標百分比表示。 會視需要 AWS Batch 有效率地啟動、管理和終止運算類型。您也可以管理自己的運算環境。因此，您有責任在為您 AWS Batch 建立的 Amazon ECS 叢集中設定和擴展執行個體。如需詳細資訊，請參閱的運算環境 AWS Batch。

任務佇列

當您提交 AWS Batch 任務時，您會將其提交至特定任務佇列，在任務排程到運算環境之前，該任務位於該佇列中。您可以將一或多個運算環境與任務佇列建立關聯。您也可以指派這些運算環境的優先順序值，甚至是跨任務佇列本身。例如，您可以有一個提交時間敏感任務的高優先順序佇列，以及當運算資源較便宜時隨時可執行之任務的低優先順序佇列。如需詳細資訊，請參閱任務佇列。

任務定義

任務定義會指定任務的執行方式。您可以將任務定義視為任務中資源的藍圖。您可以為任務提供 IAM 角色，以提供對其他 AWS 資源的存取權。您也可以同時指定記憶體和 CPU 需求。任務定義也可以為持久性儲存控制容器屬性、環境變數和掛載點。當提交單一任務時，在任務定義中的許多規格，可以指定新的值予以覆寫。如需詳細資訊，請參閱任務定義

任務

您提交到 AWS Batch的工作單位 (如 shell 指令碼、Linux 可執行檔，或 Docker 容器映像)。它具有名稱，並使用您在任務定義中指定的參數，在運算環境中的 AWS Fargate 或 Amazon EC2 資源上做為容器化應用程式執行。任務可以依名稱或 ID 參考其他任務，並且可以取決於其他任務成功完成或您指定的資源可用性。如需詳細資訊，請參閱任務。

排程政策

您可以使用排程政策來設定如何在使用者或工作負載之間配置任務佇列中的運算資源。使用公平共用排程政策，您可以將不同的共用識別符指派給工作負載或使用者。 AWS Batch 任務排程器預設為先進先出 (FIFO) 策略。如需詳細資訊，請參閱公平共用排程政策。

消耗性資源

消耗性資源是執行任務所需的資源，例如第三方授權字符、資料庫存取頻寬、調節對第三方 API 的呼叫等。您可以指定執行任務所需的消耗性資源，而 Batch 在排程任務時會將這些資源相依性納入考量。您可以只配置具有所有必要資源的任務，以減少運算資源的使用不足。如需詳細資訊，請參閱資源感知排程。

服務環境

服務環境定義 如何與 SageMaker AWS Batch 整合以執行任務。服務環境 AWS Batch 可讓 在 SageMaker 上提交和管理任務，同時提供 的佇列、排程和優先順序管理功能 AWS Batch。服務環境會定義特定服務類型的容量限制，例如 SageMaker Training 任務。容量限制控制 環境中服務任務可以使用的最大資源。如需詳細資訊，請參閱的服務環境 AWS Batch。

服務任務

服務任務是您提交 AWS Batch 至 在服務環境中執行的工作單位。服務任務會利用 AWS Batch的佇列和排程功能，同時將實際執行委派給外部服務。例如，做為服務任務提交的 SageMaker Training 任務會排入佇列並排定優先順序 AWS Batch，但 SageMaker Training 任務執行會在 SageMaker AI 基礎設施內進行。此整合可讓資料科學家和機器學習工程師受益於其 SageMaker AI Training 工作負載 AWS Batch的自動化工作負載管理和優先順序佇列。服務任務可以依名稱或 ID 參考其他任務，並支援任務相依性。如需詳細資訊，請參閱中的服務任務 AWS Batch。

設定 AWS Batch

如果您已經註冊 Amazon Web Services (AWS) 並使用 Amazon Elastic Compute Cloud (Amazon EC2) 或 Amazon Elastic Container Service (Amazon ECS)，您很快就可以使用 AWS Batch。這些服務的設定程序類似。這是因為 在其運算環境中 AWS Batch 使用 Amazon ECS 容器執行個體。若要 AWS CLI 搭配 使用 AWS Batch，您必須使用支援最新 AWS Batch 功能的 AWS CLI 版本。如果您在 中沒有看到 AWS Batch 功能的支援 AWS CLI，請升級到最新版本。如需詳細資訊，請參閱 https：//http://aws.amazon.com/cli/。

> **Note**
>
> 由於 AWS Batch 使用 Amazon EC2 的元件，因此您會將 Amazon EC2 主控台用於許多這些步驟。

完成下列任務以設定 AWS Batch。

主題

- 建立 IAM 帳戶和管理使用者
- 為您的運算環境和容器執行個體建立 IAM 角色
- 為您的執行個體建立金鑰對
- 建立 VPC
- 建立安全群組
- 安裝 AWS CLI

建立 IAM 帳戶和管理使用者

若要開始使用，您需要建立 AWS 帳戶和通常獲得管理權限的單一使用者。若要完成此操作，請完成下列教學課程：

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 管理主控台身分登入 。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 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。

為您的運算環境和容器執行個體建立 IAM 角色

您的 AWS Batch 運算環境和容器執行個體需要 AWS 帳戶 登入資料，才能代表您呼叫其他 AWS APIs。建立 AWS Identity and Access Management 角色，將這些登入資料提供給運算環境和容器執行個體，然後將該角色與您的運算環境建立關聯。

> **Note**
>
> 若要確認您的 AWS 帳戶 具有必要的許可，請參閱您帳戶的初始 IAM 服務設定。
> 在主控台初次執行體驗中，會自動為您建立 AWS Batch 運算環境和容器執行個體角色。因此，如果您想要使用 AWS Batch 主控台，您可以繼續進行下一節。如果您打算 AWS CLI 改用 ，請先完成 使用 AWS Batch的服務連結角色、 和 中的程序Amazon ECS 執行個體角色，教學課程：建立 IAM 執行角色再建立您的第一個運算環境。

為您的執行個體建立金鑰對

AWS 使用公有金鑰密碼編譯來保護執行個體的登入資訊。Linux 執行個體，例如 AWS Batch 運算環境容器執行個體，沒有用於 SSH 存取的密碼。您需要使用金鑰對，以安全地登入執行個體。當建立運算環境時，您會先指定金鑰對名稱，然後再提供使用 SSH 登入時的私有金鑰。

如果您尚未建立金鑰對，您可以使用 Amazon EC2 主控台建立金鑰對。請注意，如果您計劃在多個 中啟動執行個體 AWS 區域，請在每個區域中建立金鑰對。如需區域的詳細資訊，請參閱《Amazon EC2 使用者指南》中的區域和可用區域。

建立一組金鑰對

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 從導覽列中， AWS 區域 選取金鑰對的 。您可以選取任何可供您使用的區域，無論您的位置為何：不過，金鑰對專屬於區域。例如，如果您計劃在美國西部（奧勒岡） 區域中啟動執行個體，請為相同區域中的執行個體建立金鑰對。
3. 在導覽窗格中，選擇 Key Pairs (金鑰對)、Create Key Pair (建立金鑰對)。
4. 在 Create Key Pair (建立金鑰對) 對話方塊中，在 Key pair name (金鑰對名稱) 為新的金鑰對輸入名稱，然後選擇 Create (建立)。選擇您可以記住的名稱，例如您的使用者名稱，後面接著 `-key-pair`，加上區域名稱。例如，me-key-pair-uswest2。
5. 您的瀏覽器會自動下載私有金鑰檔案。基礎檔案名稱為您所指定的金鑰對名稱，副檔名為 `.pem`。將私有金鑰檔案存放在安全的地方。

   > **Important**
   >
   > 這是您儲存私有金鑰檔案的唯一機會。每次您連線到執行個體時，都需要在啟動執行個體時提供金鑰對的名稱，以及對應的私有金鑰。

6. 如果您在 Mac 或 Linux 電腦上使用 SSH 用戶端連線到 Linux 執行個體，請使用下列命令來設定私有金鑰檔案的許可。如此一來，只有您可以讀取它。

   ```
   $ chmod 400 your_user_name-key-pair-region_name.pem
   ```

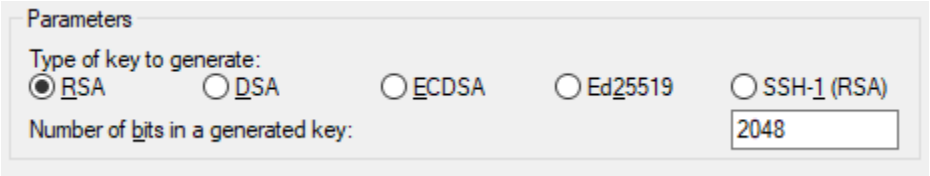
如需詳細資訊，請參閱《Amazon EC2 使用者指南》中的 Amazon EC2 金鑰對。 Amazon EC2

使用金鑰對連線至執行個體

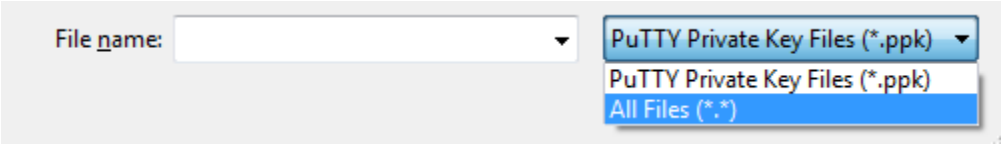
若要從執行 Mac 或 Linux 的電腦連線至您的 Linux 執行個體，請使用 `.pem` 選項與您私有金鑰的路徑，將 `-i` 檔案指定給您的 SSH 用戶端。若要從執行 Windows 的電腦連線至 Linux 執行個體，請使用 MindTerm 或 PuTTY。如果您計劃使用 PuTTY，請安裝它並使用下列程序將`.pem`檔案轉換為`.ppk`檔案。

（選用）準備使用 PuTTY 從 Windows 連線至 Linux 執行個體

1. 從 http://www.chiark.greenend.org.uk/~sgtatham/putty/ 下載並安裝 PuTTY。務必安裝整個套件。
2. 啟動 PuTTYgen（例如，從開始功能表中，選擇所有程式、PuTTY 和 PuTTYgen)。
3. 在 Type of key to generate (要產生的金鑰類型) 下，選擇 RSA (SSH-2 RSA)。如果您使用的是舊版 PuTTYgen，請選擇 SSH-2 RSA。



4. 選擇 Load (載入)。根據預設，PuTTYgen 只會顯示副檔名為 `.ppk` 的檔案。若要尋找您的 `.pem` 檔案，請選擇顯示所有類型之檔案的選項。



5. 選取您在先前程序中建立的私有金鑰檔案，然後選擇 Open (開啟)。選擇 OK (確定) 關閉確認對話方塊。
6. 選擇 Save private key (儲存私有金鑰)。PuTTYgen 會顯示有關儲存沒有密碼短語之金鑰的警告。選擇 Yes (是)。
7. 為您用於金鑰對的金鑰指定相同名稱。PuTTY 會自動新增 `.ppk` 副檔名。

建立 VPC

使用 Amazon Virtual Private Cloud (Amazon VPC)，您可以在您定義的虛擬網路中啟動 AWS 資源。我們強烈建議您在 VPC 中啟動容器執行個體。

如果您有預設 VPC，您也可以略過本節並移至下一個任務 建立安全群組。若要判斷您是否擁有預設 VPC，請參閱《Amazon EC2 使用者指南》中的 Amazon EC2 主控台中支援的平台 Amazon EC2

如需有關如何建立 Amazon VPC 的資訊，請參閱《Amazon VPC 使用者指南》中的僅建立 VPC。請參閱下表以決定要選取的選項。

選項	Value	
要建立的資源	僅 VPC	

選項	Value
名稱	可以選擇為 VPC 提供名稱。
IPv4 CIDR 區塊	IPv4 CIDR 手動輸入 CIDR 區塊大小必須為介於 /16 和 /28 之間的大小。
IPv6 CIDR 區塊	無 IPv6 CIDR 區塊
租用	預設

如需有關 Amazon VPC 的詳細資訊，請參閱《Amazon VPC 使用者指南》中的什麼是 Amazon VPC？。

建立安全群組

安全群組就像是防火牆，用於關聯的運算環境容器執行個體，可在容器執行個體層級控制傳入及傳出流量。安全群組只能在建立該群組的 VPC 中使用。

您可以新增規則至安全群組，讓您從您的 IP 地址使用 SSH 連接到您的容器執行個體。您也可以新增允許任何位置之傳入和傳出 HTTP 和 HTTPS 存取的規則。依您的任務所需在開放連接埠新增任何規則。

請注意，如果您計劃在多個區域中啟動容器執行個體，則需要在每個區域中建立安全群組。如需詳細資訊，請參閱《Amazon EC2 使用者指南》中的區域與可用區域。

> **Note**
>
> 您需要本機電腦的公有 IP 地址 (可以使用服務來取得)。例如，我們提供下列服務：http://checkip.amazonaws.com/ 或 https://checkip.amazonaws.com/。若要尋找其他能夠提供您 IP 地址的服務，請使用搜尋片語 "what is my IP address" (我的 IP 地址為何)。如果您是透過網際網路服務供應商 (ISP) 或從沒有靜態 IP 地址的防火牆後方連線，請了解用戶端電腦使用的 IP 地址範圍。

使用主控台建立安全群組

1. 在 https://console.aws.amazon.com/vpc/ 開啟 Amazon VPC 主控台。
2. 在導覽窗格中，選擇 Security Groups (安全群組)。
3. 選擇 Create Security Group (建立安全群組)。
4. 輸入安全群組的名稱和說明。您無法在建立安全群組之後變更安全群組的名稱和說明。
5. 從 VPC 中選擇 VPC。
6. （選用）根據預設，新的安全群組只會從允許所有流量離開資源的傳出規則開始。您必須新增規則啟用任何傳入流量，或是限制傳出流量。

   AWS Batch 容器執行個體不需要開啟任何傳入連接埠。不過，您可能想要新增 SSH 規則。如此一來，您就可以登入容器執行個體，並使用 Docker 命令檢查任務中的容器。如果您希望容器執行個體託管執行 Web 伺服器的任務，您也可以新增 HTTP 的規則。完成下列步驟，以新增這些選用的安全群組規則。

   在 Inbound (內送) 標籤，建立以下規則然後選擇 Create (建立)：

   - 選擇 Add Rule (新增規則)。針對 Type (類型)，選擇 HTTP。針對 Source (來源)，選擇 Anywhere (隨處) (`0.0.0.0/0`)。
   - 選擇 Add Rule (新增規則)。針對 Type (類型)，選擇 SSH。針對來源，選擇自訂 IP，並以無類別網域間路由 (CIDR) 表示法指定電腦或網路的公有 IP 地址。如果您的公司會分配某個範圍的地址，請指定整個範圍 (例如 `203.0.113.0/24`)。若要在 CIDR 表示法中指定個別 IP 地址，請選擇我的 IP。這會將路由字首新增至公/32有 IP 地址。

   > **Note**
   >
   > 基於安全考量，我們不建議您允許從所有 IP 地址 (`0.0.0.0/0`) 到執行個體的 SSH 存取，但僅用於測試目的，且僅限於短時間內。

7. 您可以立即新增標籤，也可以稍後再新增。若要新增標籤，請選擇 Add new tag (新增標籤)，然後輸入標籤的索引鍵和值。
8. 選擇 Create Security Group (建立安全群組)。

若要使用命令列建立安全群組，請參閱 >create-security-group (AWS CLI)

如需安全群組的詳細資訊，請參閱使用安全群組。

安裝 AWS CLI

若要 AWS CLI 搭配 使用 AWS Batch，請安裝 AWS CLI 最新版本。如需安裝 AWS CLI 或將其升級至最新版本的詳細資訊，請參閱AWS Command Line Interface 《 使用者指南》中的安裝 AWS 命令列界面。

AWS Batch 教學課程入門

您可以使用 AWS Batch 初次執行精靈快速開始使用 AWS Batch。完成先決條件後，您可以使用初次執行精靈來建立運算環境、任務定義和任務佇列。

您也可以使用 AWS Batch 初次執行精靈來測試您的組態，提交範例「Hello World」任務。如果您已有要在其中啟動的 Docker 映像 AWS Batch，您可以使用該映像來建立任務定義。

之後，您可以使用 AWS Batch 初次執行精靈來建立運算環境、任務佇列，並提交範例 Hello World 任務。

使用 精靈開始使用 Amazon EC2 協同運作

Amazon Elastic Compute Cloud (Amazon EC2) – 在 AWS 雲端中提供可擴展的運算容量。使用 Amazon EC2 可減少前期所需的硬體投資，讓您更快速開發並部署應用程式。

您可使用 Amazon EC2 按需要啟動任意數量的虛擬伺服器，設定安全性和聯網功能以及管理儲存。使用 Amazon EC2 可擴展與縮減規模，以處理需求或熱門峰值的變更，從而降低您預測流量的需求。

概觀

本教學課程示範如何使用 AWS Batch 精靈設定 以設定 Amazon EC2 並執行 `Hello World`。

目標對象

本教學課程專為負責設定、測試和部署的系統管理員和開發人員而設計 AWS Batch。

使用的功能

本教學課程說明如何使用 AWS Batch 主控台精靈：

- 建立和設定 Amazon EC2 運算環境
- 建立任務佇列。
- 建立任務定義
- 建立並提交要執行的任務
- 在 CloudWatch 中檢視任務的輸出

所需時間

完成本教學課程大約需要 10-15 分鐘。

區域限制

使用此解決方案沒有相關聯的國家或地區限制。

資源用量成本

建立 AWS 帳戶無需付費。不過，透過實作此解決方案，您可能會產生下表中所列的部分或全部費用。

描述	費用 (美元)
Amazon EC2 執行個體	您需為建立的每個 Amazon EC2 執行個體付費。如需定價的相關資訊，請參閱 Amazon EC2 定價。

先決條件

開始之前：

- 如果您沒有 AWS 帳戶 ，請建立 。
- 建立ecsInstanceRole執行個體角色。

步驟 1：建立運算環境

> **Important**
>
> 為了盡可能簡單快速地開始使用，本教學課程包含具有預設設定的步驟。為生產用途建立 之前，建議您先熟悉所有設定，並使用符合您需求的設定進行部署。

若要為 Amazon EC2 協同運作建立運算環境，請執行下列動作：

1. 開啟AWS Batch 主控台初次執行精靈。
2. 針對設定任務和協同運作類型，選擇 Amazon Elastic Compute Cloud(Amazon EC2)。
3. 選擇下一步。
4. 在名稱的運算環境組態區段中，指定運算環境的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。

5. 針對執行個體角色，選擇已連接必要 IAM 許可的現有執行個體角色。此執行個體角色可讓運算環境中的 Amazon ECS 容器執行個體呼叫所需的 AWS API 操作。如需詳細資訊，請參閱Amazon ECS 執行個體角色。

   執行個體角色的預設名稱為 `ecsInstanceRole`。

6. 對於執行個體組態，您可以保留預設設定。
7. 對於網路組態，請使用 的預設 VPC AWS 區域。
8. 選擇下一步。

步驟 2：建立任務佇列

任務佇列會儲存您提交的任務，直到 AWS Batch 排程器在運算環境中的資源上執行任務為止。如需詳細資訊，請參閱任務佇列

若要為 Amazon EC2 協同運作建立任務佇列，請執行下列動作：

1. 針對名稱的任務佇列組態，指定任務佇列的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
2. 對於所有其他組態選項，您可以保留預設值。
3. 選擇下一步。

步驟 3：建立任務定義

AWS Batch 任務定義會指定任務的執行方式。即使每個任務都必須參考任務定義，任務定義中指定的許多參數都可以在執行時間覆寫。

若要建立任務定義：

1. 對於建立任務定義

   a. 針對名稱，指定任務佇列的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。

   b. 對於命令 - 選用，您可以`hello world`變更為自訂訊息或保持原狀。

2. 對於所有其他組態選項，您可以保留預設值。
3. 選擇下一步。

步驟 4：建立任務

若要建立任務，請執行下列動作：

1. 在名稱的任務組態區段中，指定任務的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
2. 對於所有其他組態選項，您可以保留預設值。
3. 選擇下一步。

步驟 4：檢閱和建立

在檢閱和建立頁面上，檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立資源。

1. 針對檢閱和建立，選擇建立資源。
2. 視窗會在 AWS Batch 開始配置您的 資源時開啟。完成後，請選擇前往儀表板。在儀表板上，您應該會看到所有已配置的資源，以及任務處於 `Runnable` 狀態。您的任務已排定執行，應該會在 2-3 個小節中完成。

步驟 6：檢視任務的輸出

若要檢視任務的輸出，請執行下列動作：

1. 在導覽窗格中，選擇任務。
2. 在任務佇列下拉式清單中，選擇您為教學課程建立的任務佇列。
3. 任務表格列出您的所有任務及其目前狀態。一旦任務狀態成功，請選擇任務名稱以檢視任務的詳細資訊。
4. 在詳細資訊窗格中，選擇日誌串流名稱。任務的 CloudWatch 主控台將開啟，並且應該有一個事件具有 Message of `hello world`或您的自訂訊息。

步驟 7：清除您的教學課程資源

您需支付啟用 Amazon EC2 執行個體的費用。您可以刪除執行個體以停止產生費用。

若要刪除您建立的資源，請執行下列動作：

1. 在導覽窗格中，選擇任務佇列。

2. 在任務佇列表格中，選擇您為教學課程建立的任務佇列。
3. 選擇停用。任務佇列狀態停用後，您可以選擇刪除。
4. 刪除任務佇列後，在導覽窗格中選擇運算環境。
5. 選擇您為此教學課程建立的運算環境，然後選擇停用。運算環境可能需要 1-2 分鐘才能完成停用。
6. 一旦運算環境的狀態停用，請選擇刪除。可能需要 1-2 個小節才能刪除運算環境。

其他資源

完成教學課程後，您可能想要探索下列主題：

- 探索 AWS Batch 核心元件。如需詳細資訊，請參閱的元件 AWS Batch。
- 進一步了解 中可用的不同運算環境 AWS Batch。
- 進一步了解任務佇列及其不同的排程選項。
- 進一步了解任務定義和不同的組態選項。
- 進一步了解不同類型的任務。

使用精靈開始使用 AWS Batch 和 Fargate 協同運作

AWS Fargate 會啟動和擴展運算，以符合您為容器指定的資源需求。使用 Fargate，您不需要過度佈建或支付額外的伺服器。如需詳細資訊，請參閱 Fargate。

概觀

本教學課程示範如何使用 AWS Batch 精靈設定 ，以設定 AWS Fargate 並執行 `Hello World`。

目標對象

本教學課程專為負責設定、測試和部署的系統管理員和開發人員而設計 AWS Batch。

使用的功能

本教學課程說明如何使用 AWS Batch 主控台精靈：

- 建立和設定 AWS Fargate 運算環境
- 建立任務佇列。
- 建立任務定義
- 建立並提交要執行的任務

- 在 CloudWatch 中檢視任務的輸出

所需時間

完成本教學課程約需 10-15 分鐘。

區域限制

使用此解決方案沒有相關聯的國家或地區限制。

資源用量成本

建立 AWS 帳戶無需付費。不過，透過實作此解決方案，您可能會產生下表中所列的部分或全部費用。

Description	費用 (美元)
定價是以請求的 vCPU、記憶體、作業系統、CPU 架構和任務或 Pod 的儲存資源為基礎。	如需定價的詳細資訊，請參閱 Fargate 定價。

先決條件

開始之前：

- 如果您沒有 ， AWS 帳戶 請建立 。
- 建立任務執行角色。如果您尚未建立任務執行角色，則可以建立它做為本教學課程的一部分。

步驟 1：建立運算環境

> **Important**
>
> 為了盡可能簡單快速地開始使用，本教學課程包含具有預設設定的步驟。為生產用途建立 之前，建議您先熟悉所有設定，並使用符合您需求的設定進行部署。

若要為 Fargate 協同運作建立運算環境，請執行下列動作：

1. 開啟AWS Batch 主控台初次執行精靈。
2. 針對設定任務和協同運作類型，選擇 Fargate。

3. 選擇下一步。
4. 在名稱的運算環境組態區段中，指定運算環境的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
5. 對於所有其他組態選項，您可以保留預設值。
6. 選擇下一步。

步驟 2：建立任務佇列

任務佇列會儲存您提交的任務，直到 AWS Batch 排程器在運算環境中的資源上執行任務為止。若要建立任務佇列：

若要為 Fargate 協同運作建立任務佇列，請執行下列動作：

1. 在名稱的任務佇列組態區段中，指定運算環境的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
2. 針對優先順序，輸入任務佇列的 900。
3. 對於所有其他組態選項，您可以保留預設值。
4. 選擇下一步。

步驟 3：建立任務定義

若要建立任務定義：

1. 在一般組態區段中：
    - 在名稱的一般組態區段中，指定運算環境的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
2. 在 Fargate 平台組態區段中：
    a. 開啟指派公有 IP 以指派公有 IP 地址。您需要公有 IP 才能下載容器映像，除非您已設定私有映像儲存庫。
    b. 針對執行角色，選擇可讓 Amazon Elastic Container Service (Amazon ECS) 代理程式代您 AWS 呼叫的任務執行角色。選擇 ecsTaskExecutionRole 或 BatchEcsTaskExecutionRole。

        若要建立執行角色，請選擇建立執行角色。在建立 IAM 角色模態中，選擇建立 IAM 角色。

        i. IAM 主控台已設定用於建立執行角色的許可設定。

ii. 對於信任的實體類型，請確認已選取AWS 服務。

iii. 針對服務或使用者案例，確認已選取 Elastic Container Service。

iv. 選擇下一步。

v. 針對許可政策，確認已選取 AmazonECSTaskExecutionRolePolicy。

vi. 選擇下一步。

vii. 針對名稱，檢閱並建立 ，確認角色名稱為 BatchEcsTaskExecutionRole。

viii. 選擇建立角色。

ix. 在 AWS Batch 主控台中，選擇執行角色旁的重新整理按鈕。選擇 BatchEcsTaskExecutionRole 執行角色。

3. 在容器組態區段中：

   - 對於 命令，您可以`hello world`變更為自訂訊息或保持原狀。

4. 對於所有其他組態選項，您可以保留預設值。

5. 選擇下一步。

步驟 4：建立任務

若要建立 Fargate 任務，請執行下列動作：

1. 在名稱的任務組態區段中，指定任務的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
2. 對於所有其他組態選項，您可以保留預設值。
3. 選擇下一步。

步驟 4：檢閱和建立

在檢閱和建立頁面上，檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立資源。

步驟 6：檢視任務的輸出

若要檢視任務的輸出，請執行下列動作：

1. 在導覽窗格中，選擇任務。
2. 在任務佇列下拉式清單中，選擇您為教學課程建立的任務佇列。

3. 任務表格列出所有任務及其目前狀態。一旦任務狀態成功，請選擇任務名稱以檢視任務的詳細資訊。
4. 在詳細資訊窗格中，選擇日誌串流名稱。任務的 CloudWatch 主控台將開啟，並且應該有一個事件具有 的訊息`hello world`或您的自訂訊息。

步驟 7：清除您的教學課程資源

您需支付啟用 Amazon EC2 執行個體的費用。您可以刪除執行個體以停止產生費用。

若要刪除您建立的資源，請執行下列動作：

1. 在導覽窗格中，選擇任務佇列。
2. 在任務佇列表格中，選擇您為教學課程建立的任務佇列。
3. 選擇停用。任務佇列狀態停用後，您可以選擇刪除。
4. 刪除任務佇列後，在導覽窗格中選擇運算環境。
5. 選擇您為此教學課程建立的運算環境，然後選擇停用。運算環境可能需要 1-2 分鐘才能完成停用。
6. 一旦運算環境的狀態停用，請選擇刪除。可能需要 1-2 個小節才能刪除運算環境。

其他資源

完成教學課程後，您可能想要探索下列主題：

- 進一步了解 最佳實務。
- 探索 AWS Batch 核心元件。如需詳細資訊，請參閱的元件 AWS Batch。
- 進一步了解 中可用的不同運算環境 AWS Batch。
- 進一步了解任務佇列及其不同的排程選項。
- 進一步了解任務定義和不同的組態選項。
- 進一步了解不同類型的任務。

使用 開始使用 AWS Batch 和 Fargate AWS CLI

本教學課程示範如何使用 AWS Fargate 協調 AWS Batch 設定 ，並使用 AWS Command Line Interface () 執行簡單的「Hello World」任務AWS CLI。您將了解如何建立運算環境、任務佇列、任務定義，以及向其提交任務 AWS Batch。

主題

- 先決條件
- 建立 IAM 執行角色
- 建立運算環境
- 建立任務佇列
- 建立任務定義
- 提交和監控任務
- 檢視任務輸出
- 清除資源
- 生產環境部署須知
- 後續步驟

先決條件

開始本教學課程之前，請確定您有下列項目。

1. AWS CLI。若需安裝，請根據 AWS CLI 安裝指南進行操作。您也可以使用 AWS CloudShell ，其中包含 AWS CLI。
2. AWS CLI 使用適當的登入資料設定您的 。若尚未設定憑證，請執行 `aws configure`。
3. 基本熟悉命令列界面和容器化概念。
4. AWS Batch 如何使用 IAM 在 中建立和管理 AWS Batch 資源、IAM 角色和 VPC 資源 AWS 帳戶。
5. 來自 中 VPC 的子網路 ID 和安全群組 ID AWS 帳戶。如果您沒有 VPC，您可以建立一個。如需使用 AWS CLI 擷取這些資源 IDs的詳細資訊，請參閱《 AWS CLI 命令參考》中的 describe-subnets 和 describe-security-groups。

所需時間：完成本教學課程約需 15-20 分鐘。

成本：本教學課程使用 Fargate 運算資源。完成本教學課程的預估成本低於 0.01 USD，假設您遵循清除指示，在完成後立即刪除資源。Fargate 定價是以使用的 vCPU 和記憶體資源為基礎，最低收費為每秒 1 分鐘。如需目前定價資訊，請參閱 AWS Fargate 定價。

建立 IAM 執行角色

AWS Batch 需要一個執行角色，允許 Amazon Elastic Container Service (Amazon ECS) 代理程式代表您進行 AWS API 呼叫。Fargate 任務需要此角色，才能提取容器映像並將日誌寫入 Amazon CloudWatch。

建立信任政策文件

首先，建立允許 Amazon ECS 任務服務擔任角色的信任政策。

```
cat > batch-execution-role-trust-policy.json << EOF
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "ecs-tasks.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
EOF
```

建立執行角色

下列命令`BatchEcsTaskExecutionRoleTutorial`會使用您剛建立的信任政策來建立名為 的 IAM 角色。

```
aws iam create-role \
    --role-name BatchEcsTaskExecutionRoleTutorial \
    --assume-role-policy-document file://batch-execution-role-trust-policy.json
```

連接必要的政策

連接提供 Amazon ECS 任務執行必要許可的 AWS 受管政策。

```
aws iam attach-role-policy \
    --role-name BatchEcsTaskExecutionRoleTutorial \
```

```
    --policy-arn arn:aws:iam::aws:policy/service-role/AmazonECSTaskExecutionRolePolicy
```

角色現在已準備好供 AWS Batch 用於 Fargate 任務執行。

建立運算環境

運算環境會定義批次任務將執行的運算資源。在本教學課程中，您將建立受管 Fargate 運算環境，根據任務需求自動佈建和擴展資源。

建立運算環境

下列命令會建立 Fargate 運算環境。根據 將範例子網路和安全群組 IDs 取代為您自己的先決條件。

```
aws batch create-compute-environment \
    --compute-environment-name my-fargate-compute-env \
    --type MANAGED \
    --state ENABLED \
    --compute-resources type=FARGATE,maxvCpus=128,subnets=subnet-
a123456b,securityGroupIds=sg-a12b3456
```

以下顯示當命令成功執行時，輸出的外觀。

```
{
    "computeEnvironmentName": "my-fargate-compute-env",
    "computeEnvironmentArn": "arn:aws:batch:us-west-2:123456789012:compute-environment/
my-fargate-compute-env"
}
```

等待運算環境就緒

檢查運算環境的狀態，確保在繼續之前已準備就緒。

```
aws batch describe-compute-environments \
    --compute-environments my-fargate-compute-env \
    --query 'computeEnvironments[0].status'
```

```
"VALID"
```

當狀態顯示 時VALID，您的運算環境已準備好接受任務。

建立任務佇列

任務佇列會儲存提交的任務，直到 AWS Batch 排程器在運算環境中的資源上執行它們為止。任務會在佇列中依優先順序處理。

建立任務佇列

下列命令會建立優先順序為 900 的任務佇列，其使用您的 Fargate 運算環境。

```
aws batch create-job-queue \
    --job-queue-name my-fargate-job-queue \
    --state ENABLED \
    --priority 900 \
    --compute-environment-order order=1,computeEnvironment=my-fargate-compute-env
```

以下顯示當命令成功執行時，輸出的外觀。

```
{
    "jobQueueName": "my-fargate-job-queue",
    "jobQueueArn": "arn:aws:batch:us-west-2:123456789012:job-queue/my-fargate-job-
queue"
}
```

確認任務佇列已就緒

檢查您的任務佇列是否處於 ENABLED 狀態，並準備好接受任務。

```
aws batch describe-job-queues \
    --job-queues my-fargate-job-queue \
    --query 'jobQueues[0].state' "ENABLED"
```

建立任務定義

任務定義會指定任務的執行方式，包括要使用的 Docker 映像、資源需求和其他參數。對於 Fargate，您將使用資源需求，而不是傳統的 vCPU 和記憶體參數。

建立任務定義

下列命令會建立使用 busybox 容器映像執行簡單 "hello world" 命令的任務定義。123456789012 將取代為您實際 AWS 帳戶 的 ID，並將範例取代 AWS 區域 為您自己的 ID。

```
aws batch register-job-definition \
```

```
    --job-definition-name my-fargate-job-def \
    --type container \
    --platform-capabilities FARGATE \
    --container-properties '{
        "image": "busybox",
        "resourceRequirements": [
            {"type": "VCPU", "value": "0.25"},
            {"type": "MEMORY", "value": "512"}
        ],
        "command": ["echo", "hello world"],
        "networkConfiguration": {
            "assignPublicIp": "ENABLED"
        },
        "executionRoleArn": "arn:aws:iam::123456789012:role/
BatchEcsTaskExecutionRoleTutorial"
    },
{
    "jobDefinitionName": "my-fargate-job-def",
    "jobDefinitionArn": "arn:aws:batch:us-west-2:123456789012:job-definition/my-
fargate-job-def:1",
    "revision": 1
}'
```

任務定義指定 0.25 vCPU 和 512 MB 的記憶體，這是 Fargate 任務的最低資源。設定`assignPublicIp`已啟用，因此容器可以從 Docker Hub 提取忙碌方塊映像。

提交和監控任務

現在您已擁有所有必要的元件，您可以將任務提交至佇列並監控其進度。

提交任務

下列命令會使用您建立的任務定義，將任務提交到您的佇列。

```
aws batch submit-job \
    --job-name my-hello-world-job \
    --job-queue my-fargate-job-queue \
    --job-definition my-fargate-job-def
```

以下顯示當命令成功執行時，輸出的外觀。

```
{
    "jobArn": "arn:aws:batch:us-west-2:123456789012:job/my-hello-world-job",
```

```
    "jobName": "my-hello-world-job",
    "jobId": "1509xmpl-4224-4da6-9ba9-1d1acc96431a"
}
```

請記下回應中jobId傳回的 ，因為您會使用它來監控任務的進度。

監控任務狀態

使用任務 ID 來檢查任務的狀態。任務將繼續進行數個狀態：SUBMITTED、PENDING、RUNNING、、RUNNABLE STARTING和 最後 SUCCEEDED或 FAILED。

```
aws batch describe-jobs --jobs 1509xmpl-4224-4da6-9ba9-1d1acc96431a
```

以下顯示當命令成功執行時，輸出的外觀。

```
{
    "jobs": [
        {
            "jobArn": "arn:aws:batch:us-west-2:123456789012:job/my-hello-world-job",
            "jobName": "my-hello-world-job",
            "jobId": "1509xmpl-4224-4da6-9ba9-1d1acc96431a",
            "jobQueue": "arn:aws:batch:us-west-2:123456789012:job-queue/my-fargate-job-
queue",
            "status": "SUCCEEDED",
            "createdAt": 1705161908000,
            "jobDefinition": "arn:aws:batch:us-west-2:123456789012:job-definition/my-
fargate-job-def:1"
        }
    ]
}
```

當狀態顯示 時SUCCEEDED，您的任務已成功完成。

檢視任務輸出

任務完成後，您可以在 Amazon CloudWatch Logs 中檢視其輸出。

取得日誌串流名稱

首先，從任務詳細資訊擷取日誌串流名稱。將範例任務 ID 取代為您自己的任務 ID。

```
aws batch describe-jobs --jobs 1509xmpl-4224-4da6-9ba9-1d1acc96431a \
    --query 'jobs[0].attempts[0].containers[0].logStreamName' \
```

```
    --output text
```

```
my-fargate-job-def/default/1509xmpl-4224-4da6-9ba9-1d1acc96431a
```

檢視任務日誌

使用日誌串流名稱從 CloudWatch Logs 擷取任務的輸出。

```
aws logs get-log-events \
    --log-group-name /aws/batch/job \
    --log-stream-name my-fargate-job-def/default/1509xmpl-4224-4da6-9ba9-1d1acc96431a \
    --query 'events[*].message' \
    --output text
```

輸出會顯示「hello world」，確認您的任務已成功執行。

清除資源

為了避免持續收費，請清除您在本教學課程中建立的資源。由於相依性，您必須以正確的順序刪除資源。

停用和刪除任務佇列

首先，停用任務佇列，然後刪除它。

```
aws batch update-job-queue \
    --job-queue my-fargate-job-queue \
    --state DISABLED
```

```
aws batch delete-job-queue \
    --job-queue my-fargate-job-queue
```

停用和刪除運算環境

刪除任務佇列後，請停用並刪除運算環境。

```
aws batch update-compute-environment \
    --compute-environment my-fargate-compute-env \
    --state DISABLED
```

```
aws batch delete-compute-environment \
```

```
    --compute-environment my-fargate-compute-env
```

清除 IAM 角色

移除政策附件並刪除 IAM 角色。

```
aws iam detach-role-policy \
    --role-name BatchEcsTaskExecutionRoleTutorial \
    --policy-arn arn:aws:iam::aws:policy/service-role/AmazonECSTaskExecutionRolePolicy
```

```
aws iam delete-role \
    --role-name BatchEcsTaskExecutionRoleTutorial
```

移除暫存檔案

刪除您建立的信任政策檔案。

```
rm batch-execution-role-trust-policy.json
```

已成功清除所有資源。

生產環境部署須知

本教學課程旨在協助您了解 如何與 Fargate AWS Batch 搭配使用。對於生產部署，請考慮下列其他要求：

安全考量：

- 建立具有最少必要存取權的專用安全群組，而不是使用預設安全群組
- 使用私有子網路搭配 NAT Gateway，而非容器的公有 IP 指派
- 將容器映像儲存在 Amazon ECR 中，而不是使用公有儲存庫
- 實作 VPC 端點進行 AWS 服務通訊，以避免網際網路流量

架構考量事項：

- 跨多個可用區域部署以獲得高可用性
- 實作任務重試策略和無效字母佇列以進行錯誤處理
- 使用具有不同優先順序的多個任務佇列進行工作負載管理

- 根據佇列深度和資源使用率設定自動擴展政策
- 實作任務失敗和資源使用率的監控與提醒

操作考量：

- 設定 CloudWatch 儀表板和警示以進行監控
- 實作適當的記錄和稽核追蹤
- 針對基礎設施使用 CloudFormation 或 AWS CDK 做為程式碼
- 建立備份和災難復原程序

如需生產就緒架構的完整指引，請參閱 AWS Well-Architected Framework 和AWS 安全最佳實務。

後續步驟

現在您已完成本教學課程，您可以探索更進階 AWS Batch 的功能：

- 任務佇列 – 了解任務佇列排程和優先順序管理
- 任務定義 – 探索進階任務定義組態，包括環境變數、磁碟區和重試策略
- 的運算環境 AWS Batch – 了解不同的運算環境類型和擴展選項
- 多節點平行任務 – 執行跨越多個運算節點的任務
- 陣列任務 – 有效率地提交大量類似的任務
- 的最佳實務 AWS Batch – 了解生產工作負載的最佳化技術

在 Amazon EKS AWS Batch 上開始使用

AWS Batch on Amazon EKS 是一項受管服務，可將批次工作負載排程和擴展至現有的 Amazon EKS 叢集。 AWS Batch 不會代表您建立、管理或執行 Amazon EKS 叢集的生命週期操作。 AWS Batch 協調會擴展和縮減由 管理的節點 AWS Batch ，並在這些節點上執行 Pod。

AWS Batch 不會觸控與 Amazon EKS 叢集內 AWS Batch 運算環境無關的節點、自動擴展節點群組或 Pod 生命週期。為了 AWS Batch 讓 有效運作，其服務連結角色需要現有 Amazon EKS 叢集中Kubernetes的角色型存取控制 (RBAC) 許可。如需詳細資訊，請參閱 Kubernetes 文件中的使用 RBAC 授權。

AWS Batch 需要一個Kubernetes命名空間，其可將 Pod 範圍限定為 AWS Batch 任務。我們建議使用專用命名空間來隔離 AWS Batch Pod 與其他叢集工作負載。

在 AWS Batch 獲得 RBAC 存取權並建立命名空間之後，您可以使用 CreateComputeEnvironment API 操作將該 Amazon EKS 叢集與 AWS Batch 運算環境建立關聯。任務佇列可以與此新的 Amazon EKS 運算環境相關聯。任務 AWS Batch 會根據 Amazon EKS 任務定義，使用 SubmitJob API 操作提交至任務佇列。 AWS Batch 然後， 會啟動 AWS Batch 受管節點，並將任務從任務佇列做為 Pod Kubernetes 放入與 AWS Batch 運算環境相關聯的 EKS 叢集。

下列各節說明如何在 Amazon EKS AWS Batch 上設定 。

內容

- 概觀
- 先決條件
- 步驟 1：建立 的 Amazon EKS 叢集 AWS Batch
- 步驟 2：準備您的 Amazon EKS 叢集 AWS Batch
- 步驟 3：建立 Amazon EKS 運算環境
- 步驟 4：建立任務佇列並連接運算環境
- 步驟 5：建立任務定義
- 步驟 6：提交任務
- 步驟 7：檢視任務的輸出
- 步驟 8：（選用）提交具有覆寫的任務
- 步驟 9：清除您的教學課程資源
- 其他資源

概觀

本教學課程示範如何使用 AWS CLI、 `kubectl`和 AWS Batch 設定 Amazon EKS`eksctl`。

目標對象

本教學課程專為負責設定、測試和部署的系統管理員和開發人員而設計 AWS Batch。

使用的功能

本教學課程說明如何使用 AWS CLI來：

- 建立和設定 Amazon EKS 運算環境
- 建立任務佇列。
- 建立任務定義

- 建立並提交要執行的任務
- 使用覆寫提交任務

所需時間

完成本教學課程約需 30-40 分鐘。

區域限制

使用此解決方案沒有相關聯的國家或地區限制。

資源用量成本

建立 AWS 帳戶無需付費。不過，透過實作此解決方案，您可能會產生下表中所列的部分或全部費用。

描述	費用 (美元)
您需要按叢集小時付費	視執行個體而異，請參閱 Amazon EKS 定價

先決條件

開始本教學課程之前，您必須安裝並設定建立和管理 和 AWS Batch Amazon EKS 資源所需的下列工具和資源。

- AWS CLI：適用於使用 AWS 服務 (包括 Amazon EKS) 的命令列工具。本指南要求您使用 2.8.6 版或更新版本，或 1.26.0 版或更新版本。如需詳細資訊，請參閱AWS Command Line Interface 《 使用者指南》中的安裝、更新和解除安裝 AWS CLI 。安裝 之後 AWS CLI，建議您也進行設定。如需詳細資訊，請參閱AWS Command Line Interface 《 使用者指南》中的使用 進行快速組態`aws configure`。
- **`kubectl`**：命令列工具，適用於使用 Kubernetes 叢集。本指南要求您使用版本 `1.23` 或更新版本。如需詳細資訊，請參閱 Amazon EKS 使用者指南中的安裝或更新 `kubectl`。
- **`eksctl`** – 用於使用 Amazon EKS 叢集的命令列工具，可自動化許多個別任務。本指南要求您使用版本 `0.115.0` 或更新版本。如需詳細資訊，請參閱 Amazon EKS 使用者指南中的安裝或更新 `eksctl`。
- 必要的 IAM 許可 – 您使用的 IAM 安全主體必須具有使用 Amazon EKS IAM 角色和服務連結角色 CloudFormation，以及 VPC 和相關資源的許可。如需詳細資訊，請參閱《IAM 使用者指南》中的 Amazon Elastic Kubernetes Service 的動作、資源和條件金鑰和使用服務連結角色。您必須以同一位使用者的身分完成本指南中的所有步驟。

- 許可 – 呼叫 CreateComputeEnvironment API 操作的使用者建立使用 Amazon EKS 資源的運算環境需要 `eks:DescribeCluster` API 操作的許可。
- AWS 帳戶 number – 您需要知道您的 AWS 帳戶 ID。請遵循檢視 AWS 帳戶 ID 中的指示。
- （選用）CloudWatch – 若要檢查（選用）提交具有覆寫的任務，必須設定記錄。如需詳細資訊，請參閱使用 CloudWatch Logs 監控 AWS Batch Amazon EKS 任務。

步驟 1：建立 的 Amazon EKS 叢集 AWS Batch

> **Important**
>
> 為了盡可能簡單快速地開始使用，本教學課程包含具有預設設定的步驟。為生產用途建立 之前，建議您先熟悉所有設定，並使用符合您需求的設定進行部署。

安裝先決條件之後，您需要使用 建立叢集`eksctl`。建立叢集可能需要 10-15 分鐘。

```
$  eksctl create cluster --name my-cluster-name --region region-code
```

在上述命令中取代：

- 將 *my-cluster-name* 取代為您要用於叢集的名稱。
- 將 *region-code* 取代為 AWS 區域 以在其中建立叢集，例如 `us-west-2`。

本教學課程稍後需要叢集名稱和區域。

步驟 2：準備您的 Amazon EKS 叢集 AWS Batch

所有步驟都是必要的。

1. 建立 AWS Batch 任務的專用命名空間

   使用 `kubectl` 建立新的命名空間。

   ```
   $ namespace=my-aws-batch-namespace
   ```

   ```
   $ cat - <<EOF | kubectl create -f -
   {
   ```

```
  "apiVersion": "v1",
  "kind": "Namespace",
  "metadata": {
    "name": "${namespace}",
    "labels": {
      "name": "${namespace}"
    }
  }
}
EOF
```

輸出：

```
namespace/my-aws-batch-namespace created
```

2. 透過角色型存取控制 (RBAC) 啟用存取

使用 為叢集kubectl建立Kubernetes角色，以允許 AWS Batch 監看節點和 Pod，以及繫結角色。您必須為每個 EKS 叢集執行此操作一次。

```
$ cat - <<EOF | kubectl apply -f -
apiVersion: rbac.authorization.k8s.io/v1
kind: ClusterRole
metadata:
  name: aws-batch-cluster-role
rules:
  - apiGroups: [""]
    resources: ["namespaces"]
    verbs: ["get"]
  - apiGroups: [""]
    resources: ["nodes"]
    verbs: ["get", "list", "watch"]
  - apiGroups: [""]
    resources: ["pods"]
    verbs: ["get", "list", "watch"]
  - apiGroups: [""]
    resources: ["events"]
    verbs: ["list"]
  - apiGroups: [""]
    resources: ["configmaps"]
    verbs: ["get", "list", "watch"]
  - apiGroups: ["apps"]
    resources: ["daemonsets", "deployments", "statefulsets", "replicasets"]
```

```
    verbs: ["get", "list", "watch"]
  - apiGroups: ["rbac.authorization.k8s.io"]
    resources: ["clusterroles", "clusterrolebindings"]
    verbs: ["get", "list"]
---
apiVersion: rbac.authorization.k8s.io/v1
kind: ClusterRoleBinding
metadata:
  name: aws-batch-cluster-role-binding
subjects:
- kind: User
  name: aws-batch
  apiGroup: rbac.authorization.k8s.io
roleRef:
  kind: ClusterRole
  name: aws-batch-cluster-role
  apiGroup: rbac.authorization.k8s.io
EOF
```

輸出：

```
clusterrole.rbac.authorization.k8s.io/aws-batch-cluster-role created
clusterrolebinding.rbac.authorization.k8s.io/aws-batch-cluster-role-binding created
```

3. 為 建立命名空間範圍Kubernetes角色 AWS Batch ，以管理和生命週期 Pod 並將其繫結。您必須為每個唯一的命名空間執行此操作一次。

```
$ namespace=my-aws-batch-namespace
```

```
$ cat - <<EOF | kubectl apply -f - --namespace "${namespace}"
apiVersion: rbac.authorization.k8s.io/v1
kind: Role
metadata:
  name: aws-batch-compute-environment-role
  namespace: ${namespace}
rules:
  - apiGroups: [""]
    resources: ["pods"]
    verbs: ["create", "get", "list", "watch", "delete", "patch"]
  - apiGroups: [""]
    resources: ["serviceaccounts"]
    verbs: ["get", "list"]
```

```
  - apiGroups: ["rbac.authorization.k8s.io"]
    resources: ["roles", "rolebindings"]
    verbs: ["get", "list"]
---
apiVersion: rbac.authorization.k8s.io/v1
kind: RoleBinding
metadata:
  name: aws-batch-compute-environment-role-binding
  namespace: ${namespace}
subjects:
- kind: User
  name: aws-batch
  apiGroup: rbac.authorization.k8s.io
roleRef:
  kind: Role
  name: aws-batch-compute-environment-role
  apiGroup: rbac.authorization.k8s.io
EOF
```

輸出：

```
role.rbac.authorization.k8s.io/aws-batch-compute-environment-role created
rolebinding.rbac.authorization.k8s.io/aws-batch-compute-environment-role-binding
 created
```

4. 更新Kubernetesaws-auth組態映射，將先前的 RBAC 許可映射至 AWS Batch 服務連結角色。

   在下列命令中取代：

   - 將 *<your-account-number>* 取代為您的 AWS 帳戶 數字。

```
$ eksctl create iamidentitymapping \
    --cluster my-cluster-name \
    --arn "arn:aws:iam::<your-account-number>:role/AWSServiceRoleForBatch" \
    --username aws-batch
```

   輸出：

```
2022-10-25 20:19:57 [#]  adding identity "arn:aws:iam::<your-account-number>:role/
AWSServiceRoleForBatch" to auth ConfigMap
```

> **Note**
>
> 路徑aws-service-role/batch.amazonaws.com/已從服務連結角色的 ARN 中移除。這是因為aws-auth組態映射發生問題。如需詳細資訊，請參閱 中的具有路徑的角色在路徑包含在其 ARN 中時無法運作aws-authconfigmap。

步驟 3：建立 Amazon EKS 運算環境

AWS Batch 運算環境會定義運算資源參數，以符合您的批次工作負載需求。在受管運算環境中， AWS Batch 可協助您管理 Amazon EKS 叢集內運算資源 (Kubernetes節點） 的容量和執行個體類型。這是根據您在建立運算環境時定義的運算資源規格。您可以使用 EC2 隨需執行個體或 EC2 Spot 執行個體。

現在AWSServiceRoleForBatch服務連結角色可存取您的 Amazon EKS 叢集，您可以建立 AWS Batch 資源。首先，建立指向 Amazon EKS 叢集的運算環境。

- 若要讓 subnets執行eksctl get cluster *my-cluster-name*以取得叢集使用的子網路。
- 對於 securityGroupIds 參數，您可以使用與 Amazon EKS 叢集相同的安全群組。此命令會擷取叢集的安全群組 ID。

```
$ aws eks describe-cluster \
    --name my-cluster-name \
    --query cluster.resourcesVpcConfig.clusterSecurityGroupId
```

- instanceRole 會在您建立叢集時建立 。若要尋找instanceRole清單所有使用該AmazonEKSWorkerNodePolicy政策的實體：

```
$  aws iam list-entities-for-policy --policy-arn arn:aws:iam::aws:policy/
AmazonEKSWorkerNodePolicy
```

  政策角色的名稱包含您建立的叢集名稱eksctl-*my-cluster-name*-nodegroup-example。

  若要尋找 instanceRole arn，請執行下列命令：

```
$  aws iam list-instance-profiles-for-role --role-name eksctl-my-cluster-name-
nodegroup-example
```

輸出：

```
INSTANCEPROFILES        arn:aws:iam::<your-account-number>:instance-profile/
eks-04cb2200-94b9-c297-8dbe-87f12example
```

如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的建立 Amazon EKS 節點 IAM 角色和啟用叢集的 IAM 主體存取權。 如果您使用的是 Pod 網路，請參閱《Amazon EKS 使用者指南》中的設定 的 Amazon VPC CNI 外掛程式Kubernetes以使用服務帳戶的 IAM 角色。

```
$ cat <<EOF > ./batch-eks-compute-environment.json
{
  "computeEnvironmentName": "My-Eks-CE1",
  "type": "MANAGED",
  "state": "ENABLED",
  "eksConfiguration": {
    "eksClusterArn": "arn:aws:eks:region-code:your-account-number:cluster/my-cluster-
name",
    "kubernetesNamespace": "my-aws-batch-namespace"
  },
  "computeResources": {
    "type": "EC2",
    "allocationStrategy": "BEST_FIT_PROGRESSIVE",
    "minvCpus": 0,
    "maxvCpus": 128,
    "instanceTypes": [
        "m5"
    ],
    "subnets": [
        "<eks-cluster-subnets-with-access-to-internet-for-image-pull>"
    ],
    "securityGroupIds": [
        "<eks-cluster-sg>"
    ],
    "instanceRole": "<eks-instance-profile>"
  }
}
EOF
```

```
$ aws batch create-compute-environment --cli-input-json file://./batch-eks-compute-environment.json
```

備註

- 維護 Amazon EKS 運算環境是共同的責任。如需詳細資訊，請參閱Kubernetes 節點的共同責任。

步驟 4：建立任務佇列並連接運算環境

> **Important**
>
> 在繼續之前，請務必確認運算環境運作狀態良好。DescribeComputeEnvironments API 操作可用來執行此操作。
>
> ```
> $ aws batch describe-compute-environments --compute-environments My-Eks-CE1
> ```
>
> 確認 `status` 參數不是 `INVALID`。如果是，請查看原因的 `statusReason` 參數。如需詳細資訊，請參閱故障診斷 AWS Batch。

提交至此新任務佇列的任務會在加入與您運算環境相關聯之 Amazon EKS 叢集的 AWS Batch 受管節點上，以 Pod 的形式執行。

```
$ cat <<EOF > ./batch-eks-job-queue.json
 {
    "jobQueueName": "My-Eks-JQ1",
    "priority": 10,
    "computeEnvironmentOrder": [
      {
        "order": 1,
        "computeEnvironment": "My-Eks-CE1"
      }
    ]
  }
EOF
```

```
$ aws batch create-job-queue --cli-input-json file://./batch-eks-job-queue.json
```

步驟 5：建立任務定義

下列任務定義會指示 Pod 休眠 60 秒。

```
$ cat <<EOF > ./batch-eks-job-definition.json
{
  "jobDefinitionName": "MyJobOnEks_Sleep",
  "type": "container",
  "eksProperties": {
    "podProperties": {
      "hostNetwork": true,
      "containers": [
        {
          "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
          "command": [
            "sleep",
            "60"
          ],
          "resources": {
            "limits": {
              "cpu": "1",
              "memory": "1024Mi"
            }
          }
        }
      ],
      "metadata": {
        "labels": {
          "environment": "test"
        }
      }
    }
  }
}
EOF
```

```
$ aws batch register-job-definition --cli-input-json file://./batch-eks-job-
definition.json
```

備註

- cpu 和 memory 參數有考量。如需詳細資訊，請參閱Amazon EKS AWS Batch 上的 記憶體和 vCPU 考量事項。

步驟 6：提交任務

執行下列 AWS CLI 命令以提交新的任務。

```
$ aws batch submit-job --job-queue My-Eks-JQ1 \
    --job-definition MyJobOnEks_Sleep --job-name My-Eks-Job1
```

若要檢查任務的狀態：

```
$ aws batch describe-jobs --job <jobId-from-submit-response>
```

備註

- 如需在 Amazon EKS 資源上執行任務的詳細資訊，請參閱 Amazon EKS 任務。

步驟 7：檢視任務的輸出

若要檢視任務的輸出，請執行下列動作：

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇任務。
3. 在任務佇列下拉式清單中，選擇您為教學課程建立的任務佇列。
4. 任務表格列出所有任務及其目前狀態。一旦任務的狀態成功，請選擇任務的名稱 *My-Eks-JQ1*，以檢視任務的詳細資訊。
5. 在詳細資訊窗格中，從 開始和有時停止應該相隔一分鐘。

步驟 8：（選用）提交具有覆寫的任務

此任務會覆寫傳遞至 container 的命令。任務完成後， AWS Batch 會積極地清除 Pod，以減少對 的負載Kubernetes。若要檢查任務的詳細資訊，必須設定記錄。如需詳細資訊，請參閱使用 CloudWatch Logs 監控 AWS Batch Amazon EKS 任務。

```
$ cat <<EOF > ./submit-job-override.json
{
  "jobName": "EksWithOverrides",
  "jobQueue": "My-Eks-JQ1",
  "jobDefinition": "MyJobOnEks_Sleep",
  "eksPropertiesOverride": {
    "podProperties": {
      "containers": [
        {
          "command": [
            "/bin/sh"
          ],
          "args": [
            "-c",
            "echo hello world"
          ]
        }
      ]
    }
  }
}
EOF
```

```
$ aws batch submit-job --cli-input-json file://./submit-job-override.json
```

備註

- 為了改善對操作詳細資訊的可見性，請啟用 Amazon EKS 控制平面記錄。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS 控制平面記錄。
- Daemonsets 和 kubelets 額外負荷會影響可用的 vCPU 和記憶體資源，特別是擴展和任務配置。如需詳細資訊，請參閱Amazon EKS AWS Batch 上的 記憶體和 vCPU 考量事項。

若要檢視任務的輸出，請執行下列動作：

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇任務。
3. 在任務佇列下拉式清單中，選擇您為教學課程建立的任務佇列。
4. 任務表格列出所有任務及其目前狀態。一旦任務狀態成功，請選擇任務名稱以檢視任務的詳細資訊。

5. 在詳細資訊窗格中，選擇日誌串流名稱。任務的 CloudWatch 主控台將開啟，並且應該有一個事件具有 Message of `hello world`或您的自訂訊息。

步驟 9：清除您的教學課程資源

您需支付啟用 Amazon EC2 執行個體的費用。您可以刪除執行個體以停止產生費用。

若要刪除您建立的資源，請執行下列動作：

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇任務佇列。
3. 在任務佇列表格中，選擇您為教學課程建立的任務佇列。
4. 選擇停用。任務佇列狀態停用後，您可以選擇刪除。
5. 刪除任務佇列後，在導覽窗格中選擇運算環境。
6. 選擇您為此教學課程建立的運算環境，然後選擇停用。運算環境可能需要 1-2 分鐘才能完成停用。
7. 一旦運算環境的狀態停用，請選擇刪除。可能需要 1 到 2 分鐘才能刪除運算環境。

其他資源

完成教學課程後，您可能想要探索下列主題：

- 進一步了解 最佳實務。
- 探索 AWS Batch 核心元件。如需詳細資訊，請參閱的元件 AWS Batch。
- 進一步了解 中可用的不同運算環境 AWS Batch。
- 進一步了解任務佇列及其不同的排程選項。
- 進一步了解任務定義和不同的組態選項。
- 進一步了解不同類型的任務。

在 Amazon EKS 私有叢集 AWS Batch 上開始使用

AWS Batch 是一種受管服務，可協調 Amazon Elastic Kubernetes Service (Amazon EKS) 叢集中的批次工作負載。這包括佇列、相依性追蹤、受管任務重試和優先順序、Pod 管理和節點擴展。此功能會將現有的私有 Amazon EKS 叢集與 連線 AWS Batch ，以大規模執行任務。您可以使用

eksctl(Amazon EKS 的命令列界面）、 AWS 主控台或 AWS Command Line Interface來建立具有所有其他必要資源的私有 Amazon EKS 叢集。

根據預設，Amazon EKS 僅限私有叢集沒有傳入/傳出網際網路存取權，而且您只能從 VPC 或連線網路中存取 API 伺服器。Amazon VPC 端點用於啟用對其他服務 AWS 的私有存取。 eksctl支援使用預先存在的 Amazon VPC 和子網路建立全私有叢集。 eksctl也會在提供的 Amazon VPC 中建立 Amazon VPC 端點，並修改所提供子網路的路由表。

每個子網路都應有一個與其相關聯的明確路由表，因為 eksctl 不會修改主路由表。您的叢集必須從 Amazon VPC 中的容器登錄檔提取映像。您也可以在 Amazon VPC 中建立 Amazon Elastic Container Registry，並將容器映像複製到其中，以供節點提取。如需詳細資訊，請參閱將容器映像從一個儲存庫複製到另一個儲存庫。若要開始使用 Amazon ECR 私有儲存庫，請參閱 Amazon ECR 私有儲存庫。

您可以選擇性地使用 Amazon ECR 建立提取快取規則。為外部公有登錄檔建立提取快取規則後，您可以使用 Amazon ECR 私有登錄檔統一資源識別符 (URI) 從該外部公有登錄檔提取映像。然後，Amazon ECR 會建立儲存庫並快取映像。使用 Amazon ECR 私有登錄 URI 提取快取映像時，Amazon ECR 會檢查遠端登錄檔，以查看是否有映像的新版本，並最多每 24 小時更新您的私有登錄檔一次。

內容

- 概觀
- 先決條件
- 步驟 1：建立 的 EKS 叢集 AWS Batch
- 步驟 2：準備您的 EKS 叢集 AWS Batch
- 步驟 3：建立 Amazon EKS 運算環境
- 步驟 4：建立任務佇列並連接運算環境
- 步驟 5：使用提取快取建立 Amazon ECR
- 步驟 6：註冊任務定義
- 步驟 7：提交要執行的任務
- 步驟 8：檢視任務的輸出
- 步驟 9：（選用）提交具有覆寫的任務
- 步驟 10：清除您的教學課程資源
- 其他資源
- 故障診斷

概觀

本教學課程示範如何使用 AWS CloudShell、 `kubectl`和 AWS Batch 設定私有 Amazon EKS`eksctl`。

目標對象

本教學課程專為負責設定、測試和部署的系統管理員和開發人員而設計 AWS Batch。

使用的功能

本教學課程說明如何使用 AWS CLI來：

- 使用 Amazon Elastic Container Registry (Amazon ECR) 存放容器映像
- 建立和設定 Amazon EKS 運算環境
- 建立任務佇列。
- 建立任務定義
- 建立並提交要執行的任務
- 使用覆寫提交任務

所需時間

完成本教學課程約需 40-50 分鐘。

區域限制

使用此解決方案沒有相關聯的國家或地區限制。

資源用量成本

建立 AWS 帳戶無需付費。不過，透過實作此解決方案，您可能會產生下表中所列的部分或全部費用。

描述	費用 (美元)
您需要按叢集小時付費	視執行個體而異，請參閱 Amazon EKS 定價
Amazon EC2 執行個體	您需為建立的每個 Amazon EC2 執行個體付費。如需定價的相關資訊，請參閱 Amazon EC2 定價。

先決條件

本教學使用 AWS CloudShell ，這是您直接從 啟動的瀏覽器型預先驗證 Shell AWS 管理主控台。這允許在叢集不再具有公有網際網路存取權時存取叢集。 AWS CLI、 kubectl和 可能eksctl已安裝為的一部分 AWS CloudShell。如需 的詳細資訊 AWS CloudShell，請參閱AWS CloudShell《 使用者指南》。的替代方案 AWS CloudShell 是連線到叢集的 VPC 或連線的網路。

若要執行 kubectl 命令，您將需要 Amazon EKS 叢集的私有存取權。這表示通往叢集 API 伺服器的所有流量都必須來自叢集的 VPC 或連線的網路。

- AWS CLI – 使用 AWS 服務的命令列工具，包括 Amazon EKS。本指南要求您使用 2.8.6 版或更新版本，或 1.26.0 版或更新版本。如需詳細資訊，請參閱AWS Command Line Interface 《 使用者指南》中的安裝、更新和解除安裝 AWS CLI 。安裝 之後 AWS CLI，建議您也進行設定。如需詳細資訊，請參閱AWS Command Line Interface 《 使用者指南》中的使用 進行快速組態aws configure。
- **kubectl**：命令列工具，適用於使用 Kubernetes 叢集。本指南要求您使用版本 1.23 或更新版本。如需詳細資訊，請參閱 Amazon EKS 使用者指南中的安裝或更新 kubectl。
- **eksctl** – 使用 Amazon EKS 叢集的命令列工具，可自動化許多個別任務。本指南要求您使用版本 0.115.0 或更新版本。如需詳細資訊，請參閱 Amazon EKS 使用者指南中的安裝或更新 eksctl。
- 許可 – 呼叫 CreateComputeEnvironment API 操作以建立使用 Amazon EKS 資源的運算環境的使用者需要 eks:DescribeCluster和 eks:ListClusters API 操作的許可。您可以依照 IAM 使用者指南中新增和移除 IAM 身分許可的指示，將 AWSBatchFullAccess 受管政策連接至您的使用者帳戶。
- InstanceRole – 您需要InstanceRole為具有 AmazonEKSWorkerNodePolicy和 AmazonEC2ContainerRegistryPullOnly政策的 Amazon EKS 節點建立 。如需如何建立 的指示，InstanceRole請參閱建立 Amazon EKS 節點 IAM 角色。您將需要 的 ARNInstanceRole。
- AWS 帳戶 ID – 您需要知道您的 AWS 帳戶 ID。請遵循檢視 AWS 帳戶 ID 中的指示。
- （選用） CloudWatch – 若要檢查 （選用） 提交具有覆寫的任務，必須設定記錄。如需詳細資訊，請參閱使用 CloudWatch Logs 監控 AWS Batch Amazon EKS 任務。

步驟 1：建立 的 EKS 叢集 AWS Batch

> **Important**
>
> 為了盡可能簡單快速地開始使用，本教學課程包含具有預設設定的步驟。為生產用途建立 之前，建議您先熟悉所有設定，並使用符合您需求的設定進行部署。

我們建議您使用 `eksctl`和下列組態檔案來建立叢集。若要手動設定叢集，請遵循《Amazon EKS 使用者指南》中以有限的網際網路存取部署私有叢集中的指示。

1. 開啟 AWS CloudShell 主控台，並將區域設定為 `us-east-1`。對於教學課程的其餘部分，請確定您使用 `us-east-1`。
2. `us-east-1` 使用範例組態檔案在 區域中建立私有 EKS `eksctl` 叢集。將 yaml 檔案儲存到您的 AWS CloudShell 環境並命名為 `clusterConfig.yaml` 。您可以使用您要用於叢集的名稱來變更 *`my-test-cluster`*。

```
kind: ClusterConfig
apiVersion: eksctl.io/v1alpha5
metadata:
    name: my-test-cluster
    region: us-east-1
availabilityZones:
    - us-east-1a
    - us-east-1b
    - us-east-1c
managedNodeGroups:
    - name: ng-1
      privateNetworking: true
privateCluster:
    enabled: true
    skipEndpointCreation: false
```

3. 使用 命令建立您的 資源：`eksctl create cluster -f clusterConfig.yaml`。叢集建立可能需要 10-15 分鐘。
4. 叢集建立完成後，您必須將 AWS CloudShell IP 地址新增至允許清單。若要尋找您的 AWS CloudShell IP 地址，請執行下列命令：

```
curl http://checkip.amazonaws.com
```

擁有公有 IP 地址後，您必須建立允許清單規則：

```
aws eks update-cluster-config \
  --name my-test-cluster \
  --region us-east-1 \
  --resources-vpc-config
 endpointPublicAccess=true,endpointPrivateAccess=true,publicAccessCidrs=["<Public
 IP>/32"]
```

然後將更新套用至 kubectl 組態檔案：

```
aws eks update-kubeconfig --name my-test-cluster --region us-east-1
```

5. 若要測試您是否可存取節點，請執行下列命令：

```
kubectl get nodes
```

命令的輸出為：

```
NAME                                STATUS   ROLES    AGE     VERSION
ip-192-168-107-235.ec2.internal   Ready    none     1h   v1.32.3-eks-473151a
ip-192-168-165-40.ec2.internal    Ready    none     1h   v1.32.3-eks-473151a
ip-192-168-98-54.ec2.internal     Ready    none     1h   v1.32.1-eks-5d632ec
```

步驟 2：準備您的 EKS 叢集 AWS Batch

所有步驟都是必要步驟，且必須在其中完成 AWS CloudShell。

1. 建立 AWS Batch 任務的專用命名空間

使用 kubectl 建立新的命名空間。

```
$ namespace=my-aws-batch-namespace
```

```
$ cat - <<EOF | kubectl create -f -
{
  "apiVersion": "v1",
  "kind": "Namespace",
```

```
  "metadata": {
    "name": "${namespace}",
    "labels": {
      "name": "${namespace}"
    }
  }
}
EOF
```

輸出：

```
namespace/my-aws-batch-namespace created
```

2. 透過角色型存取控制 (RBAC) 啟用存取

使用 為叢集kubectl建立Kubernetes角色，以允許 AWS Batch 監看節點和 Pod，以及繫結角色。您必須為每個 Amazon EKS 叢集執行此操作一次。

```
$ cat - <<EOF | kubectl apply -f -
apiVersion: rbac.authorization.k8s.io/v1
kind: ClusterRole
metadata:
  name: aws-batch-cluster-role
rules:
  - apiGroups: [""]
    resources: ["namespaces"]
    verbs: ["get"]
  - apiGroups: [""]
    resources: ["nodes"]
    verbs: ["get", "list", "watch"]
  - apiGroups: [""]
    resources: ["pods"]
    verbs: ["get", "list", "watch"]
  - apiGroups: [""]
    resources: ["events"]
    verbs: ["list"]
  - apiGroups: [""]
    resources: ["configmaps"]
    verbs: ["get", "list", "watch"]
  - apiGroups: ["apps"]
    resources: ["daemonsets", "deployments", "statefulsets", "replicasets"]
    verbs: ["get", "list", "watch"]
  - apiGroups: ["rbac.authorization.k8s.io"]
```

```
    resources: ["clusterroles", "clusterrolebindings"]
    verbs: ["get", "list"]
---
apiVersion: rbac.authorization.k8s.io/v1
kind: ClusterRoleBinding
metadata:
  name: aws-batch-cluster-role-binding
subjects:
- kind: User
  name: aws-batch
  apiGroup: rbac.authorization.k8s.io
roleRef:
  kind: ClusterRole
  name: aws-batch-cluster-role
  apiGroup: rbac.authorization.k8s.io
EOF
```

輸出：

```
clusterrole.rbac.authorization.k8s.io/aws-batch-cluster-role created
clusterrolebinding.rbac.authorization.k8s.io/aws-batch-cluster-role-binding created
```

為 建立命名空間範圍Kubernetes角色 AWS Batch ，以管理和生命週期 Pod 並將其繫結。您必須為每個唯一的命名空間執行此操作一次。

```
$ namespace=my-aws-batch-namespace
```

```
$ cat - <<EOF | kubectl apply -f - --namespace "${namespace}"
apiVersion: rbac.authorization.k8s.io/v1
kind: Role
metadata:
  name: aws-batch-compute-environment-role
  namespace: ${namespace}
rules:
  - apiGroups: [""]
    resources: ["pods"]
    verbs: ["create", "get", "list", "watch", "delete", "patch"]
  - apiGroups: [""]
    resources: ["serviceaccounts"]
    verbs: ["get", "list"]
  - apiGroups: ["rbac.authorization.k8s.io"]
```

```
    resources: ["roles", "rolebindings"]
    verbs: ["get", "list"]
---
apiVersion: rbac.authorization.k8s.io/v1
kind: RoleBinding
metadata:
  name: aws-batch-compute-environment-role-binding
  namespace: ${namespace}
subjects:
- kind: User
  name: aws-batch
  apiGroup: rbac.authorization.k8s.io
roleRef:
  kind: Role
  name: aws-batch-compute-environment-role
  apiGroup: rbac.authorization.k8s.io
EOF
```

輸出：

```
role.rbac.authorization.k8s.io/aws-batch-compute-environment-role created
rolebinding.rbac.authorization.k8s.io/aws-batch-compute-environment-role-binding
 created
```

更新Kubernetesaws-auth組態映射，將先前的 RBAC 許可映射至 AWS Batch 服務連結角色。

```
$ eksctl create iamidentitymapping \
    --cluster my-test-cluster \
    --arn "arn:aws:iam::<your-account-ID>:role/AWSServiceRoleForBatch" \
    --username aws-batch
```

輸出：

```
2022-10-25 20:19:57 [#]  adding identity "arn:aws:iam::<your-account-ID>:role/
AWSServiceRoleForBatch" to auth ConfigMap
```

> **Note**
>
> 路徑aws-service-role/batch.amazonaws.com/已從服務連結角色的 ARN 中移除。這是因為aws-auth組態映射發生問題。如需詳細資訊，請參閱 中的具有路徑的角色在路徑包含在其 ARN 中時無法運作aws-authconfigmap。

步驟 3：建立 Amazon EKS 運算環境

AWS Batch 運算環境定義運算資源參數，以滿足批次工作負載需求。在受管運算環境中， AWS Batch 可協助您管理 Amazon EKS 叢集內運算資源 (Kubernetes節點） 的容量和執行個體類型。這是以您在建立運算環境時定義的運算資源規格為基礎。您可以使用 EC2 隨需執行個體或 EC2 Spot 執行個體。

現在AWSServiceRoleForBatch服務連結角色可存取您的 Amazon EKS 叢集，您可以建立 AWS Batch 資源。首先，建立指向 Amazon EKS 叢集的運算環境。

- 若要讓 subnets執行eksctl get cluster *my-test-cluster*以取得叢集使用的子網路。
- 對於 securityGroupIds 參數，您可以使用與 Amazon EKS 叢集相同的安全群組。此命令會擷取叢集的安全群組 ID。

```
$ aws eks describe-cluster \
    --name my-test-cluster \
    --query cluster.resourcesVpcConfig.clusterSecurityGroupId
```

- 使用instanceRole您在先決條件中建立的 ARN。

```
$ cat <<EOF > ./batch-eks-compute-environment.json
{
  "computeEnvironmentName": "My-Eks-CE1",
  "type": "MANAGED",
  "state": "ENABLED",
  "eksConfiguration": {
    "eksClusterArn": "arn:aws:eks:us-east-1:<your-account-ID>:cluster/my-test-cluster",
    "kubernetesNamespace": "my-aws-batch-namespace"
  },
  "computeResources": {
    "type": "EC2",
    "allocationStrategy": "BEST_FIT_PROGRESSIVE",
    "minvCpus": 0,
```

```
    "maxvCpus": 128,
    "instanceTypes": [
        "m5"
    ],
    "subnets": [
        "<eks-cluster-subnets-with-access-to-the-image-for-image-pull>"
    ],
    "securityGroupIds": [
        "<eks-cluster-sg>"
    ],
    "instanceRole": "<eks-instance-profile>"
  }
}
EOF
```

```
$ aws batch create-compute-environment --cli-input-json file://./batch-eks-compute-environment.json
```

備註

- 維護 Amazon EKS 運算環境是共同的責任。如需詳細資訊，請參閱 Amazon EKS 中的安全性。

步驟 4：建立任務佇列並連接運算環境

> **Important**
>
> 在繼續之前，請務必確認運算環境運作狀態良好。DescribeComputeEnvironments API 操作可用來執行此操作。
>
> ```
> $ aws batch describe-compute-environments --compute-environments My-Eks-CE1
> ```
>
> 確認 `status` 參數不是 `INVALID`。如果是，請查看原因的 `statusReason` 參數。如需詳細資訊，請參閱故障診斷 AWS Batch。

提交至此新任務佇列的任務會在加入與您運算環境相關聯之 Amazon EKS 叢集的 AWS Batch 受管節點上，以 Pod 的形式執行。

```
$ cat <<EOF > ./batch-eks-job-queue.json
```

```
 {
    "jobQueueName": "My-Eks-JQ1",
    "priority": 10,
    "computeEnvironmentOrder": [
      {
        "order": 1,
        "computeEnvironment": "My-Eks-CE1"
      }
    ]
  }
EOF
```

```
$ aws batch create-job-queue --cli-input-json file://./batch-eks-job-queue.json
```

步驟 5：使用提取快取建立 Amazon ECR

由於叢集沒有公有網際網路存取，您必須為容器映像建立 Amazon ECR。以下指示會建立具有提取快取規則的 Amazon ECR 來存放映像。

1. 下列命令會建立提取快取規則。您可以使用不同的######課程字首。

   ```
   aws ecr create-pull-through-cache-rule \
       --ecr-repository-prefix "my-prefix" \
       --upstream-registry-url "public.ecr.aws" \
       --region us-east-1
   ```

2. 使用公有 ECR 驗證。

   ```
   aws ecr get-login-password --region us-east-1 | docker login --username AWS --
   password-stdin <your-account-ID>.dkr.ecr.us-east-1.amazonaws.com
   ```

   現在您可以提取映像。

   ```
   docker pull <your-account-ID>.dkr.ecr.us-east-1.amazonaws.com/my-prefix/
   amazonlinux/amazonlinux:2
   ```

3. 您可以執行下列命令來驗證儲存庫和映像：

   ```
   aws ecr describe-repositories
   ```

```
aws ecr describe-images --repository-name my-prefix/amazonlinux/amazonlinux
```

4. 用於提取容器的影像字串格式如下：

```
<your-account-ID>.dkr.ecr.us-east-1.amazonaws.com/my-prefix/amazonlinux/
amazonlinux:2
```

步驟 6：註冊任務定義

下列任務定義會指示 Pod 休眠 60 秒。

在任務定義的映像欄位中，提供儲存在私有 ECR 儲存庫中的映像連結，而不是提供公有 ECR 儲存庫中的映像連結。請參閱下列範例任務定義：

```
$ cat <<EOF > ./batch-eks-job-definition.json
{
  "jobDefinitionName": "MyJobOnEks_Sleep",
  "type": "container",
  "eksProperties": {
    "podProperties": {
      "hostNetwork": true,
      "containers": [
        {
          "image": "<your-account-ID>.dkr.ecr.us-east-1.amazonaws.com/my-prefix/
amazonlinux/amazonlinux:2",
          "command": [
            "sleep",
            "60"
          ],
          "resources": {
            "limits": {
              "cpu": "1",
              "memory": "1024Mi"
            }
          }
        }
      ],
      "metadata": {
        "labels": {
          "environment": "test"
        }
```

```
      }
    }
  }
}
EOF
```

```
$ aws batch register-job-definition --cli-input-json file://./batch-eks-job-definition.json
```

備註

- cpu 和 memory 參數有考量事項。如需詳細資訊，請參閱Amazon EKS AWS Batch 上的 記憶體和 vCPU 考量事項。

步驟 7：提交要執行的任務

在 中執行下列 AWS CLI 命令 AWS CloudShell 以提交新的任務，並傳回唯一的 JobID。

```
$ aws batch submit-job --job-queue My-Eks-JQ1 \
    --job-definition MyJobOnEks_Sleep - -job-name My-Eks-Job1
```

備註

- 如需在 Amazon EKS 資源上執行任務的詳細資訊，請參閱 Amazon EKS 任務。

步驟 8：檢視任務的輸出

若要檢查任務的狀態：

```
$ aws batch describe-jobs --job <JobID-from-submit-response>
```

startedAt 和 stoppedAt應該相隔一分鐘。

步驟 9：（選用）提交具有覆寫的任務

此任務會覆寫傳遞至容器的命令。

```
$ cat <<EOF > ./submit-job-override.json
{
```

```
  "jobName": "EksWithOverrides",
  "jobQueue": "My-Eks-JQ1",
  "jobDefinition": "MyJobOnEks_Sleep",
  "eksPropertiesOverride": {
    "podProperties": {
      "containers": [
        {
          "command": [
            "/bin/sh"
          ],
          "args": [
            "-c",
            "echo hello world"
          ]
        }
      ]
    }
  }
}
EOF
```

```
$ aws batch submit-job - -cli-input-json file://./submit-job-override.json
```

備註

- 為了改善對操作詳細資訊的可見性，請啟用 Amazon EKS 控制平面記錄。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS 控制平面記錄。
- Daemonsets 和 kubelets 額外負荷會影響可用的 vCPU 和記憶體資源，特別是擴展和任務配置。如需詳細資訊，請參閱Amazon EKS AWS Batch 上的 記憶體和 vCPU 考量事項。

步驟 10：清除您的教學課程資源

您需支付啟用 Amazon EC2 執行個體的費用。您可以刪除執行個體以停止產生費用。

若要刪除您建立的資源，請執行下列動作：

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇任務佇列。
3. 在任務佇列表格中，選擇您為教學課程建立的任務佇列。

4. 從動作選擇停用。任務佇列狀態停用後，您可以選擇刪除。
5. 刪除任務佇列後，在導覽窗格中選擇運算環境。
6. 選擇您為此教學課程建立的運算環境，然後選擇停用動作。運算環境可能需要 1-2 分鐘才能完成停用。
7. 一旦運算環境的狀態停用，請選擇刪除。可能需要 1 到 2 分鐘才能刪除運算環境。

其他資源

完成教學課程後，您可能想要探索下列主題：

- 進一步了解 最佳實務。
- 探索 AWS Batch 核心元件。如需詳細資訊，請參閱的元件 AWS Batch。
- 進一步了解 中可用的不同運算環境 AWS Batch。
- 進一步了解任務佇列及其不同的排程選項。
- 進一步了解任務定義和不同的組態選項。
- 進一步了解不同類型的任務。

故障診斷

如果 啟動的節點 AWS Batch 無法存取存放映像的 Amazon ECR 儲存庫 （或任何其他儲存庫），則您的任務可能會保持在 STARTING 狀態。這是因為 Pod 將無法下載映像並執行您的 AWS Batch 任務。如果您按一下 啟動的 Pod 名稱 AWS Batch ，應該可以查看錯誤訊息並確認問題。錯誤訊息看起來應該類似以下內容：

```
Failed to pull image "public.ecr.aws/amazonlinux/amazonlinux:2": rpc error: code =
Unknown desc = failed to pull and unpack image
"public.ecr.aws/amazonlinux/amazonlinux:2": failed to resolve reference
"public.ecr.aws/amazonlinux/amazonlinux:2": failed to do request: Head
"https://public.ecr.aws/v2/amazonlinux/amazonlinux/manifests/2": dial tcp: i/o timeout
```

如需其他常見的故障診斷案例，請參閱故障診斷 AWS Batch。如需根據 Pod 狀態進行故障診斷，請參閱如何在 Amazon EKS 中對 Pod 狀態進行故障診斷？。

SageMaker AI AWS Batch 入門

AWS Batch 服務任務可讓您透過具有排程、優先順序和佇列功能 AWS Batch 的任務佇列提交 SageMaker 訓練任務。本教學課程示範如何使用 AWS Batch 服務任務來設定和執行簡單的 SageMaker 訓練任務。

內容

- 概觀
- 先決條件
- 步驟 1：建立 SageMaker AI 執行角色
- 步驟 2：建立您的服務環境
- 步驟 3：建立 SageMaker 任務佇列
- 步驟 4：建立並提交訓練任務
- 步驟 5：監控任務狀態
- 步驟 6：檢視任務輸出
- 步驟 7：清除您的教學課程資源
- 其他資源

概觀

本教學課程示範如何使用 設定 SageMaker Training 任務 AWS Batch 的服務任務 AWS CLI。

目標對象

本教學課程專為負責大規模設定和執行機器學習訓練任務的資料科學家和開發人員而設計。

使用的功能

本教學課程說明如何使用 AWS CLI 來：

- 為 SageMaker Training 任務建立服務環境
- 建立 SageMaker 訓練任務佇列
- 使用 `SubmitServiceJob` API 提交服務任務
- 監控任務狀態並檢視輸出
- 存取訓練任務的 CloudWatch 日誌

所需時間

此教學課程需約 15 分鐘完成。

區域限制

本教學課程可在可使用 AWS Batch 和 SageMaker AI 的任何 AWS 區域中完成。

資源用量成本

建立 AWS 帳戶無需付費。不過，透過實作此解決方案，您可能會產生下列資源的成本：

Description	費用 (美元)
SageMaker AI Training 執行個體	您需為所使用的每個 SageMaker AI Training 執行個體付費。如需定價的詳細資訊，請參閱 SageMaker AI 定價。
Amazon S3 儲存體	儲存訓練任務輸出的最低成本。如需詳細資訊，請參閱 Amazon S3 定價。

先決條件

開始本教學課程之前，您必須安裝並設定建立和管理 AWS Batch 和 SageMaker AI 資源所需的下列工具和資源。

- AWS CLI – 使用 AWS 服務的命令列工具，包括 AWS Batch 和 SageMaker AI。本指南要求您使用 2.8.6 版或更新版本。如需詳細資訊，請參閱AWS Command Line Interface 《 使用者指南》中的安裝、更新和解除安裝 AWS CLI 。安裝 之後 AWS CLI，建議您也進行設定。如需詳細資訊，請參閱AWS Command Line Interface 《 使用者指南》中的使用 進行快速組態`aws configure`。

步驟 1：建立 SageMaker AI 執行角色

SageMaker AI 使用執行角色來代表您使用其他服務 AWS 來執行操作。您必須建立執行角色，並授予 SageMaker AI 使用訓練任務所需服務和資源的許可。使用 `AmazonSageMakerFullAccess`受管政策，因為它包含 Amazon S3 的許可。

> **Note**
>
> 請使用下列指示來建立本教學課程的 SageMaker AI 執行角色。

在您為生產環境建立執行角色之前，我們建議您檢閱 SageMaker AI 開發人員指南中的如何使用 SageMaker AI 執行角色。 SageMaker

1. 建立 IAM 角色

   使用下列信任政策建立名為 sagemaker-trust-policy.json的 JSON 檔案：

   JSON

   ```
   {
       "Version":"2012-10-17",
       "Statement": [
           {
               "Effect": "Allow",
               "Principal": {
                   "Service": "sagemaker.amazonaws.com"
               },
               "Action": "sts:AssumeRole"
           }
       ]
   }
   ```

   使用信任政策建立 IAM 角色：

   ```
   aws iam create-role \
       --role-name SageMakerExecutionRole \
       --assume-role-policy-document file://sagemaker-trust-policy.json \
       --description "Execution role for SageMaker training jobs"
   ```

2. 連接 受管政策

   將所需的 受管政策連接至角色：

   ```
   aws iam attach-role-policy \
       --role-name SageMakerExecutionRole \
       --policy-arn arn:aws:iam::aws:policy/AmazonSageMakerFullAccess
   ```

   ```
   aws iam attach-role-policy \
   ```

```
    --role-name SageMakerExecutionRole \
    --policy-arn arn:aws:iam::aws:policy/AmazonS3FullAccess
```

3. 記下角色 ARN

   取得角色 ARN，您會在後續步驟中需要此角色：

   ```
   aws iam get-role --role-name SageMakerExecutionRole --query 'Role.Arn' --output
    text
   ```

   儲存此 ARN，因為您將在建立訓練任務承載時使用它。

步驟 2：建立您的服務環境

服務環境定義 SageMaker Training 任務的容量限制。服務環境會封裝可同時執行的訓練執行個體數量上限。

> **Important**
>
> 當您為 SageMaker Training 建立第一個服務環境時， AWS Batch 會自動在您的AWSServiceRoleForAWSBatchWithSagemaker帳戶中建立名為 的服務連結角色。此角色允許 代表您 AWS Batch 佇列和管理 SageMaker Training 任務。如需此服務連結角色及其許可的詳細資訊，請參閱 the section called “SageMaker 整合角色”。

建立可處理最多 5 個執行個體的服務環境：

```
aws batch create-service-environment \
    --service-environment-name TutorialServiceEnvironment \
    --service-environment-type SAGEMAKER_TRAINING \
    --capacity-limits capacityUnit=NUM_INSTANCES,maxCapacity=5
```

輸出：

```
{
    "serviceEnvironmentName": "TutorialServiceEnvironment",
    "serviceEnvironmentArn": "arn:aws:batch:your-region:your-account-id:service-
environment/TutorialServiceEnvironment"
}
```

確認您的服務環境已成功建立：

```
aws batch describe-service-environments --service-
environments TutorialServiceEnvironment
```

輸出：

```
{
    "serviceEnvironments": [
        {
            "serviceEnvironmentName": "TutorialServiceEnvironment",
            "serviceEnvironmentArn": "arn:aws:batch:your-region:your-account-
id:service-environment/TutorialServiceEnvironment",
            "serviceEnvironmentType": "SAGEMAKER_TRAINING",
            "state": "ENABLED",
            "status": "VALID",
            "capacityLimits": [
                {
                    "maxCapacity": 5,
                    "capacityUnit": "NUM_INSTANCES"
                }
            ],
            "tags": {}
        }
    ]
}
```

如需服務環境的詳細資訊，請參閱 服務環境。

步驟 3：建立 SageMaker 任務佇列

SageMaker 任務佇列會管理服務任務的排程和執行。提交至此佇列的任務將根據可用容量分派至您的服務環境。

建立 SageMaker 訓練任務佇列：

```
aws batch create-job-queue \
    --job-queue-name my-sm-training-fifo-jq \
    --job-queue-type SAGEMAKER_TRAINING \
    --priority 1 \
    --service-environment-order order=1,serviceEnvironment=TutorialServiceEnvironment
```

輸出：

```
{
    "jobQueueName": "my-sm-training-fifo-jq",
    "jobQueueArn": "arn:aws:batch:your-region:your-account-id:job-queue/my-sm-training-fifo-jq"
}
```

確認您的任務佇列已成功建立：

```
aws batch describe-job-queues --job-queues my-sm-training-fifo-jq
```

輸出：

```
{
    "jobQueues": [
        {
            "jobQueueName": "my-sm-training-fifo-jq",
            "jobQueueArn": "arn:aws:batch:your-region:your-account-id:job-queue/my-sm-training-fifo-jq",
            "state": "ENABLED",
            "status": "VALID",
            "statusReason": "JobQueue Healthy",
            "priority": 1,
            "computeEnvironmentOrder": [],
            "serviceEnvironmentOrder": [
                {
                    "order": 1,
                    "serviceEnvironment": "arn:aws:batch:your-region:your-account-id:service-environment/TutorialServiceEnvironment"
                }
            ],
            "jobQueueType": "SAGEMAKER_TRAINING",
            "tags": {}
        }
    ]
}
```

如需 SageMaker 任務佇列的詳細資訊，請參閱 the section called “建立 SageMaker 任務佇列”。

步驟 4：建立並提交訓練任務

現在，您將建立簡單的訓練任務，並將其提交到您的任務佇列。此範例使用基本的「hello world」訓練任務，示範服務任務功能。

建立名為 *my_training_job.json* 且具有下列內容的檔案。將 *your-account-id* 取代為您的帳戶 AWS ID：

> **Note**
>
> S3OutputPath 建立 SageMaker 訓練任務時需要 ，但本教學課程的結果不會存放在 Amazon S3 儲存貯體中，您可以在下列 JSON 中使用 路徑。在您的生產環境中，如果您選擇將輸出存放在該處，則需要有效的 Amazon S3 儲存貯體。

```
{
    "TrainingJobName": "my-simple-training-job",
    "RoleArn": "arn:aws:iam::your-account-id:role/SageMakerExecutionRole",
    "AlgorithmSpecification": {
        "TrainingInputMode": "File",
        "TrainingImage": "763104351884.dkr.ecr.us-west-2.amazonaws.com/pytorch-
training:2.0.0-cpu-py310",
        "ContainerEntrypoint": [
            "echo",
            "hello world"
        ]
    },
    "ResourceConfig": {
        "InstanceType": "ml.c5.xlarge",
        "InstanceCount": 1,
        "VolumeSizeInGB": 1
    },
    "OutputDataConfig": {
        "S3OutputPath": "s3://your-s3-bucket/output"
    },
    "StoppingCondition": {
        "MaxRuntimeInSeconds": 30
    }
}
```

使用 SubmitServiceJob API 提交訓練任務：

```
aws batch submit-service-job \
    --job-queue my-sm-training-fifo-jq \
    --job-name my-batch-sm-job \
    --service-job-type SAGEMAKER_TRAINING \
    --retry-strategy attempts=1 \
    --timeout-config attemptDurationSeconds=60 \
    --service-request-payload file://my_training_job.json
```

輸出：

```
{
    "jobArn": "arn:aws:batch:your-region:your-account-id:service-job/your-job-id",
    "jobName": "my-batch-sm-job",
    "jobId": "your-job-id"
}
```

如需服務任務承載的詳細資訊，請參閱 the section called “服務任務承載”。如需提交服務任務的詳細資訊，請參閱 the section called “提交服務任務”。

步驟 5：監控任務狀態

您可以使用下列 AWS Batch APIs監控訓練任務：DescribeServiceJob、ListServiceJobs 和 GetJobQueueSnapshot。本節顯示檢查任務狀態和佇列資訊的不同方法。

檢視佇列中正在執行的任務：

```
aws batch list-service-jobs \
    --job-queue my-sm-training-fifo-jq --job-status RUNNING
```

輸出：

```
{
    "jobSummaryList": [
        {
            "latestAttempt": {
                "serviceResourceId": {
                    "name": "TrainingJobArn",
                    "value": "arn:aws:sagemaker:your-region:your-account-id:training-
job/AWSBatch<my-simple-training-job><your-attempt-id>"
                }
            },
```

```
            "createdAt": 1753718760,
            "jobArn": "arn:aws:batch:your-region:your-account-id:service-job/your-job-id",
            "jobId": "your-job-id",
            "jobName": "my-batch-sm-job",
            "serviceJobType": "SAGEMAKER_TRAINING",
            "status": "RUNNING",
            "startedAt": 1753718820
        }
    ]
}
```

檢視處於 RUNNABLE 狀態的任務：

```
aws batch list-service-jobs \
    --job-queue my-sm-training-fifo-jq --job-status RUNNABLE
```

取得佇列中即將執行任務的快照：

```
aws batch get-job-queue-snapshot --job-queue my-sm-training-fifo-jq
```

輸出：

```
{
    "frontOfQueue": {
        "jobs": [
            {
                "jobArn": "arn:aws:batch:your-region:your-account-id:service-job/your-job-id",
                "earliestTimeAtPosition": 1753718880
            },
            {
                "jobArn": "arn:aws:batch:your-region:your-account-id:service-job/your-job-id-2",
                "earliestTimeAtPosition": 1753718940
            }
        ],
        "lastUpdatedAt": 1753718970
    }
}
```

依名稱搜尋任務：

```
aws batch list-service-jobs \
    --job-queue my-sm-training-fifo-jq \
    --filters name=JOB_NAME,values="my-batch-sm-job"
```

輸出：

```
{
    "jobSummaryList": [
        {
            "latestAttempt": {
                "serviceResourceId": {
                    "name": "TrainingJobArn",
                    "value": "arn:aws:sagemaker:your-region:your-account-id:training-
job/AWSBatch<my-simple-training-job><your-attempt-id>"
                }
            },
            "createdAt": 1753718760,
            "jobArn": "arn:aws:batch:your-region:your-account-id:service-job/your-job-
id",
            "jobId": "your-job-id",
            "jobName": "my-batch-sm-job",
            "serviceJobType": "SAGEMAKER_TRAINING",
            "status": "RUNNING"
        }
    ]
}
```

如需任務狀態映射的詳細資訊，請參閱 the section called “服務任務狀態”。

步驟 6：檢視任務輸出

任務完成後，您可以透過 和 SageMaker AI APIs來檢視其輸出 AWS Batch 和日誌。

從 取得任務的詳細資訊 AWS Batch：

```
aws batch describe-service-job \
    --job-id your-job-id
```

輸出：

```
{
```

```
    "attempts": [
        {
            "serviceResourceId": {
                "name": "TrainingJobArn",
                "value": "arn:aws:sagemaker:your-region:your-account-id:training-job/
AWSBatch<my-simple-training-job><your-attempt-id>"
            },
            "startedAt": 1753718820,
            "stoppedAt": 1753718880,
            "statusReason": "Received status from SageMaker: Training job completed"
        }
    ],
    "createdAt": 1753718760,
    "jobArn": "arn:aws:batch:your-region:your-account-id:service-job/your-job-id",
    "jobId": "your-job-id",
    "jobName": "my-batch-sm-job",
    "jobQueue": "arn:aws:batch:your-region:your-account-id:job-queue/my-sm-training-
fifo-jq",
    "latestAttempt": {
        "serviceResourceId": {
            "name": "TrainingJobArn",
            "value": "arn:aws:sagemaker:your-region:your-account-id:training-job/
AWSBatch<my-simple-training-job><your-attempt-id>"
        }
    },
    "retryStrategy": {
        "attempts": 1,
        "evaluateOnExit": []
    },
    "serviceRequestPayload": "your-training-job-request-json",
    "serviceJobType": "SAGEMAKER_TRAINING",
    "startedAt": 1753718820,
    "status": "SUCCEEDED",
    "statusReason": "Received status from SageMaker: Training job completed",
    "stoppedAt": 1753718880,
    "tags": {},
    "timeoutConfig": {
        "attemptDurationSeconds": 60
    }
}
```

此命令會傳回完整的任務資訊，包括 SageMaker Training 任務 ARN，您可以用來直接透過 SageMaker AI 存取任務：

```
aws sagemaker describe-training-job \
    --training-job-name AWSBatch<my-simple-training-job><your-attempt-id>
```

若要檢視訓練任務的 CloudWatch 日誌，請先取得日誌串流名稱：

```
aws logs describe-log-streams \
    --log-group-name /aws/sagemaker/TrainingJobs \
    --log-stream-name-prefix AWSBatchmy-simple-training-job
```

輸出：

```
{
    "logStreams": [
        {
            "logStreamName": "your-log-stream-name",
            "creationTime": 1753718830,
            "firstEventTimestamp": 1753718840,
            "lastEventTimestamp": 1753718850,
            "lastIngestionTime": 1753718860,
            "uploadSequenceToken": upload-sequence-token,
            "arn": "arn:aws:logs:your-region:your-account-id:log-group:/aws/sagemaker/
TrainingJobs:log-stream:AWSBatch<my-simple-training-job><your-attempt-id>/algo-1-algo-
id",
            "storedBytes": 0
        }
    ]
}
```

然後使用先前回應中的日誌串流名稱擷取日誌：

```
aws logs get-log-events \
    --log-group-name /aws/sagemaker/TrainingJobs \
    --log-stream-name your-log-stream-name
```

輸出：

```
{
    "events": [
        {
            "timestamp": 1753718845,
            "message": "hello world",
```

```
            "ingestionTime": 1753718865
        }
    ],
    "nextForwardToken": "next-forward-token",
    "nextBackwardToken": "next-backward-token"
}
```

日誌輸出會顯示來自訓練任務的「hello world」訊息，確認任務已成功執行。

步驟 7：清除您的教學課程資源

完成教學課程後，請清除您建立的資源，以避免持續收費。

首先，停用和刪除任務佇列：

```
aws batch update-job-queue \
    --job-queue my-sm-training-fifo-jq \
    --state DISABLED
```

等待任務佇列停用，然後刪除它：

```
aws batch delete-job-queue \
    --job-queue my-sm-training-fifo-jq
```

接著，停用和刪除服務環境：

```
aws batch update-service-environment \
    --service-environment TutorialServiceEnvironment \
    --state DISABLED
```

等待服務環境停用，然後刪除它：

```
aws batch delete-service-environment \
    --service-environment TutorialServiceEnvironment
```

其他資源

完成教學課程後，您可能想要探索下列主題：

- 我們建議您使用 PySDK 建立服務任務並提交到您的任務佇列，因為 PySDK 具有協助程式類別和公用程式。如需使用 PySDK 的範例，請參閱 GitHub 上的 SageMaker AI 範例。

- 進一步了解 the section called “服務任務”。
- 探索中的服務任務承載 AWS Batch更複雜的訓練任務組態。
- 了解 在 中提交服務任務 AWS Batch和 SubmitServiceJob API。
- 檢閱將 AWS Batch 服務任務狀態映射至 SageMaker AI 狀態以了解任務狀態轉換。
- 如需使用 Python 建立和提交 SageMaker 訓練任務的更豐富功能，請造訪 SageMaker AI Python SDK 文件。 SageMaker
- 探索 SageMaker 範例筆記本，以取得更複雜的機器學習工作流程。

AWS Batch Widget 儀表板

透過 AWS Batch 儀表板，您可以監控最近的任務、任務佇列和運算環境。根據預設，會顯示下列儀表板小工具：

- 任務概觀 – 如需 AWS Batch 任務的詳細資訊，請參閱 任務。
- 任務佇列概觀 – 如需 AWS Batch 任務佇列的詳細資訊，請參閱 任務佇列。
- 運算環境概觀 – 如需 AWS Batch 運算環境的詳細資訊，請參閱的運算環境 AWS Batch。

您可以自訂儀表板頁面上顯示的小工具。下列各節說明您可以新增的其他小工具。

主題

- 如何將單一任務佇列小工具新增至 AWS Batch 儀表板
- 如何將 CloudWatch Container Insights 小工具新增至 AWS Batch 儀表板
- 如何將任務日誌小工具新增至 AWS Batch 儀表板

如何將單一任務佇列小工具新增至 AWS Batch 儀表板

單一任務佇列小工具會顯示單一任務佇列的詳細資訊。

若要新增此小工具，請遵循下列步驟。

1. 開啟 AWS Batch 主控台。
2. 從導覽列中，選擇您想要 AWS 區域 的 。
3. 在導覽窗格中，選擇 Dashboard (儀表板)。
4. 選擇新增小工具。
5. 針對單一任務佇列，選擇新增小工具。
6. 針對任務佇列，選取您想要的任務佇列。
7. 針對任務狀態，選擇您要顯示的任務狀態。
8. （選用）如果您不想顯示運算環境的屬性，請關閉顯示連線的運算環境。
9. 針對運算環境屬性，選取您想要的屬性。
10. 選擇新增。

如何將 CloudWatch Container Insights 小工具新增至 AWS Batch 儀表板

此小工具會顯示 AWS Batch 運算環境和任務的彙總指標。如需更多 Container Insights 的相關資訊，請參閱 the section called “CloudWatch 容器洞見”。

若要新增此小工具，請遵循下列步驟。

1. 開啟 AWS Batch 主控台。
2. 從導覽列中，選擇您想要 AWS 區域 的 。
3. 在導覽窗格中，選擇 Dashboard (儀表板)。
4. 選擇新增小工具。
5. 如需容器洞見，請選擇新增小工具。
6. 針對運算環境，選擇您想要的運算環境。
7. 選擇新增。

如何將任務日誌小工具新增至 AWS Batch 儀表板

此小工具會在一個方便的位置顯示與您的任務不同的日誌。如需任務日誌的詳細資訊，請參閱 the section called “在 CloudWatch Logs 中檢視任務日誌”。

若要新增此小工具，請遵循下列步驟。

1. 開啟 AWS Batch 主控台。
2. 從導覽列中，選擇您想要 AWS 區域 的 。
3. 在導覽窗格中，選擇 Dashboard (儀表板)。
4. 選擇新增小工具。
5. 針對任務日誌，選擇新增小工具。
6. 在任務 ID 中，輸入所需任務的任務 ID。
7. 選擇新增。

的運算環境 AWS Batch

任務佇列會映射至一或多個運算環境。運算環境包含用於執行容器化批次任務的 Amazon ECS 容器執行個體。特定運算環境也可以對應至一或多個任務佇列。在任務佇列中，相關聯的運算環境每個都有一個由排程器用來判斷準備執行的任務執行位置的順序。如果第一個運算環境的狀態為 VALID且具有可用的資源，則任務會排程到該運算環境中的容器執行個體。如果第一個運算環境的狀態為 INVALID或無法提供適當的運算資源，排程器會嘗試在下一個運算環境中執行任務。

主題

- 受管運算環境
- 未受管的運算環境
- 建立運算環境
- 在 中更新運算環境 AWS Batch
- 運算資源 AMI
- 搭配 使用 Amazon EC2 啟動範本 AWS Batch
- 執行個體中繼資料服務 (IMDS) 組態
- EC2 組態
- 的執行個體類型配置策略 AWS Batch
- 運算資源記憶體管理
- Fargate 運算環境
- Amazon EKS 運算環境

受管運算環境

您可以使用受管運算環境來 AWS Batch 管理環境中運算資源的容量和執行個體類型。這是根據您在建立運算環境時定義的運算資源規格。您可以選擇使用 Amazon EC2 隨需執行個體和 Amazon EC2 Spot 執行個體。或者，您也可以在受管運算環境中使用 Fargate 和 Fargate Spot 容量。使用 Spot 執行個體時，您可以選擇設定最高價格。如此一來，Spot 執行個體只會在 Spot 執行個體價格低於隨需價格的指定百分比時啟動。

Important

不支援 Fargate Spot 執行個體 Windows containers on AWS Fargate。如果 FargateWindows 任務提交至僅使用 Fargate Spot 運算環境的任務佇列，則會封鎖任務佇列。

Important

AWS Batch 會代表您和在您的帳戶中建立和管理多個 AWS 資源，包括 Amazon EC2 啟動範本、Amazon EC2 Auto Scaling 群組、Amazon EC2 Spot 機群和 Amazon ECS 叢集。這些受管資源經過專門設定，以確保最佳 AWS Batch 操作。除非文件中明確說明 AWS Batch ，否則手動修改這些批次受管資源可能會導致意外行為，導致INVALID運算環境、執行個體擴展行為不佳、延遲工作負載處理或意外成本。 AWS Batch 服務無法確定是否支援這些手動修改。一律使用支援的 Batch APIs或 Batch 主控台來管理您的運算環境。

受管運算環境會在您指定的 VPC 和子網路中啟動 Amazon EC2 執行個體，然後向 Amazon ECS 叢集註冊它們。Amazon EC2 執行個體需要外部網路存取，才能與 Amazon ECS 服務端點通訊。有些子網路不會為 Amazon EC2 執行個體提供公有 IP 地址。如果您的 Amazon EC2 執行個體沒有公有 IP 地址，則必須使用網路地址轉譯 (NAT) 來取得此存取權。如需詳細資訊，請參閱 《Amazon VPC 使用者指南》中的 NAT 閘道。如需如何建立 VPC 的詳細資訊，請參閱 建立 Virtual Private Cloud。

根據預設， AWS Batch 受管運算環境會使用運算資源的 Amazon ECS 最佳化 AMI 最新核准版本。不過，由於各種原因，您可能想要建立自己的 AMI 以用於受管運算環境。如需詳細資訊，請參閱運算資源 AMI。

Note

AWS Batch 建立 AMIs。例如，發行較新版本的 Amazon ECS 最佳化 AMIs 時，不會更新運算環境中的 AMI。您負責管理訪客作業系統。這包括任何更新和安全性修補程式。您也需要負責您在運算資源上安裝的任何其他應用程式軟體或公用程式。有兩種方式可以為您的 AWS Batch 任務使用新的 AMI。原始方式是完成以下步驟：

1. 新建內有新 AMI 的運算環境。
2. 將運算環境新增至現有的任務佇列。
3. 將較早的運算環境從任務佇列移除。
4. 刪除較早的運算環境。

2022 年 4 月， AWS Batch 新增了更新運算環境的增強支援。如需詳細資訊，請參閱在 中更新運算環境 AWS Batch。若要使用運算環境的增強型更新功能來更新 AMI，請遵循下列規則：

- 請勿設定服務角色 (serviceRole) 參數，或將其設定為 AWSServiceRoleForBatch 服務連結角色。
- 將配置策略 (allocationStrategy) 參數設定為 BEST_FIT_PROGRESSIVE、SPOT_CAPACITY_OPTIMIZED或 SPOT_PRICE_CAPACITY_OPTIMIZED。
- 將更新設定為最新的映像版本 (updateToLatestImageVersion) 參數設定為 true。
- 請勿在 imageId、 imageIdOverride (在 ec2Configuration) 或啟動範本 () 中指定 AMI IDlaunchTemplate。在這種情況下， 會 AWS Batch 選取基礎設施更新啟動 AWS Batch 時 支援的最新 Amazon ECS 最佳化 AMI。或者，您可以在 imageId或 imageIdOverride 參數中指定 AMI ID，或LaunchTemplate屬性識別的啟動範本。變更任何這些屬性會啟動基礎設施更新。如果在啟動範本中指定 AMI ID，則無法在 imageId或 imageIdOverride 參數中指定 AMI ID 來取代 AMI ID。只能透過指定不同的啟動範本來取代它。或者，如果啟動範本版本設定為 $Default或 $Latest，則設定啟動範本的新預設版本 （ 如果是 $Default)，或將新版本新增至啟動範本 （ 如果是)$Latest。

如果遵循這些規則，啟動基礎設施更新的任何更新都會重新選取 AMI ID。如果啟動範本 (launchTemplate) 中的version設定設為 $Latest或 $Default，即使launchTemplate未更新，也會在基礎設施更新時評估啟動範本的最新或預設版本。

建立多節點平行任務時的考量

AWS Batch 建議建立專用運算環境，以執行多節點平行 (MNP) 任務和非 MNP 任務。這是因為運算容量在受管運算環境中的建立方式所致。建立新的受管運算環境時，如果您指定的minvCpu值大於零，則 AWS Batch 會建立執行個體集區，僅用於非 MNP 任務。如果提交多節點平行任務， 會 AWS Batch 建立新的執行個體容量來執行多節點平行任務。如果在同一運算環境中同時有單一節點和多節點平行任務在設定 minvCpus或 maxvCpus值的相同運算環境中執行，如果所需的運算資源不可用，AWS Batch 則在建立執行新任務所需的運算資源之前，會等待目前的任務完成。

未受管的運算環境

在未受管的運算環境中，您可以管理自己的運算資源。 AWS Batch 支援 Amazon ECS 和 Amazon EKS 的未受管運算環境，可讓您維持對基礎設施的控制，同時利用 Batch 的任務排程功能。

> **Note**
>
> AWS 未受管的運算環境不支援 Fargate 資源。

未受管的 Amazon ECS 運算環境

對於未受管的 Amazon ECS 運算環境，您必須驗證您用於運算資源的 AMI 是否符合 Amazon ECS 容器執行個體 AMI 規格。如需詳細資訊，請參閱運算資源 AMI 規格及教學課程：建立運算資源 AMI。

建立未受管的運算環境之後，請使用 DescribeComputeEnvironments API 操作來檢視運算環境詳細資訊。尋找與環境相關聯的 Amazon ECS 叢集，然後在該 Amazon ECS 叢集中手動啟動您的容器執行個體。

下列 AWS CLI 命令也提供 Amazon ECS 叢集 ARN。

```
$ aws batch describe-compute-environments \
    --compute-environments unmanagedCE \
    --query "computeEnvironments[].ecsClusterArn"
```

如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的啟動 Amazon ECS 容器執行個體。當您啟動運算資源時，請指定資源向下列 Amazon EC2 使用者資料註冊的 Amazon ECS 叢集 ARN。將 *ecsClusterArn* 取代為您使用上一個命令取得的叢集 ARN。

```
#!/bin/bash
echo "ECS_CLUSTER=ecsClusterArn" >> /etc/ecs/ecs.config
```

未受管的 Amazon EKS 運算環境

在未受管的 Amazon EKS 運算環境中，您可以管理自己的 Kubernetes 節點，同時 AWS Batch 處理任務排程和配置。它可讓您針對安全、合規或操作需求，直接控制 Kubernetes 基礎設施。您有責任佈建和設定 Amazon EKS 節點，同時 會與您現有的 Amazon EKS 叢集 AWS Batch 整合，以排程和執行任務。

如需詳細資訊，請參閱教學課程：使用 Amazon EKS 資源建立未受管的運算環境。

建立運算環境

您必須先建立運算環境 AWS Batch，才能在 中執行任務。您可以建立受管運算環境，其中 會根據您的規格 AWS Batch 管理環境中的 Amazon EC2 執行個體或 AWS Fargate 資源。或者，您也可以建立未受管的運算環境，在環境中處理 Amazon EC2 執行個體組態。

> **Important**
>
> 下列案例不支援 Fargate Spot 執行個體：
>
> - Windows containers on AWS Fargate
>
> 如果任務提交到只使用 Fargate Spot 運算環境的任務佇列，在這些情況下，任務佇列將會遭到封鎖。

主題

- 教學課程：使用 Fargate 資源建立受管運算環境
- 教學課程：使用 Amazon EC2 資源建立受管運算環境
- 教學課程：使用 Amazon EC2 資源建立未受管的運算環境
- 教學課程：使用 Amazon EKS 資源建立受管運算環境
- 教學課程：使用 Amazon EKS 資源建立未受管的運算環境
- 資源：運算環境範本
- 執行個體類型運算資料表

教學課程：使用 Fargate 資源建立受管運算環境

完成下列步驟，以使用 AWS Fargate 資源建立受管運算環境。

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇 Compute environments (運算環境)。
4. 選擇建立。
5. 設定運算環境。

> **Note**
>
> Windows containers on AWS Fargate 任務的運算環境必須至少有一個 vCPU。

a. 針對運算環境組態，選擇 Fargate。

b. 針對名稱，指定運算環境的唯一名稱。名稱最多可包含 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。

c. 針對服務角色，選擇服務連結角色，讓 AWS Batch 服務代表您呼叫所需的 AWS API 操作。例如，選擇 AWSServiceRoleForBatch。如需詳細資訊，請參閱使用 AWS Batch的服務連結角色。

d. （選用）展開標籤。若要新增標籤，請選擇 Add tag (新增標籤)。然後，輸入金鑰名稱和選用值。選擇 Add tag (新增標籤)。

e. 選擇下一頁。

6. 在執行個體組態區段中：

a. （選用）對於使用 Fargate Spot 容量，開啟 Fargate Spot。如需 Fargate Spot 的相關資訊，請參閱使用 Amazon EC2 Spot 和 Fargate_SPOT。

b. 針對最大 vCPUs，選擇運算環境可以向外擴展的 vCPUs 數量上限，無論任務佇列需求為何。

c. 選擇下一頁。

7. 設定網路。

> **Important**
>
> 運算資源需要存取，才可以與 Amazon ECS 服務端點通訊。可透過介面 VPC 端點或透過具備公有 IP 地址的運算資源來實現。
> 如需介面 VPC 端點的詳細資訊，請參閱 Amazon Elastic Container Service 開發人員指南中的 Amazon ECS 介面 VPC 端點 (AWS PrivateLink)。
> 如果您沒有設定介面 VPC 端點，且運算資源沒有公有 IP 地址，則它們必須使用網路地址轉譯 (NAT) 來提供此存取。如需詳細資訊，請參閱 Amazon VPC 使用者指南中的 NAT 閘道。如需詳細資訊，請參閱the section called “建立 VPC”。

a. 針對虛擬私有雲端 (VPC) ID，選擇您要啟動執行個體的 VPC。

b. 針對子網路，選擇要使用的子網路。根據預設，所選 VPC 內的所有子網路都可用。

> **Note**
>
> AWS Batch Fargate 上的 目前不支援 Local Zones。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Local Zones、Wavelength Zones 和 Amazon ECS 叢集 AWS Outposts。

c. 在 Security groups (安全群組) 中，選擇連接至您的執行個體的安全群組。根據預設，會選擇您的 VPC 預設的安全群組。

d. 選擇下一頁。

8. 針對檢閱，檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立運算環境。

教學課程：使用 Amazon EC2 資源建立受管運算環境

完成下列步驟，使用 Amazon Elastic Compute Cloud (Amazon EC2) 資源建立受管運算環境。

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇 Environments (環境)。
4. 選擇建立環境，然後選擇運算環境。
5. 設定環境。

   a. 針對運算環境組態，選擇 Amazon Elastic Compute Cloud (Amazon EC2)。

   b. 針對協調類型，選擇受管。

   c. 針對名稱，指定運算環境的唯一名稱。名稱最多可包含 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。

   d. 針對服務角色，選擇服務連結角色，讓 AWS Batch 服務代表您呼叫所需的 AWS API 操作。例如，選擇 AWSServiceRoleForBatch。如需詳細資訊，請參閱使用 AWS Batch的服務連結角色。

   e. 在 Instance role (執行個體角色) 中，選擇建立新的執行個體描述檔，或使用附有所需 IAM 許可的現有執行個體描述檔。此執行個體描述檔可讓為運算環境建立的 Amazon ECS 容器執行個體代表您呼叫所需的 AWS API 操作。如需詳細資訊，請參閱Amazon ECS 執行個體角色。如果您選擇建立新的執行個體描述檔，會為您建立所需的角色 (`ecsInstanceRole`)。

   f. （選用）展開標籤。

i. （選用）針對 EC2 標籤，選擇新增標籤，將標籤新增至在運算環境中啟動的資源。然後，輸入金鑰名稱和選用值。選擇 Add tag (新增標籤)。

ii. （選用）針對標籤，選擇新增標籤。然後，輸入金鑰名稱和選用值。選擇 Add tag (新增標籤)。

如需詳細資訊，請參閱標記您的 AWS Batch 資源。

g. 選擇下一步。

6. 在執行個體組態區段中：

a. （選用）對於使用 Spot 執行個體啟用，請開啟 Spot。如需詳細資訊，請參閱 Spot 執行個體。

b. （僅限 Spot) 針對 % 的隨需價格上限，選擇 Spot 執行個體價格與執行個體啟動之前該執行個體類型的隨需價格相比的最大百分比。例如，如果您的最高價格為 20%，則 Spot 價格必須小於該 EC2 執行個體目前隨需價格的 20%。您一律會支付最低價 (市價) 且絕不超過您的最大百分比。如果您將此欄位空，預設值是隨需價格的 100%。

c. （僅限 Spot) 針對 Spot 機群角色，選擇要套用至 Spot 運算環境的現有 Amazon EC2 Spot 機群 IAM 角色。如果您還沒有現有的 Amazon EC2 Spot Fleet IAM 角色，您必須先建立一個角色。如需詳細資訊，請參閱Amazon EC2 Spot 機群角色。

> **Important**
>
> 若要在建立時標記 Spot 執行個體，您的 Amazon EC2 Spot Fleet IAM 角色必須使用較新的 AmazonEC2SpotFleetTaggingRole 受管政策。AmazonEC2SpotFleetRole 受管政策沒有標記 Spot 執行個體所需的許可。如需詳細資訊，請參閱建立時未標記 Spot 執行個體及the section called “標記您的 資源”。

d. 針對最小 vCPUs，選擇運算環境維護的最小 vCPUs 數量，無論任務佇列需求為何。

e. 針對所需的 vCPUs，選擇運算環境啟動的 vCPUs 數量。隨著任務佇列需求增加， AWS Batch 可以增加運算環境的所需 vCPU 數，並新增 EC2 執行個體 (最多達最大 vCPU 數)。隨著需求減少， AWS Batch 可以減少運算環境的所需 vCPU 數，並移除執行個體 (最少可達最小 vCPU 數)。

f. 針對最大 vCPUs，選擇運算環境可以擴展的 vCPUs 數量上限，無論任務佇列需求為何。

g. 針對允許的執行個體類型，選擇可以啟動的 Amazon EC2 執行個體類型。您可以指定執行個體系列來啟動這些系列中的任何執行個體類型（例如 c5、 c5n或 p3)。或者，您可以指定

系列中的特定大小（例如 c5.8xlarge)。金屬執行個體類型不在執行個體系列中。例如，c5不包含 c5.metal。

AWS Batch 如果您選擇下列其中一項， 可以為您選取執行個體類型：

- optimal 選取符合您任務佇列需求的執行個體類型（來自 c4、m4、m5、、 r4 c5和 r5執行個體系列）。
- default_x86_64 選擇符合任務佇列資源需求的 x86 型執行個體類型（來自 m6i、r6i、c6i和 c7i執行個體系列）。
- default_arm64 選擇符合任務佇列資源需求的 x86 型執行個體類型（來自 m6g、r6g、c6g和 c7g執行個體系列）。

> **Note**
>
> 從 11/01/2025 開始， optimal 的行為將變更為符合 default_x86_64。在變更期間，您的執行個體系列可能會更新為較新一代。您不需要執行任何動作，即可進行升級。如需變更的詳細資訊，請參閱
>
> > **Note**
> >
> > 從 11/01/2025 開始， optimal 的行為將變更為符合 default_x86_64。在變更期間，您的執行個體系列可能會更新為較新一代。您不需要執行任何動作，即可進行升級。
>
> AWS Batch 支援 的 instanceTypes 中的單一選項optimal，以符合任務佇列的需求。我們已推出兩個新的執行個體類型選項：default_x86_64和 default_arm64。default_x86_64 如果您未選擇執行個體類型，我們將使用 。這些新選項會根據您的任務佇列需求，自動選取不同系列和世代之間具成本效益的執行個體類型，讓您快速執行工作負載。
>
> 當 中有足夠容量的新執行個體類型可用時 AWS 區域，對應的預設集區會自動更新為新的執行個體類型。現有optimal選項將繼續受到支援，而且不會遭到取代，因為基礎預設集區會支援它，以提供後續更新的執行個體。如果您使用的是 'optimal，則不需要採取任何動作。

不過，請注意，系統只會使用新的執行個體類型自動更新 ENABLED和 VALID 運算環境 (CEs)。如果您有任何 DISABLED或 INVALID CEs，它們會在重新啟用並設為 VALID 狀態時收到更新。。

Note

- 執行個體系列可用性因 而異 AWS 區域。例如，有些 AWS 區域可能沒有任何第四代執行個體系列，但有第五代和第六代執行個體系列。
- 使用 `default_x86_64`或 `default_arm64` 執行個體套件時， AWS Batch 會根據成本效益和效能的平衡來選取執行個體系列。雖然較新一代的執行個體通常提供更好的價格效能，但如果它為您的工作負載提供最佳的可用性、成本和效能組合， AWS Batch 則可以選擇較新一代的執行個體系列。例如，在同時提供 c6i 和 c7i 執行個體 AWS 區域 的 中，如果 c6i 執行個體為您的特定任務需求提供更好的成本效益， AWS Batch 則可能會選取 c6i 執行個體。如需 AWS Batch 執行個體類型和 AWS 區域 可用性的詳細資訊，請參閱執行個體類型運算資料表。
- AWS Batch 會定期將預設套件中的執行個體更新為較新的、更具成本效益的選項。更新會自動進行，而不需要您採取任何動作。您的工作負載會在更新期間繼續執行，而不會中斷。

Note

在建立運算環境時，您為其選取的執行個體類型必須共用相同架構。例如，您無法在相同的運算環境中混合使用 x86 和 ARM 執行個體。

Note

AWS Batch 會根據任務佇列中所需的數量來擴展 GPUs。若要使用 GPU 排程，運算環境必須包含來自 p3、p4、p5、p6、g3、、 g3s g4 g5或 g6系列的執行個體類型。

h. 如為配置策略，從允許的執行個體類型清單中選取執行個體類型時，選取要使用的配置策略。BEST_FIT_PROGRESSIVE 通常是 EC2 隨需運算環境、SPOT_CAPACITY_OPTIMIZED 和 SPOT_PRICE_CAPACITY_OPTIMIZED for EC2 Spot 運算環境的最佳選擇。如需詳細資訊，請參閱the section called “執行個體類型配置策略”。

i. 展開 Additional configuration (其他組態)。

i. （選用）對於置放群組，輸入置放群組名稱，以在運算環境中將資源分組。

ii. （選用）對於 EC2 金鑰對，當您連線到執行個體時，請選擇公有和私有金鑰對做為安全登入資料。如需 Amazon EC2 金鑰對的詳細資訊，請參閱 Amazon EC2 金鑰對和 Linux 執行個體。

iii. （選用）對於 EC2 組態，選擇映像類型和映像 ID 覆寫值，以提供資訊給 AWS Batch ，以為運算環境中的執行個體選取 Amazon Machine Image (AMIs)。如果未為每個映像類型指定映像 ID 覆寫， 會 AWS Batch 選取最近的 Amazon ECS 最佳化 AMI。 如果未指定映像類型，則預設為非 GPU、非 Graviton 執行個體的 Amazon Linux 2。 AWS

> **Important**
>
> 若要使用自訂 AMI，請選擇影像類型，然後在影像 ID 覆寫方塊中輸入自訂 AMI ID。

Amazon Linux 2

所有以 AWS Graviton 為基礎的執行個體系列（例如 C6g、R6g、 M6g和 T4g) 的預設值，可用於所有非 GPU 執行個體類型。

Amazon Linux 2 (GPU)

所有 GPU 執行個體系列（例如 P4和 G4) 的預設值，可用於所有非 AWS Graviton 型執行個體類型。

Amazon Linux 2023

AWS Batch 支援 Amazon Linux 2023。

> **Note**
>
> Amazon Linux 2023 不支援A1執行個體。

Amazon Linux 2023 (GPU)

所有 GPU 執行個體系列 （例如 P4和 G4) 的預設值，可用於所有非 AWS Graviton 型執行個體類型。

> **Note**
>
> 您為運算環境選擇的 AMI 必須符合您打算用於該運算環境之執行個體類型的架構。例如，如果您的運算環境使用 A1 執行個體類型，則您所選擇的運算資源 AMI 必須支援 ARM 執行個體。Amazon ECS 同時提供 Amazon ECS 最佳化 Amazon Linux 2 AMI 的 x86 和 ARM 版本。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 最佳化 Amazon Linux 2 AMI。

j. （選用）展開啟動範本

i. 針對預設啟動範本，選取現有的 Amazon EC2 啟動範本來設定您的運算資源。範本的預設版本會自動填入。如需詳細資訊，請參閱搭配 使用 Amazon EC2 啟動範本 AWS Batch。

> **Note**
>
> 在啟動範本中，您可以指定您建立的自訂 AMI。

ii. （選用）對於預設版本，輸入 `$Default`、 `$Latest`或要使用的特定版本編號。

> **Note**
>
> 注意：如果您使用替代變數 ($Default 或 $Latest)，它們會在儲存此組態時套用目前的預設或最新版本編號。如果預設或最新版本在未來變更，您必須更新資訊 - 不會自動更新。

> **Important**
>
> 如果啟動範本的版本參數為 `$Default`或 `$Latest`，則會在基礎設施更新期間評估指定啟動範本的預設或最新版本。如果預設選取不同的 AMI ID，或選取最新版本的啟動範本，則會在更新中使用該 AMI ID。如需詳細資訊，請參閱the section called “基礎設施更新期間的 AMI 選擇”。

iii. （選用）對於覆寫啟動範本，選擇新增覆寫啟動範本

A. （選用）針對啟動範本，選取要用於特定執行個體類型和系列的現有 Amazon EC2 啟動範本。

B. （選用）對於預設版本，輸入要使用的特定版本編號、 `$Default`或 `$Latest`。

> **Note**
>
> 如果您使用 `$Default`或 `$Latest`變數， AWS Batch 會在建立運算環境時套用目前資訊。如果預設或最新版本在未來變更，您必須透過 UpdateComputeEnvironment 或 AWS 管理主控台 - 更新資訊 AWS Batch。

C. （選用）針對目標執行個體類型，選取您要套用覆寫啟動範本的執行個體類型或系列。

> **Note**
>
> 如果您指定覆寫啟動範本，則需要目標執行個體類型。如需詳細資訊，請參閱 LaunchTemplateSpecificationOverride.targetInstanceTypes。

> **Note**
>
> 如果您想要選取的執行個體類型或系列未出現在此清單中，請檢閱您在 中進行的選擇`Allowed instance types`。

k. 選擇下一步。

7. 在網路組態區段中：

> **Important**
>
> 運算資源需要存取，才可以與 Amazon ECS 服務端點通訊。可透過介面 VPC 端點或透過具備公有 IP 地址的運算資源來實現。
> 如需介面 VPC 端點的詳細資訊，請參閱 Amazon Elastic Container Service 開發人員指南中的 Amazon ECS 介面 VPC 端點 (AWS PrivateLink)。
> 如果您沒有設定介面 VPC 端點，且運算資源沒有公有 IP 地址，則它們必須使用網路地址轉譯 (NAT) 來提供此存取。如需詳細資訊，請參閱 Amazon VPC 使用者指南中的 NAT 閘道。如需詳細資訊，請參閱the section called “建立 VPC”。

a. 對於虛擬私有雲端 (VPC) ID，請選擇要啟動執行個體的 VPC。

b. 對於子網路，選擇要使用的子網路。根據預設，所選 VPC 內的所有子網路都可用。

> **Note**
>
> AWS Batch Amazon EC2 上的 支援 Local Zones。如需詳細資訊，請參閱 Amazon EC2 使用者指南中的 Local Zones 和 Local Zones、Wavelength Zones 中的 Amazon ECS 叢集，以及 Amazon Elastic Container Service 開發人員指南 AWS Outposts中的 Local Zones。 Amazon EC2

c. （選用）對於安全群組，選擇要連接到執行個體的安全群組。根據預設，會選擇您的 VPC 預設的安全群組。

> **Note**
>
> 注意：如果您使用替代變數 ($Default 或 $Latest)，它們會在儲存此組態時套用目前的預設或最新版本編號。如果預設或最新版本在未來變更，您必須更新資訊 - 不會自動更新。

8. 選擇下一頁。

9. 針對檢閱，檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立運算環境。

教學課程：使用 Amazon EC2 資源建立未受管的運算環境

完成下列步驟，使用 Amazon Elastic Compute Cloud (Amazon EC2) 資源建立未受管的運算環境。

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在運算環境頁面上，選擇建立。
4. 設定環境。
   a. 針對運算環境組態，選擇 Amazon Elastic Compute Cloud (Amazon EC2)。
   b. 針對協調類型，選擇未受管。
5. 針對名稱，指定運算環境的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
6. 針對服務角色，選擇可讓 AWS Batch 服務代表您呼叫所需 AWS API 操作的角色。

   > **Note**
   >
   > 您無法將 `AWSServiceRoleForBatch` 用於未受管的運算環境。

7. 針對最大 vCPUs，選擇運算環境可以擴展的 vCPUs 數量上限，無論任務佇列需求為何。
8. （選用）展開標籤。若要新增標籤，請選擇 Add tag (新增標籤)。然後，輸入金鑰名稱和選用值。選擇 Add tag (新增標籤)。如需詳細資訊，請參閱標記您的 AWS Batch 資源。
9. 選擇下一頁。
10. 針對檢閱，請檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立運算環境。

教學課程：使用 Amazon EKS 資源建立受管運算環境

完成下列步驟，使用 Amazon Elastic Kubernetes Service (Amazon EKS) 資源建立受管運算環境。

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇 Compute environments (運算環境)。
4. 選擇建立。
5. 針對運算環境組態，選擇 Amazon Elastic Kubernetes Service (Amazon EKS)。
6. 針對名稱，指定運算環境的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
7. 針對執行個體角色，選擇已連接必要 IAM 許可的現有執行個體描述檔。

> **Note**
>
> 若要在 AWS Batch 主控台中建立運算環境，請選擇具有 `eks:ListClusters`和 `eks:DescribeCluster`許可的執行個體描述檔。

8. 針對 EKS 叢集，選擇現有的 Amazon EKS 叢集。
9. 在命名空間中，輸入Kubernetes命名空間來將叢集中的程序分組 AWS Batch 。
10. （選用）展開標籤。選擇新增標籤，然後輸入鍵/值對。
11. 選擇下一頁。
12. （選用）對於使用 EC2 Spot 執行個體，開啟啟用使用 Spot 執行個體來使用 Amazon EC2 Spot 執行個體。
13. （僅限 Spot) 針對 % 的隨需價格上限，選擇 Spot 執行個體價格與執行個體啟動之前該執行個體類型的隨需價格相比的最大百分比。例如，如果您的最高價格為 20%，則 Spot 價格必須小於該 EC2 執行個體目前隨需價格的 20%。您一律會支付最低價 (市價) 且絕不超過您的最大百分比。如果您將此欄位空，預設值是隨需價格的 100%。
14. （僅限 Spot) 針對 Spot 機群角色，選擇SPOT運算環境的 Amazon EC2 Spot 機群 IAM 角色。

> **Important**
>
> 如果配置策略設定為 `BEST_FIT` 或未指定，則需要此角色。

15. （選用）對於最小 vCPUs，選擇運算環境維護的最小 vCPUs 數量，無論任務佇列需求為何。
16. （選用）針對最大 vCPUs，選擇運算環境可以向外擴展的 vCPUs 數量上限，無論任務佇列需求為何。
17. 針對允許的執行個體類型，選擇可以啟動的 Amazon EC2 執行個體類型。您可以指定執行個體系列來啟動這些系列中的任何執行個體類型（例如 `c5`、 `c5n`或 `p3`)。或者，您可以指定系列中的特定大小（例如 `c5.8xlarge`)。金屬執行個體類型不在執行個體系列中。例如， `c5`不包含 `c5.metal`。

    AWS Batch 如果您選擇下列其中一項， 可以為您選取執行個體類型：

    - `optimal` 選取符合您任務佇列需求的執行個體類型（來自 `c4`、`m4`、`m5`、、 `r4` `c5`和 `r5`執行個體系列）。
    - `default_x86_64` 選擇符合任務佇列資源需求的 x86 型執行個體類型（來自 m6i、r6i、 c6i和 c7i執行個體系列）。

- `default_arm64` 選擇符合任務佇列資源需求的 x86 型執行個體類型 （ 來自 m6g、r6g、 c6g 和 c7g執行個體系列）。

> **Note**
>
> 從 11/01/2025 開始， `optimal` 的行為將變更為符合 `default_x86_64`。在變更期間，您的執行個體系列可能會更新為較新一代。您不需要執行任何動作，即可進行升級。如需變更的詳細資訊，請參閱 接收自動執行個體系列更新的最佳執行個體類型組態。

> **Note**
>
> - 執行個體系列可用性因 而異 AWS 區域。例如，有些 AWS 區域可能沒有任何第四代執行個體系列，但有第五代和第六代執行個體系列。
> - 使用 `default_x86_64`或 `default_arm64` 執行個體套件時， 會根據成本效益和效能的平衡來 AWS Batch 選取執行個體系列。雖然較新一代的執行個體通常提供更好的價格效能，但如果它為您的工作負載提供最佳的可用性、成本和效能組合， AWS Batch 則可以選擇較新一代的執行個體系列。例如，在同時提供 c6i 和 c7i 執行個體 AWS 區域的 中，如果 c6i 執行個體為您的特定任務需求提供更好的成本效益， AWS Batch 則可能會選取 c6i 執行個體。如需 AWS Batch 執行個體類型和 AWS 區域 可用性的詳細資訊，請參閱執行個體類型運算資料表。
> - AWS Batch 會定期將預設套件中的執行個體更新為較新的、更具成本效益的選項。更新會自動執行，而不需要您採取任何動作。您的工作負載會在更新期間繼續執行，而不會中斷

> **Note**
>
> 在建立運算環境時，您為其選取的執行個體類型必須共用相同架構。例如，您無法在相同的運算環境中混合使用 x86 和 ARM 執行個體。

> **Note**
>
> AWS Batch 會根據任務佇列中所需的數量來擴展 GPUs。若要使用 GPU 排程，運算環境必須包含來自 p3、p4、p5、、p6、g3、 g3s g4 g5或 g6系列的執行個體類型。

18. （選用）展開其他組態。

    a. （選用）對於置放群組，輸入置放群組名稱以將運算環境中的資源分組。

    b. 針對配置策略，選擇 BEST_FIT_PROGRESSIVE。

    c. （選用）對於 Amazon Machine Image AMIs) 組態，選擇新增 amazon Machine Image (amis) 組態。

       您可以使用 Amazon EKS 最佳化 Amazon Linux AMI 或自訂 AMI。

       i. 若要使用 Amazon EKS 最佳化的 Amazon Linux AMI：

          A. 針對映像類型，選擇下列其中一項：

             - Amazon Linux 2：所有 AWS 以 Graviton 為基礎的執行個體系列（例如 C6g、R6g、 M6g和 T4g) 的預設值，可用於所有非 GPU 執行個體類型。
             - Amazon Linux 2（加速）：所有 GPU 執行個體系列（例如 P4和 G4) 的預設 ，可用於所有非 AWS Graviton 型執行個體類型。
             - Amazon Linux 2023： AWS Batch 支援 Amazon Linux 2023 (AL2023)。
             - Amazon Linux 2023（加速）：GPU 執行個體系列，可用於所有非 AWS Graviton 型執行個體類型。

          B. 對於Kubernetes版本，請在Kubernetes版本編號中輸入 。

       ii. 若要使用自訂 AMI：

          A. 針對映像類型，選擇自訂 AMI 依據的 AMI 類型：

             - Amazon Linux 2：所有以 AWS Graviton 為基礎的執行個體系列（例如 C6g、R6g、 M6g和 T4g) 的預設值，可用於所有非 GPU 執行個體類型。
             - Amazon Linux 2（加速）：所有 GPU 執行個體系列（例如 P4和 G4) 的預設 ，可用於所有非 AWS Graviton 型執行個體類型。
             - Amazon Linux 2023： AWS Batch 支援 AL2023。

- Amazon Linux 2023（加速）：GPU 執行個體系列，可用於所有非 AWS Graviton 型執行個體類型。

B. 針對映像 ID 覆寫，輸入自訂 AMI ID。

C. 對於Kubernetes版本，請輸入Kubernetes版本編號。

d. （選用）針對啟動範本，選擇現有的啟動範本。

e. （選用）針對啟動範本版本，輸入 **$Default**、 **$Latest**或版本號碼。

f. （選用）對於覆寫啟動範本，若要新增覆寫，請選擇新增覆寫啟動範本：

i. （選用）針對啟動範本，選擇要新增覆寫的啟動範本。

ii. （選用）針對啟動範本版本，選擇啟動範本的版本編號 `$Default`或 `$Latest`。

iii. （選用）針對目標執行個體類型，選擇應套用此覆寫的執行個體類型或系列。這只能以允許執行個體類型中包含的執行個體類型和系列為目標。

iv. （選用）針對 userdataType，選擇 EKS 節點初始化。只有在啟動範本或啟動範本覆寫中指定了 AMI 時，才能使用此欄位。針對以 `EKS_AL2023`或 為基礎的自訂 AMIs 選擇 EKS_NODEADM，`EKS_AL2023_NVIDIA`或針對 `EKS_AL2`和 選擇 EKS_BOOSTRAP_SHEKS_AL_`NVIDIA`。預設值為 EKS_BOOSTRAP_SH。

當您在相同運算環境中同時使用 AL2 和 AL2023-based自訂 AMIs 的混合環境時，您會使用 userdataType。

19. 選擇下一頁。
20. 對於虛擬私有雲端 (VPC) ID，請選擇要啟動執行個體的 VPC。
21. 對於子網路，選擇要使用的子網路。根據預設，所選 VPC 內的所有子網路都可用。

> **Note**
>
> AWS Batch Amazon EKS 上的 支援 Local Zones。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS 和 AWS 本地區域。

22. （選用）對於安全群組，選擇要連接到執行個體的安全群組。預設會選取 VPC 的預設安全群組。
23. 選擇下一頁。
24. 針對檢閱，請檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立運算環境。

教學課程：使用 Amazon EKS 資源建立未受管的運算環境

完成下列步驟，使用 Amazon Elastic Kubernetes Service (Amazon EKS) 資源建立未受管的運算環境。

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從頁面頂端的導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇 Compute environments (運算環境)。
4. 選擇建立。
5. 設定環境。
   a. 針對運算環境組態，選擇 Amazon Elastic Kubernetes Service (Amazon EKS)。
   b. 針對協調類型，選擇未受管。
6. 針對名稱，指定運算環境的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
7. 針對 EKS 叢集，選擇現有的 Amazon EKS 叢集。若要建立新的 EKS 叢集，請遵循建立 Amazon EKS 叢集頁面上的步驟。
8. 在命名空間中，輸入Kubernetes命名空間以將叢集中的程序分組 AWS Batch 。
9. （選用）針對最大 vCPUs，指定您佈建容量中可用於任務排程的 vCPUs 數目上限。
10. （選用）展開標籤。選擇新增標籤，然後輸入鍵/值對。
11. 選擇下一頁。
12. 針對檢閱，檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立運算環境。

將 Amazon EKS 叢集節點指派給未受管的運算環境

建立未受管運算環境後，您需要使用運算環境 UUID 標記 Amazon EKS 節點。
首先，從 `DescribeComputeEnvironments` API 結果取得運算環境 UUID：

```
$ aws batch describe-compute-environments \
    --compute-environments unmanagedEksCE \
    --query "computeEnvironments[].{name: computeEnvironmentName, uuid: uuid}"
```

取得節點資訊：

```
kubectl get nodes -o name
```

使用 AWS Batch 運算環境 UUID 標記節點：

```
kubectl label <node-name> batch.amazonaws.com/compute-environment-uuid=uuid
```

資源：運算環境範本

下列範例顯示空的運算環境範本。您可以使用此範本來建立運算環境，然後將其儲存至 檔案，並搭配 AWS CLI `--cli-input-json`選項使用。如需這些參數的詳細資訊，請參閱 AWS Batch API 參考中的 CreateComputeEnvironment。

Note

您可以使用下列 AWS CLI 命令產生運算環境範本。

```
$ aws batch create-compute-environment --generate-cli-skeleton
```

```
{
    "computeEnvironmentName": "",
    "type": "UNMANAGED",
    "state": "DISABLED",
    "unmanagedvCpus": 0,
    "computeResources": {
        "type": "EC2",
        "allocationStrategy": "BEST_FIT_PROGRESSIVE",
        "minvCpus": 0,
        "maxvCpus": 0,
        "desiredvCpus": 0,
        "instanceTypes": [
            ""
        ],
        "imageId": "",
        "subnets": [
            ""
        ],
        "securityGroupIds": [
            ""
        ],
```

```
        "ec2KeyPair": "",
        "instanceRole": "",
        "tags": {
            "KeyName": ""
        },
        "placementGroup": "",
        "bidPercentage": 0,
        "spotIamFleetRole": "",
        "launchTemplate": {
            "launchTemplateId": "",
            "launchTemplateName": "",
            "version": ""
        },
        "ec2Configuration": [
            {
                "imageType": "",
                "imageIdOverride": "",
                "imageKubernetesVersion": ""
            }
        ]
    },
    "serviceRole": "",
    "tags": {
        "KeyName": ""
    },
    "eksConfiguration": {
        "eksClusterArn": "",
        "kubernetesNamespace": ""
    }
}
```

執行個體類型運算資料表

下表列出 AWS 區域、執行個體系列關鍵字和可用的執行個體系列。 AWS Batch 會嘗試從最新的系列配置執行個體，但因為執行個體系列可用性會因 AWS 區域 您取得較早的執行個體系列產生而有所不同。

default_x86_64

區域	執行個體系列
AWS 區域支援的所有 AWS Batch	m6i、c6i、r6i

區域	執行個體系列
	c7i

default_arm64

區域	執行個體系列
AWS 區域支援的所有 AWS Batch	m6g、c6g、r6g c7g

最佳

區域	執行個體系列
• ap-northeast-1 • ap-northeast-2 • ap-south-1 • ap-southeast-1 • ap-southeast-2 • ca-central-1 • cn-north-1 • cn-northwest-1 • eu-central-1 • eu-west-1 • eu-west-2 • sa-east-1 • us-east-1 • us-east-2 • us-gov-west-1 • us-west-1 • us-west-2	m4、c4、r4

區域	執行個體系列
• af-south-1 • ap-east-1 • ap-northeast-3 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ca-west-1 • eu-central-2 • eu-north-1 • eu-south-1 • eu-south-2 • eu-west-3 • il-central-1 • me-central-1 • me-south-1 • us-gov-east-1 • us-isob-east-1 • us-iso-east-1 • us-isof-south-1 • us-isof-east-1 • eu-isoe-west-1 • us-northeast-1	m5、c5、r5
• ap-southeast-5 • ap-southeast-7 • ap-east-2 • mx-central-1	m6、c6、r6

在 中更新運算環境 AWS Batch

AWS Batch 提供多種策略來更新運算環境，每個都專為特定更新案例和需求而設計。這些方法使用相同的基礎更新 API，但代表有效管理更新的不同規範方法。您可以使用 AWS Batch 主控台或 來管理這些更新 AWS CLI。了解這些策略可協助您選擇最符合您需求的方法，同時將對工作負載的干擾降至最低。

本主題提供可用更新策略的概觀，以及何時使用每種方法的指引。如需詳細程序，請參閱每個更新策略的個別區段。

> **Important**
>
> AWS Batch 會代表您和在您的帳戶中建立和管理多個 AWS 資源，包括 Amazon EC2 啟動範本、Amazon EC2 Auto Scaling 群組、Amazon EC2 Spot 機群和 Amazon ECS 叢集。這些受管資源經過專門設定，以確保最佳 AWS Batch 操作。除非 AWS Batch 文件中明確說明，否則手動修改這些 AWS Batch受管資源可能會導致意外行為，包括`INVALID`運算環境、次佳執行個體擴展行為、延遲工作負載處理或意外成本。 AWS Batch 服務無法確定是否支援這些手動修改。一律使用支援的 AWS Batch APIs或 AWS Batch 主控台來管理您的運算環境。

主題

- 運算環境更新策略
- 選擇正確的更新策略
- 執行擴展更新
- 執行基礎設施更新
- 執行運算環境的藍/綠更新

運算環境更新策略

當您使用擴展或基礎設施更新時，會更新運算環境。對於藍/綠更新策略，您要建立新的運算環境 （綠色），然後將工作負載從舊的運算環境 （藍色） 遷移到新的運算環境 （綠色）。

AWS Batch 提供三種不同的運算環境更新策略：

擴展更新

擴展更新會透過新增或移除執行個體來調整運算環境的容量，而無需取代現有的執行個體。這是最快的更新案例，不需要停機時間。當您需要變更容量設定 (vCPUs時，請使用擴展更新。這些更新通常會在幾分鐘內完成。

Fargate 更新使用與擴展更新相同的程序執行。如需詳細資訊，請參閱執行擴展更新。

基礎設施更新

基礎設施更新會將運算環境中的執行個體取代為具有更新設定的新執行個體。這些更新需要特定的服務角色和配置策略組態，但可提供最短的停機時間，執行中的任務可能會中斷。當您需要修改執行個體類型、AMI 組態、聯網設定、服務角色、環境狀態或其他基礎設施元件時，請使用基礎設施更新。這些更新通常會在 10-30 分鐘內完成，視任務完成而定。

如需詳細資訊，請參閱執行基礎設施更新。

藍/綠更新

藍/綠更新會與現有環境一起建立新的運算環境，允許漸進式工作負載轉換，無需停機。這種方法提供最安全的更新路徑，但需要暫時執行兩個環境。當您需要零停機時間、想要在完全部署之前測試變更、需要快速復原功能，或使用不支援的基礎設施更新組態時，請使用藍/綠更新。完成的時間是可變的，由您控制。

如需詳細資訊，請參閱執行運算環境的藍/綠更新。

選擇正確的更新策略

使用此決策指南來選取最符合您需求的更新策略：

在 時選擇擴展更新

當您只需要調整運算容量 (vCPUs) 時，請選擇擴展更新策略。當您不需要快速更新而不需要停機時間，也不需要基礎設施組態變更時，擴展更新是理想的選擇。

如需詳細程序，請參閱執行擴展更新。

在 時選擇基礎設施更新

當您需要修改執行個體類型、AMI 設定、服務角色、環境狀態或聯網組態時，請選擇基礎設施更新策略。您的環境必須使用 AWSServiceRoleForBatch 服務連結角色和 BEST_FIT_PROGRESSIVE、

SPOT_CAPACITY_OPTIMIZED或 的配置策略SPOT_PRICE_CAPACITY_OPTIMIZED。當在更新期間接受某些任務中斷，且您希望自動更新至最新的 Amazon ECS 最佳化 AMI 時，基礎設施更新運作良好。

如需詳細程序，請參閱執行基礎設施更新。

選擇藍/綠更新

當您的工作負載需要零停機時間，或您需要在轉換生產工作負載之前測試變更時，請選擇藍/綠更新策略。當快速復原功能很重要、您的環境使用BEST_FIT配置策略，或您的環境不使用 AWSServiceRoleForBatch 服務連結角色時，這種方法至關重要。當您使用需要手動更新或需要進行主要組態變更AMIs 時，藍/綠更新也是最佳選擇。

如需詳細程序，請參閱執行運算環境的藍/綠更新。

AMI 更新考量事項

AWS Batch 當符合下列所有條件時， 可以在基礎設施更新期間更新至最新的 Amazon ECS 最佳化 AMI：

> **Note**
>
> 基礎設施更新完成後updateToLatestImageVersion，會將 設定為 false。若要啟動另一個更新，updateToLatestImageVersion必須將 設定為 true。

- 運算環境使用 AWSServiceRoleForBatch 服務連結角色
- 配置策略設定為 BEST_FIT_PROGRESSIVE、 SPOT_CAPACITY_OPTIMIZED或 SPOT_PRICE_CAPACITY_OPTIMIZED
- imageId、 imageIdOverride或 啟動範本中未明確指定 AMI ID
- updateToLatestImageVersion 設定為 true

使用藍/綠部署的 AMI 更新

在這些情況下，您必須使用藍/綠部署來更新 AMIs：

- 使用特定版本的 Amazon ECS 最佳化 AMI 時

- 在下列任何項目中指定 AMI ID 時：
  - 啟動範本（必須更新或移除範本）
  - `imageId` 參數
  - EC2 組態中的 `imageIdOverride` 參數
- 使用`BEST_FIT`配置策略時（不支援基礎設施更新）
- 不使用 AWSServiceRoleForBatch 服務連結角色時

執行擴展更新

擴展更新會透過新增或移除執行個體來調整運算環境的容量。這是最快的更新策略，不需要取代現有的執行個體。擴展更新適用於任何服務角色類型和配置策略，使其成為最靈活的更新選項。

觸發擴展更新的變更

當您僅修改下列設定時， 會 AWS Batch 執行擴展更新。如果您將這些設定與其他運算環境設定一起修改， 會改為 AWS Batch 執行基礎設施更新。

下列設定會在專門修改時觸發擴展更新：

- `desiredvCpus` – 設定環境的 vCPUs目標數量。
- `maxvCpus` – 定義可啟動vCPUs 數量上限。
- `minvCpus` – 指定要維護的 vCPUs數目下限。

對於 Fargate 運算環境，您也可以修改這些設定以進行擴展更新：

- `securityGroupIds` – 運算環境的安全群組 IDs。
- `subnets` – 運算環境的子網路。

> **Note**
>
> 我們建議您不要使用 `desiredvCpus` 啟動擴展更新，因為 AWS Batch 會動態調整 `desiredvCpus`。反之，您應該更新 `minvCpus`。
> 更新 時`desiredvCpus`，值必須介於 `minvCpus`和 之間`maxvCpus`。新值必須大於或等於目前的 `desiredvCpus`。如需詳細資訊，請參閱the section called “更新`desiredvCpus`設定時的錯誤訊息”。

> **Important**
>
> 如果您將這些擴展設定與其他運算環境設定（例如執行個體類型、AMI IDs 或啟動範本）一起修改，會 AWS Batch 執行基礎設施更新，而非擴展更新。基礎設施更新需要更長的時間，而且可能會取代現有的執行個體。

Performing scaling updates using the AWS 管理主控台

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇環境，然後選擇運算環境索引標籤。
3. 選取要更新的運算環境。
4. 選擇動作，然後選擇編輯。
5. 修改支援擴展更新的一或多個設定。例如：
   - 針對最低 vCPUs輸入vCPUs數量。
   - 針對所需的 vCPUs，輸入所需的 vCPUs 數量。
   - 針對最大 vCPUs輸入vCPUs 數量。
6. 選擇儲存變更。
7. 監控運算環境狀態。更新應該快速完成，因為它只涉及擴展操作。

Performing scaling updates using the AWS CLI

使用 update-compute-environment命令來執行擴展更新。下列兩個範例示範常見的擴展操作。您可以修改下列一或多個支援擴展更新的設定

- 此範例會更新所需的 vCPUs、最小值和最大值：

```
aws batch update-compute-environment \
    --compute-environment your-compute-environment-name \
    --compute-resources minvCpus=2,maxvCpus=8
```

監控擴展更新

使用 AWS Batch 主控台監控您的擴展更新，以檢視運算環境狀態，並檢查執行個體計數和 vCPU 指標。您也可以使用 AWS CLI 搭配 describe-compute-environments命令來檢查狀態，並監控執行個體計數和 vCPU 值。

執行基礎設施更新

基礎設施更新會將運算環境中的執行個體取代為具有更新設定的新執行個體。此更新策略需要比擴展更新更長的時間，並且需要特定的服務角色和配置策略設定。基礎設施更新提供一種方法來修改基本運算環境組態，同時保持服務可用性。

> **Important**
>
> 基礎設施更新需要 AWSServiceRoleForBatch 服務連結角色，以及 `BEST_FIT_PROGRESSIVE`、 `SPOT_CAPACITY_OPTIMIZED`或 的配置策略`SPOT_PRICE_CAPACITY_OPTIMIZED`。如果您的環境不符合這些要求，請改用藍/綠更新。

觸發基礎設施更新的變更

當您修改以下任何設定時， 會 AWS Batch 執行基礎設施更新。當您搭配擴展更新設定修改這些設定時，也會發生基礎設施更新。

下列設定會觸發基礎設施更新：

運算組態

- `allocationStrategy` – 決定 如何 AWS Batch 選取執行個體類型。
- `instanceTypes` – 指定要使用的 EC2 執行個體類型。
- `bidPercentage` – Spot 執行個體的隨需價格百分比上限。
- `type` – 運算環境類型 (`EC2` 或 `SPOT`)。

AMI 和啟動組態

- `imageId` – 用於執行個體的特定 AMI。

- ec2Configuration – EC2 組態，包括 imageIdOverride。
- launchTemplate – EC2 啟動範本設定。
- ec2KeyPair – 執行個體存取的 SSH 金鑰對。
- updateToLatestImageVersion – 自動 AMI 更新設定。

網路和安全性

- subnets – 啟動執行個體的 VPC 子網路 （適用於 EC2 運算環境）。
- securityGroupIds – 執行個體的安全群組 （適用於 EC2 運算環境）。
- placementGroup – EC2 置放群組組態。

其他設定

- instanceRole – EC2 執行個體的 IAM 角色。
- tags – 套用至 EC2 執行個體的標籤。

> **Important**
>
> 如果您同時修改任何基礎設施更新設定和擴展更新設定 （例如 desiredvCpus、 或 minvCpus)maxvCpus， 會 AWS Batch 執行基礎設施更新。基礎設施更新需要比擴展更新更長的時間。

基礎設施更新期間的 AMI 選擇

在基礎設施更新期間，運算環境的 AMI ID 可能會變更，取決於是否在這三個設定中的任何一個中指定 AMIs。AMIs是在 imageId (在 中computeResources)、 imageIdOverride (在 中ec2Configuration) 或 中指定的啟動範本中指定launchTemplate。假設沒有在這些設定中指定 AMI IDs，且updateToLatestImageVersion設定為 true。然後， 支援的最新 Amazon ECS 最佳化 AMI AWS Batch 會用於任何基礎設施更新。

如果在其中至少一個設定中指定 AMI ID，則更新取決於提供更新之前使用的 AMI ID 的設定。當您建立運算環境時，選取 AMI ID 的優先順序是啟動範本、imageId設定，最後是imageIdOverride設定。不過，如果使用的 AMI ID 來自啟動範本，則更新 imageId或 imageIdOverride設定不會更新 AMI ID。更新從啟動範本選取的 AMI ID 的唯一方法是更新啟動範本。如果啟動範本的版本參數為

$Default或 $Latest，則會評估指定啟動範本的預設或最新版本。如果預設選取不同的 AMI ID，或選取最新版本的啟動範本，則會在更新中使用該 AMI ID。

如果啟動範本未用於選取 AMI ID，則會使用 imageId或 imageIdOverride 參數中指定的 AMI ID。如果同時指定兩者，則會使用 imageIdOverride 參數中指定的 AMI ID。

假設運算環境使用 imageId、 imageIdOverride或 launchTemplate 參數指定的 AMI ID，而且您想要使用 支援的最新 Amazon ECS 最佳化 AMI AWS Batch。然後，更新必須移除提供 AMI IDs的設定。對於 imageId，這需要為該參數指定空字串。對於 imageIdOverride，這需要為 ec2Configuration 參數指定空字串。

如果 AMI ID 來自啟動範本，您可以變更為下列 AWS Batch 其中一種方式支援的最新 Amazon ECS 最佳化 AMI：

- 為 launchTemplateId或 launchTemplateName 參數指定空字串，以移除啟動範本。這會移除整個啟動範本，而不是單獨移除 AMI ID。
- 如果更新版本的啟動範本未指定 AMI ID，則 updateToLatestImageVersion 參數必須設定為 true。

更新期間的任務處理

使用更新政策，設定在基礎設施更新期間處理執行中任務的方式。當您設定 時terminateJobsOnUpdate=true，執行中的任務會立即終止、忽略 jobExecutionTimeoutMinutes設定，而且只要可以取代執行個體，更新就會立即繼續。當您設定時terminateJobsOnUpdate=false，執行中的任務會在指定的逾時期間內繼續，預設逾時為 30 分鐘，如果超過逾時，任務就會終止。

> **Note**
>
> 若要重試在更新期間終止的任務，請設定任務重試策略。如需詳細資訊，請參閱the section called “自動化任務重試”。

Performing infrastructure updates using the AWS 管理主控台

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇環境，然後選擇運算環境索引標籤。
3. 選取要更新的運算環境。

4. 選擇動作，然後選擇編輯。
5. 在更新行為區段中，設定執行中任務的處理方式：
   - 選擇將 AMI 更新至最新版本，將 AMI 更新至最新版本。
   - 選擇更新時立即終止任務，以在執行更新程序時終止任務。
   - 針對任務執行逾時，輸入開始更新程序之前要等待的分鐘數。
6. 修改需要基礎設施更新的一或多個設定。例如：
   - 執行個體角色
   - 使用 EC2 Spot 執行個體
   - 允許的執行個體類型
   - 置放群組
   - EC2 金鑰對
   - EC2 組態
   - 啟動範本
   - 子網路
   - 安全群組
7. 選擇儲存變更。
8. 監控運算環境狀態。環境會在更新程序UPDATING期間顯示。

Performing infrastructure updates using the AWS CLI

使用 update-compute-environment命令來變更一或多個需要基礎設施更新的設定。以下三個範例是常見的基礎設施操作。

- 此範例會更新執行個體類型並設定更新政策：

```
aws batch update-compute-environment \
    --compute-environment your-compute-environment-name \
    --compute-resources instanceTypes=default_x86_64 \
    --update-policy terminateJobsOnUpdate=false,jobExecutionTimeoutMinutes=30
```

- 此範例會更新 VPC 子網路和安全群組：

```
aws batch update-compute-environment \
    --compute-environment your-compute-environment-name \
```

```
    --compute-resources subnets=subnet-abcd1234,subnet-efgh5678
 securityGroupIds=sg-abcd1234 \
    --update-policy terminateJobsOnUpdate=true
```

- 此範例可自動更新至最新的 Amazon ECS 最佳化 AMI：

```
aws batch update-compute-environment \
    --compute-environment your-compute-environment-name \
    --compute-resources updateToLatestImageVersion=true \
    --update-policy terminateJobsOnUpdate=false,jobExecutionTimeoutMinutes=60
```

監控基礎設施更新

使用 AWS Batch 主控台監控您的基礎設施更新，以監看運算環境狀態變更為 UPDATING、監控執行個體替換進度，以及檢查是否有任何失敗的更新。一旦運算環境狀態為 ，更新就會成功VAILD。您也可以使用 CloudWatch 追蹤執行個體終止事件，並在更新期間監控任務狀態。使用 AWS CLI，使用 describe-compute-environments命令來檢查狀態並監控執行個體生命週期事件。

執行運算環境的藍/綠更新

藍/綠更新是一種更新策略，可透過與現有的運算環境 （ 藍色 ） 一起建立新的運算環境 （ 綠色 ） 來減少停機時間和風險。此方法可讓您逐步將工作負載轉換至新環境，同時保持現有環境的運作。藍/綠更新提供最安全的更新路徑，並使用任何服務角色類型或配置策略。

概觀

藍/綠更新提供多種優點，使其非常適合生產環境。它們透過在更新過程中持續執行工作負載，提供零停機時間。此方法可讓您輕鬆復原功能，讓您可以在發生問題時快速還原至原始環境。您可以實作逐步轉換策略，在完全切換生產工作負載之前驗證新環境的效能。此方法也提供絕佳的風險緩解能力，因為原始環境會保持不變並正常運作，直到您選擇將其移除為止。

需要藍/綠更新時

在下列情況下，您必須使用藍/綠更新：

- 當您的運算環境使用BEST_FIT配置策略時 （ 不支援基礎設施更新 ）
- 當您的運算環境不使用 AWSServiceRoleForBatch 服務連結角色時
- 當您需要在不同的服務角色類型之間轉換時

建議藍/綠更新時

藍/綠更新特別建議用於生產環境，其中零停機時間對您的工作負載至關重要。當您在轉換生產工作負載之前需要測試新組態時，此方法可正常運作，確保變更符合您的效能和可靠性需求。當快速復原功能對您的操作很重要時，請選擇藍/綠更新，尤其是當您更新具有重大變更AMIs 時。當您想要在完全承諾變更之前驗證效能特性和行為，以提供更新程序的信心時，此方法也很理想。

先決條件

在執行藍/綠更新之前，請確定您有：

- 建立和管理運算環境的適當 IAM 許可
- 檢視和修改任務佇列設定的存取權
- 為您的任務定義設定的任務重試策略，以處理轉換期間的潛在失敗。如需詳細資訊，請參閱自動化任務重試。
- 新運算環境的 AMI ID。這可以是：
  - Amazon ECS 最佳化 AMI 的最新核准版本 （預設為使用）
  - 符合 Amazon ECS 容器執行個體 AMI 規格的自訂 AMI。使用自訂 AMI 時，您可以透過下列其中一種方式指定它：
    - 在 EC2 組態中使用映像 ID 覆寫欄位
    - 在啟動範本中指定它

    如需建立自訂 AMIs的詳細資訊，請參閱 教學課程：建立運算資源 AMI。

建立新環境之前，您需要記錄現有運算環境的組態。您可以使用 AWS 管理主控台 或 來執行此操作 AWS CLI。

> **Note**
>
> 下列程序詳細說明如何執行只變更 AMI 的藍/綠更新。您可以更新新環境的其他設定。

> **Important**
>
> 當您移除舊 （藍色） 運算環境時，這些執行個體上目前執行的任何任務都會失敗，因為執行個體將會終止。在任務定義中設定任務重試策略，以自動處理這些失敗。如需詳細資訊，請參閱自動化任務重試。

對新環境有信心之後：

1. 編輯任務佇列以移除舊的運算環境。
2. 等待舊環境中任何執行中的任務完成。
3. 刪除舊的運算環境。

Performing blue/green updates using the AWS 管理主控台

1. 複製您目前的運算環境
   a. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
   b. 選取您現有的運算環境。
   c. 選擇動作，然後選擇複製。
   d. 在名稱中，輸入新運算環境的唯一名稱。
   e. 選擇下一步。
   f. 在執行個體組態區段中，更新 AMI 設定：
      i. 展開 Additional configuration (其他組態)。
      ii. 對於 EC2 組態，請在映像類型中指定新的 AMI 類型，並在映像 ID 覆寫欄位中指定 AMI ID。
   g. 選擇下一步。
   h. 針對網路組態，選擇下一步。
   i. 檢閱從現有環境自動複製的其他設定。
   j. 選擇建立運算環境。
   k. 等待新的運算環境狀態變成 VALID。
2. 變更任務佇列順序
   a. 在導覽窗格中，選擇任務佇列。
   b. 選取與現有運算環境相關聯的任務佇列。
   c. 選擇編輯。
   d. 在連線運算環境中，新增新的運算環境：
      - 新增訂單編號高於現有環境的新運算環境，以轉換工作負載。
      - 驗證新環境是否正常運作後，您可以提供較低的訂單號碼，使其成為主要環境。

e. 選擇更新任務佇列。

3. 清除

a. 在新環境中監控任務執行，以確保一切如預期般運作。

b. 對新環境有信心之後：

1. 編輯任務佇列以移除舊的運算環境。
2. 等待舊環境中任何執行中的任務完成。
3. 刪除舊的運算環境。

Performing blue/green updates using the AWS CLI

1. 若要使用 取得組態 AWS CLI，請使用下列命令：

```
aws batch describe-compute-environments \
  --compute-environments your-compute-environment-name
```

建立新環境時，請儲存輸出以供參考。

2. 使用現有環境的組態建立新的運算環境，但使用新的 AMI。以下是範例命令結構：

將範例值取代為上一個步驟的實際組態：

```
cat <<EOF > ./blue-green-compute-environment.json
{
  "computeEnvironmentName": "your-new-compute-environment-name",
  "type": "MANAGED",
  "state": "ENABLED",
  "computeResources": {
    "instanceRole": "arn:aws:iam::012345678901:instance-profile/
ecsInstanceRole",
    "type": "EC2",
    "minvCpus": 2,
    "desiredvCpus": 2,
    "maxvCpus": 256,
    "instanceTypes": [
      "optimal"
    ],
    "allocationStrategy": "BEST_FIT_PROGRESSIVE",
    "ec2Configuration": [
```

```
      {
        "imageType": "ECS_AL2023",
        "imageIdOverride": "ami-0abcdef1234567890"
      }
    ],
    "subnets": [,
      "subnet-0abcdef1234567890"
    ],
    "securityGroupIds": [
      "sg-0abcdef1234567890"
    ]
  }
}
EOF
```

```
$ aws batch create-compute-environment --cli-input-json file://./blue-green-compute-environment.json
```

3. 等待新的環境變成可用：

```
aws batch describe-compute-environments \
  --compute-environments your-new-compute-environment-name \
  --query 'computeEnvironments[].status'
```

4. 將新的運算環境新增至您的任務佇列：

```
aws batch update-job-queue \
  --job-queue your-job-queue \
  --compute-environment-order order=1,computeEnvironment=your-existing-environment \
  order=2,computeEnvironment=your-new-compute-environment-name
```

5. 驗證後，請再次更新，讓新的環境成為主要環境：

```
aws batch update-job-queue \
  --job-queue your-job-queue \
  --compute-environment-order order=1,computeEnvironment=your-new-compute-environment-name
```

在舊環境中完成所有任務之後，請停用 ，然後刪除它：

```
aws batch update-compute-environment \
    --compute-environment your-existing-environment \
    --state DISABLED
```

```
aws batch delete-compute-environment \
  --compute-environment your-existing-environment
```

運算資源 AMI

根據預設， AWS Batch 受管運算環境會使用運算資源的 Amazon ECS 最佳化 AMI 最新核准版本。不過，您可能想要建立自己的 AMI，以用於受管和未受管的運算環境。如果您需要以下任何一項，我們建議您建立自己的 AMI：

- 增加 AMI 根磁碟區的儲存體大小或資料磁碟區
- 為支援的 Amazon EC2 執行個體類型新增執行個體儲存磁碟區
- 自訂 Amazon ECS 容器代理程式
- 自訂 Docker
- 設定 GPU 工作負載 AMI，以允許容器存取支援 Amazon EC2 執行個體類型的 GPU 硬體

> **Note**
>
> 建立運算環境之後， AWS Batch 不會升級運算環境中的 AMIs。當有較新版本的 Amazon ECS 最佳化 AMIs 可用時， AWS Batch 也不會更新運算環境中的 AMI。您負責管理訪客作業系統。這包括任何更新和安全性修補程式。您也需要負責您在運算資源上安裝的任何其他應用程式軟體或公用程式。若要為您的 AWS Batch 任務使用新的 AMI，請執行下列動作：
>
> 1. 新建內有新 AMI 的運算環境。
> 2. 將運算環境新增至現有的任務佇列。
> 3. 將較早的運算環境從任務佇列移除。
> 4. 刪除較早的運算環境。
>
> 2022 年 4 月， AWS Batch 新增了更新運算環境的增強支援。如需詳細資訊，請參閱在 中更新運算環境 AWS Batch。若要使用運算環境的增強型更新功能來更新 AMI，請遵循下列規則：

- 請勿設定服務角色 (serviceRole) 參數，或將其設定為 AWSServiceRoleForBatch 服務連結角色。
- 將配置策略 (allocationStrategy) 參數設定為 BEST_FIT_PROGRESSIVE、SPOT_CAPACITY_OPTIMIZED或 SPOT_PRICE_CAPACITY_OPTIMIZED。
- 將更新設定為最新的映像版本 (updateToLatestImageVersion) 參數設定為 true。
- 請勿在 imageId、 imageIdOverride (在 ec2Configuration) 或啟動範本 () 中指定 AMI IDlaunchTemplate。當您未指定 AMI ID 時， 會 AWS Batch 選取在啟動基礎設施更新時 AWS Batch 支援的最新 Amazon ECS 最佳化 AMI。或者，您可以在 imageId或 imageIdOverride 參數中指定 AMI ID。或者，您可以指定LaunchTemplate屬性識別的啟動範本。變更任何這些屬性會啟動基礎設施更新。如果在啟動範本中指定 AMI ID，則無法在 imageId或 imageIdOverride 參數中指定 AMI ID 來取代 AMI ID。只能透過指定不同的啟動範本來取代 AMI ID。如果啟動範本版本設定為 $Default或 $Latest，則可以透過為啟動範本設定新的預設版本 （ 如果 $Default) 或將新版本新增至啟動範本 （ 如果) 來取代 AMI ID$Latest。

如果遵循這些規則，則啟動基礎設施更新的任何更新都會重新選取 AMI ID。如果啟動範本 (launchTemplate) 中的version設定設為 $Latest或 $Default，即使 launchTemplate 未更新，也會在基礎設施更新時評估啟動範本的最新或預設版本。

主題

- 運算資源 AMI 規格
- 教學課程：建立運算資源 AMI
- 使用 GPU 工作負載 AMI
- Amazon Linux 棄用
- Amazon EKS Amazon Linux 2 AMI 棄用
- Amazon ECS Amazon Linux 2 AMI 棄用

運算資源 AMI 規格

基本 AWS Batch 運算資源 AMI 規格包含下列項目：

必要

- 在 HVM 虛擬化類型 AMI 上執行至少 3.10 版 Linux 核心的現代 Linux 發行版本。不支援 Windows 容器。

  > **Important**
  >
  > 多節點平行任務只能在已安裝 `ecs-init`套件的 Amazon Linux 執行個體上啟動的運算資源上執行。我們建議您在建立運算環境時使用預設的 Amazon ECS 最佳化 AMI。您可以不指定自訂 AMI 來執行此操作。如需詳細資訊，請參閱多節點平行任務。

- Amazon ECS 容器代理程式。建議您使用最新的 版本。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的安裝 Amazon ECS 容器代理程式。
- Amazon ECS 容器代理程式啟動時，必須將`awslogs`日誌驅動程式指定為具有 `ECS_AVAILABLE_LOGGING_DRIVERS`環境變數的可用日誌驅動程式。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 容器代理程式組態。
- 至少執行 1.9 版的 Docker 協助程式，以及任何 Docker 執行時間相依性。如需詳細資訊，請參閱 Docker 文件中的檢查執行時間相依性。

  > **Note**
  >
  > 我們建議您使用隨附的 Docker 版本，並使用您使用的對應 Amazon ECS 代理程式版本進行測試。Amazon ECS 在 GitHub 上提供了 Amazon ECS 最佳化 AMI 的 Linux 變體變更記錄。如需詳細資訊，請參閱變更記錄。

建議

- 執行和監控 Amazon ECS 代理程式的初始化和贊助程序。Amazon ECS 最佳化 AMI 使用 `ecs-init` upstart 程序，其他作業系統可能會使用 `systemd`。如需詳細資訊和範例，請參閱《Amazon Elastic Container Service 開發人員指南》中的容器執行個體使用者資料組態指令碼範例。如需有關 `ecs-init` 的詳細資訊，請參閱 `ecs-init`GitHub 上的 專案。受管運算環境至少需要 Amazon ECS 代理程式在開機時啟動。如果 Amazon ECS 代理程式未在運算資源上執行，則無法接受來自的任務 AWS Batch。

Amazon ECS 最佳化 AMI 已預先設定這些要求和建議。我們建議您使用 Amazon ECS 最佳化 AMI 或 Amazon Linux AMI 搭配為運算資源安裝的`ecs-init`套件。如果您的應用程式需要特定作業系統或

Docker 版本，但這些 AMI 尚未提供，請選擇另一個 AMIs。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 最佳化 AMI。

教學課程：建立運算資源 AMI

您可以建立自己的自訂運算資源 AMI，以用於受管和未受管的運算環境。如需說明，請參閱 運算資源 AMI 規格。然後，在建立自訂 AMI 之後，您可以建立運算環境，使用該 AMI，您可以將任務佇列與之建立關聯。最後，開始將任務提交至該佇列。

建立自訂運算資源 AMI

1. 選擇要從中開始的基本 AMI。基本 AMI 必須使用 HVM 虛擬化。基本 AMI 不能是 Windows AMI。

   > **Note**
   >
   > 您為運算環境選擇的 AMI 必須符合您打算用於該運算環境之執行個體類型的架構。例如，如果您的運算環境使用 A1 執行個體類型，則您所選擇的運算資源 AMI 必須支援 ARM 執行個體。Amazon ECS 同時提供 Amazon ECS 最佳化 Amazon Linux 2 AMI 的 x86 和 ARM 版本。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 最佳化 Amazon Linux 2 AMI。

   Amazon ECS 最佳化 Amazon Linux 2 AMI 是受管運算環境中運算資源的預設 AMI。Amazon ECS 最佳化 Amazon Linux 2 AMI 由 AWS Batch AWS 工程師預先設定並在 上進行測試。這是一個最小的 AMI，您可以開始使用 並 AWS 快速取得正在執行的運算資源。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS Optimized AMI。

   或者，您可以選擇另一個 Amazon Linux 2 變體，並使用下列命令安裝`ecs-init`套件。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的在 Amazon Linux 2 EC2 執行個體上安裝 Amazon ECS 容器代理程式：

   ```
   $ sudo amazon-linux-extras disable docker
   $ sudo amazon-linux-extras install ecs-init
   ```

   例如，如果您想要在 AWS Batch 運算資源上執行 GPU 工作負載，您可以從 Amazon Linux Deep Learning AMI 開始。然後，設定 AMI 以執行 AWS Batch 任務。如需詳細資訊，請參閱使用 GPU 工作負載 AMI。

> **Important**
>
> 您可以選擇不支援`ecs-init`套件的基本 AMI。不過，如果您這麼做，則必須設定在開機時啟動 Amazon ECS 代理程式並保持執行的方式。您也可以檢視數個用於`systemd`啟動和監控 Amazon ECS 容器代理程式的使用者資料組態指令碼範例。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的容器執行個體使用者資料組態指令碼範例。

2. 使用適用於 AMI 的適當儲存選項，從您選取的基礎 AMI 啟動執行個體。您可以設定連接 Amazon EBS 磁碟區的大小和數量，或者如果您選取的執行個體類型支援，則可以設定執行個體儲存磁碟區。如需詳細資訊，請參閱《Amazon EC2 使用者指南》中的啟動執行個體和 Amazon EC2 執行個體存放區。 Amazon EC2 Amazon EC2
3. 使用 連線至您的執行個體，SSH並執行任何必要的組態任務。這可能包括下列任何或所有步驟：

   - 安裝 Amazon ECS 容器代理程式。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的安裝 Amazon ECS 容器代理程式。
   - 設定指令碼，以設置執行個體存放區磁碟區的格式。
   - 將執行個體存放區磁碟區或 Amazon EFS 檔案系統新增至 `/etc/fstab` 檔案，以便在開機時掛載。
   - 設定 Docker 選項，例如啟用偵錯或調整基礎映像大小。
   - 安裝套件或複製檔案。

   如需詳細資訊，請參閱《Amazon EC2 使用者指南》中的使用 SSH 連線至 Linux 執行個體。
4. 如果您在執行個體上啟動 Amazon ECS 容器代理程式，您必須在建立 AMI 之前停止它並移除任何持久性資料檢查點檔案。否則，如果您不這樣做，代理程式不會在從您的 AMI 啟動的執行個體上啟動。

   a. 停用 Amazon ECS 容器代理程式。

      - Amazon ECS 最佳化 Amazon Linux 2 AMI：

        ```
        sudo systemctl stop ecs
        ```

      - Amazon ECS 最佳化 Amazon Linux AMI：

        ```
        sudo stop ecs
        ```

b. 移除持久性資料檢查點檔案。根據預設，這些檔案位於 /var/lib/ecs/data/目錄中。如果有的話，請使用下列命令來移除這些檔案。

```
sudo rm -rf /var/lib/ecs/data/*
```

5. 從執行中的執行個體建立新的 AMI。如需詳細資訊，請參閱《Amazon EC2 使用者指南》中的建立 Amazon EBS 支援的 Linux AMI。 Amazon EC2

使用新的 AMI 搭配 AWS Batch

1. 建立新的 AMI 之後，請使用新的 AMI 建立運算環境。若要這樣做，請在建立 AWS Batch 運算環境時，選擇映像類型，然後在映像 ID 覆寫方塊中輸入自訂 AMI ID。如需詳細資訊，請參閱the section called “教學課程：使用 Amazon EC2 資源建立受管運算環境”。

> **Note**
>
> 您為運算環境選擇的 AMI 必須符合您打算用於該運算環境之執行個體類型的架構。例如，如果您的運算環境使用 A1 執行個體類型，則您所選擇的運算資源 AMI 必須支援 ARM 執行個體。Amazon ECS 同時提供 Amazon ECS 最佳化 Amazon Linux 2 AMI 的 x86 和 ARM 版本。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 最佳化 Amazon Linux 2 AMI。

2. 建立任務佇列，並與新的運算環境建立關聯。如需詳細資訊，請參閱建立任務佇列。

> **Note**
>
> 與任務佇列關聯的所有運算環境皆必須共用相同的架構。 AWS Batch 不支援在單一任務佇列中混用運算環境架構類型。

3. (選用) 將範例任務提交到新的任務佇列。如需詳細資訊，請參閱任務定義範例、建立單一節點任務定義及教學課程：提交任務。

使用 GPU 工作負載 AMI

若要在 AWS Batch 運算資源上執行 GPU 工作負載，您必須使用 AMI 搭配 GPU 支援。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的在 Amazon ECS 上使用 GPUs 和 Amazon ECS 最佳化 AMI。 AMIs

在受管運算環境中，如果運算環境指定任何 p3、p4、p5、p6g3、g5、、 g3s g4或 g6執行個體類型或執行個體系列，則 AWS Batch 會使用 Amazon ECS GPU 最佳化 AMI。

在未受管的運算環境中，建議使用 Amazon ECS GPU 最佳化 AMI。您可以使用 AWS Command Line Interface 或 AWS Systems Manager 參數存放區 GetParameter、GetParameters 和 GetParametersByPath 操作來擷取建議的 Amazon ECS GPU 最佳化 AMIs。

> **Note**
>
> p5 執行個體系列僅在等於或晚於 Amazon ECS GPU 最佳化 AMI 20230912的版本上受支援，且與 p2和 g2執行個體類型不相容。如果您需要使用p5執行個體，請確保您的運算環境不包含 p2或 g2執行個體，並使用最新的預設批次 AMI。建立新的運算環境將使用最新的 AMI，但如果您要將運算環境更新為包含 p5，您可以透過在 ComputeResource 屬性true中將updateToLatestImageVersion設定為 來確保您使用的是最新的 AMI。如需 AMI 與 GPU 執行個體相容性的詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的在 Amazon ECS 上使用 GPUs。

下列範例示範如何使用 GetParameter 命令。

AWS CLI

```
$ aws ssm get-parameter --name /aws/service/ecs/optimized-ami/amazon-linux-2/gpu/
recommended \
                        --region us-east-2 --output json
```

輸出包含 Value 參數中的 AMI 資訊。

```
{
    "Parameter": {
        "Name": "/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended",
        "LastModifiedDate": 1555434128.664,
        "Value": "{\"schema_version\":1,\"image_name\":\"amzn2-ami-ecs-gpu-
hvm-2.0.20190402-x86_64-ebs\",\"image_id\":\"ami-083c800fe4211192f\",\"os\":\"Amazon
 Linux 2\",\"ecs_runtime_version\":\"Docker version 18.06.1-ce\",\"ecs_agent_version
\":\"1.27.0\"}",
        "Version": 9,
        "Type": "String",
        "ARN": "arn:aws:ssm:us-east-2::parameter/aws/service/ecs/optimized-ami/
amazon-linux-2/gpu/recommended"
```

```
    }
}
```

Python

```
from __future__ import print_function

import json
import boto3

ssm = boto3.client('ssm', 'us-east-2')

response = ssm.get_parameter(Name='/aws/service/ecs/optimized-ami/amazon-linux-2/
gpu/recommended')
jsonVal = json.loads(response['Parameter']['Value'])
print("image_id   = " + jsonVal['image_id'])
print("image_name = " + jsonVal['image_name'])
```

這時輸出只會包含 AMI ID 和 AMI 名稱：

```
image_id   = ami-083c800fe4211192f
image_name = amzn2-ami-ecs-gpu-hvm-2.0.20190402-x86_64-ebs
```

以下範例會示範 GetParameters 的使用。

AWS CLI

```
$ aws ssm get-parameters --names  /aws/service/ecs/optimized-ami/amazon-linux-2/gpu/
recommended/image_name \
                                  /aws/service/ecs/optimized-ami/amazon-linux-2/gpu/
recommended/image_id \
                          --region us-east-2 --output json
```

這時輸出會包含個別參數的完整中繼資料：

```
{
    "InvalidParameters": [],
    "Parameters": [
        {
            "Name": "/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/
image_id",
```

```
            "LastModifiedDate": 1555434128.749,
            "Value": "ami-083c800fe4211192f",
            "Version": 9,
            "Type": "String",
            "ARN": "arn:aws:ssm:us-east-2::parameter/aws/service/ecs/optimized-ami/
amazon-linux-2/gpu/recommended/image_id"
        },
        {
            "Name": "/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/
image_name",
            "LastModifiedDate": 1555434128.712,
            "Value": "amzn2-ami-ecs-gpu-hvm-2.0.20190402-x86_64-ebs",
            "Version": 9,
            "Type": "String",
            "ARN": "arn:aws:ssm:us-east-2::parameter/aws/service/ecs/optimized-ami/
amazon-linux-2/gpu/recommended/image_name"
        }
    ]
}
```

Python

```
from __future__ import print_function

import boto3

ssm = boto3.client('ssm', 'us-east-2')

response = ssm.get_parameters(
            Names=['/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/
image_name',
                   '/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/
image_id'])
for parameter in response['Parameters']:
    print(parameter['Name'] + " = " + parameter['Value'])
```

輸出包含 AMI ID 和 AMI 名稱，使用名稱的完整路徑。

```
/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/image_id =
 ami-083c800fe4211192f
/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/image_name = amzn2-
ami-ecs-gpu-hvm-2.0.20190402-x86_64-ebs
```

下列範例示範如何使用 GetParametersByPath 命令。

AWS CLI

```
$ aws ssm get-parameters-by-path --path /aws/service/ecs/optimized-ami/amazon-
linux-2/gpu/recommended \
                                  --region us-east-2 --output json
```

輸出包含指定路徑下所有參數的完整中繼資料。

```
{
    "Parameters": [
        {
            "Name": "/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/
ecs_agent_version",
            "LastModifiedDate": 1555434128.801,
            "Value": "1.27.0",
            "Version": 8,
            "Type": "String",
            "ARN": "arn:aws:ssm:us-east-2::parameter/aws/service/ecs/optimized-ami/
amazon-linux-2/gpu/recommended/ecs_agent_version"
        },
        {
            "Name": "/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/
ecs_runtime_version",
            "LastModifiedDate": 1548368308.213,
            "Value": "Docker version 18.06.1-ce",
            "Version": 1,
            "Type": "String",
            "ARN": "arn:aws:ssm:us-east-2::parameter/aws/service/ecs/optimized-ami/
amazon-linux-2/gpu/recommended/ecs_runtime_version"
        },
        {
            "Name": "/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/
image_id",
            "LastModifiedDate": 1555434128.749,
            "Value": "ami-083c800fe4211192f",
            "Version": 9,
            "Type": "String",
            "ARN": "arn:aws:ssm:us-east-2::parameter/aws/service/ecs/optimized-ami/
amazon-linux-2/gpu/recommended/image_id"
        },
        {
```

```
            "Name": "/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/
image_name",
            "LastModifiedDate": 1555434128.712,
            "Value": "amzn2-ami-ecs-gpu-hvm-2.0.20190402-x86_64-ebs",
            "Version": 9,
            "Type": "String",
            "ARN": "arn:aws:ssm:us-east-2::parameter/aws/service/ecs/optimized-ami/
amazon-linux-2/gpu/recommended/image_name"
        },
        {
            "Name": "/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/
os",
            "LastModifiedDate": 1548368308.143,
            "Value": "Amazon Linux 2",
            "Version": 1,
            "Type": "String",
            "ARN": "arn:aws:ssm:us-east-2::parameter/aws/service/ecs/optimized-ami/
amazon-linux-2/gpu/recommended/os"
        },
        {
            "Name": "/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/
schema_version",
            "LastModifiedDate": 1548368307.914,
            "Value": "1",
            "Version": 1,
            "Type": "String",
            "ARN": "arn:aws:ssm:us-east-2::parameter/aws/service/ecs/optimized-ami/
amazon-linux-2/gpu/recommended/schema_version"
        }
    ]
}
```

Python

```
from __future__ import print_function

import boto3

ssm = boto3.client('ssm', 'us-east-2')

response = ssm.get_parameters_by_path(Path='/aws/service/ecs/optimized-ami/amazon-
linux-2/gpu/recommended')
for parameter in response['Parameters']:
```

```
    print(parameter['Name'] + " = " + parameter['Value'])
```

輸出包含指定路徑上所有參數名稱的值，並使用名稱的完整路徑。

```
/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/ecs_agent_version =
 1.27.0
/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/ecs_runtime_version =
 Docker version 18.06.1-ce
/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/image_id =
 ami-083c800fe4211192f
/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/image_name = amzn2-
ami-ecs-gpu-hvm-2.0.20190402-x86_64-ebs
/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/os = Amazon Linux 2
/aws/service/ecs/optimized-ami/amazon-linux-2/gpu/recommended/schema_version = 1
```

如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的擷取 Amazon ECS 最佳化 AMI 中繼資料。

Amazon Linux 棄用

Amazon Linux AMI （也稱為 Amazon Linux 1) 已於 2023 年 12 月 31 日終止對 Amazon Linux AMI 的 AWS Batch 支援，因為自 2024 年 1 月 1 日起，它不會收到任何安全性更新或錯誤修正。如需 Amazon Linux end-of-life的詳細資訊，請參閱 AL 常見問答集。

我們建議您將現有的 Amazon Linux 型運算環境更新為 Amazon Linux 2023，以防止無法預期的工作負載中斷，並繼續接收安全性和其他更新。

使用 Amazon Linux AMI 的運算環境可能會在 2023 年 12 月 31 日end-of-life之後繼續運作。不過，這些運算環境將不再收到來自 的任何新軟體更新、安全修補程式或錯誤修正 AWS。end-of-life，您有責任在 Amazon Linux AMI 上維護這些運算環境。建議您將 AWS Batch 運算環境遷移至 Amazon Linux 2023 或 Amazon Linux 2，以維持最佳效能和安全性。

如需 AWS Batch 從 Amazon Linux AMI 遷移至 Amazon Linux 2023 或 Amazon Linux 2 的說明，請參閱更新運算環境 - AWS Batch。

Amazon EKS Amazon Linux 2 AMI 棄用

AWS 將終止對 Amazon EKS 最佳化 Amazon Linux 2 AMIs支援，自 11/26/25 起生效。我們建議在 AWS Batch 11/26/25 之前將 Amazon EKS 運算環境遷移至 Amazon Linux 2023，以維持最佳效能和安全性。

雖然您可以在 11/26/25end-of-support日期之後，在 Amazon EKS 運算環境中繼續使用批次提供的 Amazon EKS 最佳化 Amazon Linux 2 AMIs，但這些運算環境將不再收到來自 的任何新軟體更新、安全修補程式或錯誤修正 AWS。end-of-life，您有責任在 Amazon EKS 最佳化 Amazon Linux 2 AMI 上維護這些運算環境。

如需 Amazon EKS AL2 end-of-life的詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS AMI 棄用FAQs。

如需將 AWS Batch Amazon EKS 運算環境從 Amazon Linux 2 遷移至 Amazon Linux 2023 的說明，請參閱 如何從 EKS AL2 升級到 EKS AL2023。

Amazon ECS Amazon Linux 2 AMI 棄用

AWS 將結束對 Amazon Linux 2 的支援。從 2026 年 1 月開始， AWS Batch 會將新 Amazon ECS 運算環境的預設 AMI 從 Amazon Linux 2 變更為 Amazon Linux 2023。我們建議將 AWS Batch Amazon ECS 運算環境遷移至 Amazon Linux 2023，以維持最佳效能和安全性。

如需 Amazon Linux 2 end-of-life的詳細資訊，請參閱 Amazon Linux 2 FAQs。

如需 Amazon Linux 2 和 Amazon Linux 2023 之間的差異資訊，請參閱《Amazon Linux 2023 使用者指南》中的比較 Amazon Linux 2023 和 Amazon Linux 2。

如需有關 Amazon ECS 最佳化 AMI 的 Amazon Linux 2023 變更的資訊，請參閱《Amazon ECS 使用者指南》中的從 Amazon Linux 2 遷移至 Amazon Linux 2023 Amazon ECS 最佳化 AMI。

如需將 AWS Batch Amazon ECS 運算環境從 Amazon Linux 2 遷移至 Amazon Linux 2023 的說明，請參閱 如何從 ECS AL2 遷移至 ECS AL2023。

搭配 使用 Amazon EC2 啟動範本 AWS Batch

AWS Batch 支援搭配 ECAmazon EC2 EC2 啟動範本。使用啟動範本，您可以修改 AWS Batch 運算資源的預設組態，而不需要建立自訂 AMIs。

> **Note**
>
> AWS Fargate 資源不支援啟動範本。

您必須先建立啟動範本，才能將它與運算環境建立關聯。您可以在 Amazon EC2 主控台中建立啟動範本。或者，您可以使用 AWS CLI 或 AWS 開發套件。例如，下列 JSON 檔案代表啟動範本，該範本會調整預設 AWS Batch 運算資源 AMI 的 Docker 資料磁碟區大小，並將其設定為加密。

```
{
    "LaunchTemplateName": "increase-container-volume-encrypt",
    "LaunchTemplateData": {
        "BlockDeviceMappings": [
            {
                "DeviceName": "/dev/xvda",
                "Ebs": {
                    "Encrypted": true,
                    "VolumeSize": 100,
                    "VolumeType": "gp2"
                }
            }
        ]
    }
}
```

您可以將 JSON 儲存到名為 的檔案`lt-data.json`並執行下列 AWS CLI 命令，以建立先前的啟動範本。

```
aws ec2 --region <region> create-launch-template --cli-input-json file://lt-data.json
```

如需啟動範本的詳細資訊，請參閱《Amazon EC2 使用者指南》中的從啟動範本啟動執行個體。

如果您使用啟動範本建立您的運算環境，您可以將以下現有的運算環境參數移至您的啟動範本：

> **Note**
>
> 假設這些參數 (Amazon EC2 標籤除外） 都指定在啟動範本和運算環境組態中。然後，以運算環境參數為優先。Amazon EC2 標籤會在啟動範本和運算環境組態之間合併。如果標籤的金鑰發生衝突，則以運算環境組態中的值為優先。

- Amazon EC2 金鑰對
- Amazon EC2 AMI ID
- 安全群組 IDs
- Amazon EC2 標籤

下列啟動範本參數會由 忽略 AWS Batch：

- 執行個體類型 (在您建立運算環境時，指定所需的執行個體類型)
- 執行個體角色 (在您建立運算環境時，指定所需的執行個體角色)
- 網路界面子網路 (在您建立運算環境時，指定所需的子網路)
- 執行個體市場選項 (AWS Batch 必須控制 Spot 執行個體組態)
- 停用 API 終止 (AWS Batch 必須控制執行個體生命週期)

AWS Batch 只會在基礎設施更新期間使用新的啟動範本版本來更新啟動範本。如需詳細資訊，請參閱在 中更新運算環境 AWS Batch。

預設和覆寫啟動範本

您可以為運算環境定義預設啟動範本，並為特定執行個體類型和系列定義覆寫啟動範本。這對您來說非常有用，因此預設範本會用於運算環境中的大多數執行個體類型。

替代變數 `$Default`和 `$Latest`可用來取代命名特定版本。如果您未提供覆寫啟動範本，則會自動套用預設啟動範本。

如果您使用 `$Default`或 `$Latest`變數， AWS Batch 會在建立運算環境時套用目前資訊。如果預設或最新版本在未來變更，您必須透過 UpdateComputeEnvironment 或 AWS 管理主控台 - 更新資訊 AWS Batch。

若要提供額外的彈性，您可以定義覆寫啟動範本套用至特定的運算執行個體類型或系列。

> **Note**
>
> 每個運算環境最多可指定十 (10) 個覆寫啟動範本。

使用 `targetInstanceTypes` 參數來選取應該使用此覆寫啟動範本的執行個體類型或系列。執行個體類型或系列必須先由 `instanceTypes` 參數識別。

如果您定義啟動範本覆寫並決定稍後移除，您可以傳遞空陣列，以在 `UpdateComputeEnvironment` API 操作中取消設定 `overrides` 參數。您也可以在提交 `UpdateComputeEnvironment` API 操作時選擇不包含 `overrides` 參數。如需詳細資訊，請參閱`LaunchTemplateSpecification.overrides`

如需詳細資訊，請參閱《 AWS Batch API 參考指南`LaunchTemplateSpecificationOverride.targetInstanceTypes`》中的 。

啟動範本中的 Amazon EC2 使用者資料

您可以在啟動範本中提供 Amazon EC2 使用者資料，該資料會在執行個體啟動時由 cloud-init 執行。您的使用者資料可以執行常見的組態案例，包括但不限於下列項目：

- 包括使用者或群組
- 安裝套件
- 建立分區和檔案系統

啟動範本中的 Amazon EC2 使用者資料必須採用 MIME 分段封存格式。這是因為您的使用者資料會與設定運算資源所需的其他 AWS Batch 使用者資料合併。您可以將多個使用者資料區塊組合在一起成為單一 MIME 分段檔案。例如，您可能想要將設定 Docker 協助程式的雲端引導與寫入 Amazon ECS 容器代理程式組態資訊的使用者資料 shell 指令碼結合。

如果您使用的是 AWS CloudFormation，AWS::CloudFormation::Init 類型可與 cfn-init 協助程式指令碼搭配使用，以執行常見的組態案例。

MIME 分段檔案包含下列元件：

- 內容類型和部分邊界宣告：`Content-Type: multipart/mixed; boundary="==BOUNDARY=="`
- MIME 版本宣告：`MIME-Version: 1.0`
- 一或多個使用者資料區塊，其中包含下列元件：
  - 發出使用者資料區塊開頭訊號的開啟界限：`--==BOUNDARY==`。您必須在此邊界之前將行保留空白。
  - 區塊的內容類型宣告：`Content-Type: text/cloud-config; charset="us-ascii"`。如需內容類型的詳細資訊，請參閱「Cloud-Init 說明文件」。您必須在內容類型宣告後保留該行空白。
  - 使用者資料的內容，例如 shell 命令或`cloud-init`指令的清單。
- 發出 MIME 分段檔案結尾訊號的結束界限：`--==BOUNDARY==--`。您必須在關閉界限之前將行保留空白。

> **Note**
>
> 如果您在 Amazon EC2 主控台中將使用者資料新增至啟動範本，您可以將其貼上為純文字。或者，您可以從檔案上傳。如果您使用 `base64` AWS CLI 或 AWS 開發套件，則必須先編碼

使用者資料，並在呼叫 CreateLaunchTemplate 時提交該字串做為 UserData 參數的值，如此 JSON 檔案所示。

```
{
    "LaunchTemplateName": "base64-user-data",
    "LaunchTemplateData": {
        "UserData":
 "ewogICAgIkxhdW5jaFRlbXBsYXRlTmFtZSI6ICJpbmNyZWFzZS1jb250YWluZXItdm9sdW..."
    }
}
```

主題

- 參考：Amazon EC2 啟動範本範例

參考：Amazon EC2 啟動範本範例

以下是範例 MIME 分段檔案，您可以用來建立自己的範本。

範例

- 範例：掛載現有的 Amazon EFS 檔案系統
- 範例：覆寫預設 Amazon ECS 容器代理程式組態
- 範例：掛載現有的 Amazon FSx for Lustre 檔案系統

範例：掛載現有的 Amazon EFS 檔案系統

Example

此範例 MIME 分段檔案會設定運算資源來安裝amazon-efs-utils套件，並在 掛載現有的 Amazon EFS 檔案系統/mnt/efs。

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

--==MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"

packages:
```

```
- amazon-efs-utils

runcmd:
- file_system_id_01=fs-abcdef123
- efs_directory=/mnt/efs

- mkdir -p ${efs_directory}
- echo "${file_system_id_01}:/ ${efs_directory} efs tls,_netdev" >> /etc/fstab
- mount -a -t efs defaults

--==MYBOUNDARY==--
```

範例：覆寫預設 Amazon ECS 容器代理程式組態

Example

此範例 MIME 多段檔案會覆寫運算資源預設的 Docker 影像清除設定。

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

--==MYBOUNDARY==
Content-Type: text/x-shellscript; charset="us-ascii"

#!/bin/bash
echo ECS_IMAGE_CLEANUP_INTERVAL=60m >> /etc/ecs/ecs.config
echo ECS_IMAGE_MINIMUM_CLEANUP_AGE=60m >> /etc/ecs/ecs.config

--==MYBOUNDARY==--
```

範例：掛載現有的 Amazon FSx for Lustre 檔案系統

Example

此範例 MIME 分段檔案會將運算資源設定為從 Extras Library 安裝`lustre2.10`套件，並將現有的 FSx for Lustre 檔案系統掛載至 ，`/scratch`並掛載名稱為 `fsx`。此範例適用於 Amazon Linux 2。如需其他 Linux 發行版本的安裝說明，請參閱《Amazon FSx for Lustre 使用者指南》中的安裝 Lustre 用戶端。如需詳細資訊，請參閱《Amazon FSx for Lustre 使用者指南》中的自動掛載 Amazon FSx 檔案系統。 FSx

```
MIME-Version: 1.0
```

```
Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

--==MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"

runcmd:
- file_system_id_01=fs-0abcdef1234567890
- region=us-east-2
- fsx_directory=/scratch
- amazon-linux-extras install -y lustre2.10
- mkdir -p ${fsx_directory}
- mount -t lustre ${file_system_id_01}.fsx.${region}.amazonaws.com@tcp:fsx
 ${fsx_directory}

--==MYBOUNDARY==--
```

在容器內容的磁碟區和 mountPoints 成員中，掛載點必須對應到容器中。

```
{
    "volumes": [
        {
            "host": {
                "sourcePath": "/scratch"
            },
            "name": "Scratch"
        }
    ],
    "mountPoints": [
        {
            "containerPath": "/scratch",
            "sourceVolume": "Scratch"
        }
    ],
}
```

執行個體中繼資料服務 (IMDS) 組態

執行個體中繼資料服務 (IMDS) 會將 EC2 執行個體的相關中繼資料提供給在這些執行個體上執行的應用程式。針對所有新工作負載使用 IMDSv2，並將現有的工作負載從 IMDSv1 遷移至 IMDSv2，以提高安全性。如需 IMDS 和設定 IMDS 的詳細資訊，請參閱《Amazon EC2 使用者指南》中的使用執行個體中繼資料來管理 EC2 執行個體和設定新執行個體的執行個體中繼資料選項。 Amazon EC2

組態案例

根據您的運算環境設定選擇適當的組態方法：

沒有啟動範本的預設 AMI

當您使用預設 AWS Batch AMI 且未指定啟動範本時，請選擇下列其中一個選項：

1. 使用 Amazon Linux 2023 預設 AMI – Amazon Linux 2023 預設需要 IMDSv2。當您建立運算環境時，請選取 Amazon Linux 2023 做為映像類型。
2. 設定帳戶層級 IMDSv2 組態 – 將 AWS 您的帳戶設定為要求所有新執行個體使用 IMDSv2。此設定會影響您在帳戶中啟動的所有新執行個體。如需說明，請參閱《Amazon EC2 使用者指南》中的將 IMDSv2 設定為帳戶的預設值。 Amazon EC2

> **Note**
>
> 帳戶層級 IMDS 組態可由啟動範本或 AMI 組態覆寫。啟動範本設定優先於帳戶層級設定。

沒有啟動範本的自訂 AMI

當您在沒有啟動範本的情況下使用自訂 AMI 時，請選擇下列其中一個選項：

1. 使用 Amazon Linux 2023 作為基礎 – 使用 Amazon Linux 2023 作為基礎映像建置您的自訂 AMI。如需為 Batch 建立自訂 AMIs 的詳細資訊，請參閱 教學課程：建立運算資源 AMI。
2. 在自訂 AMI 中設定 IMDSv2 – 當您建立自訂 AMI 時，請將其設定為需要 IMDSv2。如需說明，請參閱《Amazon EC2 使用者指南》中的設定自訂 AMI 的執行個體中繼資料選項。
3. 設定帳戶層級 IMDSv2 組態 – 將 AWS 您的帳戶設定為要求所有新執行個體使用 IMDSv2。此設定會影響您在帳戶中啟動的所有新執行個體。如需說明，請參閱《Amazon EC2 使用者指南》中的將 IMDSv2 設定為帳戶的預設值。 Amazon EC2

> **Note**
>
> 帳戶層級 IMDS 組態可由啟動範本或 AMI 組態覆寫。啟動範本設定優先於帳戶層級設定。

使用啟動範本

當您在運算環境中使用啟動範本時，請將中繼資料選項新增至啟動範本，以要求 IMDSv2。如需搭配 Batch 使用啟動範本的詳細資訊，請參閱 搭配 使用 Amazon EC2 啟動範本 AWS Batch。

```
{
    "LaunchTemplateName": "batch-imdsv2-template",
    "VersionDescription": "IMDSv2 only template for Batch",
    "LaunchTemplateData": {
        "MetadataOptions": {
            "HttpTokens": "required"
        }
    }
}
```

使用 CLI AWS 建立啟動範本：

```
aws ec2 create-launch-template --cli-input-json file://imds-template.json
```

EC2 組態

AWS Batch 針對 EC2 和 EC2 Spot 運算環境使用 Amazon ECS 最佳化 AMIs。預設值為 Amazon Linux 2 (ECS_AL2)。從 2026 年 1 月開始，預設值將變更為 AL2023 (ECS_AL2023)。

AWS 將結束對 Amazon Linux 2 的支援。我們建議將 AWS Batch Amazon ECS 運算環境遷移至 Amazon Linux 2023，以維持最佳效能和安全性。如需詳細資訊，請參閱Amazon ECS Amazon Linux 2 AMI 棄用。

我們建議您將現有的 Amazon Linux 型運算環境更新為 Amazon Linux 2023，以防止無法預期的工作負載中斷，並繼續接收安全性和其他更新。

如需 AWS Batch 從 Amazon Linux AMI 遷移至 Amazon Linux 2023 的說明，請參閱 如何從 ECS AL2 遷移至 ECS AL2023

主題

- 如何從 ECS AL2 遷移至 ECS AL2023

如何從 ECS AL2 遷移至 ECS AL2023

AL2023 是以 Linux 為基礎的作業系統，旨在為您的雲端應用程式提供安全、穩定且高效能的環境。如需 AL2 和 AL2023 之間差異的詳細資訊，請參閱《Amazon Linux 2023 使用者指南》中的比較 Amazon Linux 2023 和 Amazon Linux 2。

從 2026 年 1 月開始， AWS Batch 會將新 Amazon ECS 運算環境的預設 AMI 從 Amazon Linux 2 變更為 Amazon Linux 2023，因為 AWS 將結束對 Amazon Linux 2 的支援。當您在建立新的運算環境時，未指定 imageType.Ec2Configuration 欄位的值時，會使用預設 AMI。我們建議將 AWS Batch Amazon ECS 運算環境遷移至 Amazon Linux 2023，以維持最佳效能和安全性。

根據運算環境的設定方式，您可以使用下列其中一個從 AL2 到 AL2023 的升級路徑。

使用 Ec2Configuration.ImageType 升級

- 如果您未使用啟動範本或啟動範本覆寫，請將 Ec2Configuration.ImageType 變更為 `ECS_AL2023`（或使用 GPU 執行個體`ECS_AL2023_NVIDIA`時），然後執行 UpdateComputeEnvironment。
- 如果您指定 Ec2Configuration.ImageIdOverride，則 Ec2Configuration.ImageType 必須符合 Ec2Configuration.ImageIdOverride 中指定的 AMI 類型。

  如果您不相符`ImageIdOverride`，`ImageType`則運算環境可能無法正常運作。

使用啟動範本進行升級

- 如果您使用根據 指定 AMI 的啟動範本`ECS_AL2023`，請確定您的啟動範本與 Amazon Linux 2023 相容。如需 Amazon ECS 最佳化 AMI 的 Amazon Linux 2023 變更相關資訊，請參閱《Amazon ECS 使用者指南》中的從 Amazon Linux 2 遷移至 Amazon Linux 2023 Amazon ECS 最佳化 AMI。
- 對於 AL2023 AMIs，請確認任何自訂使用者資料或初始化指令碼都與 AL2023 環境和套件管理系統相容。

使用 升級 CloudFormation

- 如果您使用 CloudFormation 來管理運算環境，請更新您的範本，將 中的 `ImageType` 屬性`Ec2Configuration`從 `ECS_AL2` 變更為 `ECS_AL2023`（或使用 GPU 執行個體`ECS_AL2023_NVIDIA`時）：

```
ComputeEnvironment:
```

```
  Type: AWS::Batch::ComputeEnvironment
  Properties:
    ComputeResources:
      Ec2Configuration:
        - ImageType: ECS_AL2023
```

然後更新您的 CloudFormation 堆疊以套用變更。

- 如果您的 CloudFormation 範本使用 指定自訂 AMI`ImageId0verride`，請確定 AMI ID 對應至 AL2023-based AMI，且符合 `ImageType`設定。

遷移考量事項

從 Amazon Linux 2 遷移至 Amazon Linux 2023 時，請考慮下列事項：

- 套件管理 – Amazon Linux 2023 使用 `dnf`而非 `yum`進行套件管理。
- 系統服務 – 有些系統服務及其組態在 AL2 和 AL2023 之間可能不同。
- 容器執行時間 – AL2 和 AL2023 都支援 Docker，但 AL2023 可能會有不同的預設組態。
- 安全 – AL2023 包含增強的安全功能，可能需要更新與安全相關的組態。
- 執行個體中繼資料服務第 2 版 (IMDSv2) – IMDSv2 是一種工作階段導向服務，需要字符型身分驗證才能存取 EC2 執行個體中繼資料，以提供增強的安全性。如需 IMDS 的詳細資訊，請參閱《Amazon EC2 使用者指南》中的 和執行個體中繼資料服務第 2 版的運作方式。

如需變更和遷移考量的完整清單，請參閱《Amazon ECS 使用者指南》中的從 Amazon Linux 2 遷移至 Amazon Linux 2023 Amazon ECS 最佳化 AMI。

的執行個體類型配置策略 AWS Batch

建立受管運算環境時， 會從最符合任務需求的`instanceTypes`指定 AWS Batch 中選取執行個體類型。當 AWS Batch 需要額外容量時，配置策略會定義行為。此參數不適用於在 Fargate 資源上執行的任務。請勿指定此參數。

`BEST_FIT` (default)

AWS Batch 會選取最符合任務需求的執行個體類型，並偏好成本最低的執行個體類型。如果所選執行個體類型的其他執行個體無法使用，請 AWS Batch 等待其他執行個體可用。如果沒有足夠的執行個體可用，或者如果使用者達到 Amazon EC2 服務配額，則在目前執行中的任務完成之前，

不會執行其他任務。這種配置策略可以降低成本，但可能限制擴展。如果您使用 Spot Fleets 搭配 BEST_FIT，則必須指定 Spot Fleet IAM 角色。更新運算環境時BEST_FIT不支援 。如需詳細資訊，請參閱在 中更新運算環境 AWS Batch。

> **Note**
>
> AWS Batch 會管理您帳戶中 AWS 的資源。根據預設，使用最初使用的啟動組態的 BEST_FIT 配置策略來運算環境。不過，隨著時間的推移，新 AWS 帳戶的啟動組態使用將受到限制。因此，從 2024 年 4 月底開始，新建立的 BEST_FIT 運算環境預設會啟動範本。如果您的服務角色缺少管理啟動範本的許可， AWS Batch 可能會繼續使用啟動組態。現有的運算環境將繼續使用啟動組態。

`BEST_FIT_PROGRESSIVE`

AWS Batch 會選取足夠大的其他執行個體類型，以符合佇列中任務的需求。建議使用每個單位 vCPU 成本較低的執行個體類型。如果先前所選執行個體類型的其他執行個體無法使用， AWS Batch 會選取新的執行個體類型。

> **Note**
>
> 對於多節點平行任務， AWS Batch 選擇可用的最佳執行個體類型。如果執行個體類型因為容量不足而無法使用，則不會啟動系列中的其他執行個體類型。

`SPOT_CAPACITY_OPTIMIZED`

AWS Batch 會選取一個或多個大小足以符合佇列中任務需求的執行個體類型。較不可能中斷的執行個體類型為首選。此配置策略僅適用於 Spot 執行個體運算資源。

`SPOT_PRICE_CAPACITY_OPTIMIZED`

價格和容量最佳化分配策略會考慮價格和容量，來選擇最不可能中斷且價格最低的 Spot 執行個體集區。此配置策略僅適用於 Spot 執行個體運算資源。

> **Note**
>
> 我們建議您在大多數執行個體`SPOT_CAPACITY_OPTIMIZED`中使用 `SPOT_PRICE_CAPACITY_OPTIMIZED`，而不是 。

BEST_FIT_PROGRESSIVE 和 BEST_FIT策略使用隨需或 Spot 執行個體，而 SPOT_CAPACITY_OPTIMIZED和 SPOT_PRICE_CAPACITY_OPTIMIZED策略使用 Spot 執行個體。不過， AWS Batch 可能需要超過 maxvCpus 才能滿足您的容量需求。在這種情況下， AWS Batch 絕不maxvCpus會超過單一執行個體。

運算資源記憶體管理

當 Amazon ECS 容器代理程式將運算資源註冊到運算環境時，代理程式必須判斷運算資源可以為您的任務預留多少記憶體。由於平台記憶體負荷和系統核心佔用的記憶體，此數字與 Amazon EC2 執行個體安裝的記憶體數量不同。舉例而言，m4.large 執行個體安裝了 8 GiB 的記憶體。不過，這不一定會轉譯為運算資源註冊時可用於任務的 8192 MiB 記憶體。

假設您為任務指定 8192 MiB，而且運算資源都沒有 8192 MiB 或更高的記憶體可滿足此需求。然後，任務無法放置在您的運算環境中。如果您使用的是受管運算環境， AWS Batch 必須啟動較大的執行個體類型以容納請求。

預設 AWS Batch 運算資源 AMI 也會為 Amazon ECS 容器代理程式和其他關鍵系統程序保留 32 MiB 的記憶體。此記憶體不適用於任務配置。如需詳細資訊，請參閱預留系統記憶體。

Amazon ECS 容器代理程式會使用 Docker ReadMemInfo() 函數來查詢作業系統可用的記憶體總量。Linux 提供命令列公用程式來判斷總記憶體。

Example- 判定 Linux 記憶體總量

free 命令會傳回作業系統辨識的總記憶體。

```
$ free -b
```

以下是執行 Amazon ECS 最佳化 Amazon Linux AMI 之m4.large執行個體的範例輸出。

```
             total       used       free     shared    buffers     cached
Mem:    8373026816  348180480 8024846336      90112   25534464  205418496
-/+ buffers/cache:  117227520 8255799296
```

此執行個體有 8373026816 個位元組的總記憶體。這表示有 7985 MiB 可用於任務。

主題

- 預留系統記憶體

- 教學課程：檢視運算資源記憶體
- Amazon EKS AWS Batch 上的 記憶體和 vCPU 考量事項

預留系統記憶體

如果您使用 任務佔用運算資源上的所有記憶體，您的任務可能會與記憶體的關鍵系統程序競爭，並可能導致系統故障。Amazon ECS 容器代理程式提供稱為 的組態變數`ECS_RESERVED_MEMORY`。您可以使用此組態變數，從配置給您任務MiB的集區中移除指定數量的記憶體。這可為重要系統程序有效地預留記憶體。

預設 AWS Batch 運算資源 AMI 為 Amazon ECS 容器代理程式和其他關鍵系統程序保留 32 MiB 的記憶體。我們建議為 Amazon ECS 容器代理程式和其他關鍵系統程序保留 5% 的記憶體緩衝區。

教學課程：檢視運算資源記憶體

您可以在 Amazon ECS 主控台中或使用 DescribeContainerInstances API 操作，檢視運算資源向 註冊的記憶體數量。如果您嘗試為特定執行個體類型提供盡可能多的記憶體，以最大化資源使用率，您可以觀察該運算資源可用的記憶體，然後為您的任務指派足夠的記憶體。

檢視運算資源記憶體

1. 開啟主控台，網址為 https://console.aws.amazon.com/ecs/v2。
2. 選擇叢集，然後選擇託管要檢視之運算資源的叢集。

   您運算環境的叢集名稱以您運算環境的名稱開頭。
3. 選擇基礎設施。
4. 在容器執行個體下，選擇容器執行個體。
5. 資源和聯網區段顯示運算資源的已註冊和可用記憶體。

   總容量記憶體值是第一次啟動時向 Amazon ECS 註冊的運算資源，而可用記憶體值是尚未分配給任務的值。

Amazon EKS AWS Batch 上的 記憶體和 vCPU 考量事項

在 Amazon EKS AWS Batch 的 中，您可以指定可供容器使用的資源。例如，您可以指定 vCPU 和記憶體資源的 `requests`或 `limits`值。

以下是指定 vCPU 資源的限制：

- 至少必須指定一個 vCPU `requests`或`limits`值。
- 一個 vCPU 單位等同於一個實體或虛擬核心。
- vCPU 值必須以整數或增量 0.25 輸入。
- 最小的有效 vCPU 值為 0.25。
- 如果指定兩者，則`requests`值必須小於或等於`limits`值。如此一來，您就可以同時設定軟式和硬式 vCPU 組態。
- vCPU 值無法以 milliCPU 格式指定。例如， `100m`不是有效的值。
- AWS Batch 使用 `requests`值進行擴展決策。如果未指定`requests`值，則會將該`limits`值複製到該`requests`值。

以下是指定記憶體資源的限制：

- 至少必須指定一個記憶體`requests`或`limits`值。
- 記憶體值必須位於 mebibytes() 中MiBs。
- 如果指定兩者，則`requests`值必須等於 `limits`值。
- AWS Batch 使用 `requests`值進行擴展決策。如果未指定`requests`值，則會將該`limits`值複製到該`requests`值。

以下是指定 GPU 資源的限制：

- 如果指定兩者，則`requests`值必須等於 `limits`值。
- AWS Batch 使用 `requests`值進行擴展決策。如果未指定`requests`值，則會將該`limits`值複製到該`requests`值。

範例：任務定義

Amazon EKS 任務定義 AWS Batch 上的以下內容會設定軟 vCPU 共用。這可讓 AWS Batch Amazon EKS 使用執行個體類型的所有 vCPU 容量。不過，如果有其他任務正在執行，則會分配最多 個 2 vCPUs 記憶體限制為 2 GB。

```
{
   "jobDefinitionName": "MyJobOnEks_Sleep",
   "type": "container",
```

```
    "eksProperties": {
        "podProperties": {
            "containers": [
                {
                    "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
                    "command": ["sleep", "60"],
                    "resources": {
                        "requests": {
                            "cpu": "2",
                            "memory": "2048Mi"
                        }
                    }
                }
            ]
        }
    }
}
```

在 Amazon EKS 任務定義 AWS Batch 上，下列 `request`的值為 1，並將最多 4 vCPUs分配給任務。

```
{
    "jobDefinitionName": "MyJobOnEks_Sleep",
    "type": "container",
    "eksProperties": {
        "podProperties": {
            "containers": [
                {
                    "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
                    "command": ["sleep", "60"],
                    "resources": {
                        "requests": {
                            "cpu": "1"
                        },
                        "limits": {
                            "cpu": "4",
                            "memory": "2048Mi"
                        }
                    }
                }
            ]
        }
    }
}
```

在 Amazon EKS 任務定義 AWS Batch 上，下列項目會設定 的 vCPU `limits`值1和 1 GB 的記憶體`limits`值。

```
{
   "jobDefinitionName": "MyJobOnEks_Sleep",
   "type": "container",
   "eksProperties": {
       "podProperties": {
           "containers": [
               {
                   "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
                   "command": ["sleep", "60"],
                   "resources": {
                       "limits": {
                           "cpu": "1",
                           "memory": "1024Mi"
                       }
                   }
               }
           ]
       }
   }
}
```

當 將 AWS Batch on Amazon EKS 任務 AWS Batch 轉譯為 Amazon EKS Pod 時， 會將`limits`值 AWS Batch 複製到 `requests`值。如果未指定`requests`值，即為 。當您提交上述範例任務定義時，Pod `spec`如下所示。

```
apiVersion: v1
kind: Pod
...
spec:
  ...
  containers:
    - command:
        - sleep
        - 60
      image: public.ecr.aws/amazonlinux/amazonlinux:2
      resources:
        limits:
          cpu: 1
          memory: 1024Mi
```

```
        requests:
          cpu: 1
          memory: 1024Mi
      ...
```

節點 CPU 和記憶體保留

AWS Batch 依賴 `bootstrap.sh` 檔案的預設邏輯進行 vCPU 和記憶體保留。如需 `bootstrap.sh` 檔案的詳細資訊，請參閱 https：//bootstrap.sh。當您調整 vCPU 和記憶體資源的大小時，請考慮以下範例。

> **Note**
>
> 如果沒有執行個體正在執行，vCPU 和記憶體保留最初會影響 AWS Batch 擴展邏輯和決策。執行個體執行後， 會 AWS Batch 調整初始配置。

範例：節點 CPU 保留

CPU 保留值使用執行個體可用的 vCPUs 總數，以毫秒為單位計算。

vCPU 號碼	預留百分比
1	6%
2	1%
3-4	0.5%
4 及更高版本	0.25%

使用上述值時，下列為 true：

- 具有 2 個 vCPUs 之`c5.large`執行個體的 CPU 保留值為 70 公尺。計算方式如下：(1*60) + (1*10) = 70 公尺。
- 具有 96 個 vCPUs 之`c5.24xlarge`執行個體的 CPU 保留值為 310 公尺。計算方式如下：(1*60) + (1*10) + (2*5) + (92*2.5) = 310 公尺。

在此範例中，有 1930 （計算值 2000-70) 微核心 vCPU 單位可用於在c5.large執行個體上執行任務。假設您的任務需要 2(2*1000 公尺） vCPU 單位，則該任務不適用於單一c5.large執行個體。不過，需要 1.75 vCPU 單位符合的任務。

範例：節點記憶體保留

記憶體保留值使用下列項目以 MB 為單位計算：

- 執行個體容量，以 MB 為單位。例如，8 GB 執行個體為 7，748 MiB。
- kubeReserved 值。kubeReserved 值是為系統協助程式預留的記憶體量。此kubeReserved值的計算方式如下：((11 * 執行個體類型支援的 Pod 數量上限） + 255)。如需執行個體類型支援的 Pod 數量上限的相關資訊，請參閱 eni-max-pods.txt
- HardEvictionLimit 值。當可用的記憶體低於 HardEvictionLimit值時，執行個體會嘗試移出 Pod。

計算可配置記憶體的公式如下：(*instance_capacity_in_MiB*) - (11 * (*maximum_number_of_pods*)) - 255 - (*HardEvictionLimit value.))#*

c5.large 執行個體最多支援 29 個 Pod。對於HardEvictionLimit值為 100 MiB 的 8 GB c5.large執行個體，可配置記憶體為 7074MiB。計算方式如下：(7748 - (11 * 29) -255 -100) = 7074 MiB。在此範例中，8，192 MiB任務不適用於此執行個體，即使它是 8 gibibyte(GiB) 執行個體。

DaemonSets

當您使用 時DaemonSets，請考慮下列事項：

- 如果 Amazon EKS 執行個體 AWS Batch 上沒有正在執行， 一開始DaemonSets會影響 AWS Batch 擴展邏輯和決策。 AWS Batch 一開始會為預期的 配置 0.5 個 vCPU 單位和 500 MiBDaemonSets。執行個體執行後， 會 AWS Batch 調整初始配置。
- 如果 DaemonSet定義 vCPU 或記憶體限制，則 Amazon EKS 任務 AWS Batch 的資源較少。我們建議您盡可能減少指派給 AWS Batch 任務DaemonSets的 數量。

Fargate 運算環境

Fargate 是一種技術，您可以搭配 使用 AWS Batch 來執行容器，而無需管理 Amazon EC2 執行個體的伺服器或叢集。使用 Fargate，就不再需要佈建、設定或擴展虛擬機器的叢集來執行容器。這樣一來即無須選擇伺服器類型、決定何時擴展叢集，或最佳化叢集壓縮。

當您使用 Fargate 資源執行任務時，您可以將應用程式封裝在容器中、指定 CPU 和記憶體需求、定義聯網和 IAM 政策，以及啟動應用程式。每個 Fargate 任務都有自己的隔離界限，不會與其他任務共用基礎核心、CPU 資源、記憶體資源或彈性網路界面。

主題

- 何時使用 Fargate
- Fargate 上的任務定義
- Fargate 上的任務佇列
- Fargate 上的運算環境

何時使用 Fargate

我們建議在大多數情況下使用 Fargate。Fargate 會啟動和擴展運算，以符合您為容器指定的資源需求。使用 Fargate，您不需要過度佈建或支付額外的伺服器。您也不需要擔心基礎設施相關參數的詳細資訊，例如執行個體類型。當運算環境需要向上擴展時，在 Fargate 資源上執行的任務可以更快地開始。一般而言，啟動新的 Amazon EC2 執行個體需要幾分鐘的時間。不過，在 Fargate 上執行的任務可以在大約 30 秒內佈建。所需的確切時間取決於數個因素，包括容器映像大小和任務數量。

不過，如果您的任務需要下列任何項目，建議您使用 Amazon EC2：

- 超過 16 vCPUs
- 超過 120 GB (GiB) 的記憶體
- GPU
- 自訂 Amazon Machine Image (AMI)
- 任何 linuxParameters 參數

如果您有大量任務，我們建議您使用 Amazon EC2 基礎設施。例如，如果同時執行的任務數量超過 Fargate 限流限制。這是因為使用 EC2 時，任務可以比 Fargate 資源以更高的速率分派給 EC2 資源。此外，當您使用 EC2 時，可以同時執行更多任務。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Fargate 服務配額。

Fargate 上的任務定義

AWS Batch 上的 任務 AWS Fargate 不支援所有可用的任務定義參數。有些參數完全不受支援，有些則對 Fargate 任務有不同的行為。

以下清單說明 Fargate 任務中無效或以其他方式限制的任務定義參數。

`platformCapabilities`

必須指定為 FARGATE。

```
"platformCapabilities": [ "FARGATE" ]
```

`type`

必須指定為 `container`。

```
"type": "container"
```

`containerProperties` 中的參數

`executionRoleArn`

必須指定在 Fargate 資源上執行的任務。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的任務 IAM 角色。

```
"executionRoleArn": "arn:aws:iam::123456789012:role/ecsTaskExecutionRole"
```

`fargatePlatformConfiguration`

（選用，僅適用於 Fargate 任務定義）。指定 Fargate 平台版本，或LATEST最近平台版本。的可能值`platformVersion`為 `1.3.0`、 `1.4.0`和 LATEST（預設）。

```
"fargatePlatformConfiguration": { "platformVersion": "1.4.0" }
```

`instanceType`, `ulimits`

不適用於在 Fargate 資源上執行的任務。

`memory`, `vcpus`

這些設定必須在 中指定 `resourceRequirements`

`privileged`

請不要指定此參數，或指定 `false`。

```
"privileged": false
```

`resourceRequirements`

必須使用支援的值來指定記憶體和 vCPU 需求。在 Fargate 資源上執行的任務不支援 GPU 資源。

如果您使用 GuardDuty 執行期監控，GuardDuty 安全代理程式會有些微的記憶體負荷。因此，記憶體限制必須包含 GuardDuty 安全代理程式的大小。如需有關 GuardDuty 安全代理程式記憶體限制的資訊，請參閱《GuardDuty 使用者指南》中的 CPU 和記憶體限制。如需最佳實務的相關資訊，請參閱《Amazon ECS 開發人員指南》中的如何在啟用執行期監控之後修復 Fargate 任務上的記憶體不足錯誤。

```
"resourceRequirements": [
  {"type": "MEMORY", "value": "512"},
  {"type": "VCPU",   "value": "0.25"}
]
```

`linuxParameters` 中的參數

`devices`, `maxSwap`, `sharedMemorySize`, `swappiness`, `tmpfs`

不適用於在 Fargate 資源上執行的任務。

`logConfiguration` 中的參數

`logDriver`

僅支援 `splunk` `awslogs`和 。如需詳細資訊，請參閱使用 awslogs 日誌驅動程式。

中的成員 `networkConfiguration`

`assignPublicIp`

如果私有子網路沒有連接 NAT 閘道來傳送流量到網際網路， `assignPublicIp` 必須為「`ENABLED`」。如需詳細資訊，請參閱AWS Batch IAM 執行角色。

Fargate 上的任務佇列

AWS Batch 上的任務佇列 AWS Fargate 基本上保持不變。唯一的限制是列出的運算環境`computeEnvironmentOrder`必須是 Fargate 運算環境 (`FARGATE` 或 `FARGATE_SPOT`)。EC2 和 Fargate 運算環境無法混合。

Fargate 上的運算環境

AWS Batch 上的運算環境 AWS Fargate 不支援所有可用的運算環境參數。完全不支援某些參數。其他對 Fargate 有特定要求。

以下清單說明在 Fargate 任務中無效或以其他方式受限的運算環境參數。

`type`

此參數必須是 `MANAGED`。

```
"type": "MANAGED"
```

`computeResources` 物件中的參數

`allocationStrategy`, `bidPercentage`, `desiredvCpus`, `imageId`, `instanceTypes`, `ec2Configuration`, `ec2KeyPair`, `instanceRole`, `launchTemplate`, `minvCpus`, `placementGroup`, `spotIamFleetRole`

這些不適用於 Fargate 運算環境，而且無法提供。

`subnets`

如果此參數中列出的子網路未連接 NAT 閘道，則任務定義中的 `assignPublicIp` 參數必須設定為 `ENABLED`。

`tags`

這不適用於 Fargate 運算環境，而且無法提供。若要指定 Fargate 運算環境的標籤，請使用不在 `computeResources` 物件中的 `tags` 參數。

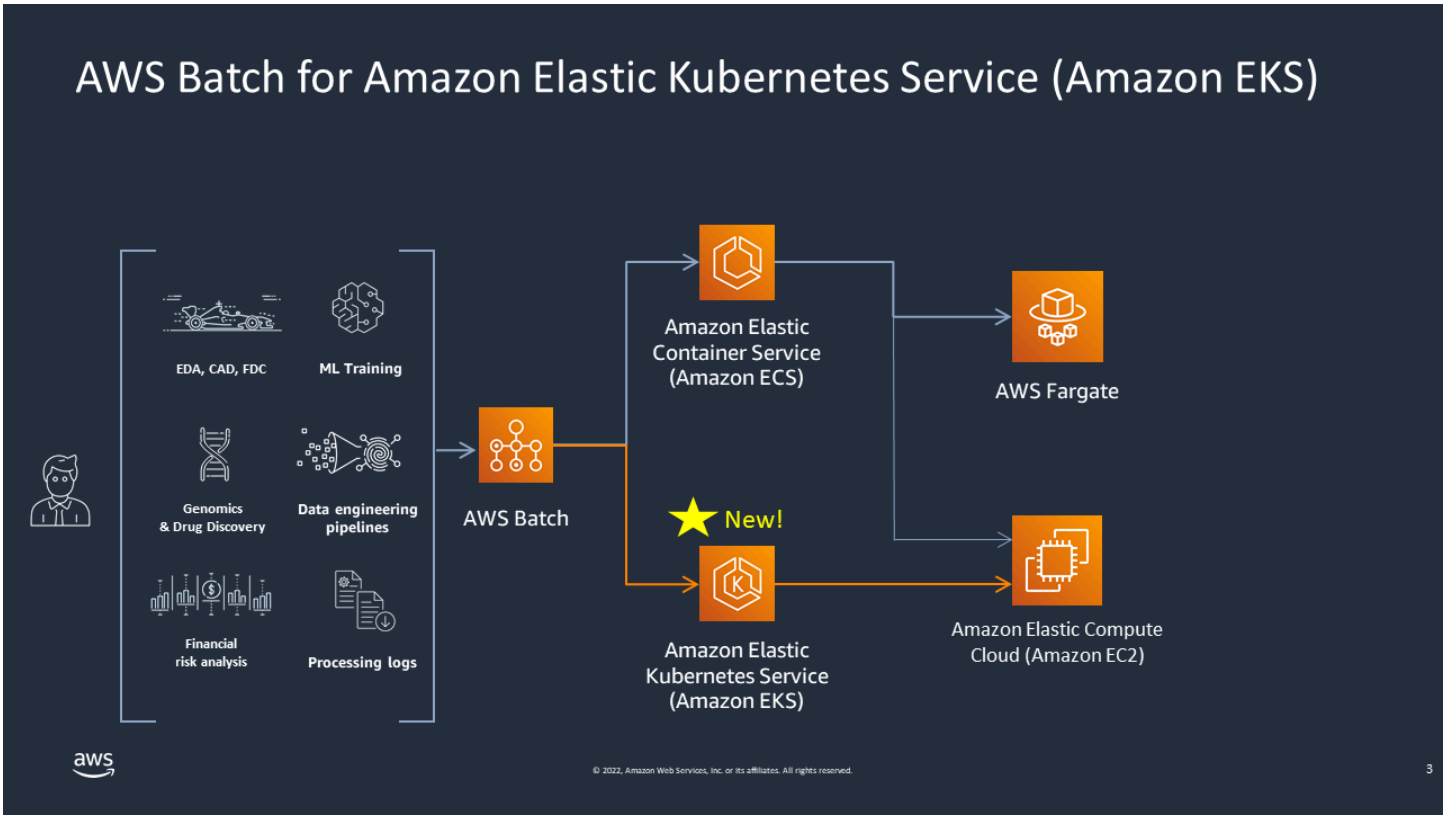
`type`

此必須為 `FARGATE` 或 `FARGATE_SPOT`。

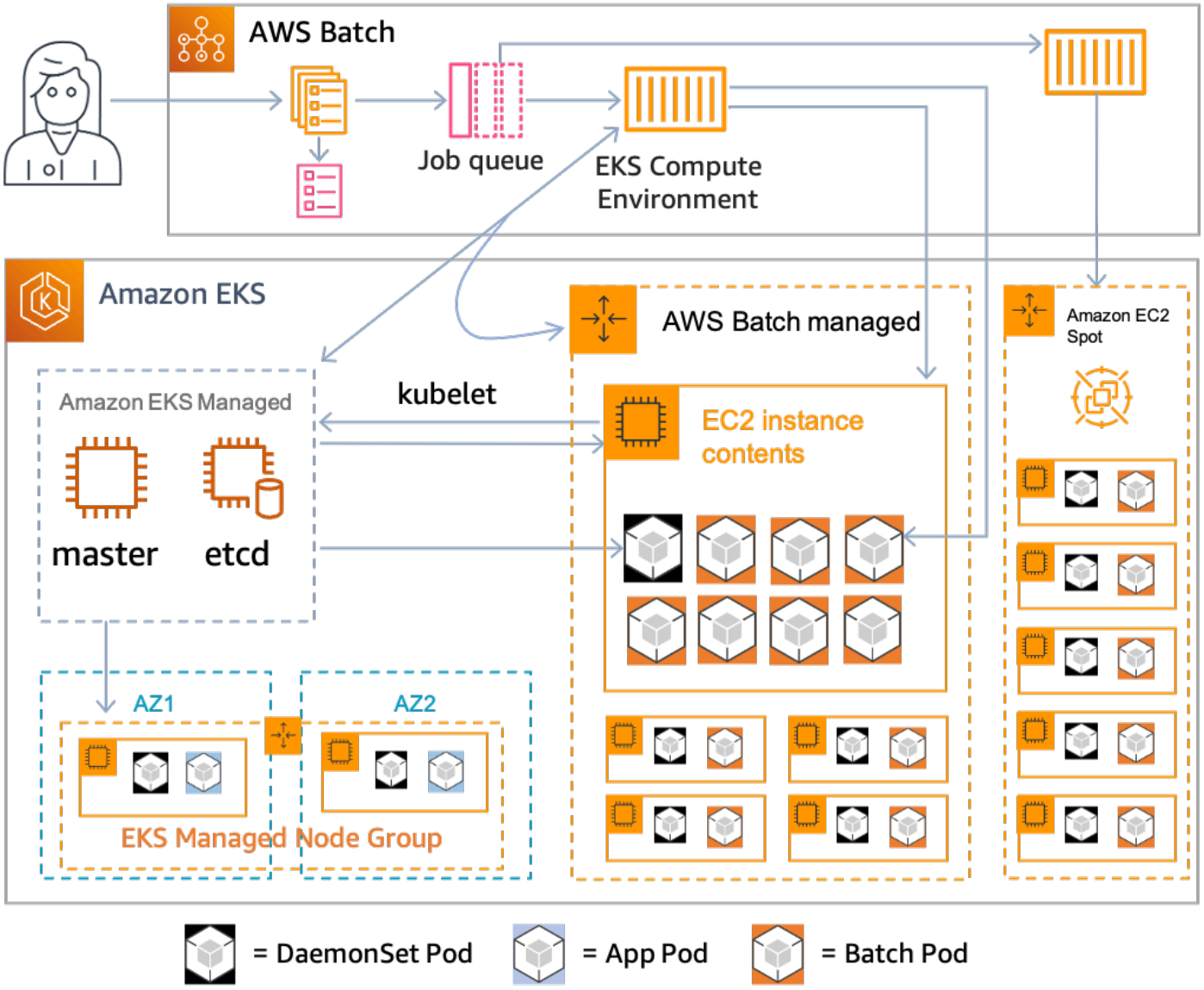
```
"type": "FARGATE_SPOT"
```

Amazon EKS 運算環境

在 Amazon EKS AWS Batch 上開始使用 提供建立 EKS 運算環境的簡短指南。本節提供 Amazon EKS 運算環境的詳細資訊。



AWS Batch 透過提供受管批次功能，簡化 Amazon EKS 叢集上的批次工作負載。這包括佇列、相依性追蹤、受管任務重試和優先順序、Pod 管理和節點擴展。 AWS Batch 可以處理多個可用區域和多個 Amazon EC2 執行個體類型和大小。 AWS Batch 整合數個 Amazon EC2 Spot 最佳實務，以容錯方式執行工作負載，減少中斷。您可以使用 AWS Batch 來執行少量的夜間任務或數百萬個任務關鍵任務。



AWS Batch 是一項受管服務，可協調Kubernetes叢集中由 Amazon Elastic Kubernetes Service (Amazon EKS) 管理的批次工作負載。 會使用「浮水印」模型，在叢集外部 AWS Batch 執行此協同運作。由於 AWS Batch 是受管服務，因此沒有Kubernetes元件 （ 例如 Operators 或 Custom Resources) 可在您的叢集中安裝或管理。 AWS Batch 只需要使用角色型存取控制 (RBAC) 來設定您的叢集， AWS Batch 以允許 與 Kubernetes API 伺服器通訊。 AWS Batch 呼叫 Kubernetes APIs Kubernetes 來建立、監控和刪除 Pod 和節點。

AWS Batch 具有內建擴展邏輯，可根據任務佇列負載擴展Kubernetes節點，並在任務容量配置方面進行最佳化。當任務佇列為空時， 會將節點 AWS Batch 縮減到您設定的最小容量，預設為零。 會 AWS Batch 管理這些節點的完整生命週期，並使用標籤和污點裝飾節點。如此一來，其他Kubernetes工作負載就不會放置在 管理的節點上 AWS Batch。例外狀況是 `DaemonSets`，其可將 AWS Batch 節點設為

目標，以提供正確執行任務所需的監控和其他功能。此外， AWS Batch 不會在叢集中未管理的節點上執行任務，特別是 Pod。如此一來，您就可以為叢集上的其他應用程式使用個別擴展邏輯和服務。

若要提交任務給 AWS Batch，您可以直接與 AWS Batch API 互動。 會將任務 AWS Batch 轉換為 ，podspecs然後建立請求，將 Pod 放置在 Amazon EKS 叢集 AWS Batch 中由 管理的節點上。您可以使用 等工具kubectl來檢視執行中的 Pod 和節點。當 Pod 完成其執行時， 會 AWS Batch 刪除其建立的 Pod，以維持Kubernetes系統較低的負載。

您可以透過連接有效的 Amazon EKS 叢集來開始使用 AWS Batch。然後將 AWS Batch 任務佇列連接到其中，並使用podspec同等屬性註冊 Amazon EKS 任務定義。最後，使用參考任務定義的 SubmitJob API 操作提交任務。如需詳細資訊，請參閱在 Amazon EKS AWS Batch 上開始使用。

Amazon EKS

主題

- Amazon EKS 預設 AMI
- 混合 AMI 環境
- 支援的 Kubernetes 版本
- 更新運算環境的Kubernetes版本
- Kubernetes 節點的共同責任
- 在 AWS Batch 受管節點DaemonSet上執行
- 自訂 Amazon EKS 啟動範本
- 如何從 EKS AL2 升級到 EKS AL2023

Amazon EKS 預設 AMI

建立 Amazon EKS 運算環境時，您不需要指定 Amazon Machine Image (AMI)。 會根據 CreateComputeEnvironment 請求中指定的Kubernetes版本和執行個體類型 AWS Batch ，選取 Amazon EKS 最佳化 AMI。一般而言，我們建議您使用預設 AMI 選擇。如需 Amazon EKS 最佳化 AMIs 的詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS 最佳化 Amazon Linux AMIs。

> **Important**
>
> Amazon Linux 2023 AMIs是 Amazon EKS AWS Batch 的預設值。
> AWS 將從 11/26/25 開始，結束對 Amazon EKS AL2-optimized和 AL2-accelerated AMIs支援。在 11/26/25end-of-support日期之後，您可以在 Amazon EKS 運算環境中繼續使用提供的

AWS Batch Amazon EKS 最佳化 Amazon Linux 2 AMIs，但這些運算環境將不再收到來自 的任何新軟體更新、安全修補程式或錯誤修正 AWS。如需從 AL2 升級到 AL2023 的詳細資訊，請參閱AWS Batch 《 使用者指南如何從 EKS AL2 升級到 EKS AL2023》中的 。

執行下列命令，以查看為您的 Amazon EKS 運算環境 AWS Batch 選取的 AMI 類型。下列範例是非 GPU 執行個體類型。

```
# compute CE example: indicates Batch has chosen the AL2 x86 or ARM EKS 1.32 AMI,
 depending on instance types
    $ aws batch describe-compute-environments --compute-environments My-Eks-CE1 \
        | jq '.computeEnvironments[].computeResources.ec2Configuration'
    [
      {
        "imageType": "EKS_AL2",
        "imageKubernetesVersion": "1.32"
      }
    ]
```

以下範例是 GPU 執行個體類型。

```
# GPU CE example: indicates Batch has choosen the AL2 x86 EKS Accelerated 1.32 AMI
    $ aws batch describe-compute-environments --compute-environments My-Eks-GPU-CE \
        | jq '.computeEnvironments[].computeResources.ec2Configuration'
    [
      {
        "imageType": "EKS_AL2_NVIDIA",
        "imageKubernetesVersion": "1.32"
      }
    ]
```

混合 AMI 環境

您可以使用啟動範本覆寫來建立具有 Amazon Linux 2 (AL2) 和 Amazon Linux 2023 (AL2023) AMIs運算環境。這適用於在從 AL2 轉換到 AL2023 時，針對不同的架構或在遷移期間使用不同的 AMIs。

> **Note**
>
> AWS 將從 11/26/25 開始，結束對 Amazon EKS AL2-optimized和 AL2-accelerated AMIs支援。雖然您可以在 11/26/25end-of-support日期之後，在 Amazon EKS 運算環境中繼續使用提

供的 AWS Batch Amazon EKS 最佳化 Amazon Linux 2 AMIs，但這些運算環境將不再收到來自 的任何新軟體更新、安全修補程式或錯誤修正 AWS。混合 AMI 環境在轉換期間非常有用，可讓您逐步將工作負載遷移至 AL2023，同時保持與現有 AL2-based工作負載的相容性。

使用兩種 AMI 類型的範例組態：

```
{
  "computeResources": {
    "launchTemplate": {
      "launchTemplateId": "TemplateId",
      "version": "1",
      "userdataType": "EKS_BOOTSTRAP_SH",
      "overrides": [
        {
          "instanceType": "c5.large",
          "imageId": "ami-al2-custom",
          "userdataType": "EKS_BOOTSTRAP_SH"
        },
        {
          "instanceType": "c6a.large",
          "imageId": "ami-al2023-custom",
          "userdataType": "EKS_NODEADM"
        }
      ]
    },
    "instanceTypes": ["c5.large", "c6a.large"]
  }
}
```

支援的 Kubernetes 版本

AWS Batch Amazon EKS 上的 目前支援下列Kubernetes版本：

- 1.34
- 1.33
- 1.32
- 1.31
- 1.30
- 1.29

當您使用 CreateComputeEnvironment API 操作或 UpdateComputeEnvironmentAPI 操作來建立或更新運算環境時，您可能會看到類似以下內容的錯誤訊息。如果您在 中指定不支援的Kubernetes版本，就會發生此問題EC2Configuration。

```
At least one imageKubernetesVersion in EC2Configuration is not supported.
```

若要解決此問題，請刪除運算環境，然後使用支援的Kubernetes版本重新建立它。

您可以在 Amazon EKS 叢集上執行次要版本升級。例如，1.yy即使不支援次要版本，您也可以將叢集從 1.xx 升級至 。

不過，在主要版本更新INVALID之後，運算環境狀態可能會變更為 。例如，如果您執行從 1.xx升級至 的主要版本升級2.yy。如果 不支援主要版本 AWS Batch，您會看到類似以下的錯誤訊息。

```
reason=CLIENT_ERROR - ... EKS Cluster version [2.yy] is unsupported
```

更新運算環境的Kubernetes版本

您可以使用 AWS Batch更新運算環境的Kubernetes版本，以支援 Amazon EKS 叢集升級。運算環境的Kubernetes版本是 AWS Batch 啟動 以執行任務之Kubernetes節點的 Amazon EKS AMI 版本。您可以在更新 Amazon EKS 叢集控制平面Kubernetes版本之前或之後，在其 Amazon EKS 節點上執行版本升級。我們建議您在升級控制平面之後更新節點。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的更新 Amazon EKS 叢集Kubernetes版本。

若要升級運算環境的Kubernetes版本，請使用 UpdateComputeEnvironment API 操作。

```
$ aws batch update-compute-environment \
    --compute-environment <compute-environment-name> \
    --compute-resources \
      'ec2Configuration=[{imageType=EKS_AL2,imageKubernetesVersion=1.32}]'
```

Kubernetes 節點的共同責任

維護運算環境是共同的責任。

- 請勿變更或移除 AWS Batch 節點、標籤、污點、命名空間、啟動範本或自動擴展群組。請勿將污點新增至 AWS Batch 受管節點。如果您進行任何這些變更，則無法支援您的運算環境，並發生失敗，包括閒置執行個體。

- 請勿將 Pod 設為 AWS Batch 受管節點的目標。如果您將 Pod 設為受管節點的目標，就會發生擴展中斷和任務佇列停滯。在自我管理節點或受管節點群組 AWS Batch 上執行不使用 的工作負載。如需詳細資訊，請參閱 Amazon EKS 使用者指南中的受管節點群組。
- 您可以DaemonSet鎖定要在 AWS Batch 受管節點上執行的 為目標。如需詳細資訊，請參閱在 AWS Batch 受管節點DaemonSet上執行。

AWS Batch 不會自動更新運算環境 AMIs。更新它們是您的責任。執行下列命令，將您的 AMIs 更新為最新的 AMI 版本。

```
$ aws batch update-compute-environment \
    --compute-environment <compute-environment-name> \
    --compute-resources 'updateToLatestImageVersion=true'
```

AWS Batch 不會自動升級Kubernetes版本。執行下列命令，將電腦環境的Kubernetes版本更新為 *1.32*。

```
$ aws batch update-compute-environment \
    --compute-environment <compute-environment-name> \
    --compute-resources \
      'ec2Configuration=[{imageType=EKS_AL2,imageKubernetesVersion=1.32}]'
```

更新至較新的 AMI 或 Kubernetes版本時，您可以指定是否要在任務更新時終止任務 (`terminateJobsOnUpdate`)，以及如果執行中的任務未完成 (`jobExecutionTimeoutMinutes`.)，則在替換執行個體之前等待多久。如需詳細資訊，請參閱 在中更新運算環境 AWS Batch和 UpdateComputeEnvironment API 操作中設定的基礎設施更新政策 (UpdatePolicy)。

在 AWS Batch 受管節點DaemonSet上執行

AWS Batch 會在 AWS Batch 受管Kubernetes節點上設定污點。您可以使用下列 DaemonSet，將要在 AWS Batch 受管節點上執行的 設為目標`tolerations`。

```
tolerations:
  - key: "batch.amazonaws.com/batch-node"
    operator: "Exists"
```

另一種做法是使用下列 `tolerations`。

```
tolerations:
  - key: "batch.amazonaws.com/batch-node"
    operator: "Exists"
    effect: "NoSchedule"
  - key: "batch.amazonaws.com/batch-node"
    operator: "Exists"
    effect: "NoExecute"
```

自訂 Amazon EKS 啟動範本

AWS Batch Amazon EKS 上的 支援啟動範本。啟動範本可以執行的操作有其限制。

Important

- 對於 EKS AL2 AMIs， AWS Batch 執行 `/etc/eks/bootstrap.sh`。請勿在啟動範本或 cloud-inituser-data指令碼`/etc/eks/bootstrap.sh`中執行 。除了 參數之外，您還可以將其他`--kubelet-extra-args`參數新增至 https：//bootstrap.sh。若要這樣做，請在 `/etc/aws-batch/batch.config` 檔案中設定 `AWS_BATCH_KUBELET_EXTRA_ARGS`變數。如需詳細資訊，請參閱下列範例。
- 對於 EKS AL2023， AWS Batch 利用來自 EKS 的 NodeConfigSpec 讓執行個體加入 EKS 叢集。 AWS Batch 會在 NodeConfigSpec 中為 EKS 叢集填入 ClusterDetails，您不需要指定它們。

Note

我們建議您不要在啟動範本中設定任何下列NodeConfigSpec設定，因為 AWS Batch 會覆寫您的值。如需詳細資訊，請參閱Kubernetes 節點的共同責任。

- `Taints`
- `Cluster Name`
- `apiServerEndpoint`
- `certificatAuthority`
- `CIDR`
- 請勿建立字首為 的標籤 `batch.amazonaws.com/`

> **Note**
>
> 如果在呼叫 CreateComputeEnvironment 之後變更啟動範本，UpdateComputeEnvironment則必須呼叫 來評估啟動範本的版本以進行取代。

主題

- 新增kubelet額外的引數
- 設定容器執行時間
- 掛載 Amazon EFS 磁碟區
- IPv6 支援

新增kubelet額外的引數

AWS Batch 支援將額外的引數新增至kubelet命令。如需支援的參數清單，請參閱 Kubernetes 文件kubelet中的 。在下列 EKS AL2 AMIs範例中， *--node-labels mylabel=helloworld* 會新增至kubelet命令列。

```
MIME-Version: 1.0
      Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

      --==MYBOUNDARY==
      Content-Type: text/x-shellscript; charset="us-ascii"

      #!/bin/bash
      mkdir -p /etc/aws-batch

      echo AWS_BATCH_KUBELET_EXTRA_ARGS=\"--node-labels mylabel=helloworld\" >> /etc/
aws-batch/batch.config

      --==MYBOUNDARY==--
```

對於 EKS AL2023 AMIs檔案格式為 YAML。如需支援的參數清單，請參閱 Kubernetes 文件NodeConfigSpec中的 。在下列 EKS AL2023 AMIs範例中， *--node-labels mylabel=helloworld* 會新增至kubelet命令列。

```
MIME-Version: 1.0
```

```
Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

--==MYBOUNDARY==
Content-Type: application/node.eks.aws

apiVersion: node.eks.aws/v1alpha1
kind: NodeConfig
spec:
  kubelet:
    flags:
    - --node-labels=mylabel=helloworld

--==MYBOUNDARY==--
```

設定容器執行時間

您可以使用 AWS Batch CONTAINER_RUNTIME環境變數，在受管節點上設定容器執行時間。下列範例會在bootstrap.sh執行containerd時將容器執行時間設定為 。如需詳細資訊，請參閱 Kubernetes 文件containerd中的 。

如果您使用的是最佳化EKS_AL2023或 EKS_AL2023_NVIDIA AMI，則不需要指定容器執行時間，因為僅支援容器化。

> **Note**
>
> CONTAINER_RUNTIME 環境變數等同於 --container-runtime的選項bootstrap.sh。如需詳細資訊，請參閱 Kubernetes 文件Options中的 。

```
MIME-Version: 1.0
      Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

      --==MYBOUNDARY==
      Content-Type: text/x-shellscript; charset="us-ascii"

      #!/bin/bash
      mkdir -p /etc/aws-batch

      echo CONTAINER_RUNTIME=containerd >> /etc/aws-batch/batch.config

      --==MYBOUNDARY==--
```

掛載 Amazon EFS 磁碟區

您可以使用啟動範本將磁碟區掛載到節點。在下列範例中，使用 `cloud-config`packages和 `runcmd`設定。如需詳細資訊，請參閱 cloud-init 文件中的雲端組態範例。

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

--==MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"

packages:
- amazon-efs-utils

runcmd:
- file_system_id_01=fs-abcdef123
- efs_directory=/mnt/efs

- mkdir -p ${efs_directory}
- echo "${file_system_id_01}:/ ${efs_directory} efs _netdev,noresvport,tls,iam 0 0"
 >> /etc/fstab
- mount -t efs -o tls ${file_system_id_01}:/ ${efs_directory}

--==MYBOUNDARY==--
```

若要在任務中使用此磁碟區，必須在 eksProperties 參數中新增至 RegisterJobDefinition。下列範例是任務定義的很大一部分。

```
{
    "jobDefinitionName": "MyJobOnEks_EFS",
    "type": "container",
    "eksProperties": {
        "podProperties": {
            "containers": [
                {
                    "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
                    "command": ["ls", "-la", "/efs"],
                    "resources": {
                        "limits": {
                            "cpu": "1",
                            "memory": "1024Mi"
                        }
                    },
```

```
                    "volumeMounts": [
                        {
                            "name": "efs-volume",
                            "mountPath": "/efs"
                        }
                    ]
                }
            ],
            "volumes": [
                {
                    "name": "efs-volume",
                    "hostPath": {
                        "path": "/mnt/efs"
                    }
                }
            ]
        }
    }
}
```

在節點中，Amazon EFS 磁碟區會掛載在 /mnt/efs目錄中。在 Amazon EKS 任務的容器中，磁碟區會掛載在 /efs目錄中。

IPv6 支援

AWS Batch 支援具有 IPv6 地址的 Amazon EKS 叢集。 AWS Batch 支援不需要自訂。不過，在開始之前，建議您檢閱《Amazon EKS 使用者指南》中將 IPv6 地址指派給 Pod 和服務中概述的考量事項和條件。

如何從 EKS AL2 升級到 EKS AL2023

Amazon EKS 最佳化 AMIs 提供兩個以 Amazon Linux 2 (AL2) 和 Amazon Linux 2023 (AL2023) 為基礎的系列。AL2023 是以 Linux 為基礎的作業系統，旨在為您的雲端應用程式提供安全、穩定且高效能的環境。如需 AL2 與 AL2023 之間差異的詳細資訊，請參閱《Amazon EKS 使用者指南》中的從 Amazon Linux 2 升級到 Amazon Linux 2023。

> **Important**
>
> AWS 將從 11/26/25 開始，結束對 Amazon EKS AL2-optimized和 AL2-accelerated AMIs 支援。我們建議在 AWS Batch 11/26/25 之前將 Amazon EKS 運算環境遷移至 Amazon Linux 2023，以維持最佳效能和安全性。雖然您可以在 11/26/25end-of-support日期之後，在

Amazon EKS 運算環境中繼續使用 AWS Batch提供的 Amazon EKS 最佳化 Amazon Linux 2 AMIs，但這些運算環境將不再收到任何新的軟體更新、安全修補程式或錯誤修正 AWS。end-of-life，您有責任在 Amazon EKS 最佳化 Amazon Linux 2 AMI 上維護這些運算環境。

根據運算環境的設定方式，您可以使用下列其中一個從 AL2 到 AL2023 的升級路徑。

使用 Ec2Configuration.ImageType 升級

- 如果您未使用啟動範本或啟動範本覆寫，請將 Ec2Configuration.ImageType 變更為 EKS_AL2023或 ，EKS_AL2023_NVIDIA然後執行 UpdateComputeEnvironment。
- 如果您指定 Ec2Configuration.ImageIdOverride，則 Ec2Configuration.ImageType 必須符合 Ec2Configuration.ImageIdOverride 中指定的 AMI 類型。

  如果您不相符ImageIdOverride，ImageType則節點不會加入叢集。

使用啟動範本進行升級

- 如果您在啟動範本或啟動範本覆寫中定義了任何kubelet額外引數，則需要將其更新為新的kubelet額外引數格式。

  如果您不符kubelet額外引數格式，則不會套用額外引數。
- 對於 AL2023 AMIs， containerd 是唯一支援的容器執行時間。您不需要在啟動範本EKS_AL2023中指定 的容器執行期。

  您無法使用 指定自訂容器執行時間EKS_AL2023。
- 如果您使用根據 指定 AMI 的啟動範本或啟動範本覆寫EKS_AL2023，則需要將 userdataType 設定為 EKS_NODEADM。

  如果您不相符 userdataType和 AMI，則節點不會加入 EKS 叢集。

的服務環境 AWS Batch

服務環境 AWS Batch 可讓 與 SageMaker AI 整合。服務環境包含 AWS Batch 提交和管理 SageMaker Training 任務所需的 SageMaker AI AWS Batch特定組態參數，同時提供佇列、排程和優先順序管理功能。

透過服務環境，資料科學家和 ML 工程師可以將具有優先順序的 SageMaker 訓練任務提交至服務任務佇列。此整合不需要手動協調 ML 工作負載、防止意外超支，並改善組織機器學習工作流程的資源使用率。

主題

- 什麼是 中的服務環境 AWS Batch
- 中的服務環境狀態和生命週期 AWS Batch
- 在 中建立服務環境 AWS Batch
- 在 中更新服務環境 AWS Batch
- 在 中刪除服務環境 AWS Batch

什麼是 中的服務環境 AWS Batch

服務環境是一種 AWS Batch 資源，其中包含 AWS Batch 與 SageMaker AI 整合所需的組態參數。服務環境 AWS Batch 可讓 AWS Batch提交和管理 SageMaker Training 任務，同時提供佇列、排程和優先順序管理功能。

服務環境可解決資料科學團隊在管理機器學習工作負載時面臨的常見挑戰。組織通常會限制訓練模型可用的執行個體數量，以防止意外超支、符合預算限制、節省預留執行個體的成本，或針對工作負載使用特定的執行個體類型。不過，資料科學家可能想要同時執行比其配置的執行個體更多工作負載，需要手動協調來決定何時執行哪些工作負載。

此協調挑戰會影響各種規模的組織，從只有少數資料科學家的團隊到大規模操作。隨著組織的成長，複雜性會增加，需要更多時間來管理工作負載協調，並且通常需要基礎設施管理員的參與。這些手動工作會浪費時間並降低執行個體效率，為客戶帶來實際成本。

透過服務環境，資料科學家和 ML 工程師可以將優先順序為可設定佇列的 SageMaker 訓練任務提交，確保工作負載在資源可用後立即自動執行，無需介入。此整合利用 AWS Batch廣泛的佇列和排程功能，讓客戶自訂其佇列和排程政策，以符合其組織的目標。

服務環境如何與其他 AWS Batch 元件搭配使用

服務環境與其他 AWS Batch 元件整合，以啟用 SageMaker Training 任務佇列：

- 任務佇列 - 服務環境與任務佇列相關聯，讓佇列能夠處理 SageMaker Training 任務的服務任務
- 服務任務 - 當您將服務任務提交至與服務環境相關聯的佇列時， AWS Batch 會使用環境的組態來提交對應的 SageMaker Training 任務
- 排程政策 - 服務環境使用 AWS Batch 排程政策來排定優先順序和管理 SageMaker Training 任務的執行順序

此整合可讓您利用 AWS Batch成熟的佇列和排程功能，同時維持 SageMaker Training 任務的完整功能和彈性。

服務環境的最佳實務

服務環境提供大規模管理 SageMaker 訓練任務的功能。遵循這些最佳實務可協助您最佳化成本、效能和營運效率，同時避免可能影響機器學習工作流程的常見組態問題。

規劃服務環境容量時，請考慮適用於 SageMaker Training 任務佇列的特定配額和限制。每個服務環境都有以執行個體數量表示的最大容量限制，可直接控制可以同時執行的 SageMaker 訓練任務數量。了解這些限制有助於防止資源爭用，並確保可預測的任務執行時間。

最佳服務環境效能取決於了解 SageMaker Training 任務排程的獨特特性。與傳統容器化任務不同，服務任務會轉換到 SCHEDULED 狀態，而 SageMaker AI 會取得並佈建必要的訓練執行個體。這表示任務開始時間可能會因執行個體可用性和區域容量而有很大差異。

> **Important**
>
> 服務環境具有特定的配額，可能會影響您擴展 SageMaker Training 工作負載的能力。每個帳戶最多可以建立 50 個服務環境，每個任務佇列僅支援一個相關聯的服務環境。此外，個別任務的服務請求承載限制為 10 KiB，而 SubmitServiceJob API 限制為每個帳戶每秒 5 次交易。在容量規劃期間了解這些限制可防止非預期的擴展限制。

有效監控服務環境需要同時注意 AWS Batch 和 SageMaker AI 服務指標。任務狀態轉換可提供有關系統效能的寶貴洞見，特別是在SCHEDULED狀態中花費的時間，這表示容量可用性模式。服務環境會維護自己的生命週期狀態，類似於運算環境，並透過 CREATING、INVALID、 VALID和 狀態進行轉換，

這些DELETING狀態應受到營運運作狀態的監控。具有成熟監控實務的組織通常會追蹤佇列深度、任務完成率和執行個體使用率模式，以隨著時間最佳化其服務環境組態。

中的服務環境狀態和生命週期 AWS Batch

服務環境會維持生命週期狀態，指出其目前的操作狀態和處理 SageMaker Training 任務的準備程度。了解這些狀態可協助您監控服務環境運作狀態、疑難排解組態問題，並確保可靠的任務處理。狀態管理系統遵循來自運算環境的已建立模式，同時適應 SageMaker Training 任務整合的獨特需求。

服務環境狀態是由 AWS Batch 根據組態驗證、資源可用性和操作運作狀態檢查自動管理。與管理實體基礎設施的運算環境不同，服務環境著重於組態驗證，以及與 SageMaker AI 服務的整合準備程度。狀態轉換可讓您了解您的服務環境是否可以成功提交和管理 SageMaker Training 任務。

服務環境狀態定義

服務環境可以處於四個可能狀態之一，指出其目前的操作狀態和處理 SageMaker Training 任務的準備程度。每個狀態代表服務環境生命週期中的特定階段，從初始建立到操作準備，再到最終刪除。下表說明每個狀態及其意義：

State	Description
CREATING	建立服務環境時的初始狀態。在此狀態下， 會 AWS Batch 驗證組態參數，並與 SageMaker AI 服務建立整合。服務環境無法處理任務，與其相關聯的任何任務佇列都不會接受服務任務提交。對於正確設定的服務環境，建立程序通常會在幾秒鐘內完成。
VALID	表示服務環境已通過所有組態驗證檢查並準備好處理 SageMaker 訓練任務的操作狀態。此狀態表示服務環境組態正確、具備所有必要許可， AWS Batch 並可代表您成功將任務提交至 SageMaker AI。服務環境會在此狀態下花費大部分的操作生命週期。
INVALID	狀態表示服務環境遇到組態或許可問題，導致無法處理 SageMaker Training 任務。在解決基礎

State	Description
	問題之前，與無效服務環境相關聯的任務佇列無法處理新的服務任務提交。
DELETING	當您請求刪除服務環境時發生的狀態。在此狀態下， AWS Batch 會確保沒有任何作用中的 SageMaker 訓練任務與環境相關聯，並執行必要的清除操作。處於此狀態的服務環境無法處理新的任務提交，刪除程序會在所有相關資源正確清除後完成。

服務環境狀態轉換

服務環境狀態轉換會根據組態變更、驗證結果和操作運作狀態監控自動進行。 AWS Batch 服務會持續監控服務環境的運作狀態，並相應地更新狀態。了解這些轉換可協助您預測組態變更何時生效，以及如何解決導致無效狀態的問題。

成功建立和驗證後，服務環境會從 轉換為 CREATING VALID。此轉換會確認所有組態參數皆正確、已正確設定必要的 IAM 許可，而且服務環境可以成功與 SageMaker AI 服務整合。一旦進入 VALID 狀態，相關聯的任務佇列就可以開始處理服務任務提交。

INVALID 當組態驗證失敗或相依性無法使用時，服務環境會從 VALID 轉換為 。這可能是由於 IAM 角色修改、違反配額的容量限制變更，或影響服務環境運作能力的外部資源修改。狀態原因欄位提供有關導致無效狀態之原因的特定詳細資訊。

解決基礎問題VALIDINVALID後，服務環境可以從 轉換回 。這可能包括更新 IAM 許可、更正容量組態，或還原對所需 AWS 資源的存取。轉換通常會在 AWS Batch 偵測到組態問題已解決時自動發生。

在 中建立服務環境 AWS Batch

您必須先建立服務環境 AWS Batch，才能在 中執行 SageMaker Training 任務。您可以建立服務環境，其中包含 與 SageMaker AI 服務 AWS Batch 整合所需的組態參數，並代表您提交 SageMaker Training 任務。

先決條件

建立服務環境之前，請確定您有：

- IAM 許可 – 建立和管理服務環境的許可。如需詳細資訊，請參閱AWS Batch IAM 政策、角色和許可。

Create a service environment (AWS Console)

使用 AWS Batch 主控台透過 Web 界面建立服務環境。

建立服務環境

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇 Environments (環境)。
3. 選擇建立環境，然後選取服務環境。
4. 針對服務環境組態，選擇 SageMaker AI。
5. 在名稱中，輸入服務環境的唯一名稱。有效字元為 a-z、A-Z、0-9、連字號 (-) 和底線 (_)。
6. 針對執行個體數量上限，輸入並行訓練執行個體數量上限
7. （選用）透過選擇新增標籤並輸入鍵/值對來新增標籤。
8. 選擇下一步。
9. 檢閱新服務環境的詳細資訊，然後選擇建立服務環境。

Create a service environment (AWS CLI)

使用 `create-service-environment`命令透過 CLI AWS 建立服務環境。

建立服務環境

1. 使用基本必要參數建立服務環境：

```
aws batch create-service-environment \
    --service-environment-name my-sagemaker-service-env \
    --service-environment-type SAGEMAKER_TRAINING \
    --capacity-limits capacityUnit=NUM_INSTANCES,maxCapacity=10
```

2. （選用）使用標籤建立服務環境：

```
aws batch create-service-environment \
    --service-environment-name my-sagemaker-service-env \
    --service-environment-type SAGEMAKER_TRAINING \
```

```
    --capacity-limits capacityUnit=NUM_INSTANCES,maxCapacity=10 \
    --tags team=data-science,project=ml-training
```

3. 確認已成功建立服務環境：

```
aws batch describe-service-environments \
    --service-environment my-sagemaker-service-env
```

服務環境會出現在具有 CREATING 狀態的環境清單中。當建立成功完成時，狀態會變更為 ，VALID且服務環境已準備好新增服務任務佇列，以便服務環境可以開始處理任務。

在 中更新服務環境 AWS Batch

您可以更新服務環境來修改其容量限制、變更其操作狀態，或更新資源標籤。服務環境更新可讓您在 SageMaker Training 工作負載需求變更或修改操作設定時調整容量，而無需重新建立環境。在更新服務環境之前，請了解哪些參數可以修改，以及變更對執行中任務的影響。

您可以變更服務環境的容量限制、狀態或標籤。

Update a service environment (AWS Console)

使用 AWS Batch 主控台透過 Web 界面更新服務環境。

更新服務環境

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇 Environments (環境)。
3. 選擇服務環境索引標籤。
4. 選擇要更新的服務環境。
5. 選擇動作，然後選擇：
   - 狀態 - 選擇啟用或停用以變更狀態。
   - 容量限制 - 修改執行個體數量上限
6. 選擇儲存變更以套用變更。

服務環境會立即更新。檢查環境詳細資訊，以確認已成功套用變更。如果您停用服務環境，相關聯的任務佇列將停止處理新的服務任務提交，直到您重新啟用為止。

Update a service environment (AWS CLI)

使用 update-service-environment命令透過 CLI AWS 修改服務環境。

更新服務環境容量限制

1. 更新服務環境的容量限制：

```
aws batch update-service-environment \
    --service-environment my-sagemaker-service-env \
    --capacity-limits capacityUnit=NUM_INSTANCES,maxCapacity=20
```

2. 確認已成功套用更新：

```
aws batch describe-service-environments \
    --service-environments my-sagemaker-service-env
```

更新服務環境狀態

1. 停用服務環境以停止處理新任務：

```
aws batch update-service-environment \
    --service-environment my-sagemaker-service-env \
    --state DISABLED
```

2. 重新啟用服務環境以繼續處理：

```
aws batch update-service-environment \
    --service-environment my-sagemaker-service-env \
    --state ENABLED
```

服務環境更新會立即生效。監控服務環境狀態，以確保更新在提交新任務之前成功完成。

在 中刪除服務環境 AWS Batch

您可以在 SageMaker Training 任務不再需要服務環境時將其刪除。刪除服務環境會移除組態，並防止進一步提交任務。刪除服務環境之前，請確定沒有任何作用中的 SageMaker Training 任務相依於它，而且沒有工作佇列與服務環境相關聯。

> **Important**
>
> 服務環境刪除是不可復原的。一旦刪除，您就無法復原服務環境或其組態。如果您未來需要類似的功能，則必須使用所需的設定建立新的服務環境。如果您稍後可能需要重新啟用服務環境，請考慮停用服務環境，而不是刪除。

> **Note**
>
> 刪除帳戶中的所有服務環境不會自動移除為 AWS Batch 和 SageMaker AI 整合建立的服務連結角色。服務連結角色仍可用於未來的服務環境建立。如果您想要移除服務連結角色，則必須在確保帳戶中不存在任何服務環境之後，使用 IAM 分別將其刪除。

刪除先決條件

您必須先取消任何服務任務佇列的關聯，然後停用服務環境，才能刪除服務環境。

刪除服務環境之前：

- 檢查作用中任務 - 確保目前沒有 SageMaker 訓練任務透過服務環境執行。
- 檢閱任務佇列 - 識別與服務環境相關聯的任務佇列，並將任務佇列與不同的服務環境建立關聯，或停用和刪除任務佇列。

任務佇列管理：與已刪除的服務環境相關聯的任務佇列仍然存在，但無法處理服務任務。在刪除原始服務環境之前，您應該刪除未使用的任務佇列，或將其與不同的服務環境建立關聯。

Delete a service environment (AWS Console)

使用 AWS Batch 主控台透過 Web 界面刪除服務環境。

刪除服務環境

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇 Environments (環境)。
3. 選擇服務環境索引標籤，然後選擇服務環境。
4. 如果服務環境已啟用，請選擇動作，然後選擇停用。
5. 一旦服務環境停用，請選擇動作，然後選擇刪除。

6. 在確認對話方塊中，選擇確認。

發生刪除時，服務環境會顯示 DELETING 狀態。刪除完成後，服務環境會從環境清單中消失。

Delete a service environment (AWS CLI)

使用 `delete-service-environment`命令透過 CLI AWS 移除服務環境。

刪除服務環境

1. 檢查與服務環境相關聯的任務佇列：

```
aws batch describe-job-queues
```

如果有任何與服務環境相關聯的任務佇列，您可以取消任務佇列與服務環境的關聯，並將其與不同的服務環境建立關聯，或刪除任務佇列。

2. 停用服務環境：

```
aws batch update-service-environment \
    --service-environment my-sagemaker-service-env \
    --state DISABLED
```

3. 刪除服務環境：

```
aws batch delete-service-environment \
    --service-environment my-sagemaker-service-env
```

4. 監控刪除程序：

```
aws batch describe-service-environments \
    --service-environment my-sagemaker-service-env
```

在刪除程序期間，服務環境會轉換為 DELETING 狀態。刪除完成後，服務環境不會再列在描述操作中。相關聯的任務佇列會保留，但在與不同的服務環境相關聯之前，無法處理服務任務。

任務佇列

任務會提交至其所在的任務佇列，直到可以排定在運算環境中執行。 AWS 帳戶可以有多個任務佇列。例如，您可以建立使用 Amazon EC2 隨需執行個體進行高優先順序任務的佇列，以及使用 Amazon EC2 Spot 執行個體進行低優先順序任務的另一個佇列。任務佇列具有排程器用來決定哪些任務應該先評估佇列以供執行的優先順序。

主題

- 建立任務佇列
- 在 中檢視任務佇列 AWS Batch
- 在 中刪除任務佇列 AWS Batch
- 公平共用排程政策
- 資源感知排程

建立任務佇列

您必須先建立任務佇列 AWS Batch，才能在 中提交任務。當您建立任務佇列時，您可以將一或多個運算環境與佇列建立關聯，並指派偏好順序。

您也可以將優先順序設定為任務佇列，以決定 AWS 批次排程器放置任務的順序。這表示，如果運算環境與多個任務佇列相關聯，則會指定優先順序較高的任務佇列。

主題

- 建立 Amazon EC2 任務佇列
- 建立 Fargate 任務佇列
- 建立 Amazon EKS 任務佇列
- 在 中建立 SageMaker 訓練任務佇列 AWS Batch
- 任務佇列範本

建立 Amazon EC2 任務佇列

請完成下列步驟，以建立 Amazon Elastic Compute Cloud (Amazon EC2) 的任務佇列。

建立 Amazon EC2 任務佇列

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇任務佇列。
4. 選擇建立。
5. 針對協調類型，選擇 Amazon Elastic Compute Cloud (Amazon EC2)。
6. 在名稱中，輸入任務佇列的唯一名稱。名稱長度上限為 128 個字元，且可包含大小寫字母、數字和底線 (_)。
7. 針對優先順序，輸入任務佇列優先順序的整數值。優先順序較高的任務佇列會在與相同運算環境相關聯的優先順序較低的任務佇列之前執行。優先順序會依遞減順序決定。例如，優先順序值為 10 的任務佇列的排程優先順序會高於優先順序值為 1 的任務佇列。
8. （選用）針對排程政策 Amazon Resource Name (ARN)，選擇現有的排程政策。
9. 對於連線的運算環境，請從清單中選擇一或多個運算環境，以與任務佇列建立關聯。依您希望佇列嘗試放置任務佇列的順序選取運算環境。任務排程器會使用您在 中選取運算環境的順序，來判斷哪個運算環境會啟動指定的任務。運算環境必須處於 `VALID` 狀態，才能將它們與任務佇列建立關聯。您可以將多達三個運算環境與工作佇列建立關聯。如果您沒有現有的運算環境，請選擇建立運算環境

   > **Note**
   >
   > 與任務佇列相關聯的所有運算環境必須共用相同的佈建模型。 AWS Batch 不支援在單一任務佇列中混合佈建模型。

10. 針對運算環境順序，選擇向上和向下箭頭以設定您想要的順序。
11. 選擇建立任務佇列以完成並建立您的任務佇列。

建立 Fargate 任務佇列

完成下列步驟以建立任務佇列 AWS Fargate。

建立 Fargate 任務佇列

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。

3. 在導覽窗格中，選擇任務佇列。
4. 選擇建立。
5. 針對協調類型，選擇 Fargate。
6. 在名稱中，輸入任務佇列的唯一名稱。名稱長度上限為 128 個字元，且可包含大小寫字母、數字和底線 (_)。
7. 針對優先順序，輸入任務佇列優先順序的整數值。優先順序較高的任務佇列會在與相同運算環境相關聯的優先順序較低的任務佇列之前執行。優先順序會依遞減順序決定。例如，優先順序值為 10 的任務佇列的排程優先順序會高於優先順序值為 1 的任務佇列。
8. （選用）針對排程政策 Amazon Resource Name (ARN)，選擇現有的排程政策。
9. 對於連線的運算環境，請從清單中選擇一或多個運算環境，以與任務佇列建立關聯。依您希望佇列嘗試放置任務佇列的順序選取運算環境。任務排程器會使用您在 中選取運算環境的順序，來判斷哪個運算環境會啟動指定的任務。運算環境必須處於 `VALID` 狀態，才能將它們與任務佇列建立關聯。您可以將多達三個運算環境與工作佇列建立關聯。

   > **Note**
   >
   > 與任務佇列相關聯的所有運算環境必須共用相同的佈建模型。 AWS Batch 不支援在單一任務佇列中混合佈建模型。

10. 針對運算環境順序，選擇向上和向下箭頭以設定您想要的順序。
11. 選擇建立任務佇列以完成並建立您的任務佇列。

建立 Amazon EKS 任務佇列

請完成下列步驟，以建立 Amazon Elastic Kubernetes Service (Amazon EKS) 的任務佇列。

建立 Amazon EKS 任務佇列

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇任務佇列。
4. 選擇建立。
5. 針對協調類型，選擇 Amazon Elastic Kubernetes Service (Amazon EKS)。
6. 在名稱中，輸入任務佇列的唯一名稱。名稱長度上限為 128 個字元，且可包含大小寫字母、數字和底線 (_)。

7. 在 Priority (優先順序)，為任務佇列的優先順序輸入整數值。優先順序較高的任務佇列會在與相同運算環境相關聯的優先順序較低的任務佇列之前執行。優先順序會依遞減順序決定。例如，優先順序值為 10 的任務佇列的排程優先順序會高於優先順序值為 1 的任務佇列。
8. （選用）針對排程政策 Amazon Resource Name (ARN)，選擇現有的排程政策。
9. 對於連線的運算環境，請從清單中選擇一或多個運算環境，以與任務佇列建立關聯。依您希望佇列嘗試放置任務佇列的順序選取運算環境。任務排程器會使用您在 中選取運算環境的順序，來判斷哪個運算環境會啟動指定的任務。運算環境必須處於 `VALID` 狀態，才能將它們與任務佇列建立關聯。您可以將多達三個運算環境與工作佇列建立關聯。

   > **Note**
   >
   > 與任務佇列相關聯的所有運算環境必須共用相同的佈建模型。 AWS Batch 不支援在單一任務佇列中混合佈建模型。

   > **Note**
   >
   > 與任務佇列相關聯的所有運算環境必須共用相同的架構。 AWS Batch 不支援在單一任務佇列中混合運算環境架構類型。

10. 針對運算環境順序，選擇向上和向下箭頭以設定您想要的順序。
11. 選擇建立任務佇列以完成並建立您的任務佇列。

在 中建立 SageMaker 訓練任務佇列 AWS Batch

SageMaker Training 任務佇列會直接與 SageMaker AI 服務整合，以提供無伺服器任務排程，而不需要您管理基礎運算基礎設施。

先決條件

在建立 SageMaker 訓練任務佇列之前，請確定您已：

- 服務環境 – 定義容量限制的服務環境。如需詳細資訊，請參閱在 中建立服務環境 AWS Batch。
- IAM 許可 – 建立和管理 AWS Batch 任務佇列和服務環境的許可。如需詳細資訊，請參閱AWS Batch IAM 政策、角色和許可。

Create a SageMaker Training job queue (AWS Batch console)

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇任務佇列和建立。
3. 針對協調類型，選擇 SageMaker Training。
4. 對於任務佇列組態：
   a. 在名稱中，輸入任務佇列的名稱。
   b. 針對 Priority，輸入介於 0 到 1000 之間的值。具有較高優先順序的任務佇列會優先於服務環境。
   c. （選用）針對排程政策 Amazon Resource Name (ARN)，選擇現有的排程政策。
   d. 對於連線的服務環境，從清單中選擇服務環境，以與任務佇列建立關聯。
5. （選用）針對任務狀態限制：
   a. 針對設定錯誤，選擇SERVICE_ENVIRONMENT_MAX_RESOURCE並輸入最長執行時間（秒）。
   b. 針對容量，選擇INSUFFICIENT_INSTANCE_CAPACITY並輸入最大執行時間（秒）。
6. 選擇建立任務佇列

Create a SageMaker Training job queue (AWS CLI)

使用 `create-job-queue`命令來建立 SageMaker Training 任務佇列。

下列範例會建立使用服務環境的基本 SageMaker Training 任務佇列：

```
aws batch create-job-queue \
  --job-queue-name my-sm-training-fifo-jq \
  --job-queue-type SAGEMAKER_TRAINING \
  --priority 1 \
  --service-environment-order order=1,serviceEnvironment=ExampleServiceEnvironment
```

以您的服務環境名稱取代 *ExampleServiceEnvironment*。

此命令會傳回類似以下的輸出：

```
{
  "jobQueueName": "my-sm-training-fifo-jq",
```

```
  "jobQueueArn": "arn:aws:batch:region:account:job-queue/my-sm-training-fifo-jq"
}
```

建立您的任務佇列後，請確認其已成功建立且處於有效狀態。

使用 describe-job-queues命令來檢視任務佇列的詳細資訊：

```
aws batch describe-job-queues --job-queues my-sm-training-fifo-jq
```

此命令會傳回類似以下的輸出：

```
{
  "jobQueues": [
    {
      "jobQueueName": "my-sm-training-fifo-jq",
      "jobQueueArn": "arn:aws:batch:region:account:job-queue/my-sm-training-fifo-
jq",
      "state": "ENABLED",
      "status": "VALID",
      "statusReason": "JobQueue Healthy",
      "priority": 1,
      "computeEnvironmentOrder": [],
      "serviceEnvironmentOrder": [
        {
          "order": 1,
          "serviceEnvironment": "arn:aws:batch:region:account:service-
environment/ExampleServiceEnvironment"
        }
      ],
      "jobQueueType": "SAGEMAKER_TRAINING",
      "tags": {},
      "jobStateTimeLimitActions": []
    }
  ]
}
```

請確定：

- state 是 ENABLED
- status 是 VALID
- statusReason 是 JobQueue Healthy

- `jobQueueType` 是 `SAGEMAKER_TRAINING`
- `serviceEnvironmentOrder` 參考您的服務環境

任務佇列範本

以下是空的任務佇列範本。您可以使用此範本來建立任務佇列。然後，您可以將此任務佇列儲存至檔案，並將其與 AWS CLI `--cli-input-json`選項搭配使用。如需這些參數的詳細資訊，請參閱 AWS Batch API 參考中的 CreateJobQueue。

> **Note**
>
> 您可以使用下列 AWS CLI 命令產生任務佇列範本。
>
> ```
> $ aws batch create-job-queue --generate-cli-skeleton
> ```

```
{
   "computeEnvironmentOrder": [
      {
         "computeEnvironment": "",
         "order": 0
      }
   ],
   "jobQueueName": "",
   "jobStateTimeLimitActions": [
      {
         "state": "RUNNABLE",
         "action": "CANCEL",
   "maxTimeSeconds": 0,
         "reason": ""

      }
   ],
   "priority": 0,
   "schedulingPolicyArn": "",
   "state": "ENABLED",
   "tags": {
      "KeyName": ""
   }
```

```
}
```

在 中檢視任務佇列 AWS Batch

在您建立任務佇列並提交任務之後，請務必能夠監控其進度。您可以使用任務詳細資訊頁面來檢閱、管理和監控您的任務佇列。

檢視任務佇列資訊

在 AWS Batch 主控台中，選取導覽窗格中的任務佇列，然後選擇所需的任務佇列以檢視其詳細資訊。在此頁面上，您可以檢閱和管理任務佇列，並查看佇列操作的其他資訊，例如任務佇列快照、任務狀態限制、環境順序、標籤和任務佇列的 JSON 程式碼。

任務佇列詳細資訊

本節提供任務佇列的概觀和維護選項。請務必注意，您可以在本節中找到 Amazon Resource Name (ARN)。

若要透過 尋找此資訊 AWS Command Line Interface，請使用 `DescribeJobQueues`操作搭配任務佇列名稱或對應的 ARN。

作用中共享

對於使用公平共用排程的任務佇列， AWS Batch 提供不同共用識別符如何使用容量的可見性。此資訊可協助您了解資源分佈，並識別可能需要調整的共享。

> **Note**
>
> 只有在任務佇列的排程演算法為公平共用時，才會顯示作用中共用索引標籤。

前 20 個作用中共享區段顯示已排程、開始和執行任務的共享識別符。此檢視包括：

- 共用識別符名稱 - 共用的唯一識別符。

  共用識別符是為公平共用排程將任務分組的標籤。當您提交具有相同共用識別符的任務時， AWS Batch 會將它們視為相同工作負載的一部分，以供資源配置之用。共用識別符有助於確保在不同團隊、專案或工作負載類型之間公平分配運算容量。如需詳細資訊，請參閱使用共用識別符來識別工作負載。
- 容量使用率 - 任務設定為使用的資源量。這是在 `instances`、 或 中測量`cpu`，vCPU取決於環境。

- 檢視任務動作 - 查看該共享所有任務的連結。

您可以同時以清單和圖表格式檢視此資訊：

- 清單檢視 - 具有確切容量數字的表格式顯示
- 圖表檢視 - 顯示相對使用率的視覺化長條圖

任務佇列快照

本節提供佇列中前 100 個RUNNABLE任務的靜態清單。您可以使用搜尋欄位，從結果區段中的任何資料欄搜尋資訊，以縮小清單範圍。快照結果區域中的任務會根據任務佇列的執行策略進行排序。對於 first-in-first-out(FIFO) 任務佇列，任務的排序是根據提交時間。對於公平共用排程任務佇列，任務的排序是根據任務優先順序和共用用量。容量必要欄位會顯示任務設定為使用的資源量。

由於結果是任務佇列的快照，因此結果清單不會自動更新。若要更新清單，請選擇區段頂端的重新整理。選擇任務的名稱超連結以導覽至任務詳細資訊，並檢視任務的狀態和其他相關資訊。

若要透過 尋找此資訊 AWS CLI，請使用 GetJobQueueSnapshot操作搭配任務佇列名稱或對應的 ARN。

```
aws batch get-job-queue-snapshot --job-queue my-sm-training-fifo-jq
```

任務狀態限制

使用此索引標籤來檢閱任務在取消之前可保持 RUNNABLE 狀態的時間組態資訊。

若要透過 尋找此資訊 AWS CLI，請使用 DescribeJobQueues操作搭配任務佇列名稱或對應的 ARN。

環境順序

如果您的任務佇列在多個環境中執行，此索引標籤會提供其順序和概觀。

若要透過 尋找此資訊 AWS CLI，請使用 DescribeJobQueues操作搭配任務佇列名稱或對應的 ARN。

Tags (標籤)

使用此索引標籤來檢閱和管理與此任務佇列相關聯的標籤。

JSON

使用此索引標籤來複製與此任務佇列相關聯的 JSON 程式碼。然後，您可以將 JSON 重複使用於 AWS CloudFormation 範本和 AWS CLI 指令碼。

在 中刪除任務佇列 AWS Batch

當您不再需要任務佇列時，您可以停用和刪除任務佇列。

Delete a job queue (AWS Batch console)

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在導覽窗格中，選擇任務佇列，然後選擇任務佇列。
3. 選擇動作，然後選擇停用。
4. 一旦任務佇列的狀態停用，請選擇動作，然後選擇刪除。
5. 在模態視窗中，選擇刪除任務佇列。

Delete a job queue (AWS CLI)

1. 停用任務佇列以防止提交新的任務：

```
aws batch update-job-queue \
  --job-queue my-sm-training-fifo-jq \
  --state DISABLED
```

2. 等待任何執行中的任務完成，然後刪除任務佇列：

```
aws batch delete-job-queue \
  --job-queue my-sm-training-fifo-jq
```

公平共用排程政策

AWS Batch 排程器會評估何時、何處以及如何執行提交至任務佇列的任務。如果您在建立任務佇列時未指定排程政策， AWS Batch 任務排程器會預設為先進先進先出 (FIFO) 策略。FIFO 策略可能會導致重要的任務「卡在」先前提交的任務後面。透過指定不同的排程政策，您可以根據您的特定需求配置運算資源。

> **Note**
>
> 如果您想要排程任務執行的特定順序，請使用 SubmitJob 中的 dependsOn 參數來指定每個任務的相依性。

如果您建立排程政策並將其連接到任務佇列，則會開啟公平共用排程。如果任務佇列具有排程政策，排程政策會決定任務執行的順序。如需詳細資訊，請參閱使用公平共用排程政策來指派共用識別符。

主題

- 使用共用識別符來識別工作負載
- 使用公平共用排程政策來指派共用識別符
- 使用公平共享排程來協助排程任務
- 教學課程：建立排程政策
- 參考：排程政策範本

使用共用識別符來識別工作負載

您可以使用共用識別符來標記任務，並區分使用者和工作負載。 AWS Batch 排程器會使用 *(T * weightFactor)*公式追蹤每個共用識別符的用量，其中 *T* 是一段時間內的 vCPU 用量。排程器會從共用識別符挑選使用量最低的任務。您可以使用共用識別符，而不覆寫它。

> **Note**
>
> 共用識別符在任務佇列中是唯一的，不會跨任務佇列彙總。

您可以設定公平共用排程優先順序，以設定任務在共用識別符上執行的順序。排程優先順序較高的任務會先排程。如果您未指定公平共享排程政策，則提交至任務佇列的所有任務都會以 FIFO 順序排程。當您提交任務時，您無法指定共用識別符或公平共用排程優先順序。

> **Note**
>
> 除非明確覆寫，否則連接的運算資源會平均分配在所有共用識別符中。

使用公平共用排程政策來指派共用識別符

您可以使用排程政策來設定任務佇列中的運算資源如何在使用者或工作負載之間配置。使用公平共用排程政策，您可以將不同的共用識別符指派給工作負載或使用者。 AWS Batch 會為每個共用識別符指派一段時間內可用資源總數的百分比。

公平共享百分比是使用 shareDecaySeconds和 shareDistribution值計算。您可以透過將共享衰減時間指派給政策，將時間新增至公平共享分析。增加時間可讓時間的權重增加，並減少定義的權重。您可以透過指定運算保留來保留未處於作用中狀態的共用識別符的運算資源。如需詳細資訊，請參閱 SchedulingPolicyDetail。

使用公平共享排程來協助排程任務

公平共用排程提供一組控制項，可協助排程任務。

> **Note**
>
> 如需排程政策參數的詳細資訊，請參閱 SchedulingPolicyDetail。

- 共用衰減秒數 – AWS Batch 排程器用來計算每個共用識別符公平共用百分比的期間 （以秒為單位）。零值表示只會測量目前的用量。較長的衰減時間會隨著時間增加權重。

  > **Note**
  >
  > 衰減的期間計算方式為：*shareDecaySeconds + OrderMinutes*其中 *OrderMinutes* 是以分鐘為單位的順序時間。

- 運算保留 – 防止單一共用識別符中的任務用完連接至任務佇列的所有資源。預留比率是 *(computeReservation/100)^ActiveFairShares*，其中 ActiveFairShares是作用中共享識別符的數量。

  > **Note**
  >
  > 如果共用識別符具有 SUBMITTED、PENDING、STARTING、 RUNNABLE或 RUNNING 狀態的任務，則視為作用中的共用識別符。在衰減的期間到期之後，共享識別符會被視為非作用中。

- 權重因數 – 共享識別符的權重因數。預設值為 1。較低的值可讓共用識別符執行的任務，或為共用識別符提供額外的執行時間。例如，使用權重因數為 0.125 (1/8) 的共享識別符的任務，會被指派為使用權重因數為 1 的共享識別符之任務的八倍運算資源。

> **Note**
>
> 只有在您需要更新預設權重因數為 1 時，才需要定義此屬性。

當任務佇列處於作用中狀態且正在處理任務時，您可以透過任務佇列快照檢閱前 100 個RUNNABLE任務的清單。如需詳細資訊，請參閱在 中檢視任務佇列 AWS Batch。

教學課程：建立排程政策

您必須建立排程政策，才能使用排程政策建立任務佇列。當您建立公平共用排程政策時，您可以將一或多個共用識別符或共用識別符字首與佇列的權重建立關聯，並選擇性地將衰減期間和運算保留指派給政策。

建立排程政策

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中選取要使用的「區域」。
3. 在導覽窗格中，選擇排程政策、建立。
4. 在名稱中，輸入排程政策的唯一名稱。可以包含最多可達 128 個字元 (大小寫)、數字、連字號和底線。
5. （選用）對於共用衰減秒，輸入公平共用排程政策的共用衰減時間的整數值。在排程任務時，較長的共用衰減時間會使用更多運算資源使用量。這可以允許使用共用識別符的任務暫時使用比該共用識別符的權重更多的運算資源，如果該共用識別符最近未使用運算資源，則允許使用。
6. （選用）針對運算保留，輸入公平共用排程政策運算保留的整數值。運算保留將保留一些 vCPUs 以用於目前非作用中的共用識別符。

   預留比率是 ActiveFairShares 是作用中共享識別符的數量(*computeReservation*/100)^*ActiveFairShares*。

   例如，computeReservation值 50 表示如果只有一個共用識別符，則 AWS Batch 應該保留最大可用 VCPU 的 50%，如果有兩個共用識別符，則保留 25%，如果有三個共用識別符，則保留 12.5%。如果只有一個共用識別符，computeReservation則值為 25 表示 AWS Batch 應該保留

最大可用 VCPU 的 25%，如果有兩個共用識別符，則保留 6.25%，如果有三個共用識別符，則保留 1.56%。

7. 在共用屬性區段中，您可以為要與公平共用排程政策建立關聯的每個共用識別符指定共用識別符和權重。

   a. 選擇新增共用識別符。

   b. 針對共用識別符，指定共用識別符。如果字串以「*」結尾，這將成為共享識別符字首，用於比對任務的共享識別符。排程政策中的所有共用識別符和共用識別符字首必須是唯一的且不能重疊。例如，您不能在相同的公平共用排程政策中具有共用識別符字首 'UserA*' 和共用識別符 'UserA1'。

   c. 針對權重因數，指定共用識別符的相對權重。預設值為 1.0。較低的值對於運算資源的優先順序較高。如果使用共用識別符字首，具有以字首開頭的共用識別符的任務將共用權重係數。這可有效地增加這些任務的權重因素，降低其個別優先順序，但維持共用識別符字首的相同權重因素。

8. （選用）在標籤區段中，您可以指定要與排程政策建立關聯的每個標籤的索引鍵和值。如需詳細資訊，請參閱標記您的 AWS Batch 資源。

9. 選擇提交以完成並建立排程政策。

參考：排程政策範本

空的排程政策範本如下所示。您可以使用此範本來建立排程政策，然後將其儲存至 檔案，並搭配 AWS CLI `--cli-input-json`選項使用。如需這些參數的詳細資訊，請參閱 AWS Batch API 參考中的 CreateSchedulingPolicy。

> **Note**
>
> 您可以使用下列 AWS CLI 命令產生任務佇列範本。
>
> ```
> $ aws batch create-scheduling-policy --generate-cli-skeleton
> ```

```
{
    "name": "",
    "fairsharePolicy": {
        "shareDecaySeconds": 0,
        "computeReservation": 0,
```

```
        "shareDistribution": [
            {
                "shareIdentifier": "",
                "weightFactor": 0.0
            }
        ]
    },
    "tags": {
        "KeyName": ""
    }
}
```

資源感知排程

AWS Batch 根據與任務佇列 (JQ) 相關聯的運算環境 (CE) 中的 vCPU、GPU 和記憶體可用性來排程任務。但有時候，只有這些 CE 資源的可用性並不保證任務會成功執行，因為它可能取決於其他必要的資源，因此這些任務會被取消或終止。這會導致運算資源的使用效率低下。為了解決此問題，資源感知排程可以在排程任務在 CE 上執行之前檢查相依、非 CE 資源的可用性。

AWS Batch 資源感知排程可讓您根據執行任務所需的消耗性資源來排程任務：第三方授權字符、資料庫存取頻寬、調節對第三方 API 的呼叫等。您可以指定執行任務所需的消耗性資源，而 Batch 在排程任務時會將這些資源相依性納入考量。您可以避免手動介入，以消除因消耗性資源不足而導致的任務失敗和長時間等待。您可以只配置具有所有必要資源的任務，以減少運算資源的使用不足。

資源感知排程適用於 FIFO 和公平共用排程政策，並可搭配 Batch 支援的所有運算平台使用，包括 EKS、ECS 和 Fargate。它可以與陣列任務、多節點平行 (MNP) 任務以及一般 Batch 任務搭配使用。

若要設定資源感知排程，請先指定執行任務所需的所有消耗性資源，以及每個資源的可用總數。然後，針對每個需要消耗性資源的任務，您可以指定每個所需資源的名稱和所需數量。批次會追蹤任務佇列中任務可用的消耗性資源數量，並確保只有在任務可成功執行所有必要的消耗性資源時，任務才會排程執行。

主題

- 建立消耗性資源
- 指定執行任務所需的資源
- 檢查有多少資源正在使用中且可用
- 更新任務正在使用的資源數量
- 尋找需要特定消耗性資源的任務

- 刪除消耗性資源

建立消耗性資源

您必須先建立消耗性資源，以代表任務執行時所使用的非 CE 資源，並且只能使用有限數量的資源。每個消耗品資源都有一個：

- 在帳戶層級必須是唯一的資源名稱 (consumableResourceName)。
- （選用）資源類型 (resourceType)，指出資源是否可在任務完成後重複使用。這可以是下列其中一項：
  - REPLENISHABLE (default)
  - NON_REPLENISHABLE
- 總數量 (totalQuantity)，指定可用消耗性資源的總數。

每個帳戶的消耗性資源數目上限為 50k。

主控台：

1. 在AWS Batch 主控台的左側導覽面板中，選擇消耗性資源。
2. 選擇建立消耗性資源。
3. 輸入唯一的資源名稱、總資源數量，然後選取資源類型是可取代還是不可取代。
4. 選擇建立消耗性資源。

API：

使用 CreateConsumableResource API 定義您想要的資源。

指定執行任務所需的資源

當您註冊任務時，您可以指定您建立的一或多個資源的名稱 (consumableResource)，以及任務每個執行個體所需的該資源數量 (quantity)。

批次會隨時追蹤每個資源的可用單位。對於任務佇列中的每個任務，批次排程器可確保您的任務僅在有指定的資源相依性可用時執行。

如果任務到達佇列前端時無法使用任務的消耗性資源，任務會等待RUNNABLE狀態，直到所有必要資源都可用或達到任務狀態時間限制（請參閱 在 中檢視任務佇列 AWS Batch)。一旦 Batch 驗證所

有資源都可用，任務就會轉換為 STARTING 狀態，然後轉換為 RUNNING。資源會在任務移至 時鎖定，STARTING然後在任務移至 SUCCEEDED或 時解除鎖定FAILED。

您也可以在提交任務時更新特定任務所需的資源數量。

主控台：

若要在定義任務時指定資源及其所需的數量：

1. 從AWS Batch 主控台使用任務定義精靈來定義任務 (任務定義 -> 建立)。
2. 在精靈的步驟 4：設定容器的取用資源下，從清單中選取所需資源的名稱。在請求的值欄位中，輸入此任務執行個體所需的此資源數量，然後選擇新增消耗性資源。
3. 對任務所需的所有消耗性資源重複上述步驟。您可以為您定義的每個任務指定最多 5 個資源。
4. 在完成任務定義精靈之後，但在選擇建立任務定義之前，您會看到已建立的消耗性資源清單。

若要在提交任務時更新所需的資源數量：

1. 在AWS Batch 主控台的左側導覽窗格中，選擇任務，然後選擇提交新任務。
2. 在精靈的步驟 2：設定覆寫中，在消耗性資源覆寫下，為要覆寫任務所需數量的任何消耗性資源輸入新的請求值。
3. 完成您要為此任務進行的所有覆寫後，請選擇下一步以繼續檢閱並提交。

API：

當您向 RegisterJobDefinition API 註冊任務時，請使用請求consumableResourceProperties部分consumableResourceList中的 來指定執行任務執行個體所需的消耗性資源，以及每個資源的數量。

當您使用 SubmitJob API 提交任務時，您可以使用請求的consumableResourcePropertiesOverride部分覆寫消耗性資源清單和每個資源的數量。請注意，這只會覆寫任務的每個執行個體所需的資源數量，而不是可用的總數量。

檢查有多少資源正在使用中且可用

Batch 可讓您查詢可用資源的數量 (availableQuantity)、使用中的資源數量 (inUseQuantity)，以及指定時間的總資源數量 (totalQuantity)。

一旦任務進入 STARTING 狀態，消耗的資源將從該資源的可用數量中減去。如果資源為 REPLENISHABLE，一旦任務移至 SUCCEEDED 或 FAILED 狀態，消耗的資源數量就會重新新增至可

用數量，而總數量將保持不變。如果資源為 NON_REPLENISHABLE，耗用的資源數量會從總和可用數量中減去，而且無論任務移至 SUCCEEDED或 FAILED 狀態，都不會新增回去。

> **Note**
>
> 此資訊可能會延遲最多 30 秒。

主控台：

1. 在AWS Batch 主控台的左側導覽面板中，選擇消耗性資源。
2. 選取可取代或不可取代索引標籤，以檢視您已建立的該類型的資源。
3. 對於每個可補充的資源，主控台會顯示名稱、資源的總數量、目前使用中的數量和剩餘數量，以及使用率的計算（使用中的資源數量除以該資源的總數量）。

   對於每個不可取代的資源，主控台會顯示名稱、目前使用中的數量，以及仍然可用的數量。

您也可以從 主控台的任務詳細資訊頁面檢視消耗性資源的目前資訊。

1. 在AWS Batch 主控台的左側導覽面板中，選擇任務，然後選取任務的名稱以開啟該任務的詳細資訊頁面。
2. 如果任務需要，則可檢視可取代資源和不可取代資源的相關資訊。對於這兩種類型，主控台會顯示資源的名稱、任務的請求數量、剩餘數量、目前使用中的數量、資源的總數量，以及目前使用率的計算（任務使用中的資源數量除以該資源的總數量）。

API：

使用傳回下列資訊的 DescribeConsumableResource API：

```
{
   "availableQuantity": number,
   "consumableResourceArn": "string",
   "consumableResourceName": "string",
   "createdAt": number,
   "inUseQuantity": number,
   "resourceType": "string",
   "tags": {
      "string" : "string"
```

```
    },
    "totalQuantity": number
}
```

ListConsumableResources API 也會報告使用中的資源數目 (inUseQuantity) 和目前可用的資源總數 (totalQuantity)，做為您在帳戶中建立的所有消耗性資源清單的一部分。此 API 也可讓您根據消耗性資源名稱篩選消耗性資源清單查詢。

更新任務正在使用的資源數量

您可以將資源的總數量重設為新值、新增至總數量或從中減去。

如果您指定的新總數量大於之前的數量，批次會相應地排程更多任務。如果新的總數量少於之前的數量，而且沒有使用此資源的單位，則 Batch 只會減少總 （或可用的） 數量。如果有使用中的單位，Batch 會立即減少可用數量，並在任務完成時，Batch 會減少總 （可用） 數量，使其最終到達新的數量。

主控台：

1. 在AWS Batch 主控台的左側導覽面板中，選擇消耗性資源。
2. 選取可取代或不可取代索引標籤，以檢視您已建立的該類型的資源。
3. 對於可取代的資源：

    1. 選擇您要更新的資源，然後選擇動作，然後選擇設定資源、新增資源或移除資源。
    2. 快顯視窗隨即出現，您可以在其中設定總值、新增資源或移除資源，具體取決於您在上一個步驟中選擇的動作。輸入要設定為新總值的數量、要新增至總數量，或要從總數量中減去的數量，然後選取確定。

    對於不可取代的資源：

    1. 選擇您要更新的資源，然後選擇動作，然後選擇設定資源、新增資源或移除資源。
    2. 快顯視窗隨即出現，您可以在其中設定可用值、新增資源或移除資源，具體取決於您在上一個步驟中選擇的動作。輸入您要設定為新可用值的數量、您要新增至可用數量，或您要從可用數量中減去的數量，然後選取確定。

API：

使用 UpdateConsumableResource API 為資源設定新的總數量，或增加或減少總數量。

尋找需要特定消耗性資源的任務

Batch 可讓您擷取需要特定消耗性資源的任務清單。

主控台：

1. 在AWS Batch 主控台的左側導覽面板中，選擇消耗性資源。
2. 在清單中，選取消耗性資源的名稱。該資源的詳細資訊頁面隨即開啟。
3. 在搜尋任務下，輸入您要套用至任務清單的任何篩選條件。您可以依任務名稱 ('equals'， 'starts with')、日期範圍 （建立任務時） 和其他條件 ('job queue'， 'job definition'， 'shared job identifier') 進行篩選。對於您要套用的每種篩選條件類型，請從下拉式清單中的可用選項中選取，然後輸入請求的任何其他資訊。

   選擇 Search (搜尋)。
4. 隨即顯示需要消耗性資源的 （篩選） 任務清單，包括任務的名稱、狀態、消耗性資源的請求單位數、其他所需的消耗性資源等。使用此清單，您可以選取要取消或終止的一或多個任務。您也可以選取任務的名稱，以開啟該任務的詳細資訊頁面。
5. 在搜尋任務下，您現在可以重新整理結果或清除搜尋並重新開始。

API：

您可以取得搭配 `ListJobsByConsumableResource` API 使用特定消耗性資源的任務清單。此 API 也可讓您使用任務狀態或任務名稱來篩選任務清單查詢。

刪除消耗性資源

您可以隨時刪除消耗性資源，即使需要資源的任務仍在執行中。刪除消耗性資源後，在收到刪除命令和任務排程器遵守刪除之間可能會有間隙，因此可能會在刪除呼叫後立即排程使用資源的任務。如果已刪除的消耗性資源具有資源類型 (`resourceType`) `REPLENISHABLE`，則會在任務完成時忽略此項目。如果您刪除消耗性資源並以相同名稱重新建立該資源，則會將其視為相同資源，並且可供`RUNNABLE`任務使用。

主控台：

1. 在AWS Batch 主控台的左側導覽面板中，選擇消耗性資源。
2. 選取可取代或不可取代索引標籤，以檢視您已建立的該類型的資源。
3. 選取您要刪除的每個資源，然後選擇刪除。隨即出現快顯視窗 刪除消耗性資源。如要確認刪除，請選擇 Delete (刪除)。

您也可以從 主控台的詳細資訊頁面刪除消耗性資源。

1. 在AWS Batch 主控台的左側導覽面板中，選擇消耗性資源。
2. 選取可取代或不可取代索引標籤，以檢視您已建立的該類型的資源。
3. 選擇您要刪除的資源名稱。消耗性資源的詳細資訊頁面隨即出現。選擇 刪除。隨即出現快顯視窗刪除消耗性資源。如要確認刪除，請選擇 Delete (刪除)。

API：

使用 `DeleteConsumableResource` API 刪除消耗性資源。

任務定義

AWS Batch 任務定義會指定任務的執行方式。雖然每個任務都必須參考任務定義，但任務定義中指定的許多參數都可以在執行時間覆寫。

任務定義中指定的部分屬性包括：

- 要與任務中的容器搭配使用的 Docker 映像。
- 要與容器搭配使用vCPUs 數量和記憶體數量。
- 容器啟動時應執行的命令。
- 啟動時，應將哪些（如果有）環境變數傳遞至容器。
- 應與容器搭配使用的任何資料磁碟區。
- 您的任務應使用哪些（如果有）IAM 角色來取得 AWS 許可。

目錄

- 建立單一節點任務定義
- 建立多節點平行任務定義
- 使用 ContainerProperties 的任務定義範本
- 使用 EcsProperties 建立任務定義
- 使用 awslogs 日誌驅動程式
- 指定敏感資料
- 任務的私有登錄檔身分驗證
- Amazon EFS 磁碟區
- 任務定義範例

建立單一節點任務定義

您必須先建立任務定義 AWS Batch，才能在 中執行任務。此程序在單節點和多節點平行任務之間略有不同。本主題特別介紹如何為 AWS Batch 非多節點平行任務（也稱為 Gang 排程) 的任務建立任務定義。

您可以在 Amazon Elastic Container Service 資源上建立多節點平行任務定義。如需詳細資訊，請參閱the section called “建立多節點平行任務定義”。

主題

- 在 Amazon EC2 資源上建立單一節點任務定義
- 在 Fargate 資源上建立單一節點任務定義
- 在 Amazon EKS 資源上建立單一節點任務定義
- 在 Amazon EC2 資源上建立具有多個容器的單一節點任務定義

在 Amazon EC2 資源上建立單一節點任務定義

請完成下列步驟，以在 Amazon Elastic Compute Cloud (Amazon EC2) 資源上建立單一節點任務定義。

若要在 Amazon EC2 資源上建立新的任務定義：

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選擇要 AWS 區域 使用的 。
3. 在左側導覽窗格中，選擇任務定義。
4. 選擇建立。
5. 針對協調類型，選擇 Amazon Elastic Compute Cloud (Amazon EC2)。
6. 對於 EC2 平台組態，請關閉啟用多節點平行處理。
7. 在名稱中，輸入任務定義的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
8. （選用）針對執行逾時，輸入逾時值（以秒為單位）。執行逾時是未完成任務終止之前的時間長度。如果嘗試超過逾時持續時間，則會停止嘗試並移至 FAILED 狀態。如需詳細資訊，請參閱任務逾時。最小值為 60 秒。
9. （選用）開啟排程優先順序。輸入介於 0 到 100 之間的排程優先順序值。較高值的優先順序較高。
10. （選用）對於任務嘗試，輸入嘗試將任務移至RUNNABLE狀態的 AWS Batch 次數。輸入介於 1 到 10 之間的數字。
11. （選用）針對重試策略條件，選擇在結束時新增評估。輸入至少一個參數值，然後選擇動作。對於每組條件，動作必須設定為重試或結束。這些動作表示下列項目：
    - 重試 – AWS Batch 重試，直到達到您指定的任務嘗試次數為止。
    - 結束 – AWS Batch 停止重試任務。

> **Important**
>
> 如果您選擇在結束時新增評估，則必須至少設定一個參數，然後選擇動作或選擇在結束時移除評估。

12. （選用）展開標籤，然後選擇新增標籤以將標籤新增至資源。輸入金鑰和選用值，然後選擇 Add tag (新增標籤)。
13. （選用）開啟傳播標籤，將標籤從任務和任務定義傳播到 Amazon ECS 任務。
14. 選擇下一頁。
15. 在容器組態區段中：

    a. 針對映像，選擇要用於任務的Docker映像。根據預設，Docker Hub 登錄檔中的映像為可用。您也可以用 *repository-url*/*image*:*tag* 指定其他儲存庫。名稱的長度上限為 225 個字元。可以包含大小寫字母、數字、連字號 (-)、底線 (_)、冒號 (:)、正斜線 (/) 和井號 (#)。此參數會映射至 Docker Remote API 的建立容器區段中的 `Image` 以及 docker run 的 `IMAGE` 參數。

    > **Note**
    >
    > Docker 映像架構必須符合執行個體排程所在之運算資源的處理器架構。例如，ARM 型 Docker 映像只能在 ARM 型運算資源上執行。

    - Amazon ECR Public 儲存庫中的映像會使用完整`registry/repository[:tag]`或命名慣例 `registry/repository[@digest]`（例如 `public.ecr.aws/`*`registry_alias`*`/`*`my-web-app`*`:`*`latest`*）。
    - Amazon ECR 儲存庫中的映像會使用完整的命名慣例 `registry/repository[:tag]`（例如 *`aws_account_id`*`.dkr.ecr.`*`region`*`.amazonaws.com/`*`my-web-app`*`:`*`latest`*）。
    - Docker Hub 上官方儲存庫中的映像，使用的是單一名稱 (例如，`ubuntu` 或 `mongo`)。
    - Docker Hub 上的其他儲存庫中的映像要求使用組織名稱 (例如，`amazon/amazon-ecs-agent`)。
    - 其他線上存放庫中的映像更進一步要求使用網域名稱 (例如，`quay.io/assemblyline/ubuntu`)。

b. 針對命令，在欄位中輸入命令，作為其 JSON 字串陣列對等項。

此參數會映射至 Docker Remote API 的建立容器區段中的 Cmd 以及 docker run 的 COMMAND 參數。如需 Docker CMD 參數的詳細資訊，請參閱 https://docs.docker.com/engine/reference/builder/#cmd。

> **Note**
>
> 您可以在命令中使用預設值來取代參數和預留位置。如需詳細資訊，請參閱參數。

c. （選用）對於執行角色，指定 IAM 角色，授予 Amazon ECS 容器代理程式代表您進行 AWS API 呼叫的許可。此功能針對任務使用 Amazon ECS IAM 角色。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 任務執行 IAM 角色。

d. 針對任務角色組態，選擇具有 AWS APIs 許可的 IAM 角色。此功能針對任務使用 Amazon ECS IAM 角色。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的任務 IAM 角色。

> **Note**
>
> 這裡只會顯示具有 Amazon Elastic Container Service 任務角色信任關係的角色。如需為您的 AWS Batch 任務建立 IAM 角色的詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的為您的任務建立 IAM 角色和政策。

16. 針對參數，選擇新增參數，將參數替換預留位置新增為鍵對和選用值對。

17. 在環境組態區段中：

a. 針對 vCPUs，輸入要保留給容器的 vCPUs 數量。此參數會映射到 Docker Remote API 的建立容器區段中的 CpuShares 以及 docker run 的 `--cpu-shares` 選項。每個 vCPU 相當於 1,024 個 CPU 共用。您必須指定至少 1 個 vCPU。

b. 針對記憶體，輸入容器可用的記憶體限制。如果您的容器嘗試超過您在此處指定的記憶體量，則容器會停止。此參數會映射到 Docker Remote API 的建立容器區段中的 Memory 以及 docker run 的 `--memory` 選項。您必須為單一工作指定至少 4 MiB 的記憶體。

> **Note**
>
> 若要最大化資源使用率，請為特定執行個體類型的任務排定記憶體的優先順序。如需詳細資訊，請參閱運算資源記憶體管理。

c. 針對 GPUs 數量，選擇要保留給容器的 GPUs 數量。

d. （選用）對於環境變數，選擇新增環境變數，將環境變數新增為名稱/值對。這些變數會傳遞至容器。

e. （選用）對於秘密，選擇新增秘密，將秘密新增為名稱/值對。這些秘密會在容器中公開。如需詳細資訊，請參閱 LogConfiguration：secretOptions。

18. 選擇下一頁。

19. 在 Linux 組態區段中：

a. 在 User (使用者) 中，輸入要在容器內使用的使用者名稱。此參數會映射到 Docker Remote API 的建立容器區段中的 `User` 以及 docker run 的 `--user` 選項。

b. （選用）若要在主機執行個體上提升任務容器的許可（類似於`root`使用者），請將特權滑桿拖曳至右側。此參數會映射到 Docker Remote API 的建立容器區段中的 `Privileged` 以及 docker run 的 `--privileged` 選項。

c. （選用）開啟啟用初始化程序以在容器內執行`init`程序。此程序會轉送訊號並接收程序。

20. （選用）在檔案系統組態區段中：

a. 開啟 Enable read only filesystem (啟用唯讀檔案系統) 以移除對磁碟區的寫入權限。

b. 針對共用記憶體大小，輸入`/dev/shm`磁碟區的大小（以 MiB 為單位）。

c. 針對最大交換大小，輸入容器可以使用的交換記憶體總量（以 MiB 為單位）。

d. 針對交換輸入介於 0 到 100 之間的值，表示容器的交換行為。如果您未指定值並啟用交換，則值預設為 60。如需詳細資訊，請參閱 LinuxParameters：swappiness。

e. （選用）展開其他組態。

f. （選用）針對 Tmpfs，選擇新增 tmpfs 以新增`tmpfs`掛載。

g. （選用）針對裝置，選擇新增裝置以新增裝置：

i. 針對 Container path (容器路徑)，指定容器執行個體中的路徑，以公開對應到主機執行個體的裝置。如果您將此保留空白，則會在容器中使用主機路徑。

ii. 針對 Host path (主機路徑)，指定主機執行個體中的裝置的路徑。

iii. 針對許可，選擇要套用至裝置的一或多個許可。可用的許可為讀取、寫入和 MKNOD。

h. （選用）針對磁碟區組態，選擇新增磁碟區以建立要傳遞至容器的磁碟區清單。輸入磁碟區的名稱和來源路徑，然後選擇新增磁碟區。您也可以選擇開啟啟用 EFS。

i. （選用）對於掛載點，選擇新增掛載點組態以新增資料磁碟區的掛載點。您必須指定來源磁碟區和容器路徑。這些掛載點會傳遞至容器執行個體Docker daemon上的 。您也可以選擇將磁碟區設為唯讀。

j. （選用）對於 Ulimits 組態，選擇新增 ulimit 以新增容器`ulimits`的值。輸入名稱、軟性限制和硬性限制值，然後選擇新增 ulimit。

21. 在任務屬性區段中：

a. 針對執行角色 - 條件式，選擇角色以允許 Amazon ECS 代理程式代表您進行 AWS API 呼叫。如需建立執行角色的詳細資訊，請參閱 教學課程：建立 IAM 執行角色。

b. 選擇啟用 ECS 執行命令，以啟用直接存取 Amazon ECS 容器殼層，並略過主機作業系統。您必須選擇任務角色。

> **Important**
>
> ECS 執行命令需要可寫入的檔案系統。

c. 針對任務角色，選擇 Amazon ECS Identity and Access Management (IAM) 角色，以允許容器代表您進行 AWS API 呼叫。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 任務 IAM 角色。

22. （選用）在記錄組態區段中：

a. 對於日誌驅動程式，選擇要使用的日誌驅動程式。如需可用日誌驅動程式的詳細資訊，請參閱 LogConfiguration：logDriver。

> **Note**
>
> 根據預設，會使用 `awslogs` 日誌驅動程式。

b. 針對選項，選擇新增選項以新增選項。輸入名稱/值對，然後選擇新增選項。

c. 針對秘密，選擇新增秘密。輸入名稱/值對，然後選擇新增秘密以新增秘密。

> **Tip**
>
> 如需詳細資訊，請參閱 LogConfiguration：secretOptions。

23. 選擇下一頁。
24. 針對任務定義檢閱，檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立任務定義。

在 Fargate 資源上建立單一節點任務定義

完成下列步驟，以在 AWS Fargate 資源上建立單一節點任務定義。

若要在 Fargate 資源上建立新的任務定義：

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從頂端導覽列中，選擇要 AWS 區域 使用的 。
3. 在左側導覽窗格中，選擇任務定義。
4. 選擇建立。
5. 對於協同運作類型，選擇 Fargate。如需詳細資訊，請參閱Fargate 運算環境。
6. 在名稱中，輸入任務定義的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
7. （選用）針對執行逾時，輸入逾時值（以秒為單位）。執行逾時是未完成任務終止之前的時間長度。如果嘗試超過逾時持續時間，則會停止嘗試並移至 FAILED 狀態。如需詳細資訊，請參閱任務逾時。最小值為 60 秒。
8. （選用）開啟排程優先順序。輸入介於 0 到 100 之間的排程優先順序值。較高值的優先順序高於較低值。
9. （選用）展開標籤，然後選擇新增標籤以將標籤新增至資源。開啟傳播標籤，從任務和任務定義傳播標籤。
10. 在 Fargate 平台組態區段中：
    a. 針對執行期平台，選擇運算環境架構。
    b. 針對作業系統系列，選擇運算環境的作業系統。
    c. 針對 CPU 架構，選擇 vCPU 架構。
    d. 對於 Fargate 平台版本，請輸入 LATEST或特定執行時間環境版本。
    e. （選用）開啟指派公有 IP，將公有 IP 地址指派給 Fargate 任務網路介面。對於在私有子網路中執行以將傳出流量傳送至網際網路的任務，私有子網路需要連接 NAT 閘道，才能將請求路由至網際網路。您可能想要這麼做，以便提取容器映像。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 任務聯網。

f. （選用）對於暫時性儲存，輸入要配置給任務的暫時性儲存量。暫時性儲存量必須介於 21 GiB 和 200 GiB 之間。根據預設，如果您未輸入值，則會配置 20 GiB 的暫時性儲存。

> **Note**
>
> 暫時性儲存需要 Fargate 平台 1.4 版或更新版本。

g. 針對執行角色，指定 IAM 角色，授予 Amazon ECS 容器和 Fargate 代理程式代表您進行 AWS API 呼叫的許可。此功能使用 Amazon ECS IAM 角色執行任務功能。如需包含組態先決條件的詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 任務執行 IAM 角色。

h. 針對任務嘗試，輸入嘗試將任務移至 RUNNABLE 狀態的 AWS Batch 次數。輸入介於 1 到 10 之間的數字。

i. 選用）針對重試策略條件，選擇在結束時新增評估。輸入至少一個參數值，然後選擇動作。對於每組條件，動作必須設定為重試或結束。這些動作表示下列項目：

- 重試 – AWS Batch 重試，直到達到您指定的任務嘗試次數為止。
- 結束 – AWS Batch 停止重試任務。

> **Important**
>
> 如果您選擇在結束時新增評估，則必須至少設定一個參數，然後選擇動作，或選擇在結束時移除評估。

11. 選擇下一頁。
12. 在容器組態區段中：

a. 在 Image (映像) 中，請選擇用於任務的 Docker 映像檔。根據預設，Docker Hub 登錄檔中的映像為可用。您也可以用 *repository-url*/*image*:*tag* 指定其他儲存庫。名稱的長度上限為 225 個字元。可包含大寫及小寫字母、數字、連字號 (-)、底線 (_)、冒號 (:)、句點 (.)、斜線 (/) 和數字符號 (#)。此參數會映射至 Docker Remote API 的建立容器區段中的 Image 以及 docker run 的 IMAGE 參數。

> **Note**
>
> Docker 映像架構必須符合執行個體排程所在之運算資源的處理器架構。例如，ARM 型 Docker 映像只能在 ARM 型運算資源上執行。

- Amazon ECR Public 儲存庫中的映像會使用完整`registry/repository[:tag]`或命名慣例 `registry/repository[@digest]` （例如 `public.ecr.aws/`*`registry_alias`*`/`*`my-web-app`*`:`*`latest`*)。
- Amazon ECR 儲存庫中的映像會使用完整的命名慣例 `registry/repository[:tag]` （例如 *`aws_account_id`*`.dkr.ecr.`*`region`*`.amazonaws.com/`*`my-web-app`*`:`*`latest`*)。
- Docker Hub 上官方儲存庫中的映像，使用的是單一名稱 (例如，`ubuntu` 或 `mongo`)。
- Docker Hub 上的其他儲存庫中的映像要求使用組織名稱 (例如，`amazon/amazon-ecs-agent`)。
- 其他線上存放庫中的映像更進一步要求使用網域名稱 (例如，`quay.io/assemblyline/ubuntu`)。

b. 在命令中，將命令輸入 欄位做為其 JSON 字串陣列對等項。

此參數會映射至 Docker Remote API 的建立容器區段中的 `Cmd` 以及 docker run 的 `COMMAND` 參數。如需 Docker `CMD` 參數的詳細資訊，請參閱 https://docs.docker.com/engine/reference/builder/#cmd。

> **Note**
>
> 您可以在命令中使用預設值來取代參數和預留位置。如需詳細資訊，請參閱參數。

c. （選用）將參數新增至任務定義做為名稱/值映射，以覆寫任務定義預設值。若要新增參數：

- 針對參數，選擇新增參數，輸入名稱/值對，然後選擇新增參數。

> **Important**
>
> 如果您選擇新增參數，您必須設定至少一個參數，或選擇移除參數

d. 在環境組態區段中：

i. 針對任務角色組態，選擇具有 AWS APIs 許可的 IAM 角色。此功能使用 Amazon ECS IAM 角色執行任務功能。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的任務 IAM 角色。

> **Note**
>
> 這裡只會顯示具有 Amazon Elastic Container Service 任務角色信任關係的角色。如需如何為您的 AWS Batch 任務建立 IAM 角色的詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的為您的任務建立 IAM 角色和政策。

ii. 針對 vCPUs，輸入要保留給容器的 vCPUs 數量。此參數會映射到 Docker Remote API 的建立容器區段中的 `CpuShares` 以及 docker run 的 `--cpu-shares` 選項。每個 vCPU 相當於 1,024 個 CPU 共用。您必須指定至少 1 個 vCPU。

iii. 針對記憶體，輸入容器可用的記憶體限制。如果您的容器嘗試超過此處指定的記憶體，則容器會停止。此參數會映射到 Docker Remote API 的建立容器區段中的 `Memory` 以及 docker run 的 `--memory` 選項。您必須為單一工作指定至少 4 MiB 的記憶體。

如果您使用 GuardDuty 執行期監控，GuardDuty 安全代理程式會有些微的記憶體負荷。因此，記憶體限制必須包含 GuardDuty 安全代理程式的大小。如需有關 GuardDuty 安全代理程式記憶體限制的資訊，請參閱《GuardDuty 使用者指南》中的 CPU 和記憶體限制。如需最佳實務的相關資訊，請參閱《Amazon ECS 開發人員指南》中的如何在啟用執行期監控之後修復 Fargate 任務上的記憶體不足錯誤。

> **Note**
>
> 若要最大化資源使用率，請為特定執行個體類型的任務排定記憶體的優先順序。如需詳細資訊，請參閱運算資源記憶體管理。

e. （選用）對於環境變數，選擇新增環境變數，將環境變數新增為名稱/值對。這些變數會傳遞至容器。

f. （選用）針對秘密，選擇新增秘密，將秘密新增為名稱/值對。這些秘密會在容器中公開。如需詳細資訊，請參閱 LogConfiguration：secretOptions。

g. 選擇下一頁。

13. （選用）在 Linux 組態區段中：

a. 針對使用者，輸入要在容器內使用的使用者名稱。

b. 開啟 Enable init process (啟用初始化處理) 以在容器內執行初始化程序。此程序會轉送訊號並接收程序。

c. 開啟 Enable read only filesystem (啟用唯讀檔案系統) 以移除對磁碟區的寫入權限。

d. （選用）展開其他組態。

e. 針對掛載點組態，選擇新增掛載點組態以新增資料磁碟區的掛載點。您必須指定來源磁碟區和容器路徑。這些掛載點會傳遞至容器執行個體Docker daemon上的 。

f. 針對磁碟區組態，選擇新增磁碟區以建立要傳遞至容器的磁碟區清單。輸入磁碟區的名稱和來源路徑，然後選擇新增磁碟區。

g. 在任務屬性區段中：

   i. 針對執行角色 - 條件式，選擇角色以允許 Amazon ECS 代理程式代表您進行 AWS API 呼叫。如需建立執行角色的詳細資訊，請參閱 教學課程：建立 IAM 執行角色。

   ii. 選擇啟用 ECS 執行命令，以啟用直接存取 Amazon ECS 容器殼層，並略過主機作業系統。您必須選擇任務角色。

   > **Important**
   >
   > ECS 執行命令需要可寫入的檔案系統。

   iii. 針對任務角色，選擇 Amazon ECS Identity and Access Management (IAM) 角色，以允許容器代表您進行 AWS API 呼叫。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 任務 IAM 角色。

h. 在記錄組態區段中：

   i. （選用）對於日誌驅動程式，選擇要使用的日誌驅動程式。如需可用日誌驅動程式的詳細資訊，請參閱 LogConfiguration：logDriver。

   > **Note**
   >
   > 根據預設，會使用 `awslogs` 日誌驅動程式。

   ii. （選用）對於選項，選擇新增選項以新增選項。輸入名稱/值對，然後選擇新增選項。

   iii. （選用）針對秘密，選擇新增秘密以新增秘密。然後，輸入名稱值對，然後選擇新增秘密。

> **Tip**
>
> 如需詳細資訊，請參閱 LogConfiguration：secretOptions。

14. 選擇下一頁。
15. 針對任務定義檢閱，檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立任務定義。

在 Amazon EKS 資源上建立單一節點任務定義

請完成下列步驟，以在 Amazon Elastic Kubernetes Service (Amazon EKS) 上建立單一節點任務定義。

若要在 Amazon EKS 資源上建立新的任務定義：

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從頂端導覽列中，選擇要 AWS 區域 使用的 。
3. 在左側導覽窗格中，選擇任務定義。
4. 選擇建立。
5. 針對協調類型，選擇 Elastic Kubernetes Service (EKS)。
6. 在名稱中，輸入任務定義的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
7. （選用）針對執行逾時，輸入逾時值（以秒為單位）。執行逾時是未完成任務終止之前的時間長度。如果嘗試超過逾時持續時間，則會停止嘗試並移至 FAILED 狀態。如需詳細資訊，請參閱任務逾時。最小值為 60 秒。
8. （選用）開啟排程優先順序。輸入介於 0 到 100 之間的排程優先順序值。較高值的優先順序高於較低值。
9. （選用）展開標籤，然後選擇新增標籤以將標籤新增至資源。
10. 選擇下一頁。
11. 在 EKS pod 屬性區段中：
    a. 針對服務帳戶名稱，輸入為在 中執行的程序提供身分的帳戶pod。
    b. 開啟主機網路以使用Kubernetespod網路模型，並為傳入連線開啟接聽連接埠。僅針對傳出通訊關閉此設定。

c. 針對 DNS 政策，選擇下列其中一項：

- 無值 (null) – pod忽略Kubernetes環境中的 DNS 設定。
- Default (預設) – pod 會繼承其執行所在節點的名稱解析度組態。

> **Note**
>
> 如果未指定 DNS 政策，Default (預設值) 不會是預設 DNS 政策。而是使用 ClusterFirst (ClusterFirst)。

- ClusterFirst – 任何與設定之叢集網域尾碼不相符的 DNS 查詢都會轉寄至繼承自節點的上游名稱伺服器。
- ClusterFirstWithHostNet – 如果 Host network (主機網路) 開啟，則使用此選項。

d. （選用）針對磁碟區，選取新增磁碟區，然後：

i. 為您的磁碟區新增名稱。

ii. （選用）新增主機上目錄的主機路徑。

iii. （選用）新增中型和大小限制以設定 Kubernetes emptyDir。

iv. （選用）提供 Pod 的秘密名稱，以及秘密是否為選用。

v. （選用）定義宣告名稱，將 Kubernetes 持久性磁碟區宣告附加至 Pod，以及是否為唯讀。

e. （選用）對於 Pod 標籤，選擇新增 Pod 標籤，然後輸入名稱/值對。

> **Important**
>
> Pod 標籤的字首不能包含 `kubernetes.io/`、`k8s.io/`或 `batch.amazonaws.com/`。

f. （選用）對於 Pod 註釋，選擇新增註釋，然後輸入名稱/值對。

> **Important**
>
> Pod 註釋的字首不能包含 `kubernetes.io/`、`k8s.io/`或 `batch.amazonaws.com/`。

g. 選擇下一頁。

h. 在容器組態區段中：

i. 在名稱中，輸入容器的唯一名稱。名稱必須以字母或數字開頭，長度最多可達 63 個字元。它可以包含大小寫字母、數字和連字號 (-)。

ii. 針對映像，選擇要用於任務的Docker映像。根據預設，Docker Hub 登錄檔中的映像為可用。您也可以用 *repository-url*/*image*:*tag* 指定其他儲存庫。名稱長度上限為 255 個字元。可包含大寫及小寫字母、數字、連字號 (-)、底線 (_)、冒號 (:)、句點 (.)、斜線 (/) 和數字符號 (#)。此參數會映射至 Docker Remote API 的建立容器區段Image中的 ，以及 的 IMAGE 參數 docker run

> **Note**
>
> Docker 映像架構必須符合執行個體排程所在之運算資源的處理器架構。例如，ARM 型 Docker 映像只能在 ARM 型運算資源上執行。

- Amazon ECR Public 儲存庫中的映像會使用完整registry/repository[:tag]或命名慣例 registry/repository[@digest] （ 例如 public.ecr.aws/*registry_alias*/*my-web-app*:*latest*)。
- Amazon ECR 儲存庫中的映像會使用完整的命名慣例 registry/repository[:tag] （ 例如 *aws_account_id*.dkr.ecr.*region*.amazonaws.com/*my-web-app*:*latest*)。
- Docker Hub 上官方儲存庫中的映像，使用的是單一名稱 (例如，ubuntu 或 mongo)。
- Docker Hub 上的其他儲存庫中的映像要求使用組織名稱 (例如，amazon/amazon-ecs-agent)。
- 其他線上存放庫中的映像更進一步要求使用網域名稱 (例如，quay.io/assemblyline/ubuntu)。

iii. （選用）對於映像提取政策，選擇提取映像的時間。

iv. （選用）針對命令，輸入要傳遞至容器的JSON命令。

v. （選用）對於引數，輸入要傳遞至容器的引數。如果未提供 引數，則會使用容器映像命令。

i. （選用）您可以將參數新增至任務定義做為名稱值映射，以覆寫任務定義預設值。若要新增參數：

- 針對參數，輸入名稱值對，然後選擇新增參數。

> **Important**
>
> 如果您選擇新增參數，則必須至少設定一個參數，或選擇移除參數

j. 在環境組態區段中：

i. 針對 vCPUs，輸入要保留給容器的 vCPUs 數量。此參數會映射到 Docker Remote API 的建立容器區段中的 `CpuShares` 以及 docker run 的 `--cpu-shares` 選項。每個 vCPU 相當於 1,024 個 CPU 共用。您必須指定至少 1 個 vCPU。

ii. 針對記憶體，輸入容器可用的記憶體限制。如果您的容器嘗試超過此處指定的記憶體，則容器會停止。此參數會映射到 Docker Remote API 的建立容器區段中的 `Memory` 以及 docker run 的 `--memory` 選項。您必須為單一工作指定至少 4 MiB 的記憶體。

> **Note**
>
> 若要最大化資源使用率，請為特定執行個體類型的任務排定記憶體的優先順序。如需詳細資訊，請參閱運算資源記憶體管理。

k. （選用）對於環境變數，選擇新增環境變數，將環境變數新增為名稱/值對。這些變數會傳遞至容器。

l. （選用）對於磁碟區掛載：

i. 選擇新增磁碟區掛載。

ii. 輸入名稱，然後在掛載磁碟區的容器中輸入掛載路徑。輸入 SubPath 以指定參考磁碟區內的子路徑，而不是其根路徑。

iii. 選擇唯讀以移除磁碟區的寫入許可。

iv. 選擇新增磁碟區掛載。

m. （選用）對於以使用者身分執行，輸入使用者 ID 以執行容器程序。

> **Note**
>
> 使用者 ID 必須存在於映像中，容器才能執行。

n. （選用）對於以群組身分執行，輸入群組 ID 以執行容器程序執行時間。

> **Note**
>
> 群組 ID 必須存在於映像中，容器才能執行。

o. （選用）若要在主機執行個體上提升您任務的容器許可（類似於`root`使用者），請將特權滑桿拖曳至右側。此參數會映射到 Docker Remote API 的建立容器區段中的 `Privileged` 以及 docker run 的 `--privileged` 選項。

p. （選用）開啟唯讀根檔案系統，以移除根檔案系統的寫入存取權。

q. （選用）開啟以非根使用者的身分執行，以非根使用者的pod身分執行 中的容器。

> **Note**
>
> 如果開啟以非根身分執行， 會在執行時間kubelet驗證映像，以確認映像未以 UID 0 執行。

r. 選擇下一頁。

12. 針對任務定義檢閱，檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立任務定義。

在 Amazon EC2 資源上建立具有多個容器的單一節點任務定義

請完成下列步驟，在 Amazon Elastic Compute Cloud (Amazon EC2) 資源上建立具有多個容器的單一節點任務定義。

若要在 Amazon EC2 資源上建立新的任務定義：

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選擇要 AWS 區域 使用的 。
3. 在左側導覽窗格中，選擇任務定義。
4. 選擇建立。
5. 針對協調類型，選擇 Amazon Elastic Compute Cloud (Amazon EC2)。
6. 對於任務定義結構，請關閉使用舊版 containerProperties 結構處理。
7. 對於 EC2 平台組態，請關閉啟用多節點平行處理。
8. 選擇下一步。
9. 在一般組態區段中，輸入下列內容：

a. 在名稱中，輸入任務定義的唯一名稱。名稱長度上限為 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。

b. 針對執行逾時 - 選用，輸入逾時值（以秒為單位）。執行逾時是未完成任務終止之前的時間長度。如果嘗試超過逾時持續時間，則會停止嘗試並移至 FAILED 狀態。如需詳細資訊，請參閱任務逾時。最小值為 60 秒。

c. 開啟排程優先順序 - 選用。輸入介於 0 到 100 之間的排程優先順序值。較高值的優先順序較高。

d. 展開標籤 - 選用，然後選擇新增標籤以將標籤新增至資源。輸入金鑰和選用值，然後選擇 Add tag (新增標籤)。

e. 開啟傳播標籤，將標籤從任務和任務定義傳播到 Amazon ECS 任務。

10. 在重試策略 - 選用區段中，輸入下列內容：

a. 針對任務嘗試，輸入嘗試將任務移至RUNNABLE狀態的 AWS Batch 次數。輸入介於 1 到 10 之間的數字。

b. 針對重試策略條件，選擇結束時新增評估。輸入至少一個參數值，然後選擇動作。對於每組條件，動作必須設定為重試或結束。這些動作代表下列項目：

- 重試 – AWS Batch 重試，直到達到您指定的任務嘗試次數為止。
- 結束 – AWS Batch 停止重試任務。

> **Important**
>
> 如果您選擇在結束時新增評估，則必須至少設定一個參數，然後選擇動作或選擇在結束時移除評估。

11. 在任務屬性區段中，輸入下列內容：

a. 針對執行角色 - 條件式，選擇角色以允許 Amazon ECS 代理程式代表您進行 AWS API 呼叫。如需建立執行角色的詳細資訊，請參閱 教學課程：建立 IAM 執行角色。

b. 選擇啟用 ECS 執行命令，以啟用直接存取 Amazon ECS 容器殼層，並略過主機作業系統。您必須選擇任務角色。

> **Important**
>
> ECS 執行命令需要可寫入的檔案系統。

c. 針對任務角色，選擇 Amazon ECS Identity and Access Management (IAM) 角色，以允許容器代表您進行 AWS API 呼叫。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 任務 IAM 角色。

d. 針對 IPC 模式task，選擇 host、 或 none。如果指定 host ，則在同一容器執行個體上指定主機 IPC 模式的任務中的所有容器都會與主機 Amazon EC2 執行個體共用相同的 IPC 資源。如果指定任務，則指定任務內的所有容器都會共用相同的 IPC 資源。如果未指定，則任務容器中的 IPC 資源為私有，不會與任務中或容器執行個體上的其他容器共用。如果沒有指定值，則 IPC 資源命名空間共用取決於容器執行個體上 Docker 常駐程式的設定。

e. 針對 PID 模式，選擇 host或 task。例如，監控附屬可能需要 pidMode 存取相同任務中執行之其他容器的相關資訊。如果指定 host ，則在同一容器執行個體上指定主機 PID 模式的任務中的所有容器都會與主機 Amazon EC2 執行個體共用相同的程序命名空間。如果已指定 task，則指定任務內的所有容器會共用相同的程序命名空間。如果未指定任何值，每個容器的預設值會是私有命名空間。

12. 在消耗性資源區段中，輸入下列內容：

    a. 輸入唯一的名稱和請求的值。

    b. 您可以選擇新增消耗性資源來新增更多消耗性資源。

13. 在儲存區段中，輸入下列內容：

    a. 輸入磁碟區的名稱和來源路徑，然後選擇新增磁碟區。您也可以選擇開啟啟用 EFS。

    b. 您可以選擇新增磁碟區來新增更多磁碟區。

14. 針對參數，選擇新增參數，將參數替換預留位置新增為鍵對和選用值對。

15. 選擇下一頁。

16. 在容器組態區段中：

    a. 在 Name (名稱) 中，輸入容器的名稱。

    b. 對於必要容器，如果容器為必要，請啟用 。

    c. 針對映像，選擇要用於任務的Docker映像。根據預設，Docker Hub 登錄檔中的映像為可用。您也可以用 *repository-url*/*image*:*tag* 指定其他儲存庫。名稱的長度上限為 225 個字元。可以包含大小寫字母、數字、連字號 (-)、底線 (_)、冒號 (:)、正斜線 (/) 和井號 (#)。此參數會映射至 Docker Remote API 的建立容器區段中的 Image 以及 docker run 的 IMAGE 參數。

> **Note**
>
> Docker 映像架構必須符合執行個體排程所在之運算資源的處理器架構。例如，ARM 型 Docker 映像只能在 ARM 型運算資源上執行。

- Amazon ECR Public 儲存庫中的映像會使用完整`registry/repository[:tag]`或命名慣例 `registry/repository[@digest]` （例如 `public.ecr.aws/`*`registry_alias`*`/`*`my-web-app`*`:`*`latest`*)。
- Amazon ECR 儲存庫中的映像會使用完整的命名慣例 `registry/repository[:tag]` （例如 *`aws_account_id`*`.dkr.ecr.`*`region`*`.amazonaws.com/`*`my-web-app`*`:`*`latest`*)。
- Docker Hub 上官方儲存庫中的映像，使用的是單一名稱 (例如，`ubuntu` 或 `mongo`)。
- Docker Hub 上的其他儲存庫中的映像要求使用組織名稱 (例如，`amazon/amazon-ecs-agent`)。
- 其他線上存放庫中的映像更進一步要求使用網域名稱 (例如，`quay.io/assemblyline/ubuntu`)。

d. 針對資源需求，請設定下列各項：

   i. 針對 vCPUs，選擇容器CPUs 數量。

   ii. 針對記憶體，選擇容器的記憶體量。

   iii. 針對 GPU - 選用，選擇容器的 GPUs 數量。

e. 在 User (使用者) 中，輸入要在容器內使用的使用者名稱。

f. 開啟 Enable read only filesystem (啟用唯讀檔案系統) 以移除對磁碟區的寫入權限。

g. 開啟特權，在主機執行個體上為任務容器提供更高的許可，類似於根使用者。

h. 針對命令，在欄位中輸入命令，作為其 JSON 字串陣列對等項。

   此參數會映射至 Docker Remote API 的建立容器區段中的 `Cmd` 以及 docker run 的 `COMMAND` 參數。如需 Docker `CMD` 參數的詳細資訊，請參閱 https://docs.docker.com/engine/reference/builder/#cmd。

> **Note**
>
> 您可以在命令中使用預設值來取代參數和預留位置。如需詳細資訊，請參閱參數。

i. 對於儲存庫登入資料 - 選用，輸入包含登入資料之秘密的 ARN。

j. 針對環境變數 - 選用，選擇新增環境變數以新增要傳遞至容器的環境變數。

k. 在 Linux 參數 - 選用區段中：

i. 開啟 Enable init process (啟用初始化處理) 以在容器內執行初始化程序。

ii. 針對共用記憶體大小，輸入 /dev/shm 磁碟區的大小 (MiB)

iii. 針對最大交換大小，輸入容器可以使用的交換記憶體總量（以 MiB 為單位）。

iv. 針對交換輸入介於 0 到 100 之間的值，以表示容器的交換行為。如果您未指定值並啟用交換，則值預設為 60。

v. 針對裝置，選擇新增裝置以新增裝置：

A. 針對 Container path (容器路徑)，指定容器執行個體中的路徑，以公開對應到主機執行個體的裝置。如果您將此保留空白，則會在容器中使用主機路徑。

B. 針對 Host path (主機路徑)，指定主機執行個體中的裝置的路徑。

C. 針對許可，選擇要套用至裝置的一或多個許可。可用的許可為讀取、寫入和 MKNOD。

vi. 針對 Tmpfs，選擇新增 tmpfs 以新增`tmpfs`掛載。

l.

> **Note**
>
> Firelens 記錄必須在專用容器中完成。若要設定 Firelens 記錄：
>
> - 在每個容器中，除了專用防火牆容器之外，請將記錄驅動程式設定為 `awsfirelens`
> - 在您的 Firelens 容器中，設定記錄目的地的 Firelens 組態 - 選用和記錄組態 - 選用

在 Firelens 組態 - 選用區段中：

> **Important**
>
> AWS Batch 在非 MNP、非 FARGATE Amazon ECS 任務上強制執行`host`網路模式。Amazon ECS Firelens 需要根使用者。執行使用 `host` 網路模式的任務時，Amazon ECS 建議不要使用根使用者 (UID 0) 執行容器，以提高安全性。因此，所有具有 Firelens 記錄的非 MNP、非 FARGATE ECS 任務都不符合安全最佳實務。

i. 針對類型，選擇 `fluentd`或 `fluentbit`。

ii. 在選項中，輸入選項的名稱/值對。您可以使用新增選項來新增更多選項。

m. 在記錄組態 - 選用區段中：

i. 對於日誌驅動程式，選擇要使用的日誌驅動程式。如需可用日誌驅動程式的詳細資訊，請參閱 LogConfiguration：logDriver。

> **Note**
>
> 根據預設，會使用 `awslogs` 日誌驅動程式。

ii. 針對選項，選擇新增選項以新增選項。輸入名稱/值對，然後選擇新增選項。

iii. 針對秘密，選擇新增秘密。輸入名稱/值對，然後選擇新增秘密以新增秘密。

> **Tip**
>
> 如需詳細資訊，請參閱 LogConfiguration：secretOptions。

n. 針對掛載點 - 選用，選擇新增掛載點以新增資料磁碟區的掛載點。您必須指定來源磁碟區和容器路徑。

o. 針對秘密 - 選用，選擇新增秘密以新增秘密。然後，輸入名稱值對，然後選擇新增秘密。

> **Tip**
>
> 如需詳細資訊，請參閱 LogConfiguration：secretOptions。

p. 針對 Ulimits - 選用，選擇新增 ulimit 以新增容器`ulimits`的值。輸入名稱、軟性限制和硬性限制值，然後選擇新增 ulimit。

q. 對於相依性 - 選用，選擇新增容器相依性。選擇容器的名稱及其狀態，以判斷此容器何時啟動。

17. 如果您只設定一個容器，則必須選擇新增容器並完成設定新容器。否則，請選擇下一步以檢閱。

建立多節點平行任務定義

您必須先建立任務定義 AWS Batch，才能在 中執行任務。此程序在單節點和多節點平行任務之間略有不同。本主題特別介紹如何為 AWS Batch 多節點平行任務 （ 也稱為 Gang 排程) 建立任務定義。如需詳細資訊，請參閱多節點平行任務。

> **Note**
>
> AWS Fargate 不支援多節點平行任務。

目錄

- 教學課程：在 Amazon EC2 資源上建立多節點平行任務定義

教學課程：在 Amazon EC2 資源上建立多節點平行任務定義

在 Amazon Elastic Compute Cloud (Amazon EC2) 資源上建立多節點平行任務定義。

> **Note**
>
> 若要建立單一節點任務定義，請參閱 在 Amazon EC2 資源上建立單一節點任務定義。

若要在 Amazon EC2 資源上建立多節點平行任務定義：

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇任務定義。
4. 選擇建立。
5. 針對協調類型，選擇 Amazon Elastic Compute Cloud (Amazon EC2)。
6. 針對啟用多節點平行，開啟多節點平行。
7. 在名稱中，輸入任務定義的唯一名稱。名稱長度上限為 128 個字元，且可包含大小寫字母、數字、連字號 (-) 和底線 (_)。
8. （選用） 針對執行逾時，指定您希望任務嘗試執行的秒數上限。如果嘗試超過逾時持續時間，則會停止嘗試並移至 `FAILED` 狀態。如需詳細資訊，請參閱任務逾時。

9. （選用）開啟排程優先順序。輸入介於 0 到 100 之間的排程優先順序值。較高值的優先順序高於較低值。
10. （選用）對於任務嘗試，輸入嘗試將任務移至RUNNABLE狀態的 AWS Batch 次數。輸入介於 1 到 10 之間的數字。
11. （選用）針對重試策略條件，選擇退出時新增評估。輸入至少一個參數值，然後選擇動作。對於每組條件，動作必須設定為重試或結束。這些動作表示下列項目：

    - 重試 – AWS Batch 重試，直到達到您指定的任務嘗試次數為止。
    - 結束 – AWS Batch 停止重試任務。

    > **Important**
    >
    > 如果您選擇在結束時新增評估，則必須至少設定一個參數，然後選擇動作或選擇在結束時移除評估。

12. （選用）展開標籤，然後選擇新增標籤以將標籤新增至資源。輸入索引鍵和選用值，然後選擇新增標籤。您也可以開啟傳播標籤，將標籤從任務和任務定義傳播到 Amazon ECS 任務。
13. 選擇下一頁。
14. 針對 Number of nodes (節點數)，請輸入要在您任務中使用的總節點數量。
15. 針對 Main node (主要節點)，請輸入要用於主要節點的節點索引。預設的主要節點索引為 0。
16. 針對執行個體類型，選擇執行個體類型。

    > **Note**
    >
    > 您選擇的執行個體類型會套用至所有節點。

17. 針對參數，選擇新增參數，將參數替換預留位置新增為鍵對和選用值對。
18. 在節點範圍區段中：

    a. 選取新增節點範圍。這會建立節點範圍區段。

    b. 針對 Target nodes (目標節點)，請使用 *range_start*:*range_end* 標記法指定您節點群組的範圍。

    您可以為您為任務指定的節點建立最多五個節點範圍。節點範圍會使用節點的索引值，且節點索引會從 0 開始。確定最終節點群組的範圍結束索引值小於您指定的節點數量。例如，假設您指定了 10 個節點，而且您想要使用單一節點群組。然後，您的結束範圍為 9。

c. 針對映像，選擇要用於任務的Docker映像。根據預設，Docker Hub 登錄檔中的映像為可用。您也可以用 *repository-url*/*image*:*tag* 指定其他儲存庫。名稱長度最多可達 225 個字元。可以包含大小寫字母、數字、連字號 (-)、底線 (_)、冒號 (:)、正斜線 (/) 和井號 (#)。此參數會映射至 Docker Remote API 的建立容器區段中的 Image 以及 docker run 的 IMAGE 參數。

> **Note**
>
> Docker 映像架構必須符合執行個體排程所在之運算資源的處理器架構。例如，ARM 型 Docker 映像只能在 ARM 型運算資源上執行。

- Amazon ECR Public 儲存庫中的映像會使用完整registry/repository[:tag]或命名慣例 registry/repository[@digest] （例如 public.ecr.aws/*registry_alias*/*my-web-app*:*latest*)。
- Amazon ECR 儲存庫中的映像會使用完整的registry/repository[:tag]命名慣例。例如 *aws_account_id*.dkr.ecr.*region*.amazonaws.com/*my-web-app*:*latest*
- Docker Hub 上官方儲存庫中的映像，使用的是單一名稱 (例如，ubuntu 或 mongo)。
- Docker Hub 上的其他儲存庫中的映像要求使用組織名稱 (例如，amazon/amazon-ecs-agent)。
- 其他線上存放庫中的映像更進一步要求使用網域名稱 (例如，quay.io/assemblyline/ubuntu)。

d. 針對命令，在欄位中輸入命令，作為其 JSON 字串陣列對等項。

此參數會映射至 Docker Remote API 的建立容器區段中的 Cmd 以及 docker run 的 COMMAND 參數。如需 Docker CMD 參數的詳細資訊，請參閱 https://docs.docker.com/engine/reference/builder/#cmd。

> **Note**
>
> 您可以在命令中使用預設值來取代參數和預留位置。如需詳細資訊，請參閱參數。

e. 在 vCPU 中，指定保留給容器的 vCPU 數量。此參數會映射到 Docker Remote API 的建立容器區段中的 CpuShares 以及 docker run 的 --cpu-shares 選項。每個 vCPU 相當於 1,024 個 CPU 共用。您必須指定至少 1 個 vCPU。

f. 在 Memory (記憶體) 中，指定提供給任務容器使用的記憶體硬性限制 (MiB)。如果您的容器嘗試超過此處指定的記憶體，則容器會停止。此參數會映射到 Docker Remote API 的建立容器區段中的 `Memory` 以及 docker run 的 `--memory` 選項。您必須為單一工作指定至少 4 MiB 的記憶體。

> **Note**
>
> 若要最大化資源使用率，您可以為特定執行個體類型提供盡可能多的記憶體。如需詳細資訊，請參閱運算資源記憶體管理。

g. （選用）針對 GPUs 數量，指定任務使用的 GPUs 數量。任務會在具有指定 GPU 數量的容器上執行，這些 GPUs 會固定到該容器。

h. （選用）對於任務角色，您可以指定 IAM 角色，為任務中的容器提供使用 AWS APIs許可。此功能使用 Amazon ECS IAM 角色執行任務功能。如需包含組態先決條件的詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的任務的 IAM 角色。

> **Note**
>
> 對於在 Fargate 資源上執行的任務，需要任務角色。

> **Note**
>
> 這裡只會顯示具有 Amazon Elastic Container Service 任務角色信任關係的角色。如需為您的 AWS Batch 任務建立 IAM 角色的詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的為您的任務建立 IAM 角色和政策。

i. （選用）對於執行角色，指定 IAM 角色，授予 Amazon ECS 容器代理程式代表您進行 AWS API 呼叫的許可。此功能使用 Amazon ECS IAM 角色執行任務功能。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 任務執行 IAM 角色。

19. （選用）展開其他組態：

a. 針對環境變數，選擇新增環境變數，將環境變數新增為名稱/值對。這些變數會傳遞至容器。

b. 對於任務角色組態，您可以指定 IAM 角色，為任務中的容器提供使用 AWS APIs許可。此功能使用 Amazon ECS IAM 角色執行任務功能。如需包含組態先決條件的詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的任務的 IAM 角色。

> **Note**
>
> 對於在 Fargate 資源上執行的任務，需要任務角色。

> **Note**
>
> 這裡只會顯示具有 Amazon Elastic Container Service 任務角色信任關係的角色。如需如何為 AWS Batch 任務建立 IAM 角色的詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的為您的任務建立 IAM 角色和政策。

c. 針對執行角色，指定 IAM 角色，授予 Amazon ECS 容器代理程式代表您進行 AWS API 呼叫的許可。此功能使用 Amazon ECS IAM 角色執行任務功能。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 任務執行 IAM 角色。

20. 在安全組態區段中：

    a. （選用）若要在主機執行個體上為任務的容器提供更高的權限（類似於`root`使用者），請開啟特權。此參數會映射到 Docker Remote API 的建立容器區段中的 `Privileged` 以及 docker run 的 `--privileged` 選項。

    b. （選用）對於使用者，輸入要在容器內使用的使用者名稱。此參數會映射到 Docker Remote API 的建立容器區段中的 `User` 以及 docker run 的 `--user` 選項。

    c. （選用）對於秘密，選擇新增秘密，將秘密新增為名稱/值對。這些秘密會在容器中公開。如需詳細資訊，請參閱 LogConfiguration：secretOptions。

21. 在 Linux 組態區段中：

    a. 開啟 Enable read only filesystem (啟用唯讀檔案系統) 以移除對磁碟區的寫入權限。

    b. （選用）開啟啟用init程序以在容器內執行`init`程序。此程序會轉送訊號並接收程序。

    c. 針對共用記憶體大小，輸入`/dev/shm`磁碟區的大小（以 MiB 為單位）。

    d. 針對最大交換大小，輸入容器可以使用的交換記憶體總量（以 MiB 為單位）。

    e. 針對交換輸入介於 0 到 100 之間的值，以表示容器的交換行為。如果您未指定值並啟用交換，則值預設為 60。如需詳細資訊，請參閱 LinuxParameters：swappiness。

    f. （選用）針對裝置，選擇新增裝置以新增裝置：

i. 針對 Container path (容器路徑)，指定容器執行個體中的路徑，以公開對應到主機執行個體的裝置。如果您將此保留空白，則會在容器中使用主機路徑。

ii. 針對 Host path (主機路徑)，指定主機執行個體中的裝置的路徑。

iii. 針對許可，選擇要套用至裝置的一或多個許可。可用的許可為讀取、寫入和 MKNOD。

22. （選用）對於掛載點，選擇新增掛載點組態以新增資料磁碟區的掛載點。您必須指定來源磁碟區和容器路徑。這些掛載點會傳遞至容器執行個體上的Docker協助程式。您也可以選擇將磁碟區設為唯讀。

23. （選用）針對 Ulimits 組態，選擇新增 ulimit 以新增容器`ulimits`的值。輸入名稱、軟性限制和硬性限制值，然後選擇新增 ulimit。

24. （選用）針對磁碟區組態，選擇新增磁碟區以建立要傳遞至容器的磁碟區清單。輸入磁碟區的名稱和來源路徑，然後選擇新增磁碟區。您也可以選擇開啟啟用 EFS。

25. （選用）針對 Tmpfs，選擇新增 tmpfs 以新增`tmpfs`掛載。

26. 在任務屬性區段中：

a. 針對執行角色 - 條件式，選擇角色以允許 Amazon ECS 代理程式代表您進行 AWS API 呼叫。如需建立執行角色的詳細資訊，請參閱 教學課程：建立 IAM 執行角色。

b.

> **Important**
>
> 若要使用 ECS 執行命令，您的運算環境必須符合多節點平行任務的運算環境考量。

選擇啟用 ECS 執行命令，以啟用直接存取 Amazon ECS 容器殼層，並略過主機作業系統。您必須選擇任務角色。

> **Important**
>
> ECS 執行命令需要可寫入的檔案系統。

c. 針對任務角色，選擇 Amazon ECS Identity and Access Management (IAM) 角色，以允許容器代表您進行 AWS API 呼叫。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 任務 IAM 角色。

27. （選用）在記錄組態區段中：

a. 對於日誌驅動程式，選擇要使用的日誌驅動程式。如需可用日誌驅動程式的詳細資訊，請參閱 LogConfiguration：logDriver。

> **Note**
>
> 根據預設，會使用 awslogs 日誌驅動程式。

b. 針對選項，選擇新增選項以新增選項。輸入名稱/值對，然後選擇新增選項。

c. 針對秘密，選擇新增秘密。輸入名稱/值對，然後選擇新增秘密以新增秘密。

> **Tip**
>
> 如需詳細資訊，請參閱 LogConfiguration：secretOptions。

28. 選擇下一頁。

29. 針對任務定義檢閱，檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立任務定義。

使用 ContainerProperties 的任務定義範本

以下是包含單一容器的空任務定義範本。您可以使用此範本來建立任務定義，然後將其儲存至 檔案，並搭配 AWS CLI --cli-input-json選項使用。如需這些參數的詳細資訊，請參閱 JobDefinition。

> **Note**
>
> 您可以使用下列 AWS CLI 命令產生單一容器任務定義範本：
>
> ```
> $ aws batch register-job-definition --generate-cli-skeleton
> ```

```
{
    "jobDefinitionName": "",
    "type": "container",
    "parameters": {
        "KeyName": ""
    },
    "schedulingPriority": 0,
    "containerProperties": {
        "image": "",
        "vcpus": 0,
```

```
        "memory": 0,
        "command": [
            ""
        ],
        "jobRoleArn": "",
        "executionRoleArn": "",
        "volumes": [
            {
                "host": {
                    "sourcePath": ""
                },
                "name": "",
                "efsVolumeConfiguration": {
                    "fileSystemId": "",
                    "rootDirectory": "",
                    "transitEncryption": "ENABLED",
                    "transitEncryptionPort": 0,
                    "authorizationConfig": {
                        "accessPointId": "",
                        "iam": "DISABLED"
                    }
                }
            }
        ],
        "environment": [
            {
                "name": "",
                "value": ""
            }
        ],
        "mountPoints": [
            {
                "containerPath": "",
                "readOnly": true,
                "sourceVolume": ""
            }
        ],
        "readonlyRootFilesystem": true,
        "privileged": true,
        "ulimits": [
            {
                "hardLimit": 0,
                "name": "",
                "softLimit": 0
```

```
            }
        ],
        "user": "",
        "instanceType": "",
        "resourceRequirements": [
            {
                "value": "",
                "type": "MEMORY"
            }
        ],
        "linuxParameters": {
            "devices": [
                {
                    "hostPath": "",
                    "containerPath": "",
                    "permissions": [
                        "WRITE"
                    ]
                }
            ],
            "initProcessEnabled": true,
            "sharedMemorySize": 0,
            "tmpfs": [
                {
                    "containerPath": "",
                    "size": 0,
                    "mountOptions": [
                        ""
                    ]
                }
            ],
            "maxSwap": 0,
            "swappiness": 0
        },
        "logConfiguration": {
            "logDriver": "syslog",
            "options": {
                "KeyName": ""
            },
            "secretOptions": [
                {
                    "name": "",
                    "valueFrom": ""
                }
```

```
            ]
        },
        "secrets": [
            {
                "name": "",
                "valueFrom": ""
            }
        ],
        "networkConfiguration": {
            "assignPublicIp": "DISABLED"
        },
        "fargatePlatformConfiguration": {
            "platformVersion": ""
        }
    },
    "nodeProperties": {
        "numNodes": 0,
        "mainNode": 0,
        "nodeRangeProperties": [
            {
                "targetNodes": "",
                "container": {
                    "image": "",
                    "vcpus": 0,
                    "memory": 0,
                    "command": [
                        ""
                    ],
                    "jobRoleArn": "",
                    "executionRoleArn": "",
                    "volumes": [
                        {
                            "host": {
                                "sourcePath": ""
                            },
                            "name": "",
                            "efsVolumeConfiguration": {
                                "fileSystemId": "",
                                "rootDirectory": "",
                                "transitEncryption": "DISABLED",
                                "transitEncryptionPort": 0,
                                "authorizationConfig": {
                                    "accessPointId": "",
                                    "iam": "ENABLED"
```

```
                }
            }
        }
    ],
    "environment": [
        {
            "name": "",
            "value": ""
        }
    ],
    "mountPoints": [
        {
            "containerPath": "",
            "readOnly": true,
            "sourceVolume": ""
        }
    ],
    "readonlyRootFilesystem": true,
    "privileged": true,
    "ulimits": [
        {
            "hardLimit": 0,
            "name": "",
            "softLimit": 0
        }
    ],
    "user": "",
    "instanceType": "",
    "resourceRequirements": [
        {
            "value": "",
            "type": "MEMORY"
        }
    ],
    "linuxParameters": {
        "devices": [
            {
                "hostPath": "",
                "containerPath": "",
                "permissions": [
                    "WRITE"
                ]
            }
        ],
```

```
                    "initProcessEnabled": true,
                    "sharedMemorySize": 0,
                    "tmpfs": [
                        {
                            "containerPath": "",
                            "size": 0,
                            "mountOptions": [
                                ""
                            ]
                        }
                    ],
                    "maxSwap": 0,
                    "swappiness": 0
                },
                "logConfiguration": {
                    "logDriver": "awslogs",
                    "options": {
                        "KeyName": ""
                    },
                    "secretOptions": [
                        {
                            "name": "",
                            "valueFrom": ""
                        }
                    ]
                },
                "secrets": [
                    {
                        "name": "",
                        "valueFrom": ""
                    }
                ],
                "networkConfiguration": {
                    "assignPublicIp": "DISABLED"
                },
                "fargatePlatformConfiguration": {
                    "platformVersion": ""
                }
            }
        }
    ]
},
"retryStrategy": {
    "attempts": 0,
```

```
        "evaluateOnExit": [
            {
                "onStatusReason": "",
                "onReason": "",
                "onExitCode": "",
                "action": "RETRY"
            }
        ]
    },
    "propagateTags": true,
    "timeout": {
        "attemptDurationSeconds": 0
    },
    "tags": {
        "KeyName": ""
    },
    "platformCapabilities": [
        "EC2"
    ],
    "eksProperties": {
        "podProperties": {
            "serviceAccountName": "",
            "hostNetwork": true,
            "dnsPolicy": "",
            "containers": [
                {
                    "name": "",
                    "image": "",
                    "imagePullPolicy": "",
                    "command": [
                        ""
                    ],
                    "args": [
                        ""
                    ],
                    "env": [
                        {
                            "name": "",
                            "value": ""
                        }
                    ],
                    "resources": {
                        "limits": {
                            "KeyName": ""
```

```
                        },
                        "requests": {
                            "KeyName": ""
                        }
                    },
                    "volumeMounts": [
                        {
                            "name": "",
                            "mountPath": "",
                            "readOnly": true
                        }
                    ],
                    "securityContext": {
                        "runAsUser": 0,
                        "runAsGroup": 0,
                        "privileged": true,
                        "readOnlyRootFilesystem": true,
                        "runAsNonRoot": true
                    }
                }
            ],
            "volumes": [
                {
                    "name": "",
                    "hostPath": {
                        "path": ""
                    },
                    "emptyDir": {
                        "medium": "",
                        "sizeLimit": ""
                    },
                    "secret": {
                        "secretName": "",
                        "optional": true
                    }
                }
            ]
        }
    }
}
```

ContainerProperties 的任務定義參數

使用 的任務定義`ContainerProperties`分為幾個部分：

- 任務定義名稱
- 任務定義的類型
- 參數替換預留位置預設值
- 任務的容器屬性
- 在 Amazon EKS 資源上執行之任務所需的任務定義 Amazon EKS 屬性
- 多節點平行任務所需的節點屬性
- 在 Fargate 資源上執行的任務所需的平台功能
- 任務定義的預設標籤傳播詳細資訊
- 任務定義的預設重試策略
- 任務定義的預設排程優先順序
- 任務定義的預設標籤
- 任務定義的預設逾時

內容

- 任務定義名稱
- 類型
- 參數
- 容器屬性
- Amazon EKS 屬性
- 平台功能
- 傳播標籤
- 節點屬性
- 重試策略
- 排程優先順序
- Tags (標籤)
- 逾時

任務定義名稱

`jobDefinitionName`

您必須在註冊任務定義時指定名稱。名稱長度上限為 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。以該名稱註冊的第一個任務定義會獲得 1 的修訂。後續使用該名稱註冊的任何任務定義，將取得遞增的修訂版號碼。

類型：字串

必要：是

類型

`type`

註冊任務定義時，需指定任務類型。如果任務在 Fargate 資源上執行，`multinode`則不支援 。如需多節點平行任務的詳細資訊，請參閱「建立多節點平行任務定義」。

類型：字串

有效值：`container` | `multinode`

必要：是

參數

`parameters`

當您提交任務時，您可以指定取代預留位置的參數，或覆寫預設任務定義參數。任務提交要求中的參數，優先於任務定義中的預設值。這表示您可以針對使用相同格式的多個任務使用相同的任務定義。您也可以在提交時以程式設計方式變更命令中的值。

類型：字串到字串映射

必要：否

註冊任務定義時，您可以在任務容器屬性的 `command` 欄位使用參數替換預留位置。語法如下。

```
"command": [
```

```
    "ffmpeg",
    "-i",
    "Ref::inputfile",
    "-c",
    "Ref::codec",
    "-o",
    "Ref::outputfile"
]
```

在上述範例中，命令中有 *Ref::inputfile*、*Ref::codec* 和 *Ref::outputfile* 參數替換預留位置。您可以使用任務定義中的 parameters 物件來設定這些預留位置的預設值。例如，若要設定 *Ref::codec* 預留位置的預設值，您應在任務定義中指定下列各項：

```
"parameters" : {"codec" : "mp4"}
```

提交此任務定義以執行時，容器 命令中的*Ref::codec*引數會取代為預設值 mp4。

容器屬性

當您註冊任務定義時，請指定在放置任務時，在容器執行個體上傳遞至 Docker 協助程式的容器屬性清單。任務定義允許使用以下的容器屬性。針對單節點任務，這些容器屬性會設定在任務定義層級。針對多節點平行任務，每個節點群組的容器屬性會設定在 節點屬性 層級。

command

傳遞至容器的命令。此參數會映射至 Docker Remote API 的建立容器區段中的 Cmd 以及 docker run 的 COMMAND 參數。如需 Docker CMD 參數的詳細資訊，請參閱 https://docs.docker.com/engine/reference/builder/#cmd。

```
"command": ["string", ...]
```

類型：字串陣列

必要：否

environment

傳遞至容器的環境變數。此參數會映射到 Docker Remote API 的建立容器區段中的 Env 以及 docker run 的 --env 選項。

> **Important**
>
> 不建議您對敏感資訊 (例如憑證) 使用純文字環境變數。

> **Note**
>
> 環境變數開頭不得為 AWS_BATCH。此命名慣例會保留給 AWS Batch 服務所設定的變數。

類型：金鑰值對的陣列

必要：否

name

環境變數的名稱。

類型：字串

必要：是，使用 environment 時。

value

環境變數的值。

類型：字串

必要：是，使用 environment 時。

```
"environment" : [
    { "name" : "envName1", "value" : "envValue1" },
    { "name" : "envName2", "value" : "envValue2" }
]
```

executionRoleArn

註冊任務定義時，您可以指定 IAM 角色。此角色為 Amazon ECS 容器代理程式提供代表您呼叫其相關聯政策中指定之 API 動作的許可。在 Fargate 資源上執行的任務必須提供執行角色。如需詳細資訊，請參閱AWS Batch IAM 執行角色。

類型：字串

必要：否

`fargatePlatformConfiguration`

在 Fargate 資源上執行之任務的平台組態。在 EC2 資源上執行的任務不得指定此參數。

類型：FargatePlatformConfiguration 物件

必要：否

`platformVersion`

AWS Fargate 平台版本用於 任務，或使用 AWS Fargate 平台LATEST的最新核准版本。

類型：字串

預設：LATEST

必要：否

`image`

用來啟動任務的映像。此字串會直接傳遞至 Docker 常駐程式。根據預設，Docker Hub 登錄檔中的映像為可用。您也可以用 *repository-url*/*image*:*tag* 指定其他儲存庫。允許最多 255 個字元 (大小寫)、數字、連字號、底線、等號、句號、正斜線、井號。此參數會映射至 Docker Remote API 的建立容器區段中的 `Image` 以及 docker run 的 `IMAGE` 參數。

> **Note**
>
> Docker 映像架構必須符合執行個體排程所在之運算資源的處理器架構。例如，ARM 型 Docker 映像只能在 ARM 型運算資源上執行。

- Amazon ECR Public 儲存庫中的映像會使用完整`registry/repository[:tag]`或命名慣例 `registry/repository[@digest]` (例如 `public.ecr.aws/`*registry_alias*`/`*my-web-app*`:`*latest*)。
- Amazon ECR 儲存庫中的映像會使用完整的`registry/repository:[tag]`命名慣例。例如 *aws_account_id*`.dkr.ecr.`*region*`.amazonaws.com/`*my-web-app*`:`*latest*。
- Docker Hub 上官方儲存庫中的映像，使用的是單一名稱 (例如，`ubuntu` 或 `mongo`)。
- Docker Hub 上的其他儲存庫中的映像要求使用組織名稱 (例如，`amazon/amazon-ecs-agent`)。

- 其他線上存放庫中的映像更進一步要求使用網域名稱 (例如，quay.io/assemblyline/ubuntu)。

類型：字串

必要：是

`instanceType`

用於多節點平行任務的執行個體類型。所有節點平行任務中的所有節點群組皆必須使用相同的執行個體類型。此參數不適用於單一節點容器任務或在 Fargate 資源上執行的任務。

類型：字串

必要：否

`jobRoleArn`

註冊任務定義時，您可以指定 IAM 角色。角色提供任務容器權限，允許其代表您呼叫相關聯政策中指定的 API 動作。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的任務 IAM 角色。

類型：字串

必要：否

`linuxParameters`

Linux 特定的修改，會套用到容器，例如用於裝置映射的詳細資訊。

```
"linuxParameters": {
    "devices": [
        {
            "hostPath": "string",
            "containerPath": "string",
            "permissions": [
                "READ", "WRITE", "MKNOD"
            ]
        }
    ],
    "initProcessEnabled": true|false,
    "sharedMemorySize": 0,
    "tmpfs": [
```

```
        {
            "containerPath": "string",
            "size": integer,
            "mountOptions": [
                "string"
            ]
        }
    ],
    "maxSwap": integer,
    "swappiness": integer
}
```

類型：LinuxParameters 物件

必要：否

devices

映射到容器的裝置列表。此參數對應到 Docker Remote API Create a container (建立容器) 一節中的 Devices，以及 Docker run 的 --device 選項。

> **Note**
>
> 此參數不適用於在 Fargate 資源上執行的任務。

類型：Device 物件的陣列

必要：否

hostPath

主機容器執行個體中可用裝置的路徑。

類型：字串

必要：是

containerPath

裝置在容器中公開的路徑。如果未指定，裝置會在與主機路徑相同的路徑中公開。

類型：字串

必要：否

`permissions`

容器中的裝置的許可。如果未指定，則許可會設為 READ、 WRITE和 MKNOD。

類型：字串陣列

必要：否

有效值：READ | WRITE | MKNOD

`initProcessEnabled`

若為 true，請在容器內執行 `init` 處理程序，該處理程序可轉寄訊號及獲得處理程序。此參數會映射到 docker run 的 `--init` 選項。在您的容器執行個體上，此參數需要 1.25 版或更新版本的 Docker Remote API。若要檢查容器執行個體的 Docker Remote API 版本，請登入容器執行個體，並執行下列命令：`sudo docker version | grep "Server API version"`

類型：布林值

必要：否

`maxSwap`

任務可以使用的交換記憶體總量（以 MiB 為單位）。此參數將會轉換為 docker run 的 `--memory-swap` 選項，其中值是容器記憶體與 `maxSwap` 值的總和。如需詳細資訊，請參閱 Docker 文件中的 `--memory-swap` 詳細資訊。

如果將 `maxSwap` 值指定為 `0`，容器不會使用交換。接受的值為 `0` 或任何正整數。如果省略 `maxSwap` 參數，容器會使用其執行所在之容器執行個體的交換組態。必須設定 `maxSwap` 值，才能使用 `swappiness` 參數。

> **Note**
>
> 此參數不適用於在 Fargate 資源上執行的任務。

類型：整數

必要：否

`sharedMemorySize`

/dev/shm 磁碟區的大小值 (以 MiB 為單位)。此參數會映射到 docker run 的 `--shm-size` 選項。

> **Note**
>
> 此參數不適用於在 Fargate 資源上執行的任務。

類型：整數

必要：否

`swappiness`

您可藉此調整容器的記憶體交換行為。除非絕對必要，否則 `swappiness`值`0`會導致交換不會發生。為 `100` 的 `swappiness` 值導致積極地交換頁面。接受的值為介於 `0` 與 `100` 之間的整數。如果未指定 `swappiness` 參數，則會使用預設值 `60`。如果未對 `maxSwap` 指定值，則會忽略此參數。如果 `maxSwap` 設定為 0，則容器不會使用交換。此參數會映射到 docker run 的 `--memory-swappiness` 選項。

當您使用每個容器交換組態時，請考量下列事項。

- 必須在容器執行個體上啟用和配置交換空間，供容器使用。

  > **Note**
  >
  > 根據預設，Amazon ECS 最佳化 AMI 沒有啟用交換功能。您必須在執行個體上啟用交換，才能使用此功能。如需詳細資訊，請參閱《Amazon EC2 使用者指南》中的執行個體存放區交換磁碟區，或如何使用交換檔案將記憶體配置為 Amazon EC2 執行個體中的交換空間？

- 交換空間參數僅針對使用 EC2 資源的任務定義提供支援。
- 如果從任務定義中省略 `maxSwap` 和 `swappiness` 參數，每個容器的預設 `swappiness` 值都為 60。總交換用量限制為容器記憶體保留的兩倍。

> **Note**
>
> 此參數不適用於在 Fargate 資源上執行的任務。

類型：整數

必要：否

`tmpfs`

tmpfs 掛載的容器路徑、掛載選項和大小。

類型：Tmpfs 物件的陣列

> **Note**
> 此參數不適用於在 Fargate 資源上執行的任務。

必要：否

`containerPath`

掛載 tmpfs 磁碟區之容器中的絕對檔案路徑。

類型：字串

必要：是

`mountOptions`

tmpfs 磁碟區掛載選項的清單。

有效值："defaults" | "ro" | "rw" | suid"nosuid" | dev"" | "nodev" | """ | "exec" noexec| "async"" | sync""" | "dirsync"" | "remount"" mand| "nomand" | "atime"" | "" | "nodiratime"" | noatimediratime""" | "bind"" | """ rbindunbindablerunbindableprivate|rprivate"" | """ | "shared"" | "rsharedslaverslaverelatime"" | "norelatime"" | "strictatime"nostrictatimemodeuidgidnr_inodesnr_blocksmpol

類型：字串陣列

必要：否

`size`

tmpfs 磁碟區大小 (以 MiB 為單位)。

類型：整數

必要：是

`logConfiguration`

任務的日誌組態規格。

此參數對應到 Docker Remote API Create a container (建立容器) 一節中的 `LogConfig`，以及 Docker run 的 `--log-driver` 選項。根據預設，容器會和 Docker 常駐程式使用一樣的日誌記錄驅動程式。不過，容器可以使用與 Docker 協助程式不同的記錄驅動程式，方法是在容器定義中指定具有此參數的日誌驅動程式。若要為容器使用不同的記錄驅動程式，必須在容器執行個體或其他日誌伺服器上設定日誌系統，以提供遠端記錄選項。如需支援的不同日誌驅動程式選項的詳細資訊，請參閱 Docker 文件中的 Configure logging drivers (設定日誌驅動程式)。

> **Note**
>
> AWS Batch 目前支援 Docker 協助程式可用的部分記錄驅動程式 （ 如 LogConfiguration 資料類型所示 ）。

在您的容器執行個體上，此參數需要 1.18 版或更新版本的 Docker Remote API。若要檢查容器執行個體的 Docker Remote API 版本，請登入容器執行個體，並執行下列命令：`sudo docker version | grep "Server API version"`

```
"logConfiguration": {
    "devices": [
        {
            "logDriver": "string",
            "options": {
                "optionName1" : "optionValue1",
                "optionName2" : "optionValue2"
            }
            "secretOptions": [
              {
                  "name" : "secretOptionName1",
                  "valueFrom" : "secretOptionArn1"
              },
              {
                  "name" : "secretOptionName2",
                  "valueFrom" : "secretOptionArn2"
              }
```

```
            ]
        }
    ]
}
```

類型：LogConfiguration 物件

必要：否

`logDriver`

用於任務的日誌驅動程式。根據預設， 會 AWS Batch 啟用`awslogs`日誌驅動程式。根據預設，針對此參數列出的有效值是 Amazon ECS 容器代理程式可與之通訊的日誌驅動程式。

此參數對應到 Docker Remote API Create a container (建立容器) 一節中的 `LogConfig`，以及 Docker run 的 `--log-driver` 選項。根據預設，任務會使用與 Docker 協助程式相同的記錄驅動程式。不過，任務可以使用與 Docker 協助程式不同的記錄驅動程式，方法是在任務定義中指定具有此參數的日誌驅動程式。如果您想要為任務指定另一個記錄驅動程式，則必須在運算環境中的容器執行個體上設定日誌系統。或者，在另一個日誌伺服器上進行設定，以提供遠端記錄選項。如需支援的不同日誌驅動程式選項的詳細資訊，請參閱 Docker 文件中的 Configure logging drivers (設定日誌驅動程式)。

> **Note**
>
> AWS Batch 目前支援 Docker 協助程式可用的部分記錄驅動程式。未來的 Amazon ECS 容器代理程式版本可能會提供更多可用的其他日誌驅動程式。

支援的記錄驅動程式為 `awslogs`、`fluentd`、`gelf`、`json-file`、`journald`、`logentries`、`syslog` 和 `splunk`。

> **Note**
>
> 在 Fargate 資源上執行的任務僅限於 `awslogs`和 `splunk` 日誌驅動程式。

在您的容器執行個體上，此參數需要 1.18 版或更新版本的 Docker Remote API。若要檢查容器執行個體的 Docker Remote API 版本，請登入容器執行個體，並執行下列命令：`sudo docker version | grep "Server API version"`

> **Note**
>
> 在容器執行個體上執行的 Amazon ECS 容器代理程式必須使用 `ECS_AVAILABLE_LOGGING_DRIVERS`環境變數註冊該執行個體上可用的記錄驅動程式。否則，放置在該執行個體上的容器無法使用這些日誌組態選項。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 容器代理程式組態。

`awslogs`

指定 Amazon CloudWatch Logs 記錄驅動程式。如需詳細資訊，請參閱 Docker 文件中的 使用 awslogs 日誌驅動程式和 Amazon CloudWatch Logs 記錄驅動程式。

`fluentd`

指定 Fluentd 記錄驅動程式。如需包括用量和選項的詳細資訊，請參閱 Docker 文件中的 Fluentd 記錄驅動程式。

`gelf`

指定 Graylog 延伸格式 (GELF) 記錄驅動程式。如需包括用量和選項的詳細資訊，請參閱 Docker 文件中的 Graylog 擴充格式記錄驅動程式。

`journald`

指定 journald 記錄驅動程式。如需包括用量和選項的詳細資訊，請參閱 Docker 文件中的日誌記錄驅動程式。

`json-file`

指定 JSON 檔案記錄驅動程式。如需包括用量和選項的詳細資訊，請參閱 Docker 文件中的 JSON 檔案記錄驅動程式。

`splunk`

指定 Splunk 記錄驅動程式。如需包括用量和選項的詳細資訊，請參閱 Docker 文件中的 Splunk 記錄驅動程式。

`syslog`

指定 syslog 記錄驅動程式。如需包括用量和選項的詳細資訊，請參閱 Docker 文件中的 Syslog 記錄驅動程式。

類型：字串

必要：是

有效值：awslogs | fluentd | gelf | journald | json-file | splunk | syslog

> **Note**
>
> 如果您有先前未列出的自訂驅動程式，想要使用 Amazon ECS 容器代理程式，您可以放棄 GitHub 上可用的 Amazon ECS 容器代理程式專案，並自訂它以使用該驅動程式。我們鼓勵您為想要進行的變更提交提取請求。不過，Amazon Web Services 目前不支援執行已修改本軟體複本的請求。

`options`

要傳送至任務日誌驅動程式的日誌組態選項。

在您的容器執行個體上，此參數需要 1.19 版或更新版本的 Docker Remote API。

類型：字串到字串映射

必要：否

`secretOptions`

此物件代表要傳送至日誌組態的秘密。如需詳細資訊，請參閱指定敏感資料。

類型：物件陣列

必要：否

`name`

要在任務中設定的日誌驅動程式選項名稱。

類型：字串

必要：是

`valueFrom`

要公開給容器日誌組態之秘密的 Amazon Resource Name (ARN)。支援的值為 Secrets Manager 秘密的完整 ARN 或 SSM 參數存放區中參數的完整 ARN。

> **Note**
>
> 如果 SSM 參數存放區參數與您啟動 AWS 區域 的任務位於相同的 中，則您可以使用參數的完整 ARN 或名稱。如果參數存在於不同區域，則必須指定完整 ARN。

類型：字串

必要：是

`memory`

此參數已棄用，請*resourceRequirements*改用 。

為任務預留的記憶體 MiB 數量。

做為如何使用 的範例`resourceRequirements`，如果您的任務定義包含類似以下的語法。

```
"containerProperties": {
  "memory": 512
}
```

使用 的同等語法`resourceRequirements`如下所示。

```
"containerProperties": {
  "resourceRequirements": [
    {
      "type": "MEMORY",
      "value": "512"
    }
  ]
}
```

類型：整數

必要：是

`mountPoints`

容器中資料磁碟區的掛載點。此參數會映射到 Docker Remote API 的建立容器區段中的 `Volumes` 以及 docker run 的 `--volume` 選項。

```
"mountPoints": [
```

```
        {
          "sourceVolume": "string",
          "containerPath": "string",
          "readOnly": true|false
        }
      ]
```

類型：物件陣列

必要：否

`sourceVolume`

要掛載的磁碟區名稱。

類型：字串

必要：是，使用 `mountPoints` 時。

`containerPath`

容器上要掛載主機磁碟區的路徑。

類型：字串

必要：是，使用 `mountPoints` 時。

`readOnly`

如果此數值為 `true`，容器擁有磁碟區的唯讀存取權。如果此值為 `false`，則容器可寫入磁碟區。

類型：布林值

必要：否

預設：False

`networkConfiguration`

在 Fargate 資源上執行之任務的網路組態。在 EC2 資源上執行的任務不得指定此參數。

```
"networkConfiguration": {
   "assignPublicIp": "string"
}
```

類型：物件陣列

必要：否

assignPublicIp

指示任務是否有公有 IP 地址。如果任務需要傳出網路存取，這是必要的。

類型：字串

有效值：ENABLED | DISABLED

必要：否

預設：DISABLED

privileged

此參數為 true 時，容器便會取得主機容器執行個體的更高許可 (類似 root 使用者)。此參數會映射到 Docker Remote API 的建立容器區段中的 Privileged 以及 docker run 的 --privileged 選項。此參數不適用於在 Fargate 資源上執行的任務。請勿提供或指定為 false。

```
"privileged": true|false
```

類型：布林值

必要：否

readonlyRootFilesystem

此參數為 true 時，容器會取得根檔案系統的唯讀存取權。此參數會映射到 Docker Remote API 的建立容器區段中的 ReadonlyRootfs 以及 docker run 的 --read-only 選項。

```
"readonlyRootFilesystem": true|false
```

類型：布林值

必要：否

resourceRequirements

指派給容器的資源類型和數量。支援的資源包括 GPU MEMORY 和 VCPU。

```
"resourceRequirements" : [
```

```
  {
    "type": "GPU",
    "value": "number"
  }
]
```

類型：物件陣列

必要：否

`type`

要指派給容器的資源類型。支援的資源包括 `GPU` `MEMORY` 和 `VCPU`。

類型：字串

必要：是，使用 `resourceRequirements` 時。

`value`

為容器預留的指定資源數量。這些值根據指定的 `type` 而有所差異。

type="GPU"

為容器保留的記實體 GPU 數量。為任務中所有容器預留的 GPUs 數量不能超過啟動任務的運算資源上可用的 GPUs 數量。

type="MEMORY"

提供給容器使用的記憶體硬性限制 (MiB)。如果您的容器嘗試使用超過此處指定的記憶體，容器便會終止。此參數對應到 Docker Remote API Create a container (建立容器) 一節中的 `Memory`，以及 Docker run 的 `--memory` 選項。您必須為單一工作指定至少 4 MiB 的記憶體。這是必要的，但可以在多個地方為多節點平行 (MNP) 任務指定。必須至少為每個節點指定一次。此參數對應到 Docker Remote API Create a container (建立容器) 一節中的 `Memory`，以及 Docker run 的 `--memory` 選項。

> **Note**
>
> 如果您嘗試為特定執行個體類型提供盡可能多的記憶體，以最大化資源使用率，請參閱 運算資源記憶體管理。

對於在 Fargate 資源上執行的任務， `value`必須符合其中一個支援的值。此外，`VCPU`這些值必須是該記憶體值支援的其中一個值。

VCPU	MEMORY
0.25 vCPU	512、1024 和 2048 MiB
0.5 vCPU	1024-4096 MiB，以 1024 MiB 遞增
1 vCPU	2048-8192 MiB，以 1024 MiB 遞增
2 vCPU	4096-16384 MiB，以 1024 MiB 遞增
4 vCPU	8192-30720 MiB，以 1024 MiB 遞增
8 vCPU	16384-61440 MiB，以 4096 MiB 遞增
16 vCPU	32768-122880 MiB，以 8192 MiB 遞增

type="VCPU"

為任務保留的 vCPU 數量。此參數對應到 Docker Remote API Create a container (建立容器) 一節中的 `CpuShares`，以及 Docker run 的 `--cpu-shares` 選項。每個 vCPU 相當於 1,024 個 CPU 共用。對於在 EC2 資源上執行的任務，您必須至少指定一個 vCPU。這是必要項目，但可以在幾個地方指定。必須至少為每個節點指定一次。

對於在 Fargate 資源上執行的任務， `value` 必須符合其中一個支援的值，且`MEMORY`值必須是該 VCPU 值支援的其中一個值。支援的值為 0.25、0.5、1、2、4、8 和 16。

Fargate 隨需 vCPU 資源計數配額的預設值為 6 個 vCPU。如需 Fargate 配額的詳細資訊，請參閱《》中的 AWS Fargate 配額Amazon Web Services 一般參考。

類型：字串

必要：是，使用 `resourceRequirements` 時。

`secrets`

公開為環境變數之任務的秘密。如需詳細資訊，請參閱指定敏感資料。

```
"secrets": [
    {
      "name": "secretName1",
      "valueFrom": "secretArn1"
    },
```

```
    {
      "name": "secretName2",
      "valueFrom": "secretArn2"
    }
    ...
]
```

類型：物件陣列

必要：否

`name`

包含秘密的環境變數名稱。

類型：字串

必要：是，使用 `secrets` 時。

`valueFrom`

公開給容器的秘密。支援的值為 Secrets Manager 秘密的完整 Amazon Resource Name (ARN)，或 SSM 參數存放區中參數的完整 ARN。

> **Note**
>
> 如果 SSM 參數存放區參數與您啟動 AWS 區域 的任務位於相同的 中，則您可以使用參數的完整 ARN 或名稱。如果參數存在於不同區域，則必須指定完整 ARN。

類型：字串

必要：是，使用 `secrets` 時。

`ulimits`

容器中要設定的 `ulimits` 值的清單。此參數會映射到 Docker Remote API 的建立容器區段中的 `Ulimits` 以及 docker run 的 `--ulimit` 選項。

```
"ulimits": [
  {
    "name": string,
    "softLimit": integer,
    "hardLimit": integer
```

```
  }
  ...
]
```

類型：物件陣列

必要：否

name

ulimit 的 type。

類型：字串

必要：是，使用 ulimits 時。

hardLimit

ulimit 類型的硬性限制。

類型：整數

必要：是，使用 ulimits 時。

softLimit

ulimit 類型的軟性限制。

類型：整數

必要：是，使用 ulimits 時。

user

要在容器內使用的使用者名稱。此參數會映射到 Docker Remote API 的建立容器區段中的 User 以及 docker run 的 --user 選項。

```
"user": "string"
```

類型：字串

必要：否

vcpus

此參數已棄用，請*resourceRequirements*改用 。

為容器保留的 vCPU 數量。

如果您的任務定義包含類似以下的行resourceRequirements，則做為如何使用 的範例：

```
"containerProperties": {
  "vcpus": 2
}
```

使用 resourceRequirements的同等行如下所示。

```
"containerProperties": {
  "resourceRequirements": [
    {
      "type": "VCPU",
      "value": "2"
    }
  ]
}
```

類型：整數

必要：是

volumes

註冊任務定義時，您可指定磁碟區清單，那些磁碟區會傳送到容器執行個體上的 Docker 協助程式。容器屬性允許使用以下參數：

```
"volumes": [
  {
    "name": "string",
    "host": {
      "sourcePath": "string"
    },
    "efsVolumeConfiguration": {
      "authorizationConfig": {
        "accessPointId": "string",
        "iam": "string"
      },
      "fileSystemId": "string",
      "rootDirectory": "string",
      "transitEncryption": "string",
      "transitEncryptionPort": number
```

```
    }
  }
]
```

`name`

磁碟區名稱。可以包含最多可達 255 個字元 (大小寫)、數字、連字號和底線。此名稱是參考容器定義 `sourceVolume` 中的 `mountPoints` 參數。

類型：字串

必要：否

`host`

`host` 參數內容決定資料磁碟區是否在主機容器執行個體和儲存位置中保留。如果 `host` 參數是空的，則 Docker 協助程式會為您的資料磁碟區指派主機路徑。不過，資料不保證會在與其相關聯的容器停止執行後持續存在。

> **Note**
>
> 此參數不適用於在 Fargate 資源上執行的任務。

類型：物件

必要：否

`sourcePath`

提供給容器的主機容器執行個體上的路徑。如果此參數是空的，則 Docker 常駐程式會為您指派主機路徑。

如果 `host` 參數包含 `sourcePath` 檔案位置，資料磁碟區將保留在主機容器執行個體上的指定位置，直到您手動將其刪除為止。如果 `sourcePath` 值不存在於主機容器執行個體上，Docker 常駐程式將建立該值。如果位置存在，將匯出來源路徑資料夾的內容。

類型：字串

必要：否

`efsVolumeConfiguration`

當您使用適用於任務儲存體的 Amazon Elastic File System 檔案系統時，會指定此參數。如需詳細資訊，請參閱Amazon EFS 磁碟區。

類型：物件

必要：否

`authorizationConfig`

Amazon EFS 檔案系統的授權組態詳細資訊。

類型：字串

必要：否

`accessPointId`

要使用的 Amazon EFS 存取點 ID。如果指定存取點，則必須省略 中指定的根目錄值`EFSVolumeConfiguration`，或將 設定為 `/`。這會強制執行在 EFS 存取點上設定的路徑。如果使用存取點，則必須在 `EFSVolumeConfiguration` 中啟用傳輸加密。如需詳細資訊，請參閱《Amazon Elastic File System 使用者指南》中的使用 Amazon EFS 存取點。

類型：字串

必要：否

`iam`

決定是否在掛載 Amazon EFS 檔案系統時使用 AWS Batch 任務定義中定義的任務 IAM 角色。如果已啟用，必須在 `EFSVolumeConfiguration` 中啟用傳輸加密。如果省略此參數，系統會使用 `DISABLED` 的預設值。如需詳細資訊，請參閱使用 Amazon EFS 存取點。

類型：字串

有效值：`ENABLED` | `DISABLED`

必要：否

`fileSystemId`

要使用的 Amazon EFS 檔案系統識別碼。

類型：字串

必要：否

`rootDirectory`

在 Amazon EFS 檔案系統中的目錄，其將掛載作為主機內的根目錄。如果省略此參數，使用 Amazon EFS 磁碟區的根目錄。如果您指定 `/`，它具有與省略此參數相同的效果。長度上限為 4,096 個字元。

> **Important**
>
> 如果在 中指定了 EFS 存取點`authorizationConfig`，則必須省略根目錄參數或將其設定為 `/`。這會強制執行在 Amazon EFS 存取點上設定的路徑。

類型：字串

必要：否

`transitEncryption`

確定是否要對 Amazon ECS 主機和 Amazon EFS 伺服器之間 Amazon EFS 傳輸中的資料啟用加密功能。若使用 Amazon EFS IAM 授權，則必須啟用傳輸加密。如果省略此參數，系統會使用 `DISABLED` 的預設值。如需詳細資訊，請參閱《Amazon Elastic File System 使用者指南》中的加密傳輸中的資料。

類型：字串

有效值：`ENABLED` | `DISABLED`

必要：否

`transitEncryptionPort`

在 Amazon ECS 主機和 Amazon EFS 伺服器之間傳送加密資料時所使用的連接埠。如果您未指定傳輸加密連接埠，它會使用 Amazon EFS 掛載協助程式使用的連接埠選擇策略。該值必須介於 0 到 65,535 之間。如需詳細資訊，請參閱《Amazon Elastic File System 使用者指南》中的 EFS 掛載協助程式。

類型：整數

必要：否

Amazon EKS 屬性

有各種 Amazon ECS 型任務特定屬性的物件。這不得為 Amazon ECS 型任務定義指定。

`podProperties`

任務 Kubernetes Pod 資源的屬性。

類型：EksPodProperties 物件

必要：否

`containers`

Amazon EKS Pod 上所使用容器的屬性。

類型：EksContainer 物件

必要：否

`args`

進入點的引數陣列。如果未指定，系統會使用容器映像的 CMD。這對應至 `args` 中 Pod 進入點部分的成員Kubernetes。環境變數參考使用容器的環境擴展。

如果沒有參考的環境變數，不會變更命令中的參考。例如，如果參考為 "$(NAME1)"，且沒有 NAME1 環境變數，命令字串會保持 "$(NAME1)"。$$ 會替換為 $，且產生的字串不會擴展。例如，$$(VAR_NAME) 會以 $(VAR_NAME) 傳遞，無論是否有 VAR_NAME 環境變數。如需詳細資訊，請參閱 Dockerfile 參考中的 CMD，以及 Kubernetes 文件中的定義 Pod 的命令和引數。

類型：字串陣列

必要：否

`command`

容器的進入點。這不是在 Shell 中執行。如果未指定，系統會使用容器映像的 ENTRYPOINT。環境變數參考使用容器的環境擴展。

如果沒有參考的環境變數，不會變更命令中的參考。例如，如果參考為 "$(NAME1)"，且沒有 NAME1 環境變數，命令字串會保持 "$(NAME1)"。$$ 會替換為 $，且產生的字串不會擴展。例如，$$(VAR_NAME) 會以 $(VAR_NAME) 傳遞，無論是否有 VAR_NAME 環境變數。無法更新進入點。如需詳細資訊，請參閱 Dockerfile 參考中的 ENTRYPOINT，以及 Kubernetes 文件中的定義容器和進入點的命令和引數。 https://kubernetes.io/docs/reference/kubernetes-api/workload-resources/pod-v1/#entrypoint

類型：字串陣列

必要：否

env

傳遞至容器的環境變數。

> **Note**
>
> 環境變數不得以 "AWS_BATCH" 開頭。此命名慣例會保留給 AWS Batch 設定的變數。

類型：EksContainerEnvironmentVariable 物件的陣列

必要：否

name

環境變數的名稱。

類型：字串

必要：是

value

環境變數的值。

類型：字串

必要：否

image

用來啟動容器的 Docker 映像檔。

類型：字串

必要：是

imagePullPolicy

容器的映像提取政策。支援的值為 Always、IfNotPresent 和 Never。此參數預設為 IfNotPresent。但如果指定 :latest 標籤，預設為 Always。如需詳細資訊，請參閱 Kubernetes 文件中的更新映像。

類型：字串

必要：否

name

容器的名稱。如果未指定名稱，系統會使用預設名稱 "Default"。Pod 中的每個容器都必須有唯一名稱。

類型：字串

必要：否

resources

指派給容器的資源類型和數量。支援的資源包括 memory cpu 和 nvidia.com/gpu。如需詳細資訊，請參閱 Kubernetes 文件中的 Pod 和容器的資源管理。

類型：EksContainerResourceRequirements 物件

必要：否

limits

為容器預留的資源類型和數量。這些值根據指定的 name 而有所差異。可以使用 limits 或 requests 物件請求資源。

memory

容器的記憶體硬性限制 (以 MiB 為單位)，使用整數，具有 "Mi" 字尾。如果您的容器嘗試使用超過指定的記憶體，容器便會終止。您必須為任務指定至少 4 MiB 的記憶體。可以在 limits、requests 或兩者中指定 memory。如果同時在這兩個位置指定 memory，則 limits 中指定的值必須等於 requests 中指定的值。

> **Note**
>
> 若要將資源使用率最大化，請為您正在使用的特定執行個體類型的任務，提供盡可能多的記憶體。如要瞭解如何作業，請參閱運算資源記憶體管理。

cpu

為容器預留的 CPU 數量。值必須是 0.25 的偶數倍數。可以在 limits、requests 或兩者中指定 cpu。如果同時在這兩個位置指定 cpu，則 limits 中指定的值至少須與 requests 中指定的值一樣大。

nvidia.com/gpu

為容器預留的 GPU 數量。值必須為整數。可以在 `limits`、`requests` 或兩者中指定 `memory`。如果同時在這兩個位置指定 `memory`，則 `limits` 中指定的值必須等於 `requests` 中指定的值。

類型：字串到字串映射

值長度限制：長度下限為 1。長度上限為 256。

必要：否

`requests`

為容器請求的資源類型和數量。這些值根據指定的 `name` 而有所差異。可以使用 `limits` 或 `requests` 物件請求資源。

memory

容器的記憶體硬性限制 (以 MiB 為單位)，使用整數，具有 "Mi" 字尾。如果您的容器嘗試使用超過指定的記憶體，容器便會終止。您必須為任務指定至少 4 MiB 的記憶體。可以在 `limits`、`requests` 或兩者中指定 `memory`。如果同時在兩者中指定 `memory`，則 `limits` 中指定的值必須等於 `requests` 中指定的值。

> **Note**
>
> 如果您嘗試為特定執行個體類型提供盡可能多的記憶體，以最大化資源使用率，請參閱 運算資源記憶體管理。

cpu

為容器預留的 CPU 數量。值必須是 `0.25` 的偶數倍數。可以在 `limits`、`requests` 或兩者中指定 `cpu`。如果同時在兩者中指定 `cpu`，則 `limits` 中指定的值至少須與 `requests` 中指定的值一樣大。

nvidia.com/gpu

為容器預留的 GPU 數量。值必須為整數。可以在 `limits`、`requests` 或兩者中指定 `nvidia.com/gpu`。如果同時在兩者中指定 `nvidia.com/gpu`，則 `limits` 中指定的值必須等於 `requests` 中指定的值。

類型：字串到字串映射

值長度限制：長度下限為 1。長度上限為 256。

必要：否

`securityContext`

任務的安全性內容。如需詳細資訊，請參閱 Kubernetes 文件中的設定 Pod 或容器的安全內容。

類型：EksContainerSecurityContext 物件

必要：否

`privileged`

此參數為 `true` 時，容器便會取得主機容器執行個體的更高許可。許可層級類似於`root`使用者許可。預設值為 `false`。此參數會映射至 Kubernetes 文件`privileged`中特權 Pod 安全政策中的政策。

類型：布林值

必要：否

`readOnlyRootFilesystem`

此參數為 `true` 時，容器會取得根檔案系統的唯讀存取權。預設值為 `false`。此參數會映射至 Kubernetes 文件`ReadOnlyRootFilesystem`中磁碟區和檔案系統 Pod 安全政策中的政策。

類型：布林值

必要：否

`runAsGroup`

指定此參數時，容器會以指定的群組 ID (`gid`) 執行。如果未指定此參數，預設值為映像中繼資料中指定的群組。此參數會映射至 Kubernetes 文件中使用者和群組 Pod 安全`RunAsGroupMustRunAs`政策中的 和 政策。 https://kubernetes.io/docs/concepts/security/pod-security-policy/#users-and-groups

類型：Long

必要：否

`runAsNonRoot`

指定此參數時，容器會以 uid 非 0 的使用者身分執行。如果未指定此參數，系統會強制執行此規則。此參數會映射至 Kubernetes 文件中使用者和群組 Pod 安全RunAsUserMustRunAsNonRoot政策中的 和 政策。 https://kubernetes.io/docs/concepts/security/pod-security-policy/#users-and-groups

類型：Long

必要：否

`runAsUser`

指定此參數時，容器會以指定的使用者 ID (uid) 執行。如果未指定此參數，預設值為映像中繼資料中指定的使用者。此參數會映射至 Kubernetes 文件中使用者和群組 Pod 安全RunAsUserMustRanAs政策中的 和 政策。 https://kubernetes.io/docs/concepts/security/pod-security-policy/#users-and-groups

類型：Long

必要：否

`volumeMounts`

磁碟區會針對適用於 Amazon EKS 任務的容器掛載。如需 中磁碟區和磁碟區掛載的詳細資訊Kubernetes，請參閱 Kubernetes 文件中的磁碟區。

類型：EksContainerVolumeMount 物件陣列

必要：否

`mountPath`

掛載磁碟區之容器上的路徑。

類型：字串

必要：否

`name`

掛載的磁碟區名稱。這必須符合 Pod 中任一磁碟區的名稱。

類型：字串

必要：否

`readOnly`

如果此數值為 `true`，容器擁有磁碟區的唯讀存取權。否則，容器可以寫入磁碟區。預設值為 `false`。

類型：布林值

必要：否

`dnsPolicy`

Pod 的 DNS 政策。預設值為 `ClusterFirst`。如果未指定 `hostNetwork` 參數，預設值為 `ClusterFirstWithHostNet`。`ClusterFirst` 指示任何與設定之叢集網域字尾不相符的 DNS 查詢，都會轉寄至繼承自節點的上游名稱伺服器。如果 RegisterJobDefinition API 操作`dnsPolicy`中未指定 的值，則 `dnsPolicy` DescribeJobDefinitions 或 DescribeJobs API 操作不會傳回 的值。視 `hostNetwork` 參數的值而定，Pod 規格設定會包含 `ClusterFirst` 或 `ClusterFirstWithHostNet`。如需詳細資訊，請參閱 Kubernetes 文件中的 Pod 的 DNS 政策。

有效值：`Default` | `ClusterFirst` | `ClusterFirstWithHostNet`

類型：字串

必要：否

`hostNetwork`

指示 Pod 是否使用主機的網路 IP 地址。預設值為 `true`。將此設定為`false`啟用 Kubernetes Pod 網路模型。大多數 AWS Batch 工作負載都是輸出限定的，不需要每個 Pod 傳入連線的 IP 配置額外負荷。如需詳細資訊，請參閱 Kubernetes 文件中的主機命名空間和 Pod 聯網。

類型：布林值

必要：否

`serviceAccountName`

用來執行 Pod 的服務帳戶名稱。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的Kubernetes服務帳戶和設定Kubernetes服務帳戶以擔任 IAM 角色和在Kubernetes文件中設定 Pod 的服務帳戶。

類型：字串

必要：否

volumes

為使用 Amazon EKS 資源的任務定義指定磁碟區。

類型：EksVolume 物件的陣列

必要：否

emptyDir

指定Kubernetes`emptyDir`磁碟區的組態。將 Pod 指派給節點時，會先建立 `emptyDir` 磁碟區。只要該 Pod 在該節點上執行，就會存在。`emptyDir` 磁碟區最初是空的。Pod 中的所有容器都可以讀取和寫入 `emptyDir` 磁碟區中的檔案。但 `emptyDir` 磁碟區可以掛載在每個容器中相同或不同路徑上。因任何原因將 Pod 從節點移除時，系統會永久刪除 `emptyDir` 中的資料。如需詳細資訊，請參閱 Kubernetes 文件中的 emptyDir。

類型：EksEmptyDir 物件

必要：否

中型

存放磁碟區的媒體。預設值為空白字串，這會使用節點的儲存空間。

""

(預設) 使用節點的磁碟儲存空間。

"Memory"

使用節點 RAM 支援的 `tmpfs` 磁碟區。節點重新開機時，磁碟區的內容會遺失，且磁碟區上的任何儲存空間都會計入容器的記憶體限制。

類型：字串

必要：否

sizeLimit

磁碟區的大小上限。預設未定義大小上限。

類型：字串

長度限制：長度下限為 1。長度上限為 256。

必要：否

hostPath

指定Kubernetes`hostPath`磁碟區的組態。`hostPath` 磁碟區會將現有檔案或目錄，從主機節點的檔案系統掛載到您的 Pod。如需詳細資訊，請參閱 Kubernetes 文件中的 hostPath。

類型：EksHostPath 物件

必要：否

path

要掛載至 Pod 上容器的主機檔案或目錄的路徑。

類型：字串

必要：否

name

磁碟區名稱。此名稱必須可當作 DNS 子網域名稱。如需詳細資訊，請參閱 Kubernetes 文件中的 DNS 子網域名稱。

類型：字串

必要：是

秘密

指定Kubernetes`secret`磁碟區的組態。如需詳細資訊，請參閱 Kubernetes 文件中的秘密。

類型：EksSecret 物件

必要：否

選擇性

指定是否必須定義秘密或秘密的金鑰。

類型：布林值

必要：否

secretName

秘密的名稱。此名稱必須可當作 DNS 子網域名稱。如需詳細資訊，請參閱 Kubernetes 文件中的 DNS 子網域名稱。

類型：字串

必要：是

平台功能

`platformCapabilities`

任務定義所需的平台功能。如果未指定任何值，則預設為 EC2。對於在 Fargate 資源上執行的任務，FARGATE會指定 。

> **Note**
>
> 如果任務在 Amazon EKS 資源上執行，則不得指定 `platformCapabilities`。

類型：字串

有效值：EC2 | FARGATE

必要：否

傳播標籤

`propagateTags`

指定是否要將標籤從任務或任務定義傳播到對應的 Amazon ECS 任務。如果沒有指定值，則不會傳播標籤。標籤只能在建立任務時傳播至任務。對於具有相同名稱的標籤，任務標籤優先於任務定義標籤。如果任務和任務定義的合併標籤總數超過 50，任務會移至 FAILED 狀態。

> **Note**
>
> 如果任務在 Amazon EKS 資源上執行，則您不得指定 `propagateTags`。

類型：布林值

必要：否

節點屬性

`nodeProperties`

註冊多節點平行任務定義時，您必須指定節點屬性的清單。這些節點屬性會定義要在任務中使用的節點數量、主節點索引，以及要使用的不同節點範圍。如果任務在 Fargate 資源上執行，則您無法指定 `nodeProperties`。請改用 `containerProperties`。任務定義允許使用以下的節點屬性。如需詳細資訊，請參閱多節點平行任務。

> **Note**
>
> 如果任務在 Amazon EKS 資源上執行，則不得指定 `nodeProperties`。

類型：NodeProperties 物件

必要：否

`mainNode`

為多節點平行任務指定主要節點的節點索引。此節點索引值必須小於節點的數量。

類型：整數

必要：是

`numNodes`

與多節點平行任務關聯的節點數量。

類型：整數

必要：是

`nodeRangeProperties`

與多節點平行任務關聯的節點範圍和其屬性清單。

> **Note**
>
> 節點群組是共用相同容器屬性的相同任務節點群組。您可以使用 AWS Batch 為每個任務指定最多五個不同的節點群組。

類型：NodeRangeProperty 物件陣列

必要：是

`targetNodes`

使用節點索引值的節點範圍。0:3 的範圍表示節點具有 0 到 3 的索引值。如果省略開始範圍值 (:n)，則會使用 0 來開始範圍。如果省略了結束範圍值 (n:)，則會使用最高可能的節點索引來結束範圍。您的累積節點範圍必須將所有節點納入考量 (0:n)。您可以巢狀節點範圍，例如 0:10和 4:5。在這種情況下，4:5範圍屬性會覆寫0:10屬性。

類型：字串

必要：否

`container`

節點範圍的容器詳細資訊。如需詳細資訊，請參閱容器屬性。

類型：ContainerProperties 物件

必要：否

重試策略

`retryStrategy`

註冊任務定義，您可以選擇性指定任務失敗後要使用的重試策略，隨此任務定義提交。在 SubmitJob 操作期間指定的任何重試策略都會覆寫此處定義的重試策略。根據預設，每個任務將嘗試一次。如果您指定多個嘗試，則會在任務失敗時重試任務。失敗嘗試的範例包括任務傳回非零結束碼，或容器執行個體終止。如需詳細資訊，請參閱自動化任務重試。

類型：RetryStrategy 物件

必要：否

`attempts`

將任務移至 RUNNABLE 狀態的次數。您可以指定嘗試 1 至 10 次。如果 attempts 超過 1 次，任務失敗後將重試該次數，直到其狀態移至 RUNNABLE。

```
"attempts": integer
```

類型：整數

必要：否

`evaluateOnExit`

最多 5 個物件的陣列，指定任務重試或失敗的條件。如果指定此參數，則也必須指定 `attempts` 參數。如果已指定 `evaluateOnExit` ，但沒有項目相符，則會重試任務。

```
"evaluateOnExit": [
   {
      "action": "string",
      "onExitCode": "string",
      "onReason": "string",
      "onStatusReason": "string"
   }
]
```

類型： EvaluateOnExit 物件的陣列

必要：否

`action`

指定符合所有指定條件 (`onStatusReason` `onReason` 和 `onExitCode`) 時要採取的動作。這些值不區分大小寫。

類型：字串

必要：是

有效值：`RETRY` | `EXIT`

`onExitCode`

包含 glob 模式，以比對針對任務`ExitCode`傳回之 的小數表示法。模式的長度上限為 512 個字元。它只能包含數字。它不能包含字母或特殊字元。可以選擇以星號 (*) 結束，以便只有字串的開頭需要完全相符。

類型：字串

必要：否

onReason

包含 glob 模式，以比對為任務傳回Reason的 。模式的長度上限為 512 個字元。它可以包含字母、數字、句點 (.)、冒號 (:) 和空格 （ 空格、標籤 ）。可以選擇以星號 (*) 結束，以便只有字串的開頭需要完全相符。

類型：字串

必要：否

onStatusReason

包含 glob 模式，以比對為任務傳回StatusReason的 。模式的長度上限為 512 個字元。它可以包含字母、數字、句點 (.)、冒號 (:) 和空格 （ 空格、標籤 ）。可以選擇以星號 (*) 結束，以便只有字串的開頭需要完全相符。

類型：字串

必要：否

排程優先順序

schedulingPriority

與此任務定義一起提交之任務的排程優先順序。這只會影響具有公平共用政策的任務佇列中的任務。排程優先順序較高的任務會排在排程優先順序較低的任務之前。

支援的最小值為 0，支援的最大值為 9999。

類型：整數

必要：否

Tags (標籤)

tags

要與任務定義建立關聯的鍵值對標籤。如需詳細資訊，請參閱標記您的 AWS Batch 資源。

類型：字串到字串映射

必要：否

逾時

`timeout`

您可以設定任務的逾時持續時間，以便在任務執行時間超過此時間時 AWS Batch 終止任務。如需詳細資訊，請參閱任務逾時。如果任務因逾時而終止，則不會重試。SubmitJob 操作期間指定的任何逾時組態都會覆寫此處定義的逾時組態。如需詳細資訊，請參閱任務逾時。

類型：JobTimeout 物件

必要：否

`attemptDurationSeconds`

AWS Batch 終止未完成任務後的持續時間，以秒為單位（從任務嘗試的`startedAt`時間戳記測量）。逾時最小值為 60 秒。

若為陣列任務，逾時會套用至子任務，而不是父陣列任務。

若為多節點平行 (MNP) 任務，逾時會套用至整個工作，而非個別節點。

類型：整數

必要：否

使用 EcsProperties 建立任務定義

透過使用 AWS Batch 的任務定義`EcsProperties`，您可以在個別容器中建立硬體、感應器、3D 環境和其他模擬的模型。您可以使用此功能以邏輯方式組織工作負載元件，並將其與主要應用程式分開。此功能可在 Amazon Elastic Container Service (Amazon ECS)、Amazon Elastic Kubernetes Service (Amazon EKS) 和 AWS Batch 上使用 AWS Fargate。

`ContainerProperties` 與`EcsProperties`任務定義

您可以選擇使用 `ContainerProperties`或 `EcsProperties`任務定義做為您的使用案例指示。在高階，使用 執行 AWS Batch 中的任務`EcsProperties`類似於使用 執行中的任務`ContainerProperties`。

使用 的舊版任務定義結構`ContainerProperties`仍受支援。如果您目前有使用此結構的工作流程，您可以繼續執行它們。

主要差別在於有新物件新增至任務定義，以容納以 `EcsProperties`為基礎的定義。

例如，在 Amazon ECS 和 Fargate ContainerProperties上使用 的任務定義具有下列結構：

```
{
   "containerProperties": {
      ...
      "image": "my_ecr_image1",
      ...
  },
...
}
```

在 Amazon ECS 和 Fargate EcsProperties上使用的任務定義具有下列結構：

```
{
  "ecsProperties": {
    "taskProperties": [{
      "containers": [
        {
          ...
          "image": "my_ecr_image1",
          ...
        },
        {
          ...
          "image": "my_ecr_image2",
          ...
        },
```

AWS Batch APIs的一般變更

以下進一步概述使用 ContainerProperties和 EcsProperties API 資料類型時的一些主要差異：

- 在 中使用的許多參數都ContainerProperties會顯示在 中TaskContainerProperties。一些範例包括 、commandimage、secrets、 privileged和 users。它們都可以在 TaskContainerProperties 中找到。
- 某些TaskContainerProperties參數在舊版結構中沒有功能同等項目。一些範例包括 、dependsOnessential、name、 ipcMode和 pidMode。如需詳細資訊，請參閱 EcsTaskDetails 和 TaskContainerProperties。

  此外，某些ContainerProperties參數在 EcsProperties 結構中沒有對等參數或應用程式。在 中taskProperties， container 已取代為 ，containers因此新物件最多可接受 10 個元

素。如需詳細資訊，請參閱 RegisterJobDefinition：containerProperties 和 EcsTaskProperties：containers。

- taskRoleArn 在功能上等同於 jobRoleArn。如需詳細資訊，請參閱 EcsTaskProperties：taskRoleArn 和 ContainerProperties：jobRoleArn。
- 您可以在 EcsProperties結構中包含一 (1) 到十 (10) 個容器。如需詳細資訊，請參閱 EcsTaskProperties：containers。
- taskProperties 和 instanceTypes 物件是陣列，但目前只接受一個元素。例如，EcsProperties：taskProperties 和 NodeRangeProperty：instanceTypes。

Amazon ECS 的多容器任務定義

為了容納 Amazon ECS 的多容器結構，某些 API 資料類型不同。例如

- ecsProperties 與單一容器定義containerProperties中的層級相同。如需詳細資訊，請參閱《 AWS Batch API 參考指南》中的 EcsProperties。
- taskProperties 包含為 Amazon ECS 任務定義的屬性。如需詳細資訊，請參閱《 AWS Batch API 參考指南》中的 EcsProperties。
- containers 在單一容器定義containerProperties中包含與 類似的資訊。主要差別在於containers可讓您定義最多十個容器。如需詳細資訊，請參閱AWS Batch 《 API 參考指南》中的 ECSTaskProperties：containers。
- essential 參數指出容器如何影響任務。所有基本容器都必須成功完成（以 0 結束），任務才能繼續。如果標示為基本的容器失敗（結束為非 0)，則任務會失敗。

  預設值為 true，且至少有一個容器必須標記為 essential。如需詳細資訊，請參閱 essential API 參考指南中的「AWS Batch 」。
- 使用 dependsOn 參數，您可以定義容器相依性的清單。如需詳細資訊，請參閱 dependsOn API 參考指南中的「AWS Batch 」。

> **Note**
>
> dependsOn 清單的複雜性和相關聯的容器執行時間可能會影響任務的開始時間。如果相依性需要很長時間才能執行，任務會保持 STARTING 狀態，直到完成為止。

如需 ecsProperties和 結構的詳細資訊，請參閱 ecsProperties 的 RegisterJobDefinition 請求語法。

Amazon EKS 的多容器任務定義

為了容納 Amazon EKS 的多容器結構，某些 API 資料類型不同。例如

- `name` 是容器的唯一識別符。單一容器不需要此物件，但在 Pod 中定義多個容器時需要此物件。如果name未針對單一容器定義 ，則會default套用預設名稱 。
- `initContainers` 在 eksPodProperties 資料類型中定義。它們會在應用程式容器之前執行，一律會執行到完成，而且必須在下一個容器啟動之前成功完成。

  這些容器已向 Amazon EKS Connector 代理程式註冊，並在 Amazon Elastic Kubernetes Service 後端資料存放區中保留註冊資訊。initContainers 物件最多可接受十 (10) 個元素。如需詳細資訊，請參閱 Kubernetes 文件中的初始化容器。

  > **Note**
  >
  > initContainers 物件可能會影響任務的開始時間。如果 initContainers需要很長時間才能執行，任務將保持 STARTING 狀態，直到完成為止。

- `shareProcessNamespace` 指出 Pod 中的容器是否可以共用相同的程序命名空間。預設值為 false。將此設定為 true，以啟用容器查看位於相同 Pod 中其他容器中的程序並發出訊號。
- 每個容器都很重要。所有容器都必須成功完成（以 0 結束），任務才能成功。如果一個容器失敗（除了 0 之外結束），則任務會失敗。

如需 eksProperties和 結構的詳細資訊，請參閱 eksProperties 的 RegisterJobDefinition 請求語法。

使用 EcsProperties 的 Reference： AWS Batch job 案例

為了說明如何根據您的需求EcsProperties來建構 使用 AWS Batch 的任務定義，本主題提供下列RegisterJobDefinition承載。您可以將這些範例複製到檔案中，根據您的需求自訂範例，然後使用 AWS Command Line Interface (AWS CLI) 呼叫 RegisterJobDefinition。

AWS Batch Amazon EC2 上的 Amazon ECS 任務

以下是 Amazon Elastic Compute Cloud 上的 Amazon Elastic Container Service AWS Batch 任務範例：

```
{
    "jobDefinitionName": "multicontainer-ecs-ec2",
```

```
    "type": "container",
    "ecsProperties": {
        "taskProperties": [
          {
            "containers": [
              {
                "name": "c1",
                "essential": false,
                "command": [
                  "echo",
                  "hello world"
                ],
                "image": "public.ecr.aws/amazonlinux/amazonlinux:latest",
                "resourceRequirements": [
                  {
                    "type": "VCPU",
                    "value": "2"
                  },
                  {
                    "type": "MEMORY",
                    "value": "4096"
                  }
                ]
              },
              {
                "name": "c2",
                "essential": false,
                "command": [
                  "echo",
                  "hello world"
                ],
                "image": "public.ecr.aws/amazonlinux/amazonlinux:latest",
                "resourceRequirements": [
                  {
                    "type": "VCPU",
                    "value": "2"
                  },
                  {
                    "type": "MEMORY",
                    "value": "4096"
                  }
                ]
              },
              {
```

```
                "name": "c3",
                "essential": true,
                "command": [
                  "echo",
                  "hello world"
                ],
                "image": "public.ecr.aws/amazonlinux/amazonlinux:latest",
                "firelensConfiguration": {
                  "type": "fluentbit",
                  "options": {
                    "enable-ecs-log-metadata": "true"
                  }
                 },
                "resourceRequirements": [
                  {
                    "type": "VCPU",
                    "value": "6"
                  },
                  {
                    "type": "MEMORY",
                    "value": "12288"
                  }
                ]
              }
            ]
          }
        ]
  }
}
```

AWS Batch Fargate 上的 Amazon ECS 任務

以下是 上的 Amazon Elastic Container Service AWS Batch 任務範例 AWS Fargate：

```
{
    "jobDefinitionName": "multicontainer-ecs-fargate",
    "type": "container",
    "platformCapabilities": [
        "FARGATE"
    ],
    "ecsProperties": {
        "taskProperties": [
          {
```

```
            "containers": [
              {
                "name": "c1",
                "command": [
                  "echo",
                  "hello world"
                ],
                "image": "public.ecr.aws/amazonlinux/amazonlinux:latest",
                "resourceRequirements": [
                  {
                    "type": "VCPU",
                    "value": "2"
                  },
                  {
                    "type": "MEMORY",
                    "value": "4096"
                  }
                ]
              },
              {
                "name": "c2",
                "essential": true,
                "command": [
                  "echo",
                  "hello world"
                ],
                "image": "public.ecr.aws/amazonlinux/amazonlinux:latest",
                "resourceRequirements": [
                  {
                    "type": "VCPU",
                    "value": "6"
                  },
                  {
                    "type": "MEMORY",
                    "value": "12288"
                  }
                ]
              }
            ],
            "executionRoleArn": "arn:aws:iam::1112223333:role/ecsTaskExecutionRole"
          }
        ]
}
```

```
}
```

AWS Batch Amazon EKS 的任務

以下是 Amazon Elastic Kubernetes Service AWS Batch 任務的範例：

```
{
  "jobDefinitionName": "multicontainer-eks",
  "type": "container",
  "eksProperties": {
    "podProperties": {
      "shareProcessNamespace": true,
      "initContainers": [
        {
          "name": "init-container",
          "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
          "command": [
            "echo"
          ],
          "args": [
            "hello world"
          ],
          "resources": {
            "requests": {
              "cpu": "1",
              "memory": "512Mi"
            }
          }
        },
        {
          "name": "init-container-2",
          "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
          "command": [
            "echo",
            "my second init container"
          ],
          "resources": {
            "requests": {
              "cpu": "1",
              "memory": "512Mi"
            }
          }
        }
```

```
      ],
      "containers": [
        {
          "name": "c1",
          "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
          "command": [
            "echo world"
         ],
          "resources": {
            "requests": {
              "cpu": "1",
              "memory": "512Mi"
            }
          }
        },
        {
          "name": "sleep-container",
          "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
          "command": [
            "sleep",
            "20"
          ],
          "resources": {
            "requests": {
              "cpu": "1",
              "memory": "512Mi"
            }
          }
        }
      ]
    }
  }
}
```

每個節點具有多個容器的 MNP AWS Batch 任務

以下是每個節點具有多個容器的多節點平行 (MNP) AWS Batch 任務範例：

```
{
  "jobDefinitionName": "multicontainer-mnp",
  "type": "multinode",
  "nodeProperties": {
    "numNodes": 6,
```

```
    "mainNode": 0,
    "nodeRangeProperties": [
      {
        "targetNodes": "0:5",
        "ecsProperties": {
          "taskProperties": [
            {
              "containers": [
                {
                  "name": "range05-c1",
                  "command": [
                    "echo",
                    "hello world"
                  ],
                  "image": "public.ecr.aws/amazonlinux/amazonlinux:latest",
                  "resourceRequirements": [
                    {
                      "type": "VCPU",
                      "value": "2"
                    },
                    {
                      "type": "MEMORY",
                      "value": "4096"
                    }
                  ]
                },
                {
                  "name": "range05-c2",
                  "command": [
                    "echo",
                    "hello world"
                  ],
                  "image": "public.ecr.aws/amazonlinux/amazonlinux:latest",
                  "resourceRequirements": [
                    {
                      "type": "VCPU",
                      "value": "2"
                    },
                    {
                      "type": "MEMORY",
                      "value": "4096"
                    }
                  ]
                }
```

```
              ]
            }
          ]
        }
      }
    ]
  }
}
```

使用 awslogs 日誌驅動程式

根據預設， 會 AWS Batch 啟用awslogs日誌驅動程式，將日誌資訊傳送至 CloudWatch Logs。您可以使用此功能，在一個方便的位置檢視與容器不同的日誌，並防止容器日誌佔用容器執行個體上的磁碟空間。本主題可協助您在任務定義中設定awslogs日誌驅動程式。

> **Note**
>
> 在 AWS Batch 主控台中，您可以在建立任務定義時，於記錄組態區段中設定awslogs日誌驅動程式。

> **Note**
>
> 您任務中的容器所記錄的資訊類型，主要取決於其ENTRYPOINT命令。根據預設，擷取的日誌會顯示您在本機執行容器時通常在互動式終端機中看到的命令輸出，也就是 STDOUT和 STDERR I/O 串流。awslogs 日誌驅動程式只會將這些日誌從 Docker 傳遞至 CloudWatch Logs。如需 Docker 日誌處理方式 (包括擷取不同檔案資料或串流的替代方法) 的詳細資訊，請參閱 Docker 文件中的檢視容器或服務的日誌。

若要將系統日誌從容器執行個體傳送至 CloudWatch Logs，請參閱 搭配 使用 CloudWatch Logs AWS Batch。如需 CloudWatch Logs 的詳細資訊，請參閱 《Amazon CloudWatch Logs 使用者指南》中的監控日誌檔案和 CloudWatch Logs 配額。

JobDefiniton 資料類型中的 AWS Batch awslogs 日誌驅動程式選項

awslogs 日誌驅動程式支援 AWS Batch 任務定義中的下列選項。如需詳細資訊，請參閱 Docker 文件中的 CloudWatch Logs 記錄驅動程式。

awslogs-region

必要：否

指定awslogs日誌驅動程式應傳送 Docker 日誌的區域。根據預設，使用的 區域與任務的區域相同。您可以選擇將所有日誌從不同區域中的任務傳送到 CloudWatch Logs 中的單一區域。這樣做可讓它們從一個位置全部可見。或者，您可以依區域分隔它們，以取得更精細的方法。不過，當您選擇此選項時，請確定指定的日誌群組存在於您指定的區域中。

awslogs-group

必要：選用

使用 awslogs-group選項，您可以指定awslogs日誌驅動程式傳送其日誌串流的目標日誌群組。如果未指定，aws/batch/job則會使用 。

awslogs-stream-prefix

必要：選用

使用 awslogs-stream-prefix選項，您可以將日誌串流與指定的字首，以及容器所屬 AWS Batch 任務的 Amazon ECS 任務 ID 建立關聯。如果您使用此選項指定前綴，則日誌串流會使用下列格式：

```
prefix-name/default/ecs-task-id
```

awslogs-datetime-format

必要：否

此選項會以 Python strftime 格式定義多行開始模式。日誌訊息包含符合模式的行，以及任何不符合模式的下列行。因此，符合的行是日誌訊息之間的分隔符號。

使用此格式的一個使用案例範例是用於剖析輸出，例如堆疊傾印，在其他情形下這可能會記錄在多個項目中。正確的模式可允許將它擷取在單一項目中。

如需詳細資訊，請參閱 awslogs-datetime-format。

如果 awslogs-datetime-format 和 awslogs-multiline-pattern 都設定，則一律以此選項優先。

> **Note**
> 多行記錄會執行常規表達式剖析並比對所有日誌訊息。這可能會對記錄效能造成負面影響。

`awslogs-multiline-pattern`

必要：否

此選項使用規則表達式來定義多行開始模式。日誌訊息包含符合模式的行，以及任何不符合模式的下列行。因此，相符的行是日誌訊息之間的分隔符號。

如需詳細資訊，請參閱 Docker 文件中的 awslogs-multiline-pattern。

如果同時設定 `awslogs-datetime-format`，會忽略此選項。

> **Note**
> 多行記錄會執行常規表達式剖析並比對所有日誌訊息。這可能會對記錄效能造成負面影響。

`awslogs-create-group`

必要：否

指定您是否希望自動建立日誌群組。若未指定此選項，則預設為 `false`。

> **Warning**
> 不建議使用此選項。我們建議您在每個任務嘗試建立日誌群組時，事先使用 CloudWatch Logs CreateLogGroup API 動作建立日誌群組，增加任務失敗的機會。

> **Note**
> 執行角色的 IAM 政策必須包含 `logs:CreateLogGroup`許可，才能嘗試使用 `awslogs-create-group`。

在任務定義中指定日誌組態

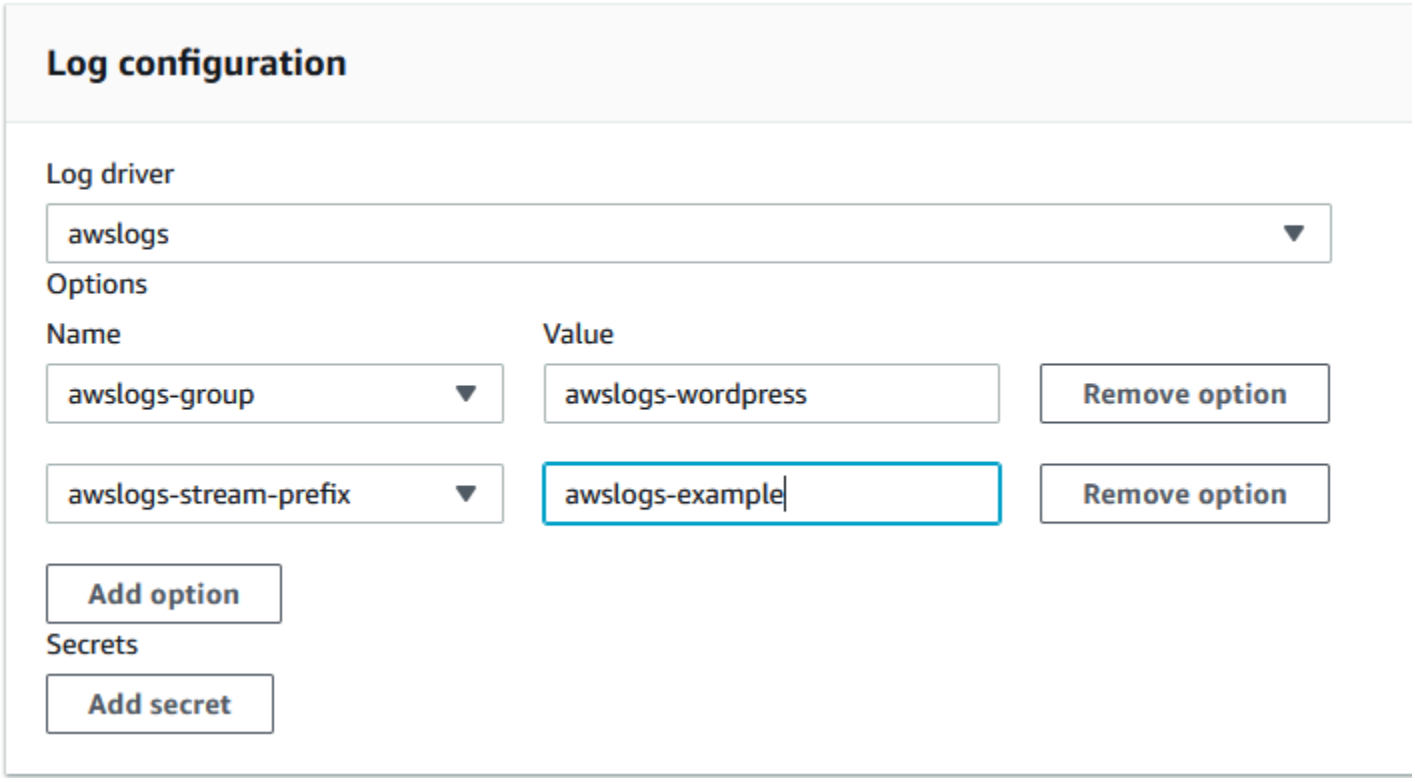
根據預設， 會 AWS Batch 啟用awslogs日誌驅動程式。本節說明如何自訂任務的awslogs日誌組態。如需詳細資訊，請參閱建立單一節點任務定義。

下列日誌組態 JSON 程式碼片段具有為每個任務指定的logConfiguration物件。一個用於將日誌傳送至名為 之日誌群組的 WordPress 任務awslogs-wordpress，另一個用於將日誌傳送至名為 之日誌群組的 MySQL 容器awslogs-mysql。兩個容器使用的日誌串流前綴皆為 awslogs-example。

```
"logConfiguration": {
    "logDriver": "awslogs",
    "options": {
        "awslogs-group": "awslogs-wordpress",
        "awslogs-stream-prefix": "awslogs-example"
    }
}
```

```
"logConfiguration": {
    "logDriver": "awslogs",
    "options": {
        "awslogs-group": "awslogs-mysql",
        "awslogs-stream-prefix": "awslogs-example"
    }
}
```

在 AWS Batch 主控台中，wordpress任務定義的日誌組態指定如下圖所示。



在任務定義日誌組態中向awslogs日誌驅動程式註冊任務定義後，您可以使用該任務定義提交任務，以開始將日誌傳送至 CloudWatch Logs。如需詳細資訊，請參閱教學課程：提交任務。

指定敏感資料

使用 時 AWS Batch，您可以將敏感資料存放在秘密或 AWS Systems Manager 參數存放區參數中 AWS Secrets Manager ，然後在任務定義中參考它們，藉此將敏感資料注入任務。

秘密可以透過以下方式公開至任務：

- 若要將敏感資料插入容器做為環境變數，請使用 secrets任務定義參數。
- 若要在任務的日誌組態中參考敏感資訊，請使用secretOptions任務定義參數。

主題

- 使用 Secrets Manager 指定敏感資料
- 使用 Systems Manager 參數存放區指定敏感資料

使用 Secrets Manager 指定敏感資料

透過 AWS Batch，您可以將敏感資料存放在 AWS Secrets Manager 秘密中，然後在任務定義中參考它們，藉此將敏感資料注入任務。存放在 Secrets Manager 秘密中的敏感資料可以做為環境變數或日誌組態的一部分公開給任務。

當您將秘密做為環境變數插入時，可以指定 JSON 金鑰或要插入的秘密版本。此程序可協助您控制向任務公開的敏感資料。如需有關秘密版本控制的詳細資訊，請參閱《AWS Secrets Manager 使用者指南》中的 AWS Secrets Manager重要術語和概念。

使用 Secrets Manager 指定敏感資料的考量事項

使用 Secrets Manager 為任務指定敏感資料時，應考慮下列事項。

- 若要使用特定 JSON 金鑰或版本的秘密注入秘密，運算環境中的容器執行個體必須安裝 1.37.0 版或更新版本的 Amazon ECS 容器代理程式。不過，我們建議您使用最新版的容器代理。如需有關檢查代理程式版本和更新至最新版本的資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的更新 Amazon ECS 容器代理程式。

  若要將秘密的完整內容插入為環境變數，或在日誌組態中插入秘密，您的容器執行個體必須具有 1.23.0 版或更新版本的容器代理程式。
- 僅支援存放文字資料的秘密，這些秘密是使用 CreateSecret API 的 `SecretString` 參數建立的。不支援存放二進位資料的秘密，這是使用 CreateSecret API `SecretBinary` 參數建立的秘密。
- 使用參考 Secrets Manager 秘密的任務定義來擷取任務的敏感資料時，如果您也使用介面 VPC 端點，則必須為 Secrets Manager 建立介面 VPC 端點。如需詳細資訊，請參閱 《AWS Secrets Manager 使用者指南》中的搭配使用 Secrets Manager 與 VPC 端點。
- 當任務一開始啟動時，會將敏感資料注入您的任務。如果隨後更新或輪換秘密，任務不會自動收到更新的值。您必須啟動新的任務，才能強制服務啟動具有更新秘密值的新任務。

AWS Batch 秘密所需的 IAM 許可

若要使用此功能，您必須擁有 執行角色，並在您的任務定義中參考它。這可讓容器代理程式提取必要的 Secrets Manager 資源。如需詳細資訊，請參閱AWS Batch IAM 執行角色。

若要讓您存取您建立的 Secrets Manager 秘密，請手動將下列許可新增為執行角色的內嵌政策。如需詳細資訊，請參閱 IAM User Guide 中的 Adding and Removing IAM Policies。

- `secretsmanager:GetSecretValue` - 如果您要參考 Secrets Manager 秘密，則需要此項目。

- `kms:Decrypt` - 只有在您的秘密使用自訂的 KMS 金鑰而非預設金鑰時，才需要此項目。您的自訂金鑰的 ARN 應該新增為資源。

下列內嵌政策範例新增必要許可。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetSecretValue",
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:secretsmanager:us-
east-2:777777777777:secret:<secret_name>",
                "arn:aws:kms:us-east-2:777777777777:key/<key_id>"
            ]
        }
    ]
}
```

插入敏感資料作為環境變數

您可以在任務定義中指定下列項目：

- 包含要在任務中設定之環境變數名稱的`secrets`物件
- Secrets Manager 秘密的 Amazon Resource Name (ARN)
- 包含要呈現給任務之敏感資料的其他參數

下列範例示範了必須為 Secrets Manager 秘密指定的完整語法。

```
arn:aws:secretsmanager:region:aws_account_id:secret:secret-name:json-key:version-
stage:version-id
```

以下部分說明其他參數。這些參數是選用的。不過，如果您不使用它們，則必須包含冒號:才能使用預設值。以下提供範例深入說明。

`json-key`

使用您要設為環境變數值的值，來指定金鑰/值對中的金鑰名稱。僅支援 JSON 格式的值。如果您沒有指定 JSON 金鑰，則會使用秘密的完整內容。

`version-stage`

指定您要使用之秘密版本的預備標籤。如果指定了版本預備標籤，就無法指定版本 ID。如果未指定版本階段，則預設會擷取具有 AWSCURRENT 階段標籤的秘密。

預備標籤會用來在不同版本的秘密更新或輪換時加以追蹤。每個版本的秘密都有一或多個預備標籤和 ID。如需詳細資訊，請參閱AWS Secrets Manager 《 使用者指南》中的 AWS Secrets Manager 的關鍵術語和概念。

`version-id`

針對您要使用的秘密版本，指定其唯一識別符。如果指定了版本 ID，就無法指定版本預備標籤。如果未指定版本 ID，則預設會擷取具有 AWSCURRENT 階段標籤的秘密。

版本 ID 會用來在不同版本的秘密更新或輪換時加以追蹤。每個版本的秘密都有 ID。如需詳細資訊，請參閱AWS Secrets Manager 《 使用者指南》中的 AWS Secrets Manager 的關鍵術語和概念。

容器定義範例

下列範例示範您可以在容器定義中參考 Secrets Manager 秘密的方法。

Example參考完整秘密

以下是任務定義的程式碼片段，顯示參考 Secrets Manager 秘密全文時的格式。

```
{
  "containerProperties": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:secret_name-
AbCdEf"
    }]
  }]
```

```
}
```

Example參考秘密中的特定金鑰

以下顯示來自 >get-secret-value 命令的範例輸出，其中顯示秘密的內容，以及與其相關聯的版本預備標籤和版本 ID。

```
{
    "ARN": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-AbCdEf",
    "Name": "appauthexample",
    "VersionId": "871d9eca-18aa-46a9-8785-981dd39ab30c",
    "SecretString": "{\"username1\":\"password1\",\"username2\":\"password2\",
\"username3\":\"password3\"}",
    "VersionStages": [
        "AWSCURRENT"
    ],
    "CreatedDate": 1581968848.921
}
```

在 ARN 結尾指定金鑰名稱，來在容器定義中參考上一個輸出的特定金鑰。

```
{
  "containerProperties": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-
AbCdEf:username1::"
    }]
  }]
}
```

Example參考特定秘密版本

以下顯示來自 >describe-secret 命令的範例輸出，其中顯示秘密的未加密內容，以及秘密所有版本的中繼資料。

```
{
    "ARN": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-AbCdEf",
    "Name": "appauthexample",
    "Description": "Example of a secret containing application authorization data.",
    "RotationEnabled": false,
```

```
    "LastChangedDate": 1581968848.926,
    "LastAccessedDate": 1581897600.0,
    "Tags": [],
    "VersionIdsToStages": {
        "871d9eca-18aa-46a9-8785-981dd39ab30c": [
            "AWSCURRENT"
        ],
        "9d4cb84b-ad69-40c0-a0ab-cead36b967e8": [
            "AWSPREVIOUS"
        ]
    }
}
```

在 ARN 結尾指定金鑰名稱，來在容器定義中參考上一個輸出的特定版本預備標籤。

```
{
  "containerProperties": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-
AbCdEf::AWSPREVIOUS:"
    }]
  }]
}
```

在 ARN 結尾指定金鑰名稱，來在容器定義中參考上一個輸出的特定版本 ID。

```
{
  "containerProperties": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-
AbCdEf::9d4cb84b-ad69-40c0-a0ab-cead36b967e8"
    }]
  }]
}
```

Example參考秘密的特定金鑰和版本預備標籤

以下說明如何參考秘密中的特定金鑰和特定版本預備標籤。

```
{
```

```
  "containerProperties": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-
AbCdEf:username1:AWSPREVIOUS:"
    }]
  }]
}
```

若要指定特定的金鑰和版本 ID，請使用下列語法。

```
{
  "containerProperties": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-
AbCdEf:username1::9d4cb84b-ad69-40c0-a0ab-cead36b967e8"
    }]
  }]
}
```

在日誌組態中注入敏感資料

當您在任務定義logConfiguration中指定 時，您可以使用要在容器中設定的日誌驅動程式選項名稱secretOptions，以及包含要呈現給容器之敏感資料之 Secrets Manager 秘密的完整 ARN。

以下是任務定義的程式碼片段，顯示參考 Secrets Manager 秘密時的格式。

```
{
  "containerProperties": [{
    "logConfiguration": [{
      "logDriver": "splunk",
      "options": {
        "splunk-url": "https://cloud.splunk.com:8080"
      },
      "secretOptions": [{
        "name": "splunk-token",
        "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:secret_name-
AbCdEf"
      }]
    }]
  }]
```

```
}
```

建立 AWS Secrets Manager 秘密

您可以使用 Secrets Manager 主控台為您的敏感資料建立秘密。如需詳細資訊，請參閱 《AWS Secrets Manager 使用者指南》中的建立基本秘密。

建立基本秘密

使用 Secrets Manager 為您的敏感資料建立秘密。

1. 開啟位於的 Secrets Manager 主控台https://console.aws.amazon.com/secretsmanager/。
2. 選擇 Store a new secret (存放新機密)。
3. 針對 Select secret type (選取秘密類型)，選擇 Other type of secrets (其他秘密類型)。
4. 將自訂秘密詳細資訊指定為 Key (金鑰) 與 Value (值) 對。例如，您可以指定 `UserName` 的金鑰，然後提供適當的使用者名稱作為其值。新增名為 `Password` 的第二個金鑰，再輸入密碼文字作為其值。您也可以新增資料庫名稱、伺服器地址或 TCP 連接埠的項目。您可以新增任意數量的對組用以存放您需要的資訊。

   或者，您可以選擇 Plaintext (純文字) 標籤，然後以任何您喜歡的方式輸入秘密值。
5. 選擇您要用來 AWS KMS 加密秘密中受保護文字的加密金鑰。如果您未選擇一個金鑰，則 Secrets Manager 將查看帳戶是否有預設金鑰，若有便會使用該金鑰。如果預設金鑰不存在，Secrets Manager 將自動為您建立一個。您也可以選擇 Add new key (新增金鑰) 來建立專供此秘密使用的自訂 KMS 金鑰。若要建立自己的 KMS 金鑰，您必須擁有在帳戶中建立 KMS 金鑰的許可。
6. 選擇下一步。
7. 針對 Secret name (秘密名稱)，請輸入可選的路徑與名稱，例如 **`production/MyAwesomeAppSecret`** 或 **`development/TestSecret`**，然後選擇 Next (下一步)。您也可以選擇性新增描述，協助您日後回憶起此秘密的用途。

   秘密名稱必須是 ASCII 字母、數字或下列任一字元：/_+=.@-
8. (選用) 在此階段，您可以為秘密設定輪換。針對此程序，維持 Disable automatic rotation (停用自動輪換)，然後選擇 Next (下一步)。

   如需有關如何在新的或現有的秘密上設定輪換的資訊，請參閱輪換您的 AWS Secrets Manager 秘密。
9. 檢視您的設定，然後選擇 Store secret (存放秘密) 以儲存您在 Secrets Manager 中輸入作為新秘密的所有內容。

使用 Systems Manager 參數存放區指定敏感資料

使用 時 AWS Batch，您可以將敏感資料儲存在參數存放區參數中 AWS Systems Manager ，然後在容器定義中參考它們，藉此將敏感資料注入容器。

主題

- 使用 Systems Manager 參數存放區指定敏感資料的考量事項
- AWS Batch 秘密所需的 IAM 許可
- 將敏感資料注入為環境變數
- 在日誌組態中注入敏感資料
- 建立 AWS Systems Manager 參數存放區參數

使用 Systems Manager 參數存放區指定敏感資料的考量事項

使用 Systems Manager 參數存放區參數來指定容器的敏感資料時，應考慮以下事項。

- 此功能需要您的容器執行個體具有 1.23.0 版或更新版本的容器代理程式。不過，我們建議您使用最新版的容器代理。如需有關檢查代理程式版本和更新至最新版本的資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的更新 Amazon ECS 容器代理程式。
- 最初啟動容器時，會將敏感資料注入您任務的容器中。如果後續更新或輪換秘密或參數存放區參數，則容器不會自動收到更新的值。您必須啟動新的任務，才能強制啟動具有更新秘密的新任務。

AWS Batch 秘密所需的 IAM 許可

若要使用此功能，您必須擁有 執行角色，並在您的任務定義中參考它。這可讓 Amazon ECS 容器代理程式提取必要的 AWS Systems Manager 資源。如需詳細資訊，請參閱AWS Batch IAM 執行角色。

若要讓您存取您建立的 AWS Systems Manager 參數存放區參數，請手動將下列許可做為內嵌政策新增至執行角色。如需詳細資訊，請參閱 IAM User Guide 中的 Adding and Removing IAM Policies。

- `ssm:GetParameters` - 如果您在任務定義中參考 Systems Manager 參數存放區參數，才需要此項目。
- `secretsmanager:GetSecretValue` - 如果您直接參考 Secrets Manager 秘密，或者您的 Systems Manager 參數存放區參數參考任務定義中的 Secrets Manager 秘密，才需要此項目。
- `kms:Decrypt` - 只有在您的秘密使用自訂 KMS 金鑰而非預設金鑰時，才需要此項目。您的自訂金鑰的 ARN 應該新增為資源。

下列內嵌政策範例新增必要許可：

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ssm:GetParameters",
                "secretsmanager:GetSecretValue",
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:ssm:us-east-2:999999999999:parameter/<parameter_name>",
                "arn:aws:secretsmanager:us-
east-2:999999999999:secret:<secret_name>",
                "arn:aws:kms:us-east-2:999999999999:key/<key_id>"
            ]
        }
    ]
}
```

將敏感資料注入為環境變數

在您的容器定義內，將 `secrets` 指定為要在容器中設定的環境變數名稱，以及 Systems Manager 參數存放區參數 (含有要呈現給容器的敏感資料) 的完整 ARN。

以下是任務定義的程式碼片段，顯示參考 Systems Manager 參數存放區參數時的格式。如果 Systems Manager 參數存放區參數與您啟動的任務位於相同的區域中，則您可以使用參數的完整 ARN 或名稱。如果參數存在於不同區域，則必須指定完整 ARN。

```
{
  "containerProperties": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:ssm:region:aws_account_id:parameter/parameter_name"
    }]
  }]
```

```
}
```

在日誌組態中注入敏感資料

在您的容器定義內，指定 `logConfiguration` 時，您可用要在容器中設定的日誌驅動程式選項名稱指定 `secretOptions`，以及 Systems Manager 參數存放區參數 (含有要呈現給容器的敏感資料) 的完整 ARN。

> **Important**
>
> 如果 Systems Manager 參數存放區參數與您啟動的任務位於相同的區域中，則您可以使用參數的完整 ARN 或名稱。如果參數存在於不同區域，則必須指定完整 ARN。

以下是任務定義的程式碼片段，顯示參考 Systems Manager 參數存放區參數時的格式。

```
{
  "containerProperties": [{
    "logConfiguration": [{
      "logDriver": "fluentd",
      "options": {
        "tag": "fluentd demo"
      },
      "secretOptions": [{
        "name": "fluentd-address",
        "valueFrom": "arn:aws:ssm:region:aws_account_id:parameter/parameter_name"
      }]
    }]
  }]
}
```

建立 AWS Systems Manager 參數存放區參數

您可以使用 AWS Systems Manager 主控台為您的敏感資料建立 Systems Manager 參數存放區參數。如需詳細資訊，請參閱 《AWS Systems Manager 使用者指南》中的演練：在命令中建立和使用參數 (主控台)。

建立參數存放區參數

1. 在 https：//https://console.aws.amazon.com/systems-manager/ 開啟 AWS Systems Manager 主控台。

2. 在導覽窗格中，選擇 Parameter Store (參數存放區)、Create parameter (建立參數)。
3. 針對 Name (名稱)，輸入階層和參數名稱。例如，輸入 `test/database_password`。
4. 針對 Description (描述)，輸入選擇性描述。
5. 在 Type (類型) 中選擇 String、StringList 或 SecureString。

> **Note**
>
> - 如果您選擇 SecureString (SecureString)，將會出現 KMS Key ID (KMS 金鑰 ID) 欄位。如果您不提供 KMS 金鑰 ID、KMS 金鑰 ARN、別名名稱或別名 ARN，系統會使用 `alias/aws/ssm`。這是 Systems Manager 的預設 KMS 金鑰。若要避免使用此金鑰，請選擇自訂金鑰。如需詳細資訊，請參閱 《AWS Systems Manager 使用者指南》中的使用安全字串參數。
> - 當您在主控台使用 `key-id` 參數及自訂 KMS 金鑰別名名稱或別名 ARN 建立安全字串參數時，您必須在別名前面指定字首 `alias/`。以下是 ARN 範例：
>
>   ```
>   arn:aws:kms:us-east-2:123456789012:alias/MyAliasName
>   ```
>
>   下列是別名名稱範例：
>
>   ```
>   alias/MyAliasName
>   ```

6. 針對 Value (值)，輸入一個值。例如 `MyFirstParameter`。如果您選擇 SecureString，則會完全按照您輸入的方式遮罩該值。
7. 選擇 Create parameter (建立參數)。

任務的私有登錄檔身分驗證

使用 之任務的私有登錄檔身分驗證 AWS Secrets Manager 可讓您安全地存放登入資料，然後在任務定義中參考登入資料。這可讓您參考存在於 外部私有登錄檔中的容器映像 AWS ，而這些登錄檔需要在您的任務定義中進行身分驗證。Amazon EC2 執行個體和 Fargate 上託管的任務支援此功能。

> **Important**
>
> 如果您的任務定義參考存放在 Amazon ECR 中的映像，則此主題不適用。如需詳細資訊，請參閱 Amazon Elastic Container Registry User Guide 中的 Using Amazon ECR Images with Amazon ECS。

對於在 Amazon EC2 執行個體上託管的任務，此功能需要版本 `1.19.0`或更新版本的容器代理程式。不過，我們建議您使用最新版的容器代理。如需有關如何檢查代理程式版本和更新至最新版本的資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的更新 Amazon ECS 容器代理程式。

對於在 Fargate 上託管的任務，此功能需要平台版本 `1.2.0` 或更新版本。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 AWS Fargate Linux 平台版本。

在您的容器定義內，使用您所建立的秘密的詳細資訊來指定 `repositoryCredentials` 物件。您參考的秘密可以是來自與使用它的任務不同的 AWS 區域 或不同的 帳戶。

> **Note**
>
> 使用 AWS Batch API AWS CLI或 AWS SDK 時，如果秘密與您啟動 AWS 區域 的任務位於相同的 中，您可以使用秘密的完整 ARN 或名稱。如果此秘密已存在於不同帳戶中，則必須指定秘密的完整 ARN。使用 時 AWS 管理主控台，一律必須指定秘密的完整 ARN。

以下是任務定義的程式碼片段，顯示所需的參數：

```
"containerProperties": [
  {
    "image": "private-repo/private-image",
    "repositoryCredentials": {
      "credentialsParameter":
 "arn:aws:secretsmanager:region:123456789012:secret:secret_name"
    }
  }
]
```

私有登錄檔身分驗證所需的 IAM 許可

需要執行角色才能使用此功能。這可讓容器代理程式提取容器映像。如需詳細資訊，請參閱AWS Batch IAM 執行角色。

若要提供對您建立之秘密的存取權，請將下列許可做為內嵌政策新增至執行角色。如需詳細資訊，請參閱新增和移除 IAM 政策。

- `secretsmanager:GetSecretValue`
- `kms:Decrypt` - 只有在您的金鑰使用自訂 KMS 金鑰而非預設金鑰時，才需要此項目。您的自訂金鑰的 Amazon Resource Name (ARN) 必須新增為資源。

下列為新增許可的內嵌政策範例。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "secretsmanager:GetSecretValue"
            ],
            "Resource": [
                "arn:aws:secretsmanager:us-
east-1:123456789012:secret:secret_name",
                "arn:aws:kms:us-east-1:123456789012:key/key_id"
            ]
        }
    ]
}
```

教學課程：建立私有登錄檔身分驗證的秘密

完成下列步驟，以使用 為您的私有登錄檔登入資料建立秘密 AWS Secrets Manager。

建立基本秘密

1. 在 https：//https://console.aws.amazon.com/secretsmanager/ 開啟 AWS Secrets Manager 主控台。
2. 選擇儲存新機密。

3. 針對 Select secret type (選取秘密類型)，選擇 Other type of secrets (其他秘密類型)。
4. 選取 Plaintext (純文字)，然後使用下列格式輸入您的私有登錄登入資料：

```
{
  "username" : "privateRegistryUsername",
  "password" : "privateRegistryPassword"
}
```

5. 選擇下一步。
6. 針對 Secret name (秘密名稱)，請輸入選用的路徑與名稱，例如 **`production/MyAwesomeAppSecret`** 或 **`development/TestSecret`**，然後選擇 Next (下一步)。您也可以選擇性新增描述，協助您日後回憶起此秘密的用途。

   秘密名稱必須是 ASCII 字母、數字或下列任一字元：`/_+=.@-`。
7. (選用) 在此階段，您可以為秘密設定輪換。針對此程序，維持 Disable automatic rotation (停用自動輪換)，然後選擇 Next (下一步)。

   如需如何在新的或現有的秘密上設定輪換的指示，請參閱輪換您的 AWS Secrets Manager 秘密。
8. 檢閱您的設定，然後選擇 Store secret (存放秘密) 以儲存您在 Secrets Manager 中輸入作為新秘密的所有內容。

註冊任務定義並在私有登錄檔下，開啟私有登錄檔身分驗證。然後，在 Secrets Manager ARN 或名稱中，輸入密碼的 Amazon Resource Name (ARN)。如需詳細資訊，請參閱私有登錄檔身分驗證所需的 IAM 許可。

Amazon EFS 磁碟區

Amazon Elastic File System (Amazon EFS) 提供簡單、可擴展的檔案儲存，可與您的 AWS Batch 任務搭配使用。利用 Amazon EFS，儲存容量即可有彈性。當您新增和移除檔案時，它會自動擴展。您的應用程式可在需要時具備所需的儲存容量。

您可以使用 Amazon EFS 檔案系統搭配 AWS Batch 在容器執行個體機群之間匯出檔案系統資料。如此一來，您的任務就可以存取相同的持久性儲存。但您必須在 Docker 常駐程式啟動之前，先設定您的容器執行個體 AMI，以掛載 Amazon EFS 檔案系統。此外，您的任務定義必須參考容器執行個體上的磁碟區掛載，才能使用檔案系統。下列各節可協助您開始使用 Amazon EFS 搭配 AWS Batch。

Amazon EFS 磁碟區考量事項

使用 Amazon EFS 磁碟區時應考慮以下事項：

- 對於使用 EC2 資源的任務，Amazon EFS 檔案系統支援已新增為使用 20191212 Amazon ECS 最佳化 AMI 版本搭配容器代理程式版本 1.35.0 的公開預覽。不過，Amazon EFS 檔案系統支援使用20200319容器代理程式 1.38.0 版的 Amazon ECS 最佳化 AMI 版本進入一般可用性，其中包含 Amazon EFS 存取點和 IAM 授權功能。我們建議您使用 Amazon ECS 最佳化 AMI 版本 20200319 或更新版本，以利用這些功能。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 最佳化 AMI 版本。

  > **Note**
  >
  > 如果您建立自己的 AMI，則必須使用容器代理程式 1.38.0 或更新版本、1.38.0-1 ecs-init版或更新版本，並在 Amazon EC2 執行個體上執行下列命令。這是為了啟用 Amazon ECS 磁碟區外掛程式。這些命令取決於您是否使用 Amazon Linux 2 或 Amazon Linux 作為基礎映像。
  >
  > Amazon Linux 2
  >
  > ```
  > $ yum install amazon-efs-utils
  > systemctl enable --now amazon-ecs-volume-plugin
  > ```
  >
  > Amazon Linux
  >
  > ```
  > $ yum install amazon-efs-utils
  > sudo shutdown -r now
  > ```

- 對於使用 Fargate 資源的任務，使用平台版本 1.4.0 或更新版本時新增了 Amazon EFS 檔案系統支援。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 AWS Fargate 平台版本。
- 使用 Fargate 資源在任務中指定 Amazon EFS 磁碟區時，Fargate 會建立主管容器，負責管理 Amazon EFS 磁碟區。主管容器會使用少量任務的記憶體。查詢任務中繼資料第 4 版端點時，可看見監督容器。如需詳細資訊，請參閱 AWS Fargate的 Amazon Elastic Container Service 使用者指南中的任務中繼資料端點版本 4。

使用 Amazon EFS 存取點

Amazon EFS 存取點是 EFS 檔案系統的應用程式特定進入點，可協助您管理共用資料集的應用程式存取。如需有關 Amazon EFS 存取點及如何控制對它們的存取的詳細資訊，請參閱《Amazon Elastic File System 使用者指南》中的使用 Amazon EFS 存取點。

存取點可以針對透過存取點提出的所有檔案系統要求，強制執行使用者身分 (包括使用者的 POSIX 群組)。存取點也可以針對檔案系統強制執行不同的根目錄，讓用戶端只能存取指定目錄或其子目錄中的資料。

> **Note**
>
> 建立 EFS 存取點時，您可以在檔案系統上指定要做為根目錄的路徑。當您在 AWS Batch 任務定義中參考具有存取點 ID 的 EFS 檔案系統時，必須省略根目錄或將其設定為 / 這會強制執行在 EFS 存取點上設定的路徑。

您可以使用 AWS Batch 任務 IAM 角色來強制執行特定應用程式使用特定存取點。透過結合 IAM 政策與存取點，您可以輕鬆地為應用程式提供特定資料集的安全存取權。此功能使用 Amazon ECS IAM 角色執行任務功能。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的任務 IAM 角色。

在任務定義中指定 Amazon EFS 檔案系統

若要為您的容器使用 Amazon EFS 檔案系統磁碟區，您必須在任務定義中指定磁碟區和掛載點組態。下列任務定義 JSON 程式碼片段顯示容器的 `volumes`和 `mountPoints` 物件的語法：

```
{
    "containerProperties": [
        {
            "image": "amazonlinux:2",
            "command": [
                "ls",
                "-la",
                "/mount/efs"
            ],
            "mountPoints": [
                {
                    "sourceVolume": "myEfsVolume",
                    "containerPath": "/mount/efs",
```

```
                    "readOnly": true
                }
            ],
            "volumes": [
                {
                    "name": "myEfsVolume",
                    "efsVolumeConfiguration": {
                        "fileSystemId": "fs-12345678",
                        "rootDirectory": "/path/to/my/data",
                        "transitEncryption": "ENABLED",
                        "transitEncryptionPort": integer,
                        "authorizationConfig": {
                            "accessPointId": "fsap-1234567890abcdef1",
                            "iam": "ENABLED"
                        }
                    }
                }
            ]
        }
    ]
}
```

`efsVolumeConfiguration`

類型：物件

必要：否

只有使用 Amazon EFS 磁碟區時才會指定此參數。

`fileSystemId`

類型：字串

必要：是

要使用的 Amazon EFS 檔案系統識別碼。

`rootDirectory`

類型：字串

必要：否

在 Amazon EFS 檔案系統中的目錄，其將掛載作為主機內的根目錄。如果省略此參數，使用 Amazon EFS 磁碟區的根目錄。指定 / 的效果與忽略此參數的效果相同。長度上限為 4，096 個字元。

> **Important**
>
> 如果在 中指定了 EFS 存取點`authorizationConfig`，則必須省略根目錄參數，或將設定為 /。這會強制執行在 EFS 存取點上設定的路徑。

`transitEncryption`

類型：字串

有效值：`ENABLED` | `DISABLED`

必要：否

決定是否為 AWS Batch 主機和 Amazon EFS 伺服器之間傳輸的 Amazon EFS 資料啟用加密。若使用 Amazon EFS IAM 授權，則必須啟用傳輸加密。如果省略此參數，系統會使用 `DISABLED` 的預設值。如需詳細資訊，請參閱《Amazon Elastic File System 使用者指南》中的加密傳輸中的資料。

`transitEncryptionPort`

類型：整數

必要：否

在 AWS Batch 主機和 Amazon EFS 伺服器之間傳送加密資料時要使用的連接埠。如果您未指定傳輸加密連接埠，它會使用 Amazon EFS 掛載協助程式使用的連接埠選擇策略。該值必須介於 0 到 65,535 之間。如需詳細資訊，請參閱《Amazon Elastic File System 使用者指南》中的 EFS 掛載協助程式。

`authorizationConfig`

類型：物件

必要：否

Amazon EFS 檔案系統的授權組態詳細資訊。

`accessPointId`

類型：字串

必要：否

要使用的存取點 ID。如果指定存取點，efsVolumeConfiguration則必須省略 中的根目錄值，或將 設定為 /。這會強制執行在 EFS 存取點上設定的路徑。如果使用存取點，則必須在 EFSVolumeConfiguration 中啟用傳輸加密。如需詳細資訊，請參閱《Amazon Elastic File System 使用者指南》中的使用 Amazon EFS 存取點。

iam

類型：字串

有效值：ENABLED | DISABLED

必要：否

決定是否在掛載 Amazon EFS 檔案系統時使用 AWS Batch 任務定義中定義的任務 IAM 角色。如果已啟用，必須在 EFSVolumeConfiguration 中啟用傳輸加密。如果省略此參數，系統會使用 DISABLED 的預設值。如需 IAM 執行角色的詳細資訊，請參閱 AWS Batch IAM 執行角色。

任務定義範例

下列主題中的任務定義範例說明如何使用常見的模式，例如環境變數、參數替換和磁碟區掛載。

目錄

- 環境變數
- 參數替換
- 測試 GPU 功能
- 多節點平行任務

環境變數

下列範例任務定義使用環境變數來指定檔案類型和 Amazon S3 URL。此特定範例來自建立簡單的「擷取和執行」 AWS Batch 任務運算部落格文章。部落格文章中描述的fetch_and_run.sh指令碼使用這些環境變數從 S3 下載myjob.sh指令碼並宣告其檔案類型。

即使命令和環境變數在此範例中硬式編碼為任務定義，您也可以指定命令和環境變數覆寫，讓任務定義更多樣化。

```
{
    "jobDefinitionName": "fetch_and_run",
    "type": "container",
    "containerProperties": {
        "image": "123456789012.dkr.ecr.us-east-1.amazonaws.com/fetch_and_run",
        "resourceRequirements": [
            {
                "type": "MEMORY",
                "value": "2000"
            },
            {
                "type": "VCPU",
                "value": "2"
            }
        ],
        "command": [
            "myjob.sh",
            "60"
        ],
        "jobRoleArn": "arn:aws:iam::123456789012:role/AWSBatchS3ReadOnly",
        "environment": [
            {
                "name": "BATCH_FILE_S3_URL",
                "value": "s3://amzn-s3-demo-source-bucket/myjob.sh"
            },
            {
                "name": "BATCH_FILE_TYPE",
                "value": "script"
            }
        ],
        "user": "nobody"
    }
}
```

參數替換

以下任務定義範例說明，如何允許替換參數和設定預設值。

Ref:: 區段中的 command 宣告用於設定的替換參數的預留位置。提交使用此任務定義的任務時，您要指定參數覆寫以填入這些值，例如 inputfile 和 outputfile。接下來的 parameters區段會設定 的預設值codec，但您可以視需要覆寫該參數。

如需詳細資訊，請參閱參數。

```
{
    "jobDefinitionName": "ffmpeg_parameters",
    "type": "container",
    "parameters": {"codec": "mp4"},
    "containerProperties": {
        "image": "my_repo/ffmpeg",
        "resourceRequirements": [
            {
                "type": "MEMORY",
                "value": "2000"
            },
            {
                "type": "VCPU",
                "value": "2"
            }
        ],
        "command": [
            "ffmpeg",
            "-i",
            "Ref::inputfile",
            "-c",
            "Ref::codec",
            "-o",
            "Ref::outputfile"
        ],
        "jobRoleArn": "arn:aws:iam::123456789012:role/ECSTask-S3FullAccess",
        "user": "nobody"
    }
}
```

測試 GPU 功能

以下工作定義範例測試 使用 GPU 工作負載 AMI 中所述的 GPU 工作負載 AMI 是否正確設定。此範例任務定義會從 GitHub 執行 TensorFlow 深度 MNIST 分類器範例。

```
{
    "containerProperties": {
        "image": "tensorflow/tensorflow:1.8.0-devel-gpu",
        "resourceRequirements": [
            {
                "type": "MEMORY",
                "value": "32000"
```

```
            },
            {
                "type": "VCPU",
                "value": "8"
            }
        ],
        "command": [
            "sh",
            "-c",
            "cd /tensorflow/tensorflow/examples/tutorials/mnist; python mnist_deep.py"
        ]
    },
    "type": "container",
    "jobDefinitionName": "tensorflow_mnist_deep"
}
```

您可以建立名為前述 JSON 文字的檔案，`tensorflow_mnist_deep.json`然後使用下列命令註冊 AWS Batch 任務定義：

```
aws batch register-job-definition --cli-input-json file://tensorflow_mnist_deep.json
```

多節點平行任務

下列任務定義範例說明多節點平行任務。如需詳細資訊，請參閱AWS 運算部落格中的在 中使用多節點平行任務建置緊密耦合的分子動態工作流程 AWS Batch。

```
{
  "jobDefinitionName": "gromacs-jobdef",
  "jobDefinitionArn": "arn:aws:batch:us-east-2:123456789012:job-definition/gromacs-
jobdef:1",
  "revision": 6,
  "status": "ACTIVE",
  "type": "multinode",
  "parameters": {},
  "nodeProperties": {
    "numNodes": 2,
    "mainNode": 0,
    "nodeRangeProperties": [
      {
        "targetNodes": "0:1",
        "container": {
          "image": "123456789012.dkr.ecr.us-east-2.amazonaws.com/gromacs_mpi:latest",
```

```
          "resourceRequirements": [
              {
                  "type": "MEMORY",
                  "value": "24000"
              },
              {
                  "type": "VCPU",
                  "value": "8"
              }
          ],
          "command": [],
          "jobRoleArn": "arn:aws:iam::123456789012:role/ecsTaskExecutionRole",
          "ulimits": [],
          "instanceType": "p3.2xlarge"
        }
      }
    ]
  }
}
```

任務

任務是由 啟動的工作單位 AWS Batch。任務可以調用為在 ECS 叢集中的 Amazon ECS 容器執行個體上執行的容器化應用程式。

容器化的任務可以參考容器映像、命令和參數。如需詳細資訊，請參閱 JobDefinition。

您可以提交大量獨立、簡單的任務

主題

- 教學課程：提交任務
- 中的服務任務 AWS Batch
- 任務狀態
- AWS Batch 任務環境變數
- 自動化任務重試
- 任務相依性
- 任務逾時
- Amazon EKS 任務
- 多節點平行任務
- Amazon EKS 上的多節點平行任務
- 陣列任務
- 執行 GPU 任務
- 檢視 AWS Batch 任務佇列中的任務
- 在任務佇列中 AWS Batch 搜尋任務
- AWS Batch 任務的網路模式
- 在 CloudWatch Logs 中檢視 AWS Batch 任務日誌
- 檢閱 AWS Batch 任務資訊

教學課程：提交任務

註冊任務定義後，您可以將其做為任務提交至 AWS Batch 任務佇列。您可以在執行時間覆寫任務定義中指定的許多參數。

提交任務

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇 Jobs (任務)。
4. 選擇提交新任務。
5. 在名稱中，輸入任務定義的唯一名稱。名稱長度最多可達 128 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。
6. 針對任務定義，為您的任務選擇現有的任務定義。如需詳細資訊，請參閱建立單一節點任務定義。
7. 針對任務佇列，選擇現有的任務佇列。如需詳細資訊，請參閱建立任務佇列。
8. 針對任務相依性，選擇新增任務相依性。
   - 在任務 ID 中，輸入任何相依性的任務 ID。然後選擇新增任務相依性。任務最多可以有 20 個相依性。如需詳細資訊，請參閱任務相依性。
9. (僅適用於陣列任務) 在 Array size (陣列大小) 中，指定 2 至 10,000 之間的陣列大小。
10. （選用）展開標籤，然後選擇新增標籤以將標籤新增至資源。輸入索引鍵和選用值，然後選擇新增標籤。
11. 選擇下一頁。
12. 在任務覆寫區段中：
    a. （選用）針對排程優先順序，輸入介於 0 到 100 之間的排程優先順序值。較高值的優先順序較高。
    b. （選用）對於任務嘗試，輸入 AWS Batch 嘗試將任務移至RUNNABLE狀態的次數上限。您可以輸入介於 1 到 10 之間的數字。如需詳細資訊，請參閱自動化任務重試。
    c. （選用）針對執行逾時，輸入逾時值（以秒為單位）。執行逾時是未完成任務終止之前的時間長度。如果嘗試超過逾時持續時間，則會停止並移至 FAILED 狀態。如需詳細資訊，請參閱任務逾時。最小值為 60 秒。

       > **Important**
       > 請勿倚賴在 Fargate 資源上執行的任務執行超過 14 天。14 天後，Fargate 資源可能無法再用於可能終止的任務。

    d. （選用）開啟傳播標籤，將標籤從任務和任務定義傳播到 Amazon ECS 任務。

13. 展開 Additional configuration (其他組態)。
14. （選用）針對重試策略條件，選擇在結束時新增評估。輸入至少一個參數值，然後選擇動作。對於每組條件，動作必須設定為重試或結束。這些動作表示下列項目：
    - 重試 – AWS Batch 重試，直到達到您指定的任務嘗試次數為止。
    - 結束 – AWS Batch 停止重試任務。

    > **Important**
    > 如果您選擇在結束時新增評估，請設定至少一個參數，然後選擇動作或選擇在結束時移除評估。

15. 針對參數，選擇新增參數以新增參數替換預留位置。然後，輸入金鑰和選用值。
16. 在容器覆寫區段中：
    a. 在命令中，將命令輸入 欄位做為其 JSON 字串陣列對等項。

       此參數會映射至 Docker Remote API 的建立容器區段中的 Cmd 以及 docker run 的 COMMAND 參數。如需 Docker CMD 參數的詳細資訊，請參閱 https://docs.docker.com/engine/reference/builder/#cmd。

       > **Note**
       > 此參數不能包含空字串。

    b. 針對 vCPUs，輸入要保留給容器的 vCPUs 數量。此參數會映射到 Docker Remote API 的建立容器區段中的 CpuShares 以及 docker run 的 `--cpu-shares` 選項。每個 vCPU 相當於 1,024 個 CPU 共用。您必須指定至少 1 個 vCPU。

    c. 針對記憶體，輸入容器可用的記憶體限制。如果容器嘗試使用超過此處指定的記憶體，容器便會終止。此參數會映射到 Docker Remote API 的建立容器區段中的 Memory 以及 docker run 的 `--memory` 選項。您必須為單一工作指定至少 4 MiB 的記憶體。

       > **Note**
       > 若要最大化資源使用率，請為特定執行個體類型的任務排定記憶體的優先順序。如需詳細資訊，請參閱運算資源記憶體管理。

d. （選用）針對 GPUs 數量，選擇要保留給容器的 GPUs 數量。

e. （選用）對於環境變數，選擇新增環境變數，將環境變數新增為名稱/值對。這些變數會傳遞至容器。

f. 選擇下一頁。

g. 針對任務檢閱，請檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立任務定義。

中的服務任務 AWS Batch

AWS Batch 服務任務可讓您透過 AWS Batch 任務佇列向 AWS 服務提交請求。目前， AWS Batch 支援 SageMaker Training 任務做為服務任務。與 AWS Batch 管理基礎容器執行的容器化任務不同，服務任務允許 在目標 AWS 服務 （例如 SageMaker AI) 處理實際任務執行時 AWS Batch 提供任務排程和佇列功能。

AWS Batch for SageMaker Training 任務可讓資料科學家以可設定佇列的優先順序提交訓練任務，確保工作負載在有資源可用時立即執行，無需介入。此功能可解決常見的挑戰，例如資源協調、防止意外超支、符合預算限制、使用預留執行個體最佳化成本，以及消除團隊成員之間手動協調的需求。

服務任務在幾個關鍵方面與容器化任務不同：

- 任務提交：必須使用 SubmitServiceJob API 提交服務任務。服務任務無法透過 AWS Batch 主控台提交。
- 任務執行： AWS Batch 排程和佇列服務任務，但目標 AWS 服務會執行實際任務工作負載。
- 資源識別符：服務任務使用包含「service-job」而非「job」ARNs，以區分它們與容器化任務。

若要開始使用 SageMaker Training AWS Batch 的服務任務，請參閱 the section called “SageMaker AI AWS Batch 入門”。

主題

- 中的服務任務承載 AWS Batch
- 在 中提交服務任務 AWS Batch
- 將 AWS Batch 服務任務狀態映射至 SageMaker AI 狀態
- 中的服務任務重試策略 AWS Batch
- 監控 AWS Batch 佇列中的服務任務
- 終止服務任務

中的服務任務承載 AWS Batch

當您使用 SubmitServiceJob 提交服務任務時，您會提供兩個定義任務的關鍵參數：`serviceJobType`和 `serviceRequestPayload`。

- `serviceJobType` 指定要執行任務 AWS 的服務。對於 SageMaker Training 任務，此值為 `SAGEMAKER_TRAINING`。
- `serviceRequestPayload` 是 JSON 編碼字串，其中包含通常直接傳送到目標服務的完整請求。對於 SageMaker Training 任務，此承載包含與 SageMaker AI CreateTrainingJob API 相同的參數。

如需所有可用參數及其說明的完整清單，請參閱 SageMaker AI CreateTrainingJob API 參考。支援的所有參數`CreateTrainingJob`都可以包含在服務任務承載中。

如需更多訓練任務組態的範例，請參閱 SageMaker AI 開發人員指南中的 APIs、CLI 和 SDKs。

我們建議您使用 PySDK 建立服務任務，因為 PySDK 具有協助程式類別和公用程式。如需使用 PySDK 的範例，請參閱 GitHub 上的 SageMaker AI 範例。

服務任務承載範例

下列範例顯示執行「hello world」訓練指令碼的 SageMaker 訓練任務的簡單服務任務承載：

呼叫 時，此承載會以 JSON 字串的形式傳遞給 `serviceRequestPayload` 參數`SubmitServiceJob`。

```
{
  "TrainingJobName": "my-simple-training-job",
  "RoleArn": "arn:aws:iam::123456789012:role/SageMakerExecutionRole",
  "AlgorithmSpecification": {
    "TrainingInputMode": "File",
    "TrainingImage": "763104351884.dkr.ecr.us-west-2.amazonaws.com/pytorch-
training:2.0.0-cpu-py310",
    "ContainerEntrypoint": [
      "echo",
      "hello world"
    ]
  },
  "ResourceConfig": {
    "InstanceType": "ml.c5.xlarge",
    "InstanceCount": 1,
    "VolumeSizeInGB": 1
```

```
  },
  "OutputDataConfig": {
    "S3OutputPath": "s3://your-output-bucket/output"
  },
  "StoppingCondition": {
    "MaxRuntimeInSeconds": 30
  }
}
```

在 中提交服務任務 AWS Batch

若要提交服務任務給 AWS Batch，您可以使用 SubmitServiceJob API。您可以使用 AWS CLI 或 SDK 提交任務。

如果您還沒有執行角色，則必須先建立一個，才能提交您的服務任務。若要建立 SageMaker AI 執行角色，請參閱 SageMaker AI 開發人員指南中的如何使用 SageMaker AI 執行角色。 SageMaker

服務任務提交工作流程

當您提交服務任務時， AWS Batch 會遵循此工作流程：

1. AWS Batch 會接收您的SubmitServiceJob請求並驗證 AWS Batch特定參數。serviceRequestPayload 會在未驗證的情況下傳遞。
2. 任務進入 SUBMITTED 狀態，並放置在指定的任務佇列中
3. AWS Batch 評估佇列前方RUNNABLE任務的服務環境中是否有可用的容量
4. 如果容量可用，任務會移至 ，SCHEDULED且任務已傳遞至 SageMaker AI
5. 當容量已取得且 SageMaker AI 已下載服務任務資料時，服務任務將開始初始化，並將任務變更為 STARTING。
6. 當 SageMaker AI 開始執行任務時，其狀態會變更為 RUNNING。
7. 當 SageMaker AI 執行任務時， AWS Batch 會監控其進度並將服務狀態映射至 AWS Batch 任務狀態。如需如何映射服務任務狀態的詳細資訊，請參閱 ???
8. 當服務任務完成時，它會移至 ，SUCCEEDED並且任何輸出都已準備好下載。

先決條件

在提交服務任務之前，請確定您有：

- 服務環境 – 定義容量限制的服務環境。如需詳細資訊，請參閱在 中建立服務環境 AWS Batch。

- SageMaker 任務佇列 – 提供任務排程的 SageMaker 任務佇列。如需詳細資訊，請參閱在 中建立 SageMaker 訓練任務佇列 AWS Batch。
- IAM 許可 – 建立和管理 AWS Batch 任務佇列和服務環境的許可。如需詳細資訊，請參閱AWS Batch IAM 政策、角色和許可。

使用 CLI AWS 提交服務任務

以下說明如何使用 CLI AWS 提交服務任務：

```
aws batch submit-service-job \
    --job-name "my-sagemaker-training-job" \
    --job-queue "my-sagemaker-job-queue" \
    --service-job-type "SAGEMAKER_TRAINING" \
    --service-request-payload '{\"TrainingJobName\": \"sagemaker-training-job-
example\", \"AlgorithmSpecification\": {\"TrainingImage\": \"123456789012.dkr.ecr.us-
east-1.amazonaws.com/pytorch-inference:1.8.0-cpu-py3\", \"TrainingInputMode
\": \"File\", \"ContainerEntrypoint\":  [\"sleep\", \"1\"]}, \"RoleArn\":
\"arn:aws:iam::123456789012:role/SageMakerExecutionRole\", \"OutputDataConfig\":
 {\"S3OutputPath\": \"s3://example-bucket/model-output/\"}, \"ResourceConfig\":
 {\"InstanceType\": \"ml.m5.large\", \"InstanceCount\": 1, \"VolumeSizeInGB\": 1}}'
    --client-token "unique-token-12345"
```

如需`serviceRequestPayload`參數的詳細資訊，請參閱 the section called “服務任務承載”。

將 AWS Batch 服務任務狀態映射至 SageMaker AI 狀態

當您使用 SubmitServiceJob 將任務提交至 SageMaker 任務佇列時， 會 AWS Batch 管理任務生命週期並將 AWS Batch 任務狀態映射至同等的 SageMaker Training 任務狀態。服務任務，例如 SageMaker Training 任務，遵循與傳統容器任務不同的狀態生命週期。雖然服務任務與容器任務共用大多數狀態，但它們會引入 SCHEDULED 狀態並展現不同的重試行為，尤其是處理來自目標服務的容量不足錯誤。

下表顯示 AWS Batch 任務狀態和對應的 SageMaker Status/SecondaryStatus：

批次狀態	SageMaker AI 主要狀態	SageMaker AI 次要狀態	Description
SUBMITTED	N/A	N/A	任務提交至佇列，等待排程器評估。

批次狀態	SageMaker AI 主要狀態	SageMaker AI 次要狀態	Description
RUNNABLE	N/A	N/A	任務已排入佇列並準備好進行排程。一旦服務環境中有足夠的資源可用，就會立即啟動處於此狀態的任務。當足夠的資源無法使用時，任務可以無限期地保持在此狀態。
SCHEDULED	InProgress	Pending	服務任務已成功提交至 SageMaker AI
STARTING	InProgress	Downloading	SageMaker Training 任務下載資料和映像。已取得訓練任務容量，並開始任務初始化。
RUNNING	InProgress	Training	SageMaker Training 任務執行演算法
RUNNING	InProgress	Uploading	訓練完成後上傳輸出成品的 SageMaker Training 任務
SUCCEEDED	Completed	Completed	SageMaker Training 任務已成功完成。輸出成品已完成上傳。
FAILED	Failed	Failed	SageMaker Training 任務遇到無法復原的錯誤。
FAILED	Stopped	Stopped	SageMaker Training 任務已使用 手動停止StopTrainingJob 。

中的服務任務重試策略 AWS Batch

服務任務重試策略 AWS Batch 允許 在特定條件下自動重試失敗的服務任務。

服務任務可能需要多次嘗試，原因如下：

- 暫時性服務問題：內部服務錯誤、限流或暫時中斷可能會導致任務在提交或執行期間失敗。
- 訓練初始化失敗：任務啟動期間的問題，例如映像提取問題或初始化錯誤，可能會在重試時解決。

透過設定適當的重試策略，您可以提高任務成功率並減少手動介入的需求，尤其是長時間執行的訓練工作負載。

> **Note**
>
> 服務任務會自動重試特定類型的失敗，例如容量不足錯誤，而不會消耗您設定的重試嘗試。您的重試策略主要處理其他類型的失敗，例如演算法錯誤或服務問題。

設定重試策略

服務任務重試策略是使用 ServiceJobRetryStrategy 設定，同時支援簡單的重試計數和條件式重試邏輯。

重試組態

最簡單的重試策略指定服務任務失敗時應進行的重試嘗試次數：

```
{
  "retryStrategy": {
    "attempts": 3
  }
}
```

此組態可讓服務任務在失敗時重試最多 3 次。

> **Important**
>
> 此attempts值代表任務可置於 RUNNABLE 狀態的總次數，包括初始嘗試次數。值為 3 表示任務一開始會嘗試一次，然後在失敗時再重試最多 2 次。

使用 evaluateOnExit 重試組態

您可以使用 evaluateOnExit 參數來指定任務應重試或允許失敗的條件。當不同類型的失敗需要不同的處理時，這非常有用。

evaluateOnExit 陣列最多可包含 5 個重試策略，每個策略會根據狀態原因指定動作 (RETRY 或 EXIT) 和條件：

```
{
  "retryStrategy": {
    "attempts": 5,
    "evaluateOnExit": [
      {
        "action": "RETRY",
        "onStatusReason": "Received status from SageMaker: InternalServerError*"
      },
      {
        "action": "EXIT",
        "onStatusReason": "Received status from SageMaker: ValidationException*"
      },
      {
        "action": "EXIT",
        "onStatusReason": "*"
      }
    ]
  }
}
```

此組態：

- 重試因 SageMaker AI 內部伺服器錯誤而失敗的任務
- 遇到驗證例外狀況的任務立即失敗（無法透過重試解決的用戶端錯誤）
- 包含針對任何其他故障類型結束的所有截獲規則

狀態原因模式比對

onStatusReason 參數支援最多 512 個字元的模式比對。模式可以使用萬用字元 (*)，並根據 SageMaker AI 傳回的狀態原因進行比對。

對於服務任務，來自 SageMaker AI 的狀態訊息字首為「從 SageMaker 接收的狀態：」，以區分它們與 AWS Batch產生的訊息。常見的模式包括：

- `Received status from SageMaker: InternalServerError*` - 比對內部服務錯誤
- `Received status from SageMaker: ValidationException*` - 比對用戶端驗證錯誤
- `Received status from SageMaker: ResourceLimitExceeded*` - 比對資源限制錯誤
- `*CapacityError*` - 比對容量相關的故障

> **Tip**
>
> 使用特定模式比對來適當處理不同的錯誤類型。例如，重試內部伺服器錯誤，但驗證錯誤立即失敗，表示任務參數發生問題。

監控 AWS Batch 佇列中的服務任務

您可以使用 `list-service-jobs`、 和 來監控 SageMaker Training 任務佇列中任務的狀態`get-job-queue-snapshot`。

檢視佇列中正在執行的任務：

```
aws batch list-service-jobs \
  --job-queue my-sm-training-fifo-jq \
  --job-status RUNNING
```

檢視佇列中等待的任務：

```
aws batch list-service-jobs \
  --job-queue my-sm-training-fifo-jq \
  --job-status RUNNABLE
```

檢視已提交至 SageMaker 但尚未執行的任務：

```
aws batch list-service-jobs \
  --job-queue my-sm-training-fifo-jq \
  --job-status SCHEDULED
```

取得佇列前面的任務快照：

```
aws batch get-job-queue-snapshot --job-queue my-sm-training-fifo-jq
```

此命令會顯示佇列中即將到來的服務任務順序。

取得詳細的服務任務資訊

使用 DescribeServiceJob操作取得特定服務任務的完整資訊，包括其目前狀態、服務資源識別符和詳細的嘗試資訊。

檢視特定任務的詳細資訊：

```
aws batch describe-service-job \
  --job-id a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d
```

此命令會傳回任務的完整資訊，包括：

- 任務 ARN 和目前狀態
- 服務資源識別符 （ 例如 SageMaker Training 任務 ARN)
- 排程優先順序和重試組態
- 包含原始服務參數的服務請求承載
- 包含開始和停止時間的詳細嘗試資訊
- 來自目標服務的狀態訊息

監控 SageMaker 訓練任務

透過 監控 SageMaker Training 任務時 AWS Batch，您可以同時存取 AWS Batch 任務資訊和基礎 SageMaker Training 任務詳細資訊。

任務詳細資訊中的服務資源識別符包含 SageMaker Training 任務 ARN：

```
{
  "latestAttempt": {
    "serviceResourceId": {
      "name": "TrainingJobArn",
      "value": "arn:aws:sagemaker:us-east-1:123456789012:training-job/my-training-job"
    }
  }
}
```

您可以使用此 ARN 直接從 SageMaker 取得其他詳細資訊：

```
aws sagemaker describe-training-job \
  --training-job-name my-training-job
```

透過檢查 AWS Batch 狀態和 SageMaker Training 任務狀態來監控任務進度。 AWS Batch 任務狀態會顯示整體任務生命週期，而 SageMaker Training 任務狀態則提供訓練程序的服務特定詳細資訊。

終止服務任務

使用 TerminateServiceJob操作停止執行中的服務任務。

終止特定服務任務：

```
aws batch terminate-service-job \
  --job-id a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d \
  --reason "Job terminated by user request"
```

當您終止服務任務時， 會 AWS Batch 停止任務並通知目標服務。對於 SageMaker 訓練任務，這也會停止 SageMaker AI 中的訓練任務。

任務狀態

當您將任務提交至 AWS Batch 任務佇列時，任務會進入 SUBMITTED 狀態。任務將經過以下狀態，直到其失敗 (以 0 代碼結束) 或失敗 (以與非零代碼結束) 為止。 AWS Batch 任務可能有以下狀態：

SUBMITTED

已提交至佇列且尚未由排程器評估的任務。排程器評估任務，判斷其是否對任何其他任務的成功完成存有任何未完成的相依性。如果有相依性，任務將移至 PENDING。如果沒有相依性，任務將移至 RUNNABLE。

PENDING

位於佇列中的任務，但由於對其他任務或資源的相依性而無法執行。如果相依性獲得滿足，任務將移至 RUNNABLE。

> **Note**
>
> 當任何子任務更新為 PENDING 時，陣列任務父會更新為 ，RUNNABLE並在子任務執行時保持為 PENDING 狀態。若要檢視這些任務，請根據PENDING狀態進行篩選，直到所有子任務達到最終狀態為止。

RUNNABLE

佇列中的某一任務沒有未完成的相依性，因此已準備好排程傳送到主機。一旦其中一個映射到任務佇列的運算環境中有足夠的資源可用，就會立即啟動處於此狀態的任務。不過，假如一直無法取得足夠的資源，任務將無限期停留在此狀態。

> **Note**
>
> 如果您的任務未進展到 STARTING，請參閱故障診斷一節任務停滯在 RUNNABLE 狀態中的 。

STARTING

這些任務已排程傳送到主機，且相關的容器初始化作業正在進行中。取出容器映像且容器設置完畢並開始執行後，該任務將轉換為 RUNNING。

映像提取持續時間、Amazon EKS initContainer 完成持續時間和 Amazon ECS containerDependency解析持續時間會在 STARTING 狀態發生。為任務提取映像所需的時間量等同於任務將處於 STARTING 狀態的時間量。

例如，如果提取任務的映像需要三分鐘，您的任務將處於 STARTING 狀態三分鐘。如果 initContainers 總共需要十分鐘才能完成，則您的 Amazon EKS 任務將在 STARTING 中十分鐘。如果您的 Amazon ECS 任務中有 Amazon ECS containerDependencies 集，則任務將處於 STARTING，直到所有容器相依性（其執行時間）解決為止。STARTING 不包含在逾時中；持續時間從 RUNNING 開始。如需詳細資訊，請參閱任務狀態。

RUNNING

任務正在運算環境中的 Amazon ECS 容器執行個體上做為容器任務執行。任務的容器結束時，處理結束代碼將判斷任務為成功或失敗。0 結束代碼表示成功，任何非零的結束代碼則表示失敗。如果與嘗試失敗有關的任務在其選用的重試策略組態中有任何剩下的嘗試，任務將再次移至 RUNNABLE。如需詳細資訊，請參閱自動化任務重試。

> **Note**
>
> RUNNING 任務的日誌可在 CloudWatch Logs 中使用。日誌群組為 /aws/batch/job，日誌串流名稱格式如下：*first200CharsOfJobDefinitionName*/default/*ecs_task_id*。此格式可能會在未來變更。
> 工作進入 RUNNING 狀態時，您可透過編寫程式的方式用 DescribeJobs API 操作擷取其日誌串流名稱。如需詳細資訊，請參閱《Amazon CloudWatch Logs 使用者指南》中的檢視傳送至 CloudWatch Logs 的日誌資料。 Amazon CloudWatch 根據預設，這些日誌永遠不會過期。不過，您可以修改保留期。如需詳細資訊，請參閱《Amazon CloudWatch Logs 使用者指南》中的變更 CloudWatch Logs 中的日誌資料保留。 Amazon CloudWatch

SUCCEEDED

已成功完成任務，結束代碼為 0。任務SUCCEEDED的任務狀態會在 中保留 AWS Batch 至少 7 天。

> **Note**
>
> SUCCEEDED 任務的日誌可在 CloudWatch Logs 中使用。日誌群組為 /aws/batch/job，日誌串流名稱格式如下：*first200CharsOfJobDefinitionName*/default/*ecs_task_id*。此格式可能會在未來變更。
> 工作進入 RUNNING 狀態時，您可透過編寫程式的方式用 DescribeJobs API 操作擷取其日誌串流名稱。如需詳細資訊，請參閱《Amazon CloudWatch Logs 使用者指南》中的檢視傳送至 CloudWatch Logs 的日誌資料。 Amazon CloudWatch 根據預設，這些日誌永遠不會過期。不過，您可以修改保留期。如需詳細資訊，請參閱《Amazon CloudWatch Logs 使用者指南》中的變更 CloudWatch Logs 中的日誌資料保留。 Amazon CloudWatch

FAILED

任務所有的可用嘗試都失敗。任務FAILED的任務狀態會在 中 AWS Batch 保留至少 7 天。

> **Note**
>
> FAILED 任務的日誌可在 CloudWatch Logs 中取得。日誌群組為 /aws/batch/job，日誌串流名稱格式如下：*first200CharsOfJobDefinitionName*/default/*ecs_task_id*。此格式可能會在未來變更。
> 工作進入 RUNNING 狀態時，您可透過編寫程式的方式用 DescribeJobs API 操作擷取其日誌串流。如需詳細資訊，請參閱《Amazon CloudWatch Logs 使用者指南》中的檢視傳送至 CloudWatch Logs 的日誌資料。 Amazon CloudWatch 根據預設，這些日誌永遠不會過期。不過，您可以修改保留期。如需詳細資訊，請參閱《Amazon CloudWatch Logs 使用者指南》中的在 CloudWatch Logs 中變更日誌資料保留。 Amazon CloudWatch

AWS Batch 任務環境變數

AWS Batch 會在容器任務中設定特定環境變數。這些環境變數提供任務內容器的自我檢查。您可以在應用程式的邏輯中使用這些變數的值。所有 AWS Batch 以 AWS_BATCH_字首開頭的變數。這是受保護的環境變數字首。您無法在任務定義或覆寫中將此字首用於您自己的變數。

以下環境變數適用於任務容器：

AWS_BATCH_CE_NAME

此變數設定為放置任務的運算環境名稱。

AWS_BATCH_JOB_ARRAY_INDEX

只會在子陣列任務中設定此變數。陣列任務索引從 0 開始，而且每個子任務會收到一個唯一的索引號碼。例如，含 10 個子系的陣列任務有 0-9 的索引值。您可以使用此索引值，控制您陣列任務子系的區分方式。如需詳細資訊，請參閱使用陣列任務索引來控制任務差異。

AWS_BATCH_JOB_ARRAY_SIZE

此變數設定為父陣列任務的大小。父陣列任務的大小會傳遞至此變數中的子陣列任務。

AWS_BATCH_JOB_ATTEMPT

會將此變數設為任務嘗試號碼。第一次嘗試的編號為 1。如需詳細資訊，請參閱自動化任務重試。

AWS_BATCH_JOB_ID

此變數設定為 AWS Batch 任務 ID。

AWS_BATCH_JOB_KUBERNETES_NODE_UID

此變數設定為 Pod 執行所在的 Kubernetes 叢集中節點物件的 Kubernetes UID。此變數只會針對在 Amazon EKS 資源上執行的任務設定。如需詳細資訊，請參閱 Kubernetes 文件中的 UIDs。

AWS_BATCH_JOB_MAIN_NODE_INDEX

只會在多節點平行任務中設定此變數。會將此變數設為任務主要節點的索引數量。您的應用程式程式碼可以將 AWS_BATCH_JOB_MAIN_NODE_INDEX與個別節點AWS_BATCH_JOB_NODE_INDEX上的 進行比較，以判斷其是否為主節點。

AWS_BATCH_JOB_MAIN_NODE_PRIVATE_IPV4_ADDRESS

此變數僅在多節點平行任務子節點中設定。此變數不存在於主節點，但會設定為任務主節點的私有 IPv4 地址。您的子節點應用程式程式碼可以使用此地址與主節點通訊。

AWS_BATCH_JOB_NODE_INDEX

只會在多節點平行任務中設定此變數。會將此變數設為節點的節點索引數量。節點索引從 0 開始，而且每個節點皆會收到一個唯一的索引號碼。例如，含 10 個子系的多節點平行任務具有 0-9 的索引值。

AWS_BATCH_JOB_NUM_NODES

只會在多節點平行任務中設定此變數。此變數設定為您為多節點平行任務請求的節點數量。

AWS_BATCH_JQ_NAME

會將此變數設為您所提交任務的任務佇列名稱。

自動化任務重試

您可將重試策略套用至任務和任務定義，讓失敗的任務自動重試。可能的失敗案例包括下列項目：

- 容器任務有任何的非零結束代碼
- Amazon EC2 執行個體失敗或終止
- 內部 AWS 服務錯誤或中斷

當任務提交至任務佇列，並進入視為嘗試RUNNING的狀態時。根據預設，每個任務會嘗試一次移至 SUCCEEDED 或 FAILED 任務狀態。不過，任務定義和任務提交工作流程都可以用來指定嘗試 1 到 10 次之間的重試策略。如果指定 evaluateOnExit，則最多可包含 5 個重試策略。如果指定 evaluateOnExit，但沒有任何重試策略相符，則會重試任務。對於不符合以退出的任務，請新增因任何原因退出的最終項目。例如，此evaluateOnExit物件有兩個具有 動作的項目，RETRY以及具有 動作的最終項目EXIT。

```
"evaluateOnExit": [
    {
        "action": "RETRY",
        "onReason": "AGENT"
    },
    {
        "action": "RETRY",
        "onStatusReason": "Task failed to start"
    },
    {
        "action": "EXIT",
        "onReason": "*"
    }
]
```

在執行時間，AWS_BATCH_JOB_ATTEMPT 環境變數設為容器的對應任務嘗試次數。第一次嘗試會編號為 1，後續嘗試會以遞增順序（例如 2、3、4)。

例如，假設任務嘗試因任何原因失敗，且重試組態中指定的嘗試次數大於AWS_BATCH_JOB_ATTEMPT數字。然後，任務會回到 RUNNABLE 狀態。如需詳細資訊，請參閱任務狀態。

> **Note**
>
> 已取消或終止的任務不會重試。此外，不會重試因為任務定義無效而失敗的任務。

如需詳細資訊，請參閱 重試策略、 建立單一節點任務定義教學課程：提交任務和 已停止任務錯誤代碼。

任務相依性

當您提交 AWS Batch 任務時，您可以指定任務所依賴的任務 IDs。當您這樣做時， AWS Batch 排程器會確保您的任務只有在指定的相依性成功完成之後才會執行。成功完成後，相依的任務將從 PENDING 轉為 RUNNABLE，然後轉為 STARTING 和 RUNNING。如果任何任務相依性失敗，相依的任務將自動從 PENDING 轉為 FAILED。

例如，A 任務可以對另外最多 20 個任務有相依性，必須等這 20 個任務成功後才能執行。接著您可以提交額外的任務，對 A 任務和最多 19 個其他的任務有相依性。

對於陣列任務，您可以指定 SEQUENTIAL 類型相依性，且不指定任務 ID，讓每個子陣列任務從索引 0 開始依序完成。您也可以使用任務 ID 指定 N_TO_N 類型相依性。如此一來，此任務的每個索引子系必須等待各相依性對應的索引子系完成後，才能開始。如需詳細資訊，請參閱陣列任務。

若要提交具有相依性 AWS Batch 的任務，請參閱 教學課程：提交任務。

資源感知排程 可讓您根據執行任務所需的消耗性資源來排程任務。您可以指定任務執行所需的消耗性資源，而 Batch 在排程任務時會將這些資源相依性納入考量。您可以只配置具有所有必要資源的任務，以減少運算資源的使用率不足。資源感知排程適用於 FIFO 和公平共用排程政策，並可搭配 Batch 支援的所有運算平台使用，包括 EKS、ECS 和 Fargate。它可以與陣列任務、多節點平行 (MNP) 任務以及一般批次任務搭配使用。

任務逾時

您可以設定任務的逾時時間，如此一來，假如任務執行超過該時間， AWS Batch 便會終止該任務。例如，您可能有一個您知道應該只需要 15 分鐘就能完成的任務。有時您的應用程式會一直卡在迴圈和執行中，因此您可以設定逾時為 30 分鐘以終止卡住的任務。

> **Important**
>
> 根據預設， AWS Batch 沒有任務逾時。如果您未定義任務逾時，任務會執行，直到容器結束為止。

您在任務定義內或是當您提交此任務時指定 `attemptDurationSeconds` 參數，該參數必須至少有 60 秒。當此秒數超過任務嘗試的`startedAt`時間戳記後， 便會 AWS Batch 終止任務。在運算資源時，您的任務容器會收到 `SIGTERM` 訊號，讓您的應用程式有機會正常關閉。如果容器在 30 秒後仍在執行中，則會傳送 `SIGKILL` 訊號以強制關閉容器。

逾時終止會依最佳作法來處理。您不應預期逾時終止會在任務嘗試逾時時發生 （可能需要多幾秒鐘的時間）。如果您的應用程式需要精確執行逾時，您應在應用程式內實作此邏輯。如果您有大量任務同時逾時，逾時終止將採用前進先出佇列，按批次終止任務。

> **Note**
>
> AWS Batch 任務沒有最大逾時值。

如果任務因超過逾時持續時間而終止，則不會重試。如果任務嘗試自行失敗，任務會在啟用重試下進行重試，且進行新嘗試時將重新開始逾時倒數。

> **Important**
>
> 在 Fargate 資源上執行的任務無法預期執行超過 14 天。如果逾時持續時間超過 14 天，則 Fargate 資源可能不再可用，且任務將終止。

對於陣列任務，子任務的逾時設定與父任務相同。

如需使用逾時組態提交 AWS Batch 任務的詳細資訊，請參閱 教學課程：提交任務。

Amazon EKS 任務

任務是工作中最小的單位 AWS Batch。Amazon EKS 上的 AWS Batch 任務具有 one-to-one映射。Kubernetes AWS Batch 任務定義是 AWS Batch 任務的範本。當您提交 AWS Batch 任務時，您可以參考任務定義、將任務佇列設為目標，並提供任務的名稱。在 Amazon EKS AWS Batch 任務

的任務定義中，eksProperties 參數會定義 Amazon EKS 任務 AWS Batch 上 支援的一組參數。在 SubmitJob 請求中，eksPropertiesOverride 參數允許覆寫某些常見參數。如此一來，您就可以針對多個任務使用任務定義的範本。當任務分派到您的 Amazon EKS 叢集時， 會將任務 AWS Batch 轉換為 podspec(Kind: Pod)。podspec 使用一些額外的 AWS Batch 參數來確保任務已正確擴展和排程。AWS Batch 會結合標籤和污點，以確保任務僅在 AWS Batch 受管節點上執行，而且其他 Pod 不會在這些節點上執行。

Important

- 如果未在 Amazon EKS 任務定義中明確設定 hostNetwork 參數，則 中的 Pod 聯網模式 AWS Batch 預設為主機模式。具體而言，會套用下列設定： hostNetwork=true和 dnsPolicy=ClusterFirstWithHostNet。
- AWS Batch 在 Pod 完成其任務後， 會立即清除任務 Pod。若要查看 Pod 應用程式日誌，請為您的叢集設定記錄服務。如需詳細資訊，請參閱使用 CloudWatch Logs 監控 AWS Batch Amazon EKS 任務。

主題

- 教學課程：將執行中的任務映射至 Pod 和節點
- 教學課程：將執行中的 Pod 映射回其任務

教學課程：將執行中的任務映射至 Pod 和節點

執行中任務podProperties的 已為目前的任務嘗試設定 podName和 nodeName 參數。使用 DescribeJobs API 操作來檢視這些參數。

以下為範例輸出。

```
$ aws batch describe-jobs --job 2d044787-c663-4ce6-a6fe-f2baf7e51b04
{
 "jobs": [
  {
   "status": "RUNNING",
   "jobArn": "arn:aws:batch:us-east-1:123456789012:job/2d044787-c663-4ce6-a6fe-
f2baf7e51b04",
   "jobDefinition": "arn:aws:batch:us-east-1:123456789012:job-definition/
MyJobOnEks_SleepWithRequestsOnly:1",
   "jobQueue": "arn:aws:batch:us-east-1:123456789012:job-queue/My-Eks-JQ1",
```

```
   "jobId": "2d044787-c663-4ce6-a6fe-f2baf7e51b04",
   "eksProperties": {
    "podProperties": {
     "nodeName": "ip-192-168-55-175.ec2.internal",
     "containers": [
      {
       "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
       "resources": {
        "requests": {
         "cpu": "1",
         "memory": "1024Mi"
        }
       }
      }
     ],
     "podName": "aws-batch.b0aca953-ba8f-3791-83e2-ed13af39428c"
    }
   }
  }
 ]
}
```

對於啟用重試的任務，每次完成嘗試nodeName的 podName和 都在 DescribeJobs API 操作的eksAttempts清單參數中。目前執行中嘗試nodeName的 podName和 位於 podProperties 物件中。

教學課程：將執行中的 Pod 映射回其任務

Pod 的標籤會指出其所屬運算環境uuid的 jobId和 。 會 AWS Batch 注入環境變數，讓任務的執行時間可以參考任務資訊。如需詳細資訊，請參閱AWS Batch 任務環境變數。您可以執行下列命令來檢視此資訊。輸出如下。

```
$ kubectl describe pod aws-batch.14638eb9-d218-372d-ba5c-1c9ab9c7f2a1 -n my-aws-batch-namespace
Name:         aws-batch.14638eb9-d218-372d-ba5c-1c9ab9c7f2a1
Namespace:    my-aws-batch-namespace
Priority:     0
Node:         ip-192-168-45-88.ec2.internal/192.168.45.88
Start Time:   Wed, 26 Oct 2022 00:30:48 +0000
Labels:       batch.amazonaws.com/compute-environment-uuid=5c19160b-
d450-31c9-8454-86cf5b30548f
              batch.amazonaws.com/job-id=f980f2cf-6309-4c77-a2b2-d83fbba0e9f0
```

```
              batch.amazonaws.com/node-uid=a4be5c1d-9881-4524-b967-587789094647
...
Status:       Running
IP:           192.168.45.88
IPs:
  IP:  192.168.45.88
Containers:
  default:
    Image:         public.ecr.aws/amazonlinux/amazonlinux:2
    ...
    Environment:
      AWS_BATCH_JOB_KUBERNETES_NODE_UID:  a4be5c1d-9881-4524-b967-587789094647
      AWS_BATCH_JOB_ID:                   f980f2cf-6309-4c77-a2b2-d83fbba0e9f0
      AWS_BATCH_JQ_NAME:                  My-Eks-JQ1
      AWS_BATCH_JOB_ATTEMPT:              1
      AWS_BATCH_CE_NAME:                  My-Eks-CE1

...
```

Amazon AWS Batch EKS 任務支援的功能

以下是在 Amazon EKS 上執行Kubernetes的任務也常見的 AWS Batch 特定功能：

- 任務相依性
- 陣列任務
- 任務逾時
- 自動化任務重試
- 使用公平共享排程來協助排程任務

Kubernetes`Secrets` 和 `ServiceAccounts`

AWS Batch 支援參考 Kubernetes`Secrets`和 `ServiceAccounts`。您可以設定 Pod 為服務帳戶使用 Amazon EKS IAM 角色。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的設定 Pod 以使用 Kubernetes服務帳戶。

相關文件

- Amazon EKS AWS Batch 上的 記憶體和 vCPU 考量事項
- 執行 GPU 任務
- 任務停滯在 RUNNABLE 狀態

多節點平行任務

您可以使用多節點平行任務來執行跨越多個 Amazon EC2 執行個體的單一任務。透過 AWS Batch 多節點平行任務 （ 也稱為 Gang 排程)，您可以執行大規模的高效能運算應用程式和分散式 GPU 模型訓練，而不需要直接啟動、設定和管理 Amazon EC2 資源。 AWS Batch 多節點平行任務與支援 IP 型節點間通訊的任何架構相容。範例包括 Apache MXNet、TensorFlow、Caffe2 或訊息傳遞界面 (MPI)。

多節點平行任務會以單一任務的形式提交。不過，您的任務定義 (或任務提交節點覆寫) 會指定要為任務或哪些節點群組建立的節點數量。每個多節點平行任務皆包含會最先啟動的主要節點。主節點啟動後，會啟動和啟動子節點。只有在主節點結束時，任務才會完成。接著會停止所有子節點。如需詳細資訊，請參閱節點群組。

多節點平行任務節點是單一租用戶。這表示每個 Amazon EC2 執行個體只會執行單一任務容器。

最終任務狀態 (SUCCEEDED 或 FAILED) 取決主要節點的最終任務狀態。若要取得多節點平行任務的狀態，請使用提交任務時傳回的任務 ID 來描述任務。如果您需要子節點的詳細資訊，請個別描述每個子節點。您可以使用 *#N* 表示法來定址節點 （ 以 0 開頭 ）。例如，若要存取任務第二個節點的詳細資訊，請使用 AWS Batch DescribeJobs *API ####* *aws_batch_job_id*#1。`started`、`stoppedAt`、`statusReason` 和 `exit` 多節點平行任務的資訊，將從主要節點填入。

如果您指定任務重試，主節點失敗會導致再次嘗試。子節點故障不會造成更多嘗試。每次新嘗試的多節點平行任務，皆會更新該嘗試所關聯的子節點。

若要在 上執行多節點平行任務 AWS Batch，您的應用程式程式碼必須包含分散式通訊所需的架構和程式庫。

主題

- 環境變數
- 節點群組
- MNP 任務的任務生命週期
- 使用 的 MNP 運算環境考量 AWS Batch

環境變數

在執行時間，每個節點都會設定所有 AWS Batch 任務接收的標準環境變數。此外，節點會設定下列環境變數，這些變數專屬於多節點平行任務：

AWS_BATCH_JOB_MAIN_NODE_INDEX

會將此變數設為任務主要節點的索引數量。您的應用程式程式碼可以將AWS_BATCH_JOB_MAIN_NODE_INDEX與個別節點AWS_BATCH_JOB_NODE_INDEX上的 進行比較，以判斷其是否為主節點。

AWS_BATCH_JOB_MAIN_NODE_PRIVATE_IPV4_ADDRESS

此變數僅在多節點平行任務子節點中設定。此變數不存在於主節點。會將此變數設為任務主要節點的私有 IPv4 地址。您的子節點應用程式程式碼可以使用此地址與主節點通訊。

AWS_BATCH_JOB_NODE_INDEX

會將此變數設為節點的節點索引數量。節點索引從 0 開始，而且每個節點皆會收到一個唯一的索引號碼。例如，含 10 個子系的多節點平行任務具有 0-9 的索引值。

AWS_BATCH_JOB_NUM_NODES

會將此變數設為您為多節點平行任務請求的節點數量。

節點群組

節點群組是一組相同的任務節點，這些節點都共用相同的容器屬性。您可以使用 AWS Batch 為每個任務指定最多五個不同的節點群組。

每個群組可以有自己的容器映像、命令、環境變數，以此類推。例如，您可以提交需要主節點和五個c5.xlarge執行個體子節點單一c5.xlarge執行個體的任務。每個不同的節點群組可以指定要為每個任務執行的不同容器映像或命令。

或者，任務中的所有節點都可以使用單一節點群組。此外，您的應用程式程式碼可以區分節點角色，例如主節點和子節點。它透過將AWS_BATCH_JOB_MAIN_NODE_INDEX環境變數與其自身的 值進行比較來執行此操作AWS_BATCH_JOB_NODE_INDEX。單一任務中最多可以有 1，000 個節點。這是 Amazon ECS 叢集中執行個體的預設限制。您可以請求提高此限制。

> **Note**
>
> 目前，多節點平行任務中的所有節點群組都必須使用相同的執行個體類型。

MNP 任務的任務生命週期

當您提交多節點平行任務時，任務會進入 SUBMITTED 狀態。然後，任務會等待任何任務相依性完成。任務也會移至 RUNNABLE 狀態。最後， AWS Batch 會佈建執行任務所需的執行個體容量，並啟動這些執行個體。

每個多節點平行任務皆包含主要節點。主節點是單一子任務，可 AWS Batch 監控 來判斷提交的多節點任務的結果。主要節點最先啟動，然後會移至 STARTING 狀態。`attemptDurationSeconds` 參數中指定的逾時值會套用至整個任務，而非節點。

當主節點在節點的容器執行後達到 RUNNING 狀態時，會啟動子節點，也會移至 STARTING 狀態。子節點會以隨機順序出現。子節點的啟動時間或順序並不固定。為了確保任務的所有節點在節點的容器執行後都處於 RUNNING 狀態，您的應用程式程式碼可以查詢 AWS Batch API 以取得主節點和子節點資訊。或者，應用程式程式碼可以等到所有節點上線，再開始任何分散式處理任務。主要節點的私有 IP 地址，在每個子節點中可做為 `AWS_BATCH_JOB_MAIN_NODE_PRIVATE_IPV4_ADDRESS` 環境變數使用。您的應用程式程式碼可以使用此資訊，對各任務之間的資料進行協調和通訊。

隨著個別節點結束，它們將移至 SUCCEEDED 或 FAILED，這取決於它們的結束程式碼。如果主要節點結束，則任務會視為完成，而所有子節點也會停止。如果子節點死亡， AWS Batch 不會對任務中的其他節點採取任何動作。如果您不希望任務繼續減少節點數量，則必須將此納入應用程式程式碼。這樣做會終止或取消任務。

使用 的 MNP 運算環境考量 AWS Batch

在設定使用 AWS Batch執行多節點平行任務的運算環境時，有幾點需要考慮。

- UNMANAGED 運算環境不支援多節點平行任務。
- 如果您想要將多節點平行任務提交至運算環境，請在單一可用區域中建立叢集置放群組，並將其與您的運算資源建立關聯。這可讓執行個體邏輯分組上的多節點平行任務接近高網路流程潛力。如需詳細資訊，請參閱《Amazon EC2 使用者指南》中的置放群組。
- 使用 Spot 執行個體的運算環境不支援多節點平行任務。
- AWS Batch 多節點平行任務使用 Amazon ECS awsvpc 網路模式，為您的多節點平行任務容器提供與 Amazon EC2 執行個體相同的聯網屬性。每個多節點平行任務容器皆會取得自己的彈性網路界面、主要私有 IP 地址及內部 DNS 主機名稱。網路界面是在與託管運算資源相同的 VPC 子網路中所建立。
- 您的運算環境可能沒有超過五個與其相關聯的安全群組。建立並連接至 MNP 任務的彈性網路介面將使用運算環境中指定的安全群組，如果您未指定安全群組，則會使用 VPC 的預設安全群組。

- awsvpc 網路模式不會為具有公有 IP 地址的多節點平行任務提供彈性網路介面。若要存取網際網路，您的運算資源必須必須在設定為使用 NAT 閘道的私有子網路中啟動。如需詳細資訊，請參閱 Amazon VPC 使用者指南中的 NAT 閘道。節點間通訊必須使用節點的私有 IP 地址或 DNS 主機名稱。在公有子網路內的運算資源上執行的多節點平行任務沒有傳出網路存取權。若要建立含私有子網路和 NAT 閘道的 VPC，請參閱建立 Virtual Private Cloud。
- 您的 帳戶無法手動分離或修改建立並連接至運算資源的彈性網路介面。這是為了防止意外刪除與執行中任務相關聯的彈性網路界面。若要釋出任務的彈性網路界面，請終止任務。
- 您的運算環境必須具有足夠的最大 vCPU，以支援您的多節點平行任務。
- 您的 Amazon EC2 執行個體配額包含執行任務所需的執行個體數量。例如，假設您的任務需要 30 個執行個體，但您的帳戶只能在區域中執行 20 個執行個體。然後，您的任務將卡在 RUNNABLE 狀態。
- 如果您在多節點平行任務中指定節點群組的執行個體類型，您的運算環境必須啟動該執行個體類型。

Amazon EKS 上的多節點平行任務

您可以在 Amazon Elastic Kubernetes Service AWS Batch 上使用 ，在受管Kubernetes叢集上執行多節點平行 (MNP) 任務 （ 也稱為 Gang 排程)。此選項通常用於無法在單一 Amazon Elastic Compute Cloud 執行個體上執行的大型、緊密耦合、高效能任務。如需詳細資訊，請參閱多節點平行任務。

您可以使用此功能來執行 Amazon EKS 受管 Kubernetes特定的高效能運算應用程式、大型語言模型訓練，以及其他人工智慧 (AI)/Machine Learning(ML) 任務。

主題

- 執行 MNP 任務
- 建立 Amazon EKS MNP 任務定義
- 提交 Amazon EKS MNP 任務
- 覆寫 Amazon EKS MNP 任務定義

執行 MNP 任務

AWS Batch 支援使用 Amazon EC2 的 Amazon Elastic Container Service 和 Amazon EKS 上的 MNP 任務。以下提供有關 功能的執行個體和容器參數的更多詳細資訊。

Amazon EKS 上 MNP 的執行個體配額

- 單一 MNP 任務最多可使用 1000 個執行個體。

- 最多 5000 個執行個體可以加入單一 Amazon EKS 叢集。
- 最多可將 5 個運算環境叢集並連接至任務佇列。

例如，您可以在任務佇列中擴展最多 5 個叢集運算環境，在每個運算環境中擴展最多 1000 個執行個體。

除了執行個體參數之外，請務必注意，您無法透過任一服務將 Fargate 用於 MNP 任務。

每個 MNP 任務只能使用一種執行個體類型。您可以透過更新運算環境或定義新的運算環境來變更執行個體類型。您也可以指定執行個體類型，並在建立任務定義時提供 vCPU 和記憶體需求。

Amazon EKS 上 MNP 的容器配額

- 多節點平行任務支援每個節點一個 Pod。
- 最多 10 個容器（或 10 個 init 容器。 如需詳細資訊，請參閱每個 Pod 中的 Init Containers。）。
- 每個 MNP 任務中最多 5 個節點範圍。
- 每個節點範圍內最多 10 個不同的容器映像。

例如，在包含 5 個節點範圍和總共 50 個唯一映像的單一 MNP 任務中，您最多可以執行 10，000 個容器。

在私有 Amazon VPC 和 Amazon EKS 叢集中執行 MNP 任務

MNP 任務可以在任何 Amazon EKS 叢集上執行，無論是否有公有網際網路。使用僅具有私有網路存取的 Amazon EKS 叢集時，請確定 AWS Batch 可以存取 Amazon EKS 控制平面和受管 Kubernetes API 伺服器。您可以透過 Amazon Virtual Private Cloud 端點授予必要的存取權。如需詳細資訊，請參閱設定端點服務。

Amazon EKS 叢集 Pod 無法從公有來源下載映像，因為私有 VPC 無法存取網際網路。Amazon EKS 叢集必須從 Amazon VPC 內的容器登錄檔提取映像。您可以在 Amazon VPC 中建立 Amazon Elastic Container Registry (Amazon ECR)，並將容器映像複製到其中以供節點存取。

您也可以使用 Amazon ECR 建立提取快取規則。為外部公有登錄檔建立提取快取規則後，您可以使用 Amazon ECR 私有登錄檔 URI 從該外部公有登錄檔提取映像。然後，Amazon ECR 會建立儲存庫並快取映像。使用 Amazon ECR 私有登錄檔 URI 提取快取的映像時，Amazon ECR 會檢查遠端登錄檔是否有新的映像版本，並且最多每 24 小時更新一次私有登錄檔。如需詳細資訊，請參閱在 Amazon ECR 中建立提取快取規則。

錯誤通知

如果您的 MNP 任務遭到封鎖，您可以透過 AWS 管理主控台 和 Amazon EventBridge 接收通知。例如，如果 MNP 任務卡在佇列的前端，您可以收到有關問題的通知，以及原因的資訊，以便您可以採取提示動作來解除封鎖任務佇列。或者，如果在不同的時間內沒有採取任何動作，您可以自動終止 MNP 任務，這可以在任務佇列範本中定義。如需詳細資訊，請參閱任務佇列封鎖事件

建立 Amazon EKS MNP 任務定義

若要在 Amazon EKS 上定義和執行 MNP 任務， RegisterJobDefinition和 SubmitJob API 操作中有新的參數。

- 在 nodeProperties區段eksProperties下使用 來定義 MNP 任務定義。
- 在提交 MNP 任務時，使用 nodePropertyOverrides區段eksPropertiesOverride下的 覆寫任務定義中定義的參數。

這些動作可以透過 API 操作和 定義 AWS 管理主控台。

參考：註冊 Amazon EKS MNP 任務定義請求承載

下列範例說明如何向兩個節點註冊 Amazon EKS MNP 任務定義。

```
{
  "jobDefinitionName": "MyEksMnpJobDefinition",
  "type": "multinode",
  "nodeProperties": {
    "numNodes": 2,
    "mainNode": 0,
    "nodeRangeProperties": [
      {
        "targetNodes" : "0:",
        "eksProperties": {
          "podProperties": {
            "containers": [
              {
                "name": "test-eks-container-1",
                "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
                "command": [
                  "sleep",
                  "60"
                ],
```

```
                "resources": {
                  "limits": {
                    "cpu": "1",
                    "memory": "1024Mi"
                  }
                },
                "securityContext":{
                  "runAsUser":1000,
                  "runAsGroup":3000,
                  "privileged":true,
                  "readOnlyRootFilesystem":true,
                  "runAsNonRoot":true
               }
              }
            ],
            "initContainers": [
               {
                  "name":"init-ekscontainer",
                  "image": "public.ecr.aws/amazonlinux/amazonlinux:2",
                  "command": [
                     "echo",
                     "helloWorld"
                   ],
                   "resources": {
                     "limits": {
                       "cpu": "1",
                       "memory": "1024Mi"
                     }
                  }
               }
            ],
            "metadata": {
               "labels": {
                  "environment" : "test"
               }
            }
          }
        }
      }
    ]
  }
}
```

若要使用 註冊任務定義 AWS CLI，請將定義複製到名為 MyEksMnpJobDefinition.json 的本機檔案，然後執行下列命令。

```
aws batch register-job-definition --cli-input-json file://MyEksMnpJobDefinition.json
```

您將會收到下列 JSON 回應。

```
{
    "jobDefinitionName": "MyEksMnpJobDefinition",
    "jobDefinitionArn": "arn:aws:batch:us-east-1:0123456789:job-definition/
MyEksMnpJobDefinition:1",
    "revision": 1
}
```

提交 Amazon EKS MNP 任務

若要使用已註冊的任務定義提交任務，請輸入下列命令。將 <EKS_JOB_QUEUE_NAME> 的值取代為與 Amazon EKS 運算環境相關聯的預先存在任務佇列的名稱或 ARN。

```
aws batch submit-job --job-queue <EKS_JOB_QUEUE_NAME> \
    --job-definition MyEksMnpJobDefinition \
    --job-name myFirstEksMnpJob
```

您將會收到下列 JSON 回應。

```
{
    "jobArn": "arn:aws:batch:region:account:job/9b979cce-9da0-446d-90e2-ffa16d52af68",
    "jobName": "myFirstEksMnpJob",
    "jobId": "<JOB_ID>"
}
```

您可以使用傳回的 jobId 搭配下列命令來檢查任務的狀態。

```
aws batch describe-jobs --jobs <JOB_ID>
```

覆寫 Amazon EKS MNP 任務定義

或者，您可以覆寫任務定義詳細資訊（例如變更 MNP 任務大小或子任務詳細資訊）。以下提供提交五個節點 MNP 任務的範例 JSON 請求承載，以及`test-eks-container-1`容器命令的變更。

```
{
  "numNodes": 5,
  "nodePropertyOverrides": [
    {
      "targetNodes": "0:",
      "eksPropertiesOverride": {
        "podProperties": {
          "containers": [
            {
              "name": "test-eks-container-1",
              "command": [
                "sleep",
                "150"
              ]
            }
          ]
        }
      }
    }
  ]
}
```

若要提交具有這些覆寫的任務，請將範例儲存至本機檔案 eks-mnp-job-nodeoverride.json，然後使用 AWS CLI 提交具有覆寫的任務。

陣列任務

陣列任務為共用常見參數的任務，例如任務定義、vCPU 和記憶體。它以一組相關但獨立的基本任務的形式執行，這些任務可能會分散在多個主機，並且可能會同時執行。陣列任務是執行超平行任務的最有效方式，例如 Monte Carlo 模擬、參數掃描或大型轉譯任務。

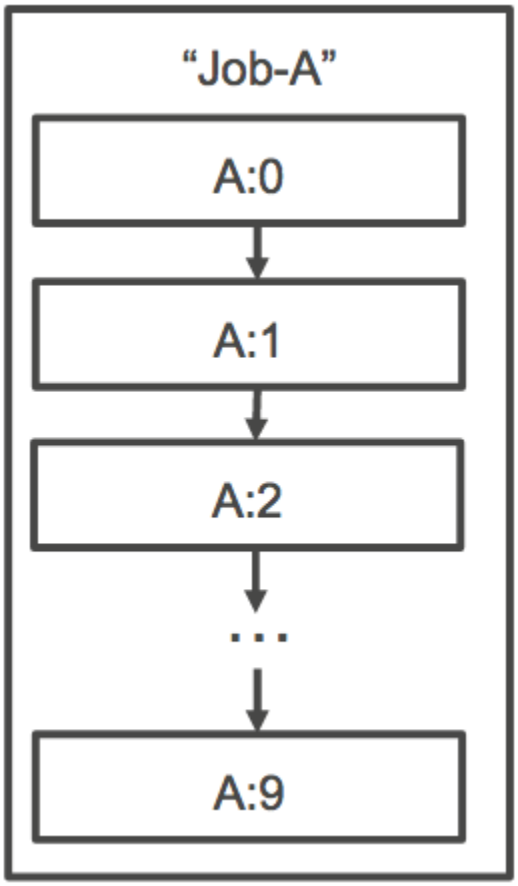
AWS Batch 陣列任務的提交方式與一般任務相同。不過，您必須指定陣列大小 (2 至 10,000)，以定義陣列內應該執行的子任務數量。如果您提交陣列大小 1000 的任務，單一任務將執行並產生 1000 個子任務。陣列任務為參考或指標，用於管理所有的子任務。如此一來，您可以使用單一查詢來提交大型工作負載。`attemptDurationSeconds` 參數中指定的逾時會套用至每個子任務。父陣列任務沒有逾時。

當您提交陣列任務時，父陣列任務會取得正常 AWS Batch 的任務 ID。每個子任務都有相同的基本 ID。不過，子任務的陣列索引會附加到父 ID 的結尾，例如*`example_job_ID`*`:0`陣列的第一個子任務。

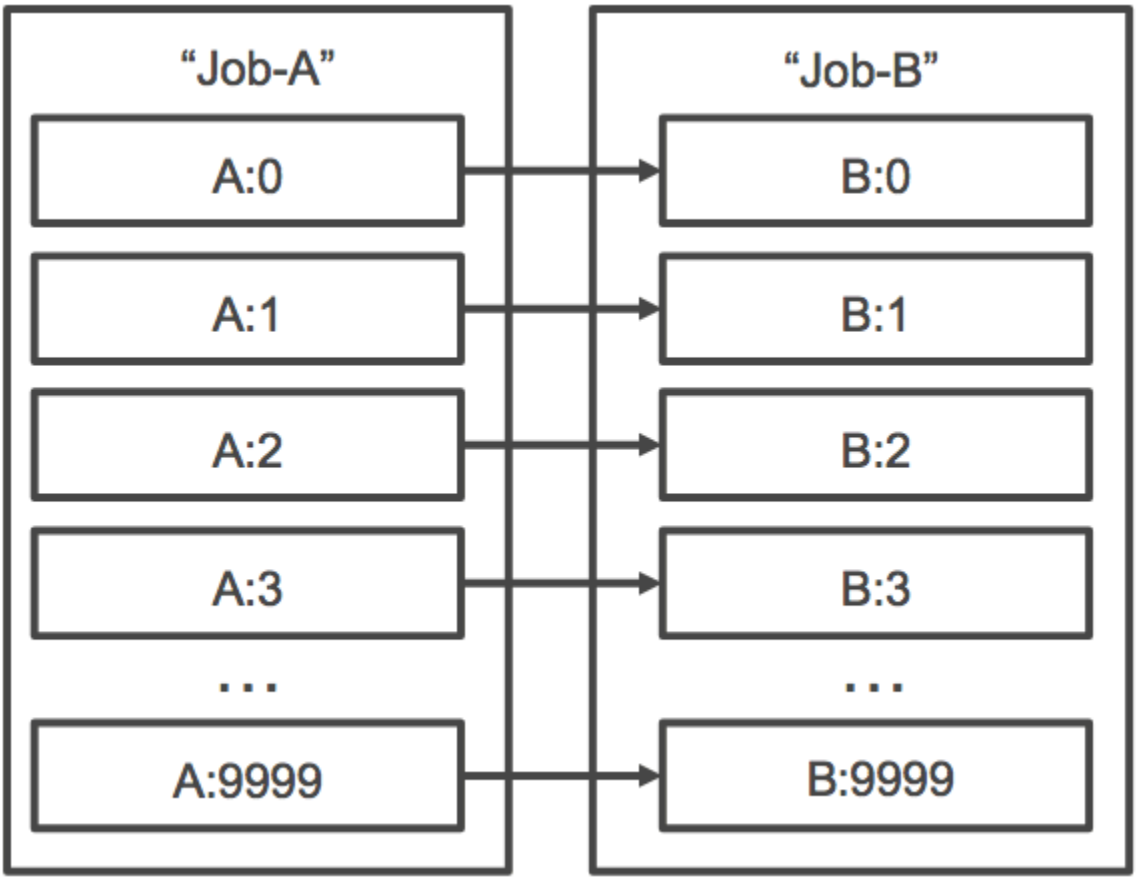
父陣列任務可以輸入 SUBMITTED、FAILED、 PENDING或 SUCCEEDED 狀態。當任何子任務更新為 PENDING時，陣列父任務會更新為 RUNNABLE。如需任務相依性的詳細資訊，請參閱 任務相依性。

在執行時間，AWS_BATCH_JOB_ARRAY_INDEX 環境變數設為容器的對應任務陣列索引編號。第一個陣列任務索引編號為 0，後續嘗試會以遞增順序 （ 例如 1、2 和 3)。您可以使用此索引值，控制您陣列任務子系的區分方式。如需詳細資訊，請參閱使用陣列任務索引來控制任務差異。

對於陣列任務的相依性，您可以指定相依性類型，例如 SEQUENTIAL 或 N_TO_N。您可以指定 SEQUENTIAL 類型相依性 (不指定任務 ID)，讓每個子陣列任務從索引 0 開始依序完成。例如，如果您提交陣列大小 100 的陣列任務，並指定 SEQUENTIAL 類型的相依性，後續將產生 100 個子任務，必須等第一個子任務完成後，下一個子任務才會開始。下圖顯示 A 任務，陣列大小 10 的陣列任務。A 任務子索引的每個任務都相依於前一個子任務。A:1 任務必須等 A:0 任務完成後才會開始。



您也可以指定 N_TO_N 類型相依性，以及陣列任務的任務 ID。如此一來，此任務的每個索引子系必須等待各相依性對應的索引子系完成後，才能開始。下圖顯示任務 A 和任務 B，兩個陣列任務的每個陣列大小為 10，000。B 任務子索引的每個任務相依於 A 任務的對應索引。B:1 任務必須等到 A:1 任務完成後才會開始。



如果您取消或終止父陣列任務，所有子任務都會隨之取消或終止。您可以取消或終止個別子任務（將它們移至 FAILED 狀態），而不會影響其他子任務。不過，如果子陣列任務失敗（單獨或透過手動取消或終止任務），父任務也會失敗。在此案例中，當所有子任務完成FAILED時，父任務會轉換為 。

如需搜尋和篩選陣列任務的詳細資訊，請參閱搜尋任務佇列中的任務。

主題

- 陣列任務工作流程的範例
- 使用陣列任務索引來控制任務差異

陣列任務工作流程的範例

AWS Batch 客戶的常見工作流程是執行先決條件設定任務、針對大量輸入任務執行一系列命令，然後使用彙總結果並將摘要資料寫入 Amazon S3、DynamoDB、Amazon Redshift 或 Aurora 的任務結束。

例如：

- JobA：執行 Amazon S3 儲存貯體中物件快速清單和中繼資料驗證的標準非陣列任務BucketA。SubmitJob JSON 語法如下所示。

```
{
    "jobName": "JobA",
    "jobQueue": "ProdQueue",
    "jobDefinition": "JobA-list-and-validate:1"
}
```

- JobB：具有 10，000 個副本的陣列任務，這些副本依賴 JobA 對 中的每個物件執行 CPU 密集型命令，BucketA並將結果上傳至 BucketB。SubmitJob JSON 語法如下所示。

```
{
    "jobName": "JobB",
    "jobQueue": "ProdQueue",
    "jobDefinition": "JobB-CPU-Intensive-Processing:1",
    "containerOverrides": {
        "resourceRequirements": [
            {
                "type": "MEMORY",
                "value": "4096"
            },
            {
                "type": "VCPU",
                "value": "32"
            }
        ]
   }
    "arrayProperties": {
        "size": 10000
    },
    "dependsOn": [
        {
            "jobId": "JobA_job_ID"
  }
    ]
}
```

- JobC：另一個JobB與N_TO_N相依性模型相依的 10，000 個複製陣列任務，對 中的每個項目執行記憶體密集型命令BucketB、將中繼資料寫入 DynamoDB，並將產生的輸出上傳到 BucketC。SubmitJob JSON 語法如下所示。

```
{
    "jobName": "JobC",
    "jobQueue": "ProdQueue",
    "jobDefinition": "JobC-Memory-Intensive-Processing:1",
    "containerOverrides": {
        "resourceRequirements": [
            {
                "type": "MEMORY",
                "value": "32768"
            },
            {
                "type": "VCPU",
                "value": "1"
            }
        ]
   }
    "arrayProperties": {
        "size": 10000
    },
    "dependsOn": [
        {
            "jobId": "JobB_job_ID",
            "type": "N_TO_N"
        }
    ]
}
```

- JobD：執行 10 個驗證步驟的陣列任務，每個步驟都需要查詢 DynamoDB，並且可能與上述任何 Amazon S3 儲存貯體互動。中的每個步驟都會JobD執行相同的命令。不過，行為會根據任務容器中AWS_BATCH_JOB_ARRAY_INDEX環境變數的值而有所不同。這些驗證步驟會依序執行（例如，JobD:0和 JobD:1)。SubmitJob JSON 語法如下所示。

```
{
    "jobName": "JobD",
    "jobQueue": "ProdQueue",
    "jobDefinition": "JobD-Sequential-Validation:1",
    "containerOverrides": {
        "resourceRequirements": [
            {
                "type": "MEMORY",
                "value": "32768"
            },
```

```
            {
                "type": "VCPU",
                "value": "1"
            }
        ]
   }
    "arrayProperties": {
        "size": 10
    },
    "dependsOn": [
        {
            "jobId": "JobC_job_ID"
        },
        {
            "type": "SEQUENTIAL"
        },

    ]
}
```

- JobE：執行一些簡單清除操作並傳送 Amazon SNS 通知的最終非陣列任務，其中包含管道已完成的訊息和輸出 URL 的連結。SubmitJob JSON 語法如下所示。

```
{
    "jobName": "JobE",
    "jobQueue": "ProdQueue",
    "jobDefinition": "JobE-Cleanup-and-Notification:1",
    "parameters": {
        "SourceBucket": "s3://amzn-s3-demo-source-bucket",
        "Recipient": "pipeline-notifications@mycompany.com"
    },
    "dependsOn": [
        {
            "jobId": "JobD_job_ID"
        }
    ]
}
```

使用陣列任務索引來控制任務差異

本教學說明如何使用 AWS_BATCH_JOB_ARRAY_INDEX環境變數來區分子任務。每個子任務都會指派給此變數。此範例使用子任務的索引號碼來讀取檔案中的特定行。然後，它會將與該行號相關聯的參數

替換為任務容器內的命令。結果是，您可以有多個執行相同 Docker 映像和命令引數 AWS Batch 的任務。不過，結果不同，因為陣列任務索引是用作修飾詞。

在此教學課程中，您可以建立一個含有彩虹中所有顏色的文字檔案，每個顏色各為一行。然後，您可以為 Docker 容器建立進入點指令碼，將索引轉換為可用於顏色檔案中行號的值。索引從零開始，但行號從一開始。建立 Dockerfile，將顏色和索引檔案複製到容器映像，並將映像`ENTRYPOINT`的設定設為進入點指令碼。Dockerfile 和資源會建置到推送到 Amazon ECR 的 Docker 映像。然後，您可以註冊使用新容器映像的任務定義、使用該任務定義提交 AWS Batch 陣列任務，以及檢視結果。

主題

- 先決條件
- 建置容器映像
- 將您的映像推送至 Amazon ECR
- 建立並註冊任務定義
- 提交 AWS Batch 陣列任務
- 檢視您的陣列任務日誌

先決條件

本教學課程工作流程具有下列先決條件：

- AWS Batch 運算環境。如需詳細資訊，請參閱建立運算環境。
- AWS Batch 任務佇列和相關聯的運算環境。如需詳細資訊，請參閱建立任務佇列。
- AWS CLI 安裝在本機系統的 。如需詳細資訊，請參閱AWS Command Line Interface 《 使用者指南》中的 >安裝 AWS Command Line Interface。
- 安裝在本機系統的 Docker。如需詳細資訊，請參閱 Docker 文件中的關於 Docker CE。

建置容器映像

您可以在 命令參數`AWS_BATCH_JOB_ARRAY_INDEX`的任務定義中使用 。不過，我們建議您建立在進入點指令碼中使用 變數的容器映像。本節說明如何建立此類容器映像。

建置 Docker 容器影像

1. 建立新的目錄做為您的 Docker 影像工作空間，然後瀏覽至該目錄。
2. 在`colors.txt`工作區目錄中建立名為 的檔案，並將以下內容貼入其中。

```
red
orange
yellow
green
blue
indigo
violet
```

3. print-color.sh 在您的工作區目錄中建立名為 的檔案，並將以下內容貼入其中。

> **Note**
>
> LINE 變數設定為 AWS_BATCH_JOB_ARRAY_INDEX + 1，因為陣列索引起始為 0，但行號從 1 開始。COLOR 變數會設定為 中與其行號colors.txt相關聯的顏色。

```
#!/bin/sh
LINE=$((AWS_BATCH_JOB_ARRAY_INDEX + 1))
COLOR=$(sed -n ${LINE}p /tmp/colors.txt)
echo My favorite color of the rainbow is $COLOR.
```

4. 在Dockerfile工作區目錄中建立名為 的檔案，並將下列內容貼入其中。此 Dockerfile 會將之前的檔案複製到您的容器，並將進入點指令碼設定為在啟動容器時執行。

```
FROM busybox
COPY print-color.sh /tmp/print-color.sh
COPY colors.txt /tmp/colors.txt
RUN chmod +x /tmp/print-color.sh
ENTRYPOINT /tmp/print-color.sh
```

5. 建置 Docker 映像。

```
$ docker build -t print-color .
```

6. 使用以下指令碼測試容器。此指令碼會在本機將AWS_BATCH_JOB_ARRAY_INDEX變數設定為 0，然後遞增以模擬具有七個子項的陣列任務。

```
$ AWS_BATCH_JOB_ARRAY_INDEX=0
while [ $AWS_BATCH_JOB_ARRAY_INDEX -le 6 ]
do
```

```
    docker run -e AWS_BATCH_JOB_ARRAY_INDEX=$AWS_BATCH_JOB_ARRAY_INDEX print-color
    AWS_BATCH_JOB_ARRAY_INDEX=$((AWS_BATCH_JOB_ARRAY_INDEX + 1))
done
```

以下為其輸出。

```
My favorite color of the rainbow is red.
My favorite color of the rainbow is orange.
My favorite color of the rainbow is yellow.
My favorite color of the rainbow is green.
My favorite color of the rainbow is blue.
My favorite color of the rainbow is indigo.
My favorite color of the rainbow is violet.
```

將您的映像推送至 Amazon ECR

現在您已建置並測試 Docker 容器，請將其推送至映像儲存庫。此範例使用 Amazon ECR，但您可以使用其他登錄檔，例如 DockerHub。

1. 建立 Amazon ECR 映像儲存庫以存放容器映像。此範例僅使用 AWS CLI，但您也可以使用 AWS 管理主控台。如需詳細資訊，請參閱《Amazon Elastic Container Registry 使用者指南》中的建立儲存庫。

   ```
   $ aws ecr create-repository --repository-name print-color
   ```

2. 使用上一個步驟傳回的 Amazon ECR 儲存庫 URI 標記您的`print-color`映像。

   ```
   $ docker tag print-color aws_account_id.dkr.ecr.region.amazonaws.com/print-color
   ```

3. 登入 Amazon ECR 登錄檔。如需詳細資訊，請參閱《Amazon Elastic Container Registry 使用者指南》中的登錄檔身分驗證。

   ```
   $ aws ecr get-login-password \
       --region region | docker login \
       --username AWS \
       --password-stdin aws_account_id.dkr.ecr.region.amazonaws.com
   ```

4. 將您的映像推送至 Amazon ECR。

   ```
   $ docker push aws_account_id.dkr.ecr.region.amazonaws.com/print-color
   ```

建立並註冊任務定義

現在您的 Docker 映像位於映像登錄檔中，您可以在 AWS Batch 任務定義中指定它。然後，您可以稍後使用它來執行陣列任務。此範例僅使用 AWS CLI。不過，您也可以使用 AWS 管理主控台。如需詳細資訊，請參閱建立單一節點任務定義。

建立任務定義

1. 在print-color-job-def.json工作區目錄中建立名為 的檔案，並將以下內容貼入其中。將映像儲存庫 URI 取代為您自己的映像 URI。

```
{
  "jobDefinitionName": "print-color",
  "type": "container",
  "containerProperties": {
    "image": "aws_account_id.dkr.ecr.region.amazonaws.com/print-color",
    "resourceRequirements": [
        {
            "type": "MEMORY",
            "value": "250"
        },
        {
            "type": "VCPU",
            "value": "1"
        }
    ]
  }
}
```

2. 向 註冊任務定義 AWS Batch。

```
$ aws batch register-job-definition --cli-input-json file://print-color-job-
def.json
```

提交 AWS Batch 陣列任務

註冊任務定義後，您可以提交使用新容器映像的 AWS Batch 陣列任務。

提交 AWS Batch 陣列任務

1. 在print-color-job.json工作區目錄中建立名為 的檔案，並將以下內容貼入其中。

> **Note**
>
> 此範例使用 the section called “先決條件”區段中提到的任務佇列。

```
{
  "jobName": "print-color",
  "jobQueue": "existing-job-queue",
  "arrayProperties": {
    "size": 7
  },
  "jobDefinition": "print-color"
}
```

2. 將任務提交到您的 AWS Batch 任務佇列。請注意輸出中傳回的任務 ID。

```
$ aws batch submit-job --cli-input-json file://print-color-job.json
```

3. 描述任務的狀態並等待任務移至 SUCCEEDED。

檢視您的陣列任務日誌

任務達到 SUCCEEDED 狀態後，您可以從任務的容器檢視 CloudWatch Logs。

在 CloudWatch Logs 中檢視任務的日誌

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 在左側導覽窗格中，選擇 Jobs (任務)。
3. 對於 Job queue (任務佇列)，請選取佇列。
4. 在 Status (狀態) 區段，選擇 succeeded (已成功)。
5. 若要顯示陣列任務的所有子任務，選取在之前的區段中傳回的任務 ID。
6. 若要查看任務容器的日誌，選取其中一個子任務，然後選擇 View logs (查看日誌)。

7. 查看其他子任務日誌。每個任務都會傳回不同的彩虹顏色。

執行 GPU 任務

GPU 任務可協助您執行使用執行個體 GPUs 的任務。

支援下列 Amazon EC2 GPU 型執行個體類型。如需詳細資訊，請參閱 Amazon EC2 G3 執行個體、Amazon EC2 G4 執行個體、Amazon EC2 G5 執行個體、Amazon EC2 G6 執行個體、Amazon EC2 P2 執行個體、Amazon EC2 P3 執行個體、Amazon EC2 P4d 執行個體、Amazon EC2 P5 執行個體、Amazon EC2 P6 執行個體、Amazon EC2 Trn1 執行個體、Amazon EC2 Trn2 執行個體、Amazon EC2 Inf1 執行個體、Amazon EC2 Inf2 執行個體、Amazon EC2 DI1 執行個體和 Amazon EC2 DI2 執行個體。

執行個體類型	GPU	記憶體	vCPUs	記憶體	網路頻寬
g3s.xlarge	1	8 GiB	4	30.5 GiB	10 Gbps
g3.4xlarge	1	8 GiB	16	122 GiB	最高 10 Gbps
g3.8xlarge	2	16 GiB	32	244 GiB	10 Gbps
g3.16xlarge	4	32 GiB	64	488 GiB	25 Gbps
g4dn.xlarge	1	16 GiB	4	16 GiB	最高 25 Gbps
g4dn.2xlarge	1	16 GiB	8	32 GiB	最高 25 Gbps
g4dn.4xlarge	1	16 GiB	16	64 GiB	最高 25 Gbps
g4dn.8xlarge	1	16 GiB	32	128 GiB	50 Gbps

執行個體類型	GPU	記憶體	vCPUs	記憶體	網路頻寬
g4dn.12xlarge	4	64 GiB	48	192 GiB	50 Gbps
g4dn.16xlarge	1	16 GiB	64	256 GiB	50 Gbps
g5.xlarge	1	24 GiB	4	16 GiB	最高 10 Gbps
g5.2xlarge	1	24 GiB	8	32 GiB	最高 10 Gbps
g5.4xlarge	1	24 GiB	16	64 GiB	最高 25 Gbps
g5.8xlarge	1	24 GiB	32	128 GiB	25 Gbps
g5.16xlarge	1	24 GiB	64	256 GiB	25 Gbps
g5.12xlarge	4	96 GiB	48	192 GiB	40Gbps
g5.24xlarge	4	96 GiB	96	384 GiB	50 Gbps
g5.48xlarge	8	192 GiB	192	768 GiB	100 Gbps
g5g.xlarge	1	16 GiB	4	8 GiB	最高 10 Gbps
g5g.2xlarge	1	16 GiB	8	16 GiB	最高 10 Gbps
g5g.4xlarge	1	16 GiB	16	32 GiB	最高 10 Gbps
g5g.8xlarge	1	16 GiB	32	64 GiB	12 Gbps
g5g.16xlarge	2	32 GiB	64	128 GiB	25 Gbps
g5g.metal	2	32 GiB	64	128 GiB	25 Gbps
g6.xlarge	1	24 GiB	4	16 GiB	最高 10 Gbps
g6.2xlarge	1	24 GiB	8	32 GiB	最高 10 Gbps
g6.4xlarge	1	24 GiB	16	64 GiB	最高 25 Gbps
g6.8xlarge	1	24 GiB	32	128 GiB	25 Gbps

執行個體類型	GPU	記憶體	vCPUs	記憶體	網路頻寬
g6.16xlarge	1	24 GiB	64	256 GiB	25 Gbps
g6.12xlarge	4	96 GiB	48	192 GiB	40Gbps
g6.24xlarge	4	96 GiB	96	384 GiB	50 Gbps
g6.48xlarge	8	192 GiB	192	768 GiB	100 Gbps
g6e.xlarge	1	48 GiB	4	32 GiB	高達 20 Gbps
g6e.2xlarge	1	48 GiB	8	64 GiB	高達 20 Gbps
g6e.4xlarge	1	48 GiB	16	128 GiB	20 Gbps
g6e.8xlarge	1	48 GiB	32	256 GiB	25 Gbps
g6e.16xlarge	1	48 GiB	64	512 GiB	35 Gbps
g6e.12xlarge	4	192 GiB	48	384 GiB	100 Gbps
g6e.24xlarge	4	192 GiB	96	768 GiB	200 Gbps
g6e.48xlarge	8	384 GiB	192	1536 GiB	400 Gbps
gr6.4xlarge	1	24 GiB	16	128 GiB	最高 25 Gbps
gr6.8xlarge	1	24 GiB	32	256 GiB	25 Gbps
p2.xlarge	1	12 GiB	4	61 GiB	高
p2.8xlarge	8	96 GiB	32	488 GiB	10 Gbps
p2.16xlarge	16	192 GiB	64	732 GiB	20 Gbps
p3.2xlarge	1	16 GiB	8	61 GiB	最高 10 Gbps
p3.8xlarge	4	64 GiB	32	244 GiB	10 Gbps
p3.16xlarge	8	128 GiB	64	488 GiB	25 Gbps

執行個體類型	GPU	記憶體	vCPUs	記憶體	網路頻寬
p3dn.24xlarge	8	256 GiB	96	768 GiB	100 Gbps
p4d.24xlarge	8	320 GiB	96	1152 GiB	400 Gbps
p4de.24xlarge	8	640 GiB	96	1152 GiB	400 Gbps
p5.48xlarge	8	640 GiB	192	2 TiB	3200 Gbps
p5e.48xlarge	8	1128 GiB	192	2 TiB	3200 Gbps
p5en.48xlarge	8	1128 GiB	192	2 TiB	3200 Gbps
p6-b200.48xlarge	8	1440 GiB	192	2 TiB	100 Gbps
trn1.2xlarge	1	32 GiB	8	32 GiB	最高 12.5 Gbps
trn1.32xlarge	16	512 GiB	128	512 GiB	800 Gbps
trn1n.32xlarge	16	512 GiB	128	512 GiB	1600 Gbps
trn2.48xlarge	16	1.5 TiB	192	2 TiB	3.2 Tbps
inf1.xlarge	1	8 GiB	4	8 GiB	最高 25 Gbps
inf1.2xlarge	1	8 GiB	8	16 GiB	最高 25 Gbps
inf1.6xlarge	4	32 GiB	24	48 GiB	25 Gbps
inf1.24xlarge	16	128 GiB	96	192 GiB	100 Gbps
inf2.xlarge	1	32 GiB	4	16 GiB	最高 15 Gbps
inf2.8xlarge	1	32 GiB	32	128 GiB	最高 25 Gbps
inf2.24xlarge	6	192 GiB	96	384 GiB	50 Gbps
inf2.48xlarge	12	384 GiB	192	768 GiB	100 Gbps
dl1.24xlarge	8	256 GiB	96	768 GiB	400 Gbps

執行個體類型	GPU	記憶體	vCPUs	記憶體	網路頻寬
dl2q.24xlarge	8	128 GiB	96	768 GiB	100 Gbps

> **Note**
>
> 對於 GPU 任務， AWS Batch 僅支援具有 NVIDIA GPUs 的執行個體類型。例如， G4ad 系列不支援 GPU 排程。您仍然可以在 G4ad 上使用 ， AWS Batch 方法是只定義任務定義中的 vcpu 和記憶體需求，然後使用 Amazon ECS 或 Amazon EKS 運算最佳化 AMI， GPUs 或透過 Amazon EC2 啟動範本使用者資料的自訂直接存取主機 GPUs，或使用 AMD GPU 的自訂 AMI。 Amazon EC2
> 提供給 AWS Batch 或 Amazon EC2 使用者資料的自訂 AMIs 上的 GPU 任務支援使用 ARM64 架構的執行個體類型，以透過自訂程式碼和組態存取 GPUs。例如，G5g執行個體系列。

任務定義的 resourceRequirements 參數會指定要固定到容器的 GPUs 數量。此 GPUs 數量不適用於在該任務期間在該執行個體上執行的任何其他任務。運算環境中執行 GPU 任務的所有執行個體類型都必須來自 p3、p4、p5、p6、g3、g5、 g3s g4或 g6執行個體系列。如果未完成此操作，GPU 任務可能會卡在 RUNNABLE 狀態。

不使用 GPUs 的任務可以在 GPU 執行個體上執行。不過，在 GPU 執行個體上執行成本可能比在類似的非 GPU 執行個體上要高。根據所需的特定 vCPU、記憶體和時間，這些非 GPU 任務可能會阻止 GPU 任務執行。

主題

- 在 Amazon EKS 上建立 GPU 型Kubernetes叢集
- 建立 Amazon EKS GPU 任務定義
- 在 Amazon EKS 叢集中執行 GPU 任務

在 Amazon EKS 上建立 GPU 型Kubernetes叢集

在 Amazon EKS 上建立 GPU 型Kubernetes叢集之前，您必須先完成 中的步驟在 Amazon EKS AWS Batch 上開始使用。此外，也請考慮下列事項：

- AWS Batch 支援使用 NVIDIA GPUs 的執行個體類型。

- 根據預設，會使用與您的 Amazon EKS 叢集控制平面Kubernetes版本相符的版本來 AWS Batch 選取 Amazon EKS 加速 AMI。

```
$ cat <<EOF > ./batch-eks-gpu-ce.json
{
  "computeEnvironmentName": "My-Eks-GPU-CE1",
  "type": "MANAGED",
  "state": "ENABLED",
  "eksConfiguration": {
    "eksClusterArn": "arn:aws:eks:<region>:<account>:cluster/<cluster-name>",
    "kubernetesNamespace": "my-aws-batch-namespace"
  },
  "computeResources": {
    "type": "EC2",
    "allocationStrategy": "BEST_FIT_PROGRESSIVE",
    "minvCpus": 0,
    "maxvCpus": 1024,
    "instanceTypes": [
      "p3dn.24xlarge",
      "p4d.24xlarge"
    ],
    "subnets": [
        "<eks-cluster-subnets-with-access-to-internet-for-image-pull>"
    ],
    "securityGroupIds": [
        "<eks-cluster-sg>"
    ],
    "instanceRole": "<eks-instance-profile>"
  }
}
EOF

$ aws batch create-compute-environment --cli-input-json file://./batch-eks-gpu-ce.json
```

AWS Batch 不會代表您管理 NVIDIA GPU 裝置外掛程式。您必須在 Amazon EKS 叢集中安裝此外掛程式，並允許它以 AWS Batch 節點為目標。如需詳細資訊，請參閱在 GitHub 上啟用 中的 GPU 支援 Kubernetes。

若要設定NVIDIA裝置外掛程式 (DaemonSet) 以鎖定 AWS Batch 節點，請執行下列命令。

```
# pull nvidia daemonset spec
```

```
$ curl -O https://raw.githubusercontent.com/NVIDIA/k8s-device-plugin/v0.12.2/nvidia-
device-plugin.yml
# using your favorite editor, add Batch node toleration
# this will allow the DaemonSet to run on Batch nodes
- key: "batch.amazonaws.com/batch-node"
  operator: "Exists"

$ kubectl apply -f nvidia-device-plugin.yml
```

我們不建議您在相同的運算環境和任務佇列配對中，將運算型 (CPU 和記憶體） 工作負載與 GPU 型工作負載混合。這是因為運算任務可能會用盡 GPU 容量。

若要連接任務佇列，請執行下列命令。

```
$ cat <<EOF > ./batch-eks-gpu-jq.json
 {
    "jobQueueName": "My-Eks-GPU-JQ1",
    "priority": 10,
    "computeEnvironmentOrder": [
      {
        "order": 1,
        "computeEnvironment": "My-Eks-GPU-CE1"
      }
    ]
  }
EOF

$ aws batch create-job-queue --cli-input-json file://./batch-eks-gpu-jq.json
```

建立 Amazon EKS GPU 任務定義

目前僅`nvidia.com/gpu`支援 ，您設定的資源值必須是整數。您無法使用 GPU 的分數。如需詳細資訊，請參閱 Kubernetes 文件中的排程 GPUs。

若要註冊 Amazon EKS 的 GPU 任務定義，請執行下列命令。

```
$ cat <<EOF > ./batch-eks-gpu-jd.json
{
    "jobDefinitionName": "MyGPUJobOnEks_Smi",
    "type": "container",
    "eksProperties": {
```

```
        "podProperties": {
            "hostNetwork": true,
            "containers": [
                {
                    "image": "nvcr.io/nvidia/cuda:10.2-runtime-centos7",
                    "command": ["nvidia-smi"],
                    "resources": {
                        "limits": {
                            "cpu": "1",
                            "memory": "1024Mi",
                            "nvidia.com/gpu": "1"
                        }
                    }
                }
            ]
        }
    }
}
EOF

$ aws batch register-job-definition --cli-input-json file://./batch-eks-gpu-jd.json
```

在 Amazon EKS 叢集中執行 GPU 任務

GPU 資源不可壓縮。 會為 GPU 任務 AWS Batch 建立 Pod 規格，其中請求的值等於限制的值。這是Kubernetes必要項目。

若要提交 GPU 任務，請執行下列命令。

```
$ aws batch submit-job --job-queue My-Eks-GPU-JQ1 --job-definition MyGPUJobOnEks_Smi --
job-name My-Eks-GPU-Job

# locate information that can help debug or find logs (if using Amazon CloudWatch Logs
 with Fluent Bit)
$ aws batch describe-jobs --job <job-id> | jq '.jobs[].eksProperties.podProperties |
 {podName, nodeName}'
{
  "podName": "aws-batch.f3d697c4-3bb5-3955-aa6c-977fcf1cb0ca",
  "nodeName": "ip-192-168-59-101.ec2.internal"
}
```

檢視 AWS Batch 任務佇列中的任務

您可以在其中檢視和篩選任務 AWS Batch。此功能提供檢視現有任務佇列的選項，並依三個選項之一篩選其任務。

搜尋和篩選能夠擷取未處於終端狀態 (SUCCEEDED 或) 的任務FAILED。一旦任務的狀態為 SUCCEEDED，或者FAILED您應該能夠擷取任務長達七天。您仍然可以檢視任務的 CloudWatch 或 Amazon EventBridge 日誌。

使用此程序在 AWS Batch 主控台中列出任務佇列中的所有任務。或者，使用篩選結果欄位，根據您指定的條件縮小結果範圍。

1. 導覽至 AWS Batch 主控台。
2. 在導覽窗格中，選擇任務。
3. 展開任務佇列下拉式清單，然後選擇您要搜尋的任務佇列。

> **Note**
>
> 您一次只能搜尋一個任務佇列內的任務。

4. 在篩選結果欄位中，輸入要包含在結果中的關鍵字。您可以使用此欄位，依任務名稱、狀態或任務 ID 進行篩選。視 屬性而定，可能會有其他運算子，例如等於 (=) 或包含您必須定義的 (:)。

> **Note**
>
> SageMaker Training 任務佇列僅支援依任務名稱和任務 ID 篩選

5. 選擇 Search (搜尋)。

在任務佇列中 AWS Batch 搜尋任務

您可以使用任務搜尋在 中搜尋和篩選 AWS Batch 任務。此功能提供在現有任務佇列中搜尋並篩選其任務的選項。

搜尋和篩選能夠擷取未處於終端狀態 (SUCCEEDED 或) 的任務FAILED。一旦任務的狀態為 SUCCEEDED，或者FAILED您應該能夠擷取任務長達七天。您仍然可以檢視任務的 CloudWatch 或 Amazon EventBridge 日誌。

若要同時使用多個條件進行搜尋，請使用進階搜尋功能。例如，您可以包含下列任何或所有篩選條件：狀態、日期範圍和其他條件（例如任務名稱、任務定義或任務 ID)。

搜尋 AWS Batch 任務AWS（主控台）

使用此程序在 AWS Batch 主控台中搜尋任務佇列中的任務。

1. 導覽至 AWS Batch 主控台。
2. 在導覽窗格中，選擇任務。
3. 開啟進階搜尋。
4. 展開任務佇列下拉式清單，然後選擇您要搜尋的任務佇列。

   > **Note**
   >
   > 您一次只能搜尋一個任務佇列內的任務。

5. 對於搜尋選項：

   a. 針對狀態下拉式清單，您可以選擇要篩選的一或多個狀態。如需詳細資訊，請參閱任務狀態及服務任務狀態。

      > **Note**
      >
      > 當任何子任務更新為 PENDING 時，陣列任務父會更新為 ，RUNNABLE並在子任務執行時保持 PENDING 狀態。若要檢視這些任務，請依PENDING狀態進行篩選，直到所有子任務都達到結束狀態為止。

   b. 選擇日期範圍，根據日期和時間範圍篩選結果。

      - 選擇相對模式，以搜尋在從目前日期和時間回溯計數的時間範圍內建立日期的任務。
      - 選擇絕對模式以搜尋在您指定的日期和時間範圍內建立日期的任務。

   c. 在其他條件欄位中，輸入要包含在搜尋結果中的關鍵字。例如，您可以使用此欄位依任務名稱、任務定義、任務 ID 或共用識別符進行搜尋。視 屬性而定，可能會有其他運算子，例如等於 (=) 或包含您必須定義的 (：)。

      > **Note**
      >
      > SageMaker Training 任務佇列僅支援依任務名稱和任務 ID 篩選

> **Note**
>
> 依共用識別符篩選時，您也可以指定任務狀態。這是其他篩選條件排除任務狀態篩選的限制的例外狀況。

6. 選擇 Search (搜尋)。

搜尋和篩選 AWS Batch 任務 (AWS CLI)

使用此程序，透過 列出任務佇列中的所有任務 AWS CLI。或者，使用 篩選條件參數，根據您指定的條件縮小結果範圍。

Search job queue (AWS CLI)

您可以使用 list-jobs 命令來搜尋和篩選任務佇列。

例如，您可以根據任務名稱搜尋任務佇列：

```
aws batch list-jobs \
    --job-queue my-job-queue \
    --filters name=JOB_NAME,values="my-job"
```

依共用識別符篩選任務：

```
aws batch list-jobs \
    --job-queue my-job-queue \
    --filters name=SHARE_IDENTIFIER,values="my-share"
```

依共用識別符篩選時，您可以包含任務狀態：

```
aws batch list-jobs \
    --job-queue my-job-queue \
    --job-status RUNNING \
    --filters name=SHARE_IDENTIFIER,values="my-share"
```

在上述命令中，進行下列變更：

- 以您的任務佇列名稱取代 *my-job-queue*。

- 將 *my-job* 取代為您的任務名稱。
- 將 *my-share* 取代為您要篩選的共用識別符。

Search service job queue (AWS CLI)

您可以使用 list-service-jobs 命令來搜尋和篩選服務任務佇列。

例如，您可以根據任務名稱搜尋服務任務佇列：

```
aws batch list-service-jobs \
    --job-queue my-sm-queue \
    --filters name=JOB_NAME,values="my-sm-job"
```

依共用識別符篩選服務任務：

```
aws batch list-service-jobs \
    --job-queue my-sm-queue \
    --filters name=SHARE_IDENTIFIER,values="my-share"
```

在上述命令中，進行下列變更：

- 以您的服務任務佇列名稱取代 *my-sm-queue*。
- 以您的服務任務名稱取代 *my-sm-job*。
- 將 *my-share* 取代為您要篩選的共用識別符。

AWS Batch 任務的網路模式

下表說明 AWS Batch 任務類型的聯網模式和典型用量。如需考量和行為的詳細資訊，請參閱「任務類型」欄中的連結。

任務類型	支援的網路模式 (s)	典型用量
ECS-EC2 簡單任務	host	用於可擴展性最高的尷尬平行批次工作負載，這些工作負載只需要輸出至運算環境中定義的 vpc。

任務類型	支援的網路模式 (s)	典型用量
ECS-EC2 多節點平行任務	awsvpc	用於緊密耦合的多主機（節點）分散式工作負載，模型化為具有任務節點之間協調通訊的單一任務。
ECS-Fargate 簡單任務	awsvpc	真正的無伺服器，適用於尷尬的平行批次工作負載。通常是最低的 TCO 和最高的容器隔離任務模型。
EKS-EC2 簡單任務	主機和 Pod	用於高度可擴展的尷尬平行批次工作負載，這些工作負載只需要輸出至運算環境中定義的 vpc。預設為主機聯網。
EKS-EC2 多節點平行任務	主機和 Pod	用於緊密耦合的多主機（節點）分散式工作負載，模型化為具有 Pod 節點之間協調通訊的單一任務。預設為主機聯網。

在 CloudWatch Logs 中檢視 AWS Batch 任務日誌

您可以設定 AWS Batch 任務將日誌資訊傳送至 Amazon CloudWatch Logs。如此一來，您可以在一個方便的位置檢視與任務不同的日誌。如需詳細資訊，請參閱搭配 使用 CloudWatch Logs AWS Batch。

您也可以在 AWS Batch 主控台中使用任務日誌來監控或疑難排解 AWS Batch 任務。

1. 開啟 AWS Batch 主控台。
2. 選擇 Jobs (任務)。如需在任務佇列中排序和篩選任務的詳細資訊，請參閱 檢視 AWS Batch 任務佇列中的任務和 搜尋任務佇列中的任務
3. 針對任務佇列，選擇您想要的任務佇列。

> **Tip**
>
> 如果任務佇列中有多個任務，您可以開啟搜尋和篩選以更快地尋找任務。如需詳細資訊，請參閱在任務佇列中 AWS Batch 搜尋任務。

4. 針對狀態，選擇您想要的任務狀態。
5. 選擇您想要的任務，便會開啟詳細資訊頁面。
6. 在詳細資訊頁面上，向下捲動至日誌串流名稱，然後選擇連結。連結會開啟任務的 Amazon CloudWatch Logs 頁面。
7. （選用）如果這是您第一次檢視日誌，您可能需要授權。

   如需授權，請輸入 **OK**，然後選擇授權以接受 Amazon CloudWatch 費用。

> **Note**
>
> 若要撤銷 CloudWatch 費用的授權：
>
> 1. 在左側導覽窗格中，選擇許可。
> 2. 針對任務日誌，選擇編輯。
> 3. 清除授權批次以使用 CloudWatch 核取方塊。
> 4. 選擇儲存變更。

檢閱 AWS Batch 任務資訊

您可以檢閱 AWS Batch 任務資訊，例如狀態、任務定義和容器資訊。

1. 開啟 AWS Batch 主控台。
2. 選擇 Jobs (任務)。
3. 針對任務佇列，選擇您想要的任務佇列。

> **Tip**
>
> 如果任務佇列中有多個任務，您可以開啟搜尋和篩選以更快地尋找任務。如需詳細資訊，請參閱在任務佇列中 AWS Batch 搜尋任務。

4. 選擇您想要的任務。

> **Note**
>
> 您也可以使用 AWS Command Line Interface (AWS CLI) 來檢視 AWS Batch 任務的詳細資訊。如需詳細資訊，請參閱《AWS CLI 命令參考》中的 describe-jobs。

中的安全性 AWS Batch

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，這些架構是為了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。共同責任模式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。在 AWS Compliance Programs 中，第三方稽核人員會定期測試和驗證我們安全的有效性。若要了解適用的合規計劃 AWS Batch，請參閱AWS 合規計劃的 服務範圍。
- 雲端內部的安全：您的責任取決於所使用的 AWS 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 時套用共同責任模型 AWS Batch。下列主題說明如何設定 AWS Batch 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 Batch 資源。

主題

- 的 Identity and Access Management AWS Batch
- AWS Batch IAM 政策、角色和許可
- AWS Batch IAM 執行角色
- 建立 Virtual Private Cloud
- 使用界面端點存取 AWS Batch
- 的合規驗證 AWS Batch
- 中的基礎設施安全 AWS Batch
- 預防跨服務混淆代理人
- 使用 記錄 AWS Batch API 呼叫 AWS CloudTrail
- IAM AWS Batch 故障診斷

的 Identity and Access Management AWS Batch

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驗證 （登入） 和授權 （具有許可） 來使用 AWS Batch 資源。IAM 是您可以免費使用 AWS 服務 的 。

主題

- 目標對象
- 使用身分驗證
- 使用政策管理存取權
- AWS Batch 如何使用 IAM
- 的身分型政策範例 AWS Batch
- AWS 的 受管政策 AWS Batch

目標對象

使用方式 AWS Identity and Access Management (IAM) 會根據您的角色而有所不同：

- 服務使用者 — 若無法存取某些功能，請向管理員申請所需許可 (請參閱 IAM AWS Batch 故障診斷)
- 服務管理員 — 負責設定使用者存取權並提交相關許可請求 (請參閱 AWS Batch 如何使用 IAM)
- IAM 管理員 — 撰寫政策以管理存取控制 (請參閱 的身分型政策範例 AWS Batch)

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、IAM 使用者或擔任 IAM 角色身分進行身分驗證。

您可以使用身分來源的登入資料，例如 AWS IAM Identity Center (IAM Identity Center)、單一登入身分驗證或 Google/Facebook 登入資料，以聯合身分的形式登入。如需有關登入的詳細資訊，請參閱《AWS 登入 使用者指南》中的如何登入您的 AWS 帳戶。

對於程式設計存取， AWS 提供 SDK 和 CLI 以密碼編譯方式簽署請求。如需詳細資訊，請參閱《IAM 使用者指南》中的 API 請求的AWS 第 4 版簽署程序。

AWS 帳戶 根使用者

當您建立 時 AWS 帳戶，您會從一個名為 AWS 帳戶 theroot 使用者的登入身分開始，該身分可完整存取所有 AWS 服務 和 資源。強烈建議不要使用根使用者來執行日常任務。有關需要根使用者憑證的任務，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使用聯合身分提供者，以 AWS 服務 使用臨時憑證存取 。

聯合身分是來自您的企業目錄、Web 身分提供者的使用者，或使用來自身分來源的 AWS 服務 憑證存取 Directory Service 。聯合身分會擔任角色，而該角色會提供臨時憑證。

若需集中化管理存取權限，建議使用 AWS IAM Identity Center。如需詳細資訊，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https://docs.aws.amazon.com/IAM/latest/UserGuide/id_users.html是一種身分具備單人或應用程式的特定許可權。建議以臨時憑證取代具備長期憑證的 IAM 使用者。如需詳細資訊，請參閱《IAM 使用者指南》中的要求人類使用者使用聯合身分提供者，以 AWS 使用臨時憑證存取 。

IAM 群組會指定 IAM 使用者集合，使管理大量使用者的許可權更加輕鬆。如需詳細資訊，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https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles.html的身分具有特定許可權，其可以提供臨時憑證。您可以透過從使用者切換到 IAM 角色（主控台）或呼叫 AWS CLI 或 AWS API 操作來擔任角色。如需詳細資訊，請參閱《IAM 使用者指南》中的擔任角色的方法。

IAM 角色適用於聯合身分使用者存取、臨時 IAM 使用者許可、跨帳戶存取權與跨服務存取，以及在 Amazon EC2 執行的應用程式。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。

使用政策管理存取權

您可以透過建立政策並將其連接到身分或資源 AWS 來控制 AWS 中的存取。政策定義與身分或資源相關聯的許可。當委託人提出請求時 AWS ， 會評估這些政策。大多數政策會以 JSON 文件 AWS 形式存放在 中。如需進一步了解 JSON 政策文件，請參閱《IAM 使用者指南》中的 JSON 政策概觀。

管理員會使用政策，透過定義哪些主體可在哪些條件下對哪些資源執行動作，以指定可存取的範圍。

預設情況下，使用者和角色沒有許可。IAM 管理員會建立 IAM 政策並將其新增至角色，供使用者後續擔任。IAM 政策定義動作的許可，無論採用何種方式執行。

身分型政策

身分型政策是附加至身分 (使用者、使用者群組或角色) 的 JSON 許可政策文件。這類政策控制身分可對哪些資源執行哪些動作，以及適用的條件。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。

身分型政策可分為內嵌政策 (直接內嵌於單一身分) 與受管政策 (可附加至多個身分的獨立政策)。如需了解如何在受管政策及內嵌政策之間做選擇，請參閱《IAM 使用者指南》中的在受管政策與內嵌政策之間選擇。

資源型政策

資源型政策是附加到資源的 JSON 政策文件。範例包括 IAM 角色信任政策與 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。您必須在資源型政策中指定主體。

資源型政策是位於該服務中的內嵌政策。您無法在以資源為基礎的政策中使用來自 IAM 的 AWS 受管政策。

其他政策類型

AWS 支援其他政策類型，可設定更多常見政策類型授予的最大許可：

- 許可界限 — 設定身分型政策可授與 IAM 實體的最大許可。如需詳細資訊，請參閱《 IAM 使用者指南》中的 IAM 實體許可界限。
- 服務控制政策 (SCP) — 為 AWS Organizations中的組織或組織單位指定最大許可。如需詳細資訊，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) — 設定您帳戶中資源可用許可的上限。如需詳細資訊，請參閱《AWS Organizations 使用者指南》中的資源控制政策 (RCP)。
- 工作階段政策 — 在以程式設計方式為角色或聯合身分使用者建立臨時工作階段時，以參數形式傳遞的進階政策。如需詳細資訊，請參《IAM 使用者指南》中的工作階段政策。

多種政策類型

當多種類型的政策適用於請求時，產生的許可會更複雜而無法理解。若要了解如何 AWS 在涉及多個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 Batch 如何使用 IAM

在您使用 IAM 管理對 的存取之前 AWS Batch，請先了解哪些 IAM 功能可與 搭配使用 AWS Batch。

您可以搭配 使用的 IAM 功能 AWS Batch

IAM 功能	AWS Batch 支援
身分型政策	是

IAM 功能	AWS Batch 支援
資源型政策	否
政策動作	是
政策資源	是
政策條件索引鍵	是
ACL	否
ABAC (政策中的標籤)	是
臨時憑證	是
主體許可	是
服務角色	是
服務連結角色	是

若要全面了解 AWS Batch 和其他 AWS 服務如何與大多數 IAM 功能搭配使用，請參閱《IAM 使用者指南》中的與 IAM AWS 搭配使用的 服務。

的身分型政策 AWS Batch

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。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。

AWS Batch的身分型政策範例

若要檢視 AWS Batch 身分型政策的範例，請參閱 的身分型政策範例 AWS Batch。

的政策動作 AWS Batch

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。政策會使用動作來授予執行相關聯動作的許可。

若要查看 AWS Batch 動作清單，請參閱《服務授權參考》中的 定義的動作 AWS Batch。

中的政策動作在動作之前 AWS Batch 使用下列字首：

```
batch
```

如需在單一陳述式中指定多個動作，請用逗號分隔。

```
"Action": [
      "batch:action1",
      "batch:action2"
         ]
```

您也可以使用萬用字元 (*) 來指定多個動作。例如，若要指定開頭是 `Describe` 文字的所有動作，請包含以下動作：

```
"Action": "batch:Describe*"
```

若要檢視 AWS Batch 身分型政策的範例，請參閱 的身分型政策範例 AWS Batch。

的政策資源 AWS Batch

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。最佳實務是使用其 Amazon Resource Name (ARN) 來指定資源。若動作不支援資源層級許可，使用萬用字元 (*) 表示該陳述式適用於所有資源。

```
"Resource": "*"
```

若要查看 AWS Batch 資源類型及其 ARNs，請參閱《服務授權參考》中的 定義的資源 AWS Batch。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 Batch定義的動作。

AWS Batch的政策條件索引鍵

支援服務特定政策條件金鑰：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。

`Condition` 元素會根據定義的條件，指定陳述式的執行時機。您可以建立使用條件運算子的條件運算式 (例如等於或小於)，來比對政策中的條件和請求中的值。若要查看所有 AWS 全域條件索引鍵，請參閱《IAM 使用者指南》中的AWS 全域條件內容索引鍵。

若要查看 AWS Batch 條件金鑰清單，請參閱《服務授權參考》中的 的條件金鑰 AWS Batch。若要了解您可以使用條件金鑰的動作和資源，請參閱 定義的動作 AWS Batch。

使用 的屬性型存取控制 (ABAC) AWS Batch

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，依據稱為標籤的屬性來定義許可。您可以將標籤連接至 IAM 實體 AWS 和資源，然後設計 ABAC 政策，以便在委託人的標籤符合資源上的標籤時允許操作。

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)。

搭配 使用暫時登入資料 AWS Batch

支援臨時憑證：是

臨時登入資料提供 AWS 資源的短期存取權，當您使用聯合身分或切換角色時會自動建立。 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 使用者指南》中的 IAM 中的臨時安全憑證與可與 IAM 搭配運作的AWS 服務。

AWS Batch的跨服務主體權限

支援轉寄存取工作階段 (FAS)：是

轉送存取工作階段 (FAS) 使用呼叫 的委託人許可 AWS 服務，並結合 AWS 服務 請求向下游服務提出請求。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。

的服務角色 AWS Batch

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給 AWS 服務。

> **Warning**
>
> 變更服務角色的許可可能會中斷 AWS Batch 功能。只有 AWS Batch 提供指引時，才能編輯服務角色。

的服務連結角色 AWS Batch

支援服務連結角色：是

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。在資料表中尋找服務，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 Batch

根據預設，使用者和角色不具備建立或修改 AWS Batch 資源的權限。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。

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 定義的動作和資源類型的詳細資訊 AWS Batch，包括每種資源類型的 ARNs 格式，請參閱《服務授權參考》中的適用於 的動作、資源和條件金鑰 AWS Batch。

主題

- 政策最佳實務
- 使用 AWS Batch 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 Batch 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 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》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 AWS Batch 主控台

若要存取 AWS Batch 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 Batch 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。

為了確保使用者和角色仍然可以使用 AWS Batch 主控台，請將 AWS Batch `ConsoleAccess`或 `ReadOnly` AWS 受管政策連接到實體。如需詳細資訊，請參閱《IAM 使用者指南》中的新增許可到使用者。

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 Batch

您可以使用 AWS 受管政策，為您的團隊和佈建的 AWS 資源進行更簡單的身分存取管理。 AWS 受管政策涵蓋各種常見的使用案例，預設可在 AWS 您的帳戶中使用，並且會代表您進行維護和更新。您無法變更 AWS 受管政策中的許可。如果您需要更大的彈性，您也可以選擇建立 IAM 客戶受管政策。如此一來，您就可以為團隊佈建的資源提供他們所需的確切許可。

如需 AWS 受管政策的詳細資訊，請參閱《IAM 使用者指南》中的 AWS 受管政策。

AWS 服務會代表您維護和更新 AWS 受管政策。 AWS 服務會定期將其他許可新增至 AWS 受管政策。當新功能啟動或操作可用時，最有可能更新 受 AWS 管政策。這些更新會自動影響附加政策的所有身分 （使用者、群組和角色）。不過，它們不會移除許可或破壞您現有的許可。

此外， AWS 支援跨多個 服務之任務函數的受管政策。例如， `ReadOnlyAccess` AWS 受管政策提供所有 AWS 服務和資源的唯讀存取權。當服務啟動新功能時， 會為新操作和資源 AWS 新增唯讀許可。如需任務職能政策的清單和說明，請參閱 IAM 使用者指南中有關任務職能的AWS 受管政策。

AWS 受管政策：BatchServiceRolePolicy

BatchServiceRolePolicy 受管 IAM 政策由`AWSServiceRoleForBatch`服務連結角色使用。這可讓代表您 AWS Batch 執行動作。您無法將此政策連接至 IAM 實體。如需詳細資訊，請參閱使用 AWS Batch的服務連結角色。

此政策允許 AWS Batch 在特定資源上完成下列動作：

- `autoscaling` – 允許 AWS Batch 建立和管理 Amazon EC2 Auto Scaling 資源。 會為大多數運算環境 AWS Batch 建立和管理 Amazon EC2 Auto Scaling 群組。
- `ec2` – 允許 AWS Batch 控制 Amazon EC2 執行個體的生命週期，以及建立和管理啟動範本和標籤。AWS Batch 會建立和管理某些 EC2 Spot 運算環境的 EC2 Spot 機群請求。
- `ecs` - 允許 為任務執行 AWS Batch 建立和管理 Amazon ECS 叢集、任務定義和任務。
- `eks` - 允許 AWS Batch 描述用於驗證的 Amazon EKS 叢集資源。

- `iam` - 允許 AWS Batch 驗證擁有者提供的角色並將其傳遞給 Amazon EC2、Amazon EC2 Auto Scaling 和 Amazon ECS。
- `logs` – 允許 AWS Batch 建立和管理 AWS Batch 任務的日誌群組和日誌串流。

若要檢視政策的 JSON，請參閱《AWS 受管政策參考指南》中的 BatchServiceRolePolicy。

AWS 受管政策：AWSBatchServiceRolePolicyForSageMaker

`AWSServiceRoleForAWSBatchWithSagemaker` 允許 代表您 AWS Batch 執行動作。您無法將此政策連接至 IAM 實體。如需詳細資訊，請參閱使用 AWS Batch的服務連結角色。

此政策允許 AWS Batch 在特定資源上完成下列動作：

- `sagemaker` – 允許 AWS Batch 管理 SageMaker AI 訓練任務和其他 SageMaker AI 資源。
- `iam:PassRole` – 允許 AWS Batch 將客戶定義的執行角色傳遞至 SageMaker AI 以進行任務執行。資源限制允許將角色傳遞至 SageMaker AI 服務。

若要檢視政策的 JSON，請參閱《AWS 受管政策參考指南》中的 AWSBatchServiceRolePolicyForSageMaker。

AWS 受管政策：AWSBatchServiceRole 政策

名為 AWSBatchServiceRole 的角色許可政策允許 AWS Batch 在特定資源上完成下列動作：

AWSBatchServiceRole 受管 IAM 政策通常由名為 AWSBatchServiceRole 的角色使用，並包含下列許可。遵循授予最低權限的標準安全建議，可使用 AWSBatchServiceRole 受管政策做為指南。如果您的使用案例不需要受管政策中授予的任何許可，請建立自訂政策並僅新增您需要的許可。此 AWS Batch 受管政策和角色可以與大多數運算環境類型搭配使用，但服務連結角色使用率是較不容易出錯、範圍更佳和改善受管體驗的首選。

- `autoscaling` – 允許 AWS Batch 建立和管理 Amazon EC2 Auto Scaling 資源。 會為大多數運算環境 AWS Batch 建立和管理 Amazon EC2 Auto Scaling 群組。
- `ec2` – 允許 AWS Batch 管理 Amazon EC2 執行個體的生命週期，以及建立和管理啟動範本和標籤。AWS Batch 會建立和管理某些 EC2 Spot 運算環境的 EC2 Spot 機群請求。
- `ecs` - 允許 為任務執行 AWS Batch 建立和管理 Amazon ECS 叢集、任務定義和任務。
- `iam` - 允許 AWS Batch 驗證擁有者提供的角色並將其傳遞給 Amazon EC2、Amazon EC2 Auto Scaling 和 Amazon ECS。
- `logs` – 允許 AWS Batch 建立和管理 AWS Batch 任務的日誌群組和日誌串流。

若要檢視政策的 JSON，請參閱《AWS 受管政策參考指南》中的 AWSBatchServiceRole。

AWS 受管政策：AWSBatchFullAccess

AWSBatchFullAccess 政策會授予 AWS Batch 動作對 AWS Batch 資源的完整存取權。它還授予描述和列出 Amazon EC2、Amazon ECS、Amazon EKS、CloudWatch 和 IAM 服務的動作存取權。如此一來，使用者或角色的 IAM 身分就可以檢視代其建立的 AWS Batch 受管資源。最後，此政策也允許將選取的 IAM 角色傳遞給這些服務。

您可以將 AWSBatchFullAccess 連接至 IAM 實體。 AWS Batch 也會將此政策連接至允許 代表您 AWS Batch 執行動作的服務角色。

若要檢視政策的 JSON，請參閱《AWS 受管政策參考指南》中的 AWSBatchFullAccess。

AWS Batch AWS 受管政策的更新

檢視自此服務開始追蹤這些變更 AWS Batch 以來， 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 AWS Batch 文件歷史記錄頁面上的 RSS 摘要。

變更	描述	Date
新增 AWSBatchServiceRolePolicyForSageMaker 政策	新增 AWSBatchServiceRolePolicyForSageMaker 服務連結角色的新 AWS 受管政策， AWS Batch 允許 代表您管理 SageMaker AI。	2025 年 7 月 31 日
BatchServiceRolePolicy 政策已更新	更新以新增描述 Spot Fleet 請求歷史記錄和 Amazon EC2 Auto Scaling 活動的支援。	2023 年 12 月 5 日
新增 AWSBatchServiceRole 政策	更新以新增陳述式 IDs，將 AWS Batch 許可授予 `ec2:DescribeSpotFleetRequestHistory` 和 `autoscaling:DescribeScalingActivities` 。	2023 年 12 月 5 日
BatchServiceRolePolicy 政策已更新	更新以新增描述 Amazon EKS 叢集的支援。	2022 年 10 月 20 日

變更	描述	Date
AWSBatchFullAccess 政策已更新	更新以新增列出和描述 Amazon EKS 叢集的支援。	2022 年 10 月 20 日
BatchServiceRolePolicy 政策已更新	更新以新增對由 管理之 Amazon EC2 容量保留群組的支援 AWS Resource Groups。如需詳細資訊，請參閱《Amazon EC2 使用者指南》中的使用容量保留群組。	2022 年 5 月 18 日
BatchServiceRolePolicy 和 AWSBatchServiceRole 政策已更新	更新以新增在 Amazon EC2 中描述 AWS Batch 受管執行個體狀態的支援，以便取代運作狀態不佳的執行個體。	2021 年 12 月 6 日
BatchServiceRolePolicy 政策已更新	更新以新增對 Amazon EC2 中置放群組、容量保留、彈性 GPU 和彈性推論資源的支援。	2021 年 3 月 26 日
已新增 BatchServiceRolePolicy 政策	透過 AWSServiceRoleForBatch 服務連結角色的 BatchServiceRolePolicy 受管政策，您可以使用由 管理的服務連結角色 AWS Batch。使用此政策，您不需要維護自己的角色，即可在運算環境中使用。	2021 年 3 月 10 日
AWSBatchFullAccess - 新增新增服務連結角色的許可	新增 IAM 許可，以允許將 AWSServiceRoleForBatch 服務連結角色新增至帳戶。	2021 年 3 月 10 日
AWS Batch 開始追蹤變更	AWS Batch 已開始追蹤其 AWS 受管政策的變更。	2021 年 3 月 10 日

AWS Batch IAM 政策、角色和許可

根據預設，使用者沒有建立或修改 AWS Batch 資源或使用 AWS Batch API、 AWS Batch 主控台或 執行任務的許可 AWS CLI。若要允許使用者執行這些動作，請建立 IAM 政策，以授予使用者特定資源和 API 操作的許可。然後，將政策連接到需要這些許可的使用者或群組。

當您將政策連接到使用者或使用者群組時，政策會允許或拒絕對特定資源執行特定任務的許可。如需詳細資訊，請參閱《IAM 使用者指南》中的許可和政策。如需管理和建立自訂 IAM 政策的詳細資訊，請參閱管理 IAM 政策。

AWS Batch AWS 服務 會代表您呼叫其他 。因此， AWS Batch 必須使用您的登入資料進行身分驗證。更具體地說，透過建立提供這些許可的 IAM 角色和政策進行 AWS Batch 身分驗證。然後，當您建立角色時，它會將角色與您的運算環境建立關聯。如需詳細資訊，請參閱《IAM 使用者指南》中的 Amazon ECS 執行個體角色、IAM 角色、使用服務連結角色，以及建立角色以委派許可給 AWS 服務。

主題

- IAM 政策結構
- 資源： 的政策範例 AWS Batch
- Resource： AWS Batch managed 政策

IAM 政策結構

下列主題說明 IAM 政策的結構。

主題

- 政策語法
- 的 API 動作 AWS Batch
- 的 Amazon Resource Name AWS Batch
- 確認使用者具有必要的許可

政策語法

IAM 政策為包含一或多個陳述式的 JSON 文件。每個陳述式的結構如下所示。

```
{
  "Statement":[{
```

```
    "Effect":"effect",
    "Action":"action",
    "Resource":"arn",
    "Condition":{
      "condition":{
    "key":"value"
    }
      }
    }
  ]
}
```

陳述式由四個主要元素組成：

- Effect (效果)：效果 可以是 Allow 或 Deny。根據預設，使用者沒有使用資源和 API 動作的許可。因此，所有請求都會遭到拒絕。明確允許覆寫預設值。明確拒絕覆寫任何允許。
- 動作：動作是您授予或拒絕許可的特定 API 動作。如需如何指定動作的指示，請參閱 的 API 動作 AWS Batch。
- Resource (資源)：受動作影響的資源。透過某些 AWS Batch API 動作，您可以在政策中包含可由動作建立或修改的特定資源。若要在陳述式中指定資源，請使用它的 Amazon Resource Name (ARN)。如需詳細資訊，請參閱AWS Batch API 動作支援的資源層級許可及的 Amazon Resource Name AWS Batch。如果 AWS Batch API 操作目前不支援資源層級許可，請包含萬用字元 (*)，以指定所有資源都可能受到 動作的影響。
- Condition (條件)：條件為選擇性。您可以使用它們來控制何時政策開始生效。

如需 IAM 政策陳述式範例的詳細資訊 AWS Batch，請參閱 資源： 的政策範例 AWS Batch。

的 API 動作 AWS Batch

在 IAM 政策陳述式中，您可以從任何支援 IAM 的服務指定任何 API 動作。對於 AWS Batch，請使用下列字首搭配 API 動作的名稱： `batch:` (例如， `batch:SubmitJob`和 `batch:CreateComputeEnvironment`)。

若要在單一陳述式中指定多個動作，請以逗號分隔每個動作。

```
"Action": ["batch:action1", "batch:action2"]
```

您也可以透過包含萬用字元 (*) 來指定多個動作。例如，您可以指定名稱開頭為「Describe」的所有動作。

```
"Action": "batch:Describe*"
```

若要指定所有 AWS Batch API 動作，請包含萬用字元 (*)。

```
"Action": "batch:*"
```

如需 AWS Batch 動作清單，請參閱 AWS Batch API 參考中的動作。

的 Amazon Resource Name AWS Batch

每個 IAM 政策陳述式都適用於您使用其 Amazon Resource Name (ARNs) 指定的資源。

Amazon Resource Name (ARN) 具有下列一般語法：

```
arn:aws:[service]:[region]:[account]:resourceType/resourcePath
```

服務

服務 (例如，batch)。

region

資源 AWS 區域 的 （ 例如，us-east-2)。

account

AWS 帳戶 ID，不含連字號 （ 例如 123456789012)。

resourceType

資源類型 (例如，compute-environment)。

resourcePath

識別資源的路徑。您可以在路徑中使用萬用字元 (*)。

AWS Batch API 操作目前支援數個 API 操作的資源層級許可。如需詳細資訊，請參閱AWS Batch API 動作支援的資源層級許可。若要指定所有資源，或如果特定 API 動作不支援 ARNs，請在 Resource元素中包含萬用字元 (*)。

```
"Resource": "*"
```

確認使用者具有必要的許可

在將 IAM 政策投入生產環境之前，請確定它授予使用者使用他們所需的特定 API 動作和資源的許可。

若要這樣做，請先建立用於測試的使用者，並將 IAM 政策連接至測試使用者。接著，以測試使用者的身分提出請求。您可以在主控台或 AWS CLI中提出測試請求。

> **Note**
>
> 您也可以使用 IAM 政策模擬器來測試您的政策。如需政策模擬器的詳細資訊，請參閱《IAM 使用者指南》中的使用 IAM 政策模擬器。

如果政策未授予使用者預期的許可，或授予過多許可，您可以視需要調整政策並重新測試。重新測試，直到您取得所要的結果。

> **Important**
>
> 政策變更的散佈可能需要幾分鐘時間才能生效。因此，我們建議您在測試政策更新之前至少等待五分鐘。

如果授權檢查失敗，請求將傳回包含診斷資訊的編碼訊息。您可使用 `DecodeAuthorizationMessage` 動作將訊息解碼。如需詳細資訊，請參閱 AWS Security Token Service API 參考中的 DecodeAuthorizationMessage 以及 AWS CLI 命令參考中的 decode-authorization-message。

資源： 的政策範例 AWS Batch

您可以建立特定的 IAM 政策，以限制您帳戶中使用者可存取的呼叫和資源。然後，您可以將這些政策連接到使用者。

當您將政策連接至使用者或使用者群組時，政策會允許或拒絕使用者對特定資源執行特定任務的許可。如需詳細資訊，請參閱《IAM 使用者指南》中的許可和政策。如需如何管理和建立自訂 IAM 政策的說明，請參閱管理 IAM 政策。

下列範例顯示政策陳述式，您可以使用這些陳述式來控制使用者擁有的 許可 AWS Batch。

範例

- 資源： 的唯讀存取 AWS Batch

- 資源：僅限於 POSIX 使用者、Docker 映像、權限層級和任務提交時的角色
- 資源：僅限於提交任務時的任務定義字首
- 資源：限制為任務佇列
- 當所有條件符合字串時拒絕動作
- 資源：當任何條件索引鍵符合字串時拒絕動作
- 資源：使用 batch:ShareIdentifier 條件索引鍵
- 使用 管理 SageMaker AI 資源 AWS Batch
- 資源：依任務定義和任務佇列上的資源標籤限制任務提交

資源： 的唯讀存取 AWS Batch

下列政策授予使用者使用名稱開頭為 `Describe`和 之所有 AWS Batch API 動作的許可`List`。

除非另一個陳述式授予他們許可，否則使用者沒有對資源執行任何動作的許可。根據預設，系統會拒絕他們使用 API 動作的許可。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "batch:Describe*",
                "batch:List*",
                "batch:Get*"
            ],
            "Resource": "*"
        }
    ]
}
```

資源：僅限於 POSIX 使用者、Docker 映像、權限層級和任務提交時的角色

下列政策允許 POSIX 使用者管理自己的一組受限任務定義。

使用第一個和第二個陳述式來註冊和取消註冊名稱字首為 *JobDefA_* 的任何任務定義名稱。

第一個陳述式也使用條件式內容金鑰來限制任務定義內 containerProperties 的 POSIX 使用者、權限狀態，以及容器映像值。如需詳細資訊，請參閱《 AWS Batch API 參考》中的 RegisterJobDefinition。在此範例中，只有在 POSIX 使用者設定為 時，才能註冊任務定義nobody。特權旗標設定為 false。最後，映像在 Amazon ECR 儲存庫myImage中設定為 。

> **Important**
>
> Docker uid 會從容器映像內將 user 參數解析給該使用者。在大多數情況下，這可在容器映像內的 /etc/passwd 檔案中找到。您可以在任務定義和任何相關聯的 IAM 政策中使用直接uid值，以避免此名稱解析。 AWS Batch API 操作和 IAM batch:User 條件索引鍵都支援數值。

使用第三個陳述式來限制只有特定角色才能執行任務定義。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "batch:RegisterJobDefinition"
            ],
            "Resource": [
                "arn:aws:batch:us-east-2:999999999999:job-definition/JobDefA_*"
            ],
            "Condition": {
                "StringEquals": {
                    "batch:User": [
                        "nobody"
                    ],
                    "batch:Image": [
                        "999999999999.dkr.ecr.us-east-2.amazonaws.com/myImage"
                    ]
                },
                "Bool": {
                    "batch:Privileged": "false"
                }
            }
```

```
        },
        {
            "Effect": "Allow",
            "Action": [
                "batch:DeregisterJobDefinition"
            ],
            "Resource": [
                "arn:aws:batch:us-east-2:999999999999:job-definition/JobDefA_*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::999999999999:role/MyBatchJobRole"
            ]
        }
    ]
}
```

資源：僅限於提交任務時的任務定義字首

使用下列政策，以任何以 *JobDefA* 開頭的任務定義名稱，將任務提交至任何任務佇列。

> **Important**
>
> 在限制任務提交的資源層級存取範圍時，您必須同時提供任務佇列和任務定義資源類型。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "batch:SubmitJob"
            ],
```

```
            "Resource": [
                "arn:aws:batch:us-east-2:111122223333:job-definition/JobDefA_*",
                "arn:aws:batch:us-east-2:111122223333:job-queue/*"
            ]
        }
    ]
}
```

資源：限制為任務佇列

使用下列政策，以任何任務定義名稱將任務提交至名為 queue1 的特定任務佇列。

> **Important**
>
> 在限制任務提交的資源層級存取範圍時，您必須同時提供任務佇列和任務定義資源類型。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "batch:SubmitJob"
            ],
            "Resource": [
                "arn:aws:batch:us-east-2:888888888888:job-definition/*",
                "arn:aws:batch:us-east-2:888888888888:job-queue/queue1"
            ]
        }
    ]
}
```

當所有條件符合字串時拒絕動作

當 `batch:Image`（容器映像 ID）條件金鑰為 "*string1*" 且 `batch:LogDriver`（容器日誌驅動程式）條件金鑰為 "*string2*." AWS Batch 時，下列政策會拒絕存取 RegisterJobDefinition API 操作。

當任務跨越多個容器時，例如多節點平行任務，容器可能會有不同的組態。如果在一個陳述式中評估多個條件索引鍵，則會使用AND邏輯合併。因此，如果多個條件索引鍵中的任何一個不符合容器，則不會套用該容器Deny的效果。反之，同一任務中的不同容器可能會遭到拒絕。

如需 的條件金鑰清單 AWS Batch，請參閱《服務授權參考》中的 的條件金鑰 AWS Batch。除了 之外batch:ShareIdentifier，所有batch條件索引鍵都可以以此方式使用。batch:ShareIdentifier 條件索引鍵是為任務定義，而非任務定義。

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "batch:RegisterJobDefinition"
      ],
      "Resource": [
        "*"
      ]
    },
    {
      "Effect": "Deny",
      "Action": "batch:RegisterJobDefinition",
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "batch:Image": "string1",
          "batch:LogDriver": "string2"
        }
      }
    }
  ]
}
```

資源：當任何條件索引鍵符合字串時拒絕動作

當 batch:Image（容器映像 ID) 條件金鑰為 "*string1*" 或 batch:LogDriver（容器日誌驅動程式）條件金鑰為 "*string2*" 時，下列政策會拒絕存取 RegisterJobDefinition API 操作。當任務跨越多

個容器時，例如多節點平行任務，容器可能會有不同的組態。如果在一個陳述式中評估多個條件索引鍵，則會使用AND邏輯合併。因此，如果多個條件索引鍵中的任何一個不符合容器，則不會套用該容器Deny的效果。反之，同一任務中的不同容器可能會遭到拒絕。

如需 的條件金鑰清單 AWS Batch，請參閱《服務授權參考》中的 的條件金鑰 AWS Batch。
除了 之外batch:ShareIdentifier，所有batch條件索引鍵都可以以此方式使用。
(batch:ShareIdentifier條件索引鍵是為任務定義，而非任務定義。)

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "batch:RegisterJobDefinition"
      ],
      "Resource": [
        "*"
      ]
    },
    {
      "Effect": "Deny",
      "Action": [
        "batch:RegisterJobDefinition"
      ],
      "Resource": [
        "*"
      ],
      "Condition": {
        "StringEquals": {
          "batch:Image": [
            "string1"
          ]
        }
      }
    },
    {
      "Effect": "Deny",
      "Action": [
        "batch:RegisterJobDefinition"
```

```
      ],
      "Resource": [
        "*"
      ],
      "Condition": {
        "StringEquals": {
          "batch:LogDriver": [
            "string2"
          ]
        }
      }
    }
  ]
}
```

資源：使用 batch:ShareIdentifier 條件索引鍵

使用下列政策，將使用任務jobDefA定義的任務提交至具有lowCpu共用識別符jobqueue1的任務佇列。

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "batch:SubmitJob"
      ],
      "Resource": [
        "arn:aws:batch:us-east-2:555555555555:job-definition/JobDefA",
        "arn:aws:batch:us-east-2:555555555555:job-queue/jobqueue1"
      ],
      "Condition": {
        "StringEquals": {
          "batch:ShareIdentifier": [
            "lowCpu"
          ]
        }
      }
```

```
    }
  ]
}
```

使用 管理 SageMaker AI 資源 AWS Batch

此政策允許 AWS Batch 管理 SageMaker AI 資源。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "batch:*"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:CreateServiceLinkedRole"
            ],
            "Resource": "arn:aws:iam::*:role/
*AWSServiceRoleForAWSBatchWithSagemaker",
            "Condition": {
                 "StringEquals": {
                     "iam:AWSServiceName": "sagemaker-
queuing.batch.amazonaws.com"
                 }
             }
         },
         {
             "Effect": "Allow",
             "Action": "iam:PassRole",
             "Resource": "*",
             "Condition": {
                 "StringEquals": {
                     "iam:PassedToService": [
                         "sagemaker.amazonaws.com"
```

```
                    ]
                }
            }
        }
    ]
}
```

資源：依任務定義和任務佇列上的資源標籤限制任務提交

只有在任務佇列具有 標籤`Environment=dev`且任務定義具有 標籤 時，才能使用下列政策來提交任務`Project=calc`。此政策示範如何在任務提交期間使用資源標籤來控制對 AWS Batch 資源的存取。

> **Important**
>
> 使用評估任務定義資源標籤的政策提交任務時，您必須使用任務定義修訂格式 () 提交任務`job-definition:revision`。如果您在未指定修訂的情況下提交任務，將不會評估任務定義標籤，這可能會繞過您預期的存取控制。資源 ARN 中的 `*:*` 模式會強制執行提交必須包含修訂，以確保標籤政策一律有效套用。

此政策使用兩個不同的陳述式，因為它會將不同的標籤條件套用至不同的資源類型。在限制任務提交的資源層級存取範圍時，您必須同時提供任務佇列和任務定義資源類型。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "batch:SubmitJob",
      "Resource": "arn:aws:batch:*:*:job-queue/*",
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/Environment": "dev"
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": "batch:SubmitJob",
      "Resource": "arn:aws:batch:*:*:job-definition/*:*",
      "Condition": {
```

```
        "StringEquals": {
          "aws:ResourceTag/Project": "calc"
        }
      }
    }
  ]
}
```

Resource： AWS Batch managed 政策

AWS Batch 提供可連接到使用者的受管政策。此政策提供使用 AWS Batch 資源和 API 操作的許可。您可以直接套用此政策，或用它做為起點來建立您自己的政策。如需這些政策中提及的每個 API 操作的詳細資訊，請參閱 AWS Batch API 參考中的動作。

AWSBatchFullAccess

此政策允許完整管理員存取 AWS Batch。

若要檢視政策的 JSON，請參閱《AWS 受管政策參考指南》中的 AWSBatchFullAccess。

AWS Batch IAM 執行角色

執行角色會授予 Amazon ECS 容器和 AWS Fargate 代理程式代表您進行 AWS API 呼叫的許可。

> **Note**
>
> Amazon ECS 容器代理程式 1.16.0 版及更新版本支援執行角色。

IAM 執行角色是必要的，視您的任務需求而定。您可以針對與您的帳戶相關聯的不同用途和服務，擁有多個執行角色。

> **Note**
>
> 如需 Amazon ECS 執行個體角色的相關資訊，請參閱 Amazon ECS 執行個體角色。如需服務角色的相關資訊，請參閱 AWS Batch 如何使用 IAM。

Amazon ECS 提供 `AmazonECSTaskExecutionRolePolicy`受管政策。此政策包含上述常見使用案例的必要許可。對於以下概述的特殊使用案例，可能需要將內嵌政策新增至您的執行角色。

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecr:GetAuthorizationToken",
        "ecr:BatchCheckLayerAvailability",
        "ecr:GetDownloadUrlForLayer",
        "ecr:BatchGetImage",
        "logs:CreateLogStream",
        "logs:PutLogEvents"
      ],
      "Resource": "*"
    }
  ]
}
```

AWS Batch API 動作支援的資源層級許可

資源層級許可一詞是指能夠指定允許使用者對資源層級許可執行動作的資源。 AWS Batch 對資源層級許可有部分支援。對於某些 AWS Batch 動作，您可以根據必須符合的條件，控制何時允許使用者使用這些動作。您也可以根據允許使用者使用的特定資源來控制 。例如，您可以授予使用者提交任務的許可，但僅限特定任務佇列，且僅能藉由特定的任務定義來達成。

如需 AWS Batch 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AWS Batch》中的適用於 的動作、資源和條件金鑰。

教學課程：建立 IAM 執行角色

如果您的帳戶還沒有 IAM 執行角色，請使用下列步驟來建立角色。

1. 在以下網址開啟 IAM 主控台：https://console.aws.amazon.com/iam/。
2. 在導覽窗格中，選擇角色。
3. 選擇 Create Role (建立角色)。
4. 針對信任的實體類型，選擇 AWS 服務。

5. 針對服務或使用案例，選擇 Elastic Container Service。然後再次選擇彈性容器服務任務。
6. 選擇下一步。
7. 針對許可政策，搜尋 AmazonECSTaskExecutionRolePolicy。
8. 選擇 AmazonECSTaskExecutionRolePolicy 政策左側的核取方塊，然後選擇下一步。
9. 針對角色名稱，輸入 `ecsTaskExecutionRole`，然後選擇建立角色。

教學課程：檢查 IAM 執行角色

使用下列程序來檢查您的帳戶是否已有 IAM 執行角色，並視需要連接 受管 IAM 政策。

1. 在以下網址開啟 IAM 主控台：https://console.aws.amazon.com/iam/。
2. 在導覽窗格中，選擇角色。
3. 搜尋 `ecsTaskExecutionRole` 的角色清單。如果您找不到角色，請參閱 教學課程：建立 IAM 執行角色。如果您找到角色，請選擇角色以檢視連接的政策。
4. 在許可索引標籤上，確認 AmazonECSTaskExecutionRolePolicy 受管政策已連接至角色。如果連接政策，則您的執行角色已正確設定。如未連接，請按照以下子步驟連接政策。
   a. 選擇新增許可，然後選擇連接政策。
   b. 搜尋 AmazonECSTaskExecutionRolePolicy。
   c. 勾選 AmazonECSTaskExecutionRolePolicy 政策左側的方塊，然後選擇連接政策。
5. 選擇 Trust relationships (信任關係)。
6. 確認信任關係包含下列政策。如果信任關係符合以下政策，則會正確設定角色。如果信任關係不相符，請選擇編輯信任政策，輸入以下內容，然後選擇更新政策。

   JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "ecs-tasks.amazonaws.com"
      },
```

```
      "Action": "sts:AssumeRole"
    }
  ]
}
```

使用 AWS Batch的服務連結角色

AWS Batch use AWS Identity and Access Management (IAM) 服務連結角色。服務連結角色是直接連結至 的唯一 IAM 角色類型 AWS Batch。服務連結角色由 預先定義， AWS Batch 並包含該服務代表您呼叫其他 AWS 服務所需的所有許可。

AWS Batch 使用兩種不同的服務連結角色：

- AWSServiceRoleForBatch - 用於包括運算環境 AWS Batch 的操作。
- AWSServiceRoleForAWSBatchWithSagemaker - 適用於 SageMaker AI 工作負載管理和佇列。

主題

- 使用 的角色 AWS Batch
- 將 的角色 AWS Batch 與 SageMaker AI 搭配使用

使用 的角色 AWS Batch

AWS Batch use AWS Identity and Access Management (IAM) 服務連結角色。服務連結角色是直接連結至 的唯一 IAM 角色類型 AWS Batch。服務連結角色由 預先定義， AWS Batch 並包含該服務代表您呼叫其他 AWS 服務所需的所有許可。

服務連結角色可讓您更 AWS Batch 輕鬆地設定，因為您不必手動新增必要的許可。 AWS Batch 會定義其服務連結角色的許可，除非另有定義，否則只能 AWS Batch 擔任其角色。定義的許可包括信任政策和許可政策，且該許可政策無法附加至其他 IAM 實體。

> **Note**
>
> 執行下列其中一項操作來指定 AWS Batch 運算環境的服務角色。
>
> - 使用服務角色的空字串。這可讓 AWS Batch 建立服務角色。
> - 以下列格式指定服務角色：`arn:aws:iam::account_number:role/aws-service-role/batch.amazonaws.com/AWSServiceRoleForBatch`。

如需詳細資訊，請參閱 AWS Batch 《 使用者指南》不正確的角色名稱或 ARN中的 。

您必須先刪除服務連結角色的相關資源，才能將其刪除。這可保護您的 AWS Batch 資源，因為您不會不小心移除存取資源的許可。

如需有關支援服務連結角色的其他 服務的資訊，請參閱AWS 使用 IAM 的服務，並在服務連結角色欄中尋找具有是的服務。選擇具有連結的是，以檢視該服務的服務連結角色文件。

的服務連結角色許可 AWS Batch

AWS Batch 使用名為 AWSServiceRoleForBatch 的服務連結角色 – 允許 代表您 AWS Batch 建立和管理 AWS 資源。

AWSServiceRoleForBatch 服務連結角色信任下列服務擔任該角色：

- `batch.amazonaws.com`

名為 BatchServiceRolePolicy 的角色許可政策允許 對指定的資源 AWS Batch 完成下列動作：

- `autoscaling` – 允許 AWS Batch 建立和管理 Amazon EC2 Auto Scaling 資源。為大多數運算環境 AWS Batch 建立和管理 Amazon EC2 Auto Scaling 群組。
- `ec2` – 允許 AWS Batch 控制 Amazon EC2 執行個體的生命週期，以及建立和管理啟動範本和標籤。AWS Batch 會建立和管理某些 EC2 Spot 運算環境的 EC2 Spot 機群請求。
- `ecs` - 允許 為任務執行 AWS Batch 建立和管理 Amazon ECS 叢集、任務定義和任務。
- `eks` - 允許 AWS Batch 描述用於驗證的 Amazon EKS 叢集資源。
- `iam` - 允許 AWS Batch 驗證擁有者提供的角色並將其傳遞給 Amazon EC2、Amazon EC2 Auto Scaling 和 Amazon ECS。
- `logs` – 允許 AWS Batch 建立和管理 AWS Batch 任務的日誌群組和日誌串流。

您必須設定許可，以允許您的使用者、群組或角色建立、編輯或刪除服務連結角色。如需詳細資訊，請參閱 IAM 使用者指南中的服務連結角色許可。

為 AWS Batch建立服務連結角色

您不需要手動建立服務連結角色，當您在 AWS 管理主控台、 AWS CLI或 AWS API 中建立運算環境時， AWS Batch 會為您建立服務連結角色。

> **Important**
>
> 此服務連結角色可以顯示在您的帳戶，如果您於其他服務中完成一項動作時，可以使用支援此角色的功能。如果您在 2021 年 3 月 10 日之前使用該 AWS Batch 服務，當服務開始支援服務連結角色時， 會在您的帳戶中 AWS Batch 建立 AWSServiceRoleForBatch 角色。若要進一步了解，請參閱我的 中出現的新角色 AWS 帳戶。

若您刪除此服務連結角色，之後需要再次建立，您可以在帳戶中使用相同程序重新建立角色。當您建立運算環境時， 會再次為您 AWS Batch 建立服務連結角色。

為 AWS Batch編輯服務連結角色

AWS Batch 不允許您編輯 AWSServiceRoleForBatch 服務連結角色。因為有各種實體可能會參考服務連結角色，所以您無法在建立角色之後變更角色名稱。然而，您可使用 IAM 來編輯角色描述。如需詳細資訊，請參閱《IAM 使用者指南》中的編輯服務連結角色。

允許 IAM 實體編輯 AWSServiceRoleForBatch 服務連結角色的描述

將下列陳述式新增至許可政策。這可讓 IAM 實體編輯服務連結角色的描述。

```
{
    "Effect": "Allow",
    "Action": [
        "iam:UpdateRoleDescription"
    ],
    "Resource": "arn:aws:iam::*:role/aws-service-role/batch.amazonaws.com/
AWSServiceRoleForBatch",
    "Condition": {"StringLike": {"iam:AWSServiceName": "batch.amazonaws.com"}}
}
```

為 AWS Batch刪除服務連結角色

若您不再使用需要服務連結角色的功能或服務，我們建議您刪除該角色。如此一來，您就沒有未主動監控或維護的未使用實體。然而，務必清除您的服務連結角色，之後才能以手動方式將其刪除。

允許 IAM 實體刪除 AWSServiceRoleForBatch 服務連結角色

將下列陳述式新增至許可政策。這可讓 IAM 實體刪除服務連結角色。

```
{
    "Effect": "Allow",
```

```
    "Action": [
        "iam:DeleteServiceLinkedRole",
        "iam:GetServiceLinkedRoleDeletionStatus"
    ],
    "Resource": "arn:aws:iam::*:role/aws-service-role/batch.amazonaws.com/
AWSServiceRoleForBatch",
    "Condition": {"StringLike": {"iam:AWSServiceName": "batch.amazonaws.com"}}
}
```

清除服務連結角色

在您可以使用 IAM 刪除服務連結角色之前，您必須先確認角色沒有作用中工作階段，並刪除單一分割區中所有 AWS 區域中使用該角色的所有 AWS Batch 運算環境。

檢查服務連結角色是否有作用中工作階段

1. 前往 https://console.aws.amazon.com/iam/ 開啟 IAM 主控台。
2. 在導覽窗格中，選擇角色，然後選擇 AWSServiceRoleForBatch 名稱 （ 而非核取方塊 ）。
3. 在 Summary (摘要) 頁面上，選擇 Access Advisor (存取 Advisor)，然後檢閱服務連結角色的近期活動。

> **Note**
>
> 如果您不知道 AWS Batch 是否使用 AWSServiceRoleForBatch 角色，您可以嘗試刪除該角色。如果服務正在使用 角色，則該角色將無法刪除。您可以檢視正在使用角色的區域。如果服務正在使用該角色，您必須先等到工作階段結束，才能刪除該角色。您無法撤銷服務連結角色的工作階段。

移除 AWSServiceRoleForBatch 服務連結角色所使用的 AWS Batch 資源

您必須刪除所有 AWS 區域中使用 AWSServiceRoleForBatch 角色的所有 AWS Batch 運算環境，才能刪除 AWSServiceRoleForBatch 角色。

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中選取要使用的「區域」。
3. 在導覽窗格中，選擇 Compute environments (運算環境)。
4. 選取運算環境。
5. 選擇停用。等待狀態變更為 DISABLED。

6. 選取運算環境。
7. 選擇 刪除。選擇刪除運算環境，確認您想要刪除運算環境。
8. 針對在所有區域中使用服務連結角色的所有運算環境，重複步驟 1–7。

在 IAM 中刪除服務連結角色 （主控台）

您可以使用 IAM 主控台刪除服務連結角色。

刪除服務連結角色 (主控台)

1. 登入 AWS 管理主控台 並開啟位於 https：//https://console.aws.amazon.com/iam/ 的 IAM 主控台。
2. 在 IAM 主控台的導覽窗格中，選擇角色。然後選取 AWSServiceRoleForBatch 旁的核取方塊，而非名稱或資料列本身。
3. 選擇 Delete role (刪除角色)。
4. 在確認對話方塊中，檢閱服務上次存取資料，以顯示每個所選取角色上次存取 AWS 服務的時間。這可協助您確認角色目前是否作用中。如果您想要繼續進行，請選擇 Yes, Delete (是，刪除) 來提交服務連結角色以進行刪除。
5. 查看 IAM 主控台通知，監視服務連結角色刪除的進度。因為 IAM 服務連結角色刪除不同步，所以在您提交角色進行刪除之後，刪除任務可能會成功或失敗。

   - 如果任務成功，則會從清單中移除角色，而且成功通知會出現在頁面頂端。
   - 如果任務失敗，您可以從通知中選擇 View details (檢視詳細資訊) 或 View Resources (檢視資源)，以了解刪除失敗的原因。如果刪除因角色使用服務資源而失敗，則服務傳回該資訊時，通知會包含資源清單。您接著可以清除資源，並重新提交刪除。

     > **Note**
     >
     > 根據服務所傳回的資訊，您可能需要重複此程序數次。例如，您的服務連結角色可能會使用六個資源，而且您的服務可能傳回其中五項的相關資訊。如果您清除五個資源，並重新提交刪除角色，則刪除會失敗，而且服務會報告還有一個資源。服務可能會傳回所有資源、其中一些資源，或未報告任何資源。

   - 如果任務失敗，而且通知未包含資源清單，則服務可能未傳回該資訊。若要瞭解如何清除該服務的資源，請參閱使用 IAM 的AWS 服務。請在表格中找到您的服務，然後選擇 Yes (是) 連結，檢視該服務的服務連結角色文件。

在 IAM 中刪除服務連結角色 (AWS CLI)

您可以從 使用 IAM 命令 AWS Command Line Interface 來刪除服務連結角色。

刪除服務連結角色 (CLI)

1. 因為無法刪除正在使用或具有相關聯資源的服務連結角色，所以您必須提交刪除要求。如果不符合這些條件，則可以拒絕該請求。您必須從回應中擷取 `deletion-task-id`，以檢查刪除任務的狀態。輸入下列命令，以提交服務連結角色刪除要求：

   ```
   $ aws iam delete-service-linked-role --role-name AWSServiceRoleForBatch
   ```

2. 使用下列命令，以檢查刪除任務的狀態：

   ```
   $ aws iam get-service-linked-role-deletion-status --deletion-task-id deletion-task-id
   ```

   刪除任務的狀態可以是 NOT_STARTED、IN_PROGRESS、SUCCEEDED 或 FAILED。如果刪除失敗，則呼叫會傳回失敗原因，以進行疑難排解。如果刪除因角色使用服務資源而失敗，則服務傳回該資訊時，通知會包含資源清單。您接著可以清除資源，並重新提交刪除。

   > **Note**
   >
   > 根據服務所傳回的資訊，您可能需要重複此程序數次。例如，您的服務連結角色可能會使用六個資源，而且您的服務可能傳回其中五項的相關資訊。如果您清除五個資源，並重新提交刪除角色，則刪除會失敗，而且服務會報告還有一個資源。服務可能會傳回所有資源，其中一些資源。或者，它可能不會報告任何資源。若要了解如何清理未報告任何資源之服務的資源，請參閱AWS 使用 IAM 的 服務。請在表格中找到您的服務，然後選擇 Yes (是) 連結，檢視該服務的服務連結角色文件。

在 IAM (AWS API) 中刪除服務連結角色

您可以使用 IAM API 刪除服務連結角色。

刪除服務連結角色 (API)

1. 如需提交服務連結名單的刪除要求，請呼叫 DeleteServiceLinkedRole。在請求中，指定 AWSServiceRoleForBatch 角色名稱。

因為無法刪除正在使用或具有相關聯資源的服務連結角色，所以您必須提交刪除要求。如果不符合這些條件，則可以拒絕該請求。您必須從回應中擷取 `DeletionTaskId`，以檢查刪除任務的狀態。

2. 若要檢查刪除的狀態，請呼叫 GetServiceLinkedRoleDeletionStatus。在請求中，指定 `DeletionTaskId`。

   刪除任務的狀態可以是 `NOT_STARTED`、`IN_PROGRESS`、`SUCCEEDED` 或 `FAILED`。如果刪除失敗，則呼叫會傳回失敗原因，以進行疑難排解。如果刪除因角色使用服務資源而失敗，則服務傳回該資訊時，通知會包含資源清單。您接著可以清除資源，並重新提交刪除。

   > **Note**
   >
   > 根據服務所傳回的資訊，您可能需要重複此程序數次。例如，您的服務連結角色可能會使用六個資源，而且您的服務可能傳回其中五項的相關資訊。如果您清除五個資源，並重新提交刪除角色，則刪除會失敗，而且服務會報告還有一個資源。服務可能會傳回所有資源、其中一些資源，或未報告任何資源。若要瞭解如何清除未報告任何資源之服務的資源，請參閱使用 IAM 的AWS 服務。請在表格中找到您的服務，然後選擇 Yes (是) 連結，檢視該服務的服務連結角色文件。

AWS Batch 服務連結角色的支援區域

AWS Batch 支援在所有提供服務的區域中使用服務連結角色。如需詳細資訊，請參閱 AWS Batch 端點。

將 的角色 AWS Batch 與 SageMaker AI 搭配使用

AWS Batch use AWS Identity and Access Management (IAM) 服務連結角色。服務連結角色是直接連結至 的唯一 IAM 角色類型 AWS Batch。服務連結角色由 預先定義， AWS Batch 並包含該服務代表您呼叫其他 AWS 服務所需的所有許可。

服務連結角色可讓您更 AWS Batch 輕鬆地設定，因為您不必手動新增必要的許可。 AWS Batch 會定義其服務連結角色的許可，除非另有定義，否則只能 AWS Batch 擔任其角色。定義的許可包括信任政策和許可政策，且該許可政策無法附加至其他 IAM 實體。

您必須先刪除服務連結角色的相關資源，才能將其刪除。這可保護您的 AWS Batch 資源，因為您不會不小心移除存取資源的許可。

如需有關支援服務連結角色的其他 服務的資訊，請參閱AWS 使用 IAM 的服務，並在服務連結角色欄中尋找具有是的服務。選擇具有連結的是，以檢視該服務的服務連結角色文件。

的服務連結角色許可 AWS Batch

AWS Batch 使用名為 AWSServiceRoleForAWSBatchWithSagemaker 的服務連結角色 – 允許 代表您 AWS Batch 佇列和管理 SageMaker 訓練任務。

AWSServiceRoleForAWSBatchWithSagemaker 服務連結角色信任下列服務擔任該角色：

- `sagemaker-queuing.batch.amazonaws.com`

角色許可政策允許 對指定的資源 AWS Batch 完成下列動作：

- `sagemaker` – 允許 AWS Batch 管理 SageMaker 訓練任務、轉換任務和其他 SageMaker AI 資源。
- `iam:PassRole` – 允許 AWS Batch 將客戶定義的執行角色傳遞至 SageMaker AI 以進行任務執行。資源限制允許將角色傳遞至 SageMaker AI 服務。

您必須設定許可，以允許您的使用者、群組或角色建立、編輯或刪除服務連結角色。如需詳細資訊，請參閱 IAM 使用者指南中的服務連結角色許可。

為 AWS Batch建立服務連結角色

您不需要手動建立服務連結角色，當您在 AWS CLI、 AWS 管理主控台或 AWS API `CreateServiceEnvironment` 中使用 建立服務環境時， AWS Batch 會為您建立服務連結角色。

若您刪除此服務連結角色，之後需要再次建立，您可以在帳戶中使用相同程序重新建立角色。當您使用建立服務環境時`CreateServiceEnvironment`， 會再次為您 AWS Batch 建立服務連結角色。

若要檢視政策的 JSON，請參閱《 AWS 受管政策參考指南》中的 AWSBatchServiceRolePolicyForSageMaker。

為 AWS Batch編輯服務連結角色

AWS Batch 不允許您編輯 AWSServiceRoleForAWSBatchWithSagemaker 服務連結角色。因為有各種實體可能會參考服務連結角色，所以您無法在建立角色之後變更角色名稱。然而，您可使用 IAM 來編輯角色描述。如需詳細資訊，請參閱《IAM 使用者指南》中的編輯服務連結角色。

為 AWS Batch刪除服務連結角色

若您不再使用需要服務連結角色的功能或服務，我們建議您刪除該角色。如此一來，您就沒有未主動監控或維護的未使用實體。然而，務必清除您的服務連結角色，之後才能以手動方式將其刪除。

清除服務連結角色

您必須先確認角色沒有作用中工作階段，並刪除單一分割區中所有 AWS 區域中使用該角色的所有服務環境，才能使用 IAM 刪除服務連結角色。

檢查服務連結角色是否有作用中工作階段

1. 前往 https://console.aws.amazon.com/iam/ 開啟 IAM 主控台。
2. 在導覽窗格中，選擇角色，然後選擇 AWSServiceRoleForAWSBatchWithSagemaker 名稱（而非核取方塊）。
3. 在 Summary (摘要) 頁面上，選擇 Access Advisor (存取 Advisor)，然後檢閱服務連結角色的近期活動。

> **Note**
>
> 如果您不知道 AWS Batch 是否使用 AWSServiceRoleForAWSBatchWithSagemaker 角色，您可以嘗試刪除該角色。如果服務使用 角色，則該角色將無法刪除。您可以檢視正在使用角色的區域。如果服務正在使用該角色，您必須先等到工作階段結束，才能刪除該角色。您無法撤銷服務連結角色的工作階段。

移除 AWSServiceRoleForAWSBatchWithSagemaker 服務連結角色所使用的 AWS Batch 資源

您必須取消所有任務佇列與所有服務環境的關聯，然後您必須刪除所有 AWS 區域中使用 AWSServiceRoleForAWSBatchWithSagemaker 角色的所有服務環境，才能刪除 AWSServiceRoleForAWSBatchWithSagemaker 角色。

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中選取要使用的「區域」。
3. 在導覽窗格中，選擇環境，然後選擇服務環境。
4. 選取所有服務環境。
5. 選擇停用。等待狀態變更為 DISABLED。
6. 選取服務環境。

7. 選擇 刪除。選擇刪除服務環境，確認您想要刪除服務環境。
8. 針對在所有區域中使用服務連結角色的所有服務環境，重複步驟 1–7。

在 IAM 中刪除服務連結角色 （主控台）

您可以使用 IAM 主控台刪除服務連結角色。

刪除服務連結角色 (主控台)

1. 登入 AWS 管理主控台 並開啟位於 https：//https://console.aws.amazon.com/iam/ 的 IAM 主控台。
2. 在 IAM 主控台的導覽窗格中，選擇角色。然後選取 AWSServiceRoleForAWSBatchWithSagemaker 旁的核取方塊，而非名稱或資料列本身。
3. 選擇 Delete role (刪除角色)。
4. 在確認對話方塊中，檢閱服務上次存取資料，以顯示每個所選取角色上次存取 AWS 服務的時間。這可協助您確認角色目前是否作用中。如果您想要繼續進行，請選擇 Yes, Delete (是，刪除) 來提交服務連結角色以進行刪除。
5. 查看 IAM 主控台通知，監視服務連結角色刪除的進度。因為 IAM 服務連結角色刪除不同步，所以在您提交角色進行刪除之後，刪除任務可能會成功或失敗。

   - 如果任務成功，則會從清單中移除角色，而且成功通知會出現在頁面頂端。
   - 如果任務失敗，您可以從通知中選擇 View details (檢視詳細資訊) 或 View Resources (檢視資源)，以了解刪除失敗的原因。如果刪除因角色使用服務資源而失敗，則服務傳回該資訊時，通知會包含資源清單。您接著可以清除資源，並重新提交刪除。

     > **Note**
     >
     > 根據服務所傳回的資訊，您可能需要重複此程序數次。例如，您的服務連結角色可能會使用六個資源，而且您的服務可能傳回其中五項的相關資訊。如果您清除五個資源，並重新提交刪除角色，則刪除會失敗，而且服務會報告還有一個資源。服務可能會傳回所有資源、其中一些資源，或未報告任何資源。

   - 如果任務失敗，而且通知未包含資源清單，則服務可能未傳回該資訊。若要瞭解如何清除該服務的資源，請參閱使用 IAM 的AWS 服務。請在表格中找到您的服務，然後選擇 Yes (是) 連結，檢視該服務的服務連結角色文件。

在 IAM 中刪除服務連結角色 (AWS CLI)

您可以從 使用 IAM 命令 AWS Command Line Interface 來刪除服務連結角色。

刪除服務連結角色 (CLI)

1. 因為無法刪除正在使用或具有相關聯資源的服務連結角色，所以您必須提交刪除要求。如果不符合這些條件，則可以拒絕該請求。您必須從回應中擷取 `deletion-task-id`，以檢查刪除任務的狀態。輸入下列命令，以提交服務連結角色刪除要求：

   ```
   $ aws iam delete-service-linked-role --role-name
    AWSServiceRoleForAWSBatchWithSagemaker
   ```

2. 使用下列命令，以檢查刪除任務的狀態：

   ```
   $ aws iam get-service-linked-role-deletion-status --deletion-task-id deletion-task-id
   ```

   刪除任務的狀態可以是 NOT_STARTED、IN_PROGRESS、SUCCEEDED 或 FAILED。如果刪除失敗，則呼叫會傳回失敗原因，以進行疑難排解。如果刪除因角色使用服務資源而失敗，則服務傳回該資訊時，通知會包含資源清單。您接著可以清除資源，並重新提交刪除。

   > **Note**
   >
   > 根據服務所傳回的資訊，您可能需要重複此程序數次。例如，您的服務連結角色可能會使用六個資源，而且您的服務可能傳回其中五項的相關資訊。如果您清除五個資源，並重新提交刪除角色，則刪除會失敗，而且服務會報告還有一個資源。服務可能會傳回所有資源，其中一些資源。或者，它可能不會報告任何資源。若要了解如何清理未報告任何資源之服務的資源，請參閱AWS 使用 IAM 的服務。請在表格中找到您的服務，然後選擇 Yes (是) 連結，檢視該服務的服務連結角色文件。

在 IAM (AWS API) 中刪除服務連結角色

您可以使用 IAM API 刪除服務連結角色。

刪除服務連結角色 (API)

1. 如需提交服務連結名單的刪除要求，請呼叫 DeleteServiceLinkedRole。在請求中，指定 AWSServiceRoleForAWSBatchWithSagemaker 角色名稱。

因為無法刪除正在使用或具有相關聯資源的服務連結角色，所以您必須提交刪除要求。如果不符合這些條件，則可以拒絕該請求。您必須從回應中擷取 `DeletionTaskId`，以檢查刪除任務的狀態。

2. 若要檢查刪除的狀態，請呼叫 GetServiceLinkedRoleDeletionStatus。在請求中，指定 `DeletionTaskId`。

   刪除任務的狀態可以是 `NOT_STARTED`、`IN_PROGRESS`、`SUCCEEDED` 或 `FAILED`。如果刪除失敗，則呼叫會傳回失敗原因，以進行疑難排解。如果刪除因角色使用服務資源而失敗，則服務傳回該資訊時，通知會包含資源清單。您接著可以清除資源，並重新提交刪除。

   > **Note**
   >
   > 根據服務所傳回的資訊，您可能需要重複此程序數次。例如，您的服務連結角色可能會使用六個資源，而且您的服務可能傳回其中五項的相關資訊。如果您清除五個資源，並重新提交刪除角色，則刪除會失敗，而且服務會報告還有一個資源。服務可能會傳回所有資源、其中一些資源，或未報告任何資源。若要瞭解如何清除未報告任何資源之服務的資源，請參閱使用 IAM 的AWS 服務。請在表格中找到您的服務，然後選擇 Yes (是) 連結，檢視該服務的服務連結角色文件。

AWS Batch 服務連結角色的支援區域

AWS Batch 支援在所有提供服務的區域中使用服務連結角色。如需詳細資訊，請參閱 AWS Batch 端點。

Amazon ECS 執行個體角色

AWS Batch 運算環境會填入 Amazon ECS 容器執行個體。它們會在本機執行 Amazon ECS 容器代理程式。Amazon ECS 容器代理程式會代表您呼叫各種 AWS API 操作。因此，執行代理程式的容器執行個體需要這些服務的 IAM 政策和角色，才能識別代理程式屬於您。您必須建立 IAM 角色和執行個體描述檔，容器執行個體才能在啟動時使用。否則，您無法建立運算環境，並在其中啟動容器執行個體。此要求適用於使用或未使用 Amazon ECS 最佳化 AMI 啟動的容器執行個體。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 執行個體角色。

主題

- 檢查您帳戶的 Amazon ECS 執行個體角色

檢查您帳戶的 Amazon ECS 執行個體角色

Amazon ECS 執行個體角色和執行個體描述檔會在主控台初次執行體驗中自動為您建立。不過，您可以依照下列步驟來檢查您的帳戶是否已有 Amazon ECS 執行個體角色和執行個體描述檔。下列步驟也涵蓋如何連接 受管 IAM 政策。

教學課程：在 IAM 主控台**`ecsInstanceRole`**中檢查

1. 在以下網址開啟 IAM 主控台：https://console.aws.amazon.com/iam/。
2. 在導覽窗格中，選擇角色。
3. 搜尋 `ecsInstanceRole` 的角色清單。如果角色不存在，請使用下列步驟來建立角色。
   a. 選擇建立角色。
   b. 對於 Trusted entity type (信任的實體類型)，請選擇 AWS 服務。
   c. 針對常用案例，選擇 EC2。
   d. 選擇下一步。
   e. 針對許可政策，搜尋 AmazonEC2ContainerServiceforEC2Role。
   f. 選擇 AmazonEC2ContainerServiceforEC2Role 旁的核取方塊，然後選擇下一步。
   g. 針對 Role Name (角色名稱)，輸入 `ecsInstanceRole`，然後選擇 Create Role (建立角色)。

   > **Note**
   > 如果您使用 AWS 管理主控台 為 Amazon EC2 建立角色，主控台會建立與角色同名的執行個體描述檔。

或者，您可以使用 AWS CLI 來建立 IAM `ecsInstanceRole` 角色。下列範例會建立具有信任政策和 AWS 受管政策的 IAM 角色。

教學課程：建立 IAM 角色和執行個體描述檔 (AWS CLI)

1. 建立下列信任政策，並將其儲存在名為 的文字檔案中`ecsInstanceRole-role-trust-policy.json`。

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": { "Service": "ec2.amazonaws.com"},
      "Action": "sts:AssumeRole"
    }
  ]
}
```

2. 使用 create-role 命令來建立ecsInstanceRole角色。在 assume-role-policy-document 參數中指定信任政策檔案位置。

```
$ aws iam create-role \
    --role-name ecsInstanceRole \
    --assume-role-policy-document file://ecsInstanceRole-role-trust-policy.json
```

3. 使用 create-instance-profile 命令來建立名為 的執行個體描述檔ecsInstanceRole。

> **Note**
>
> 您需要在 和 AWS API 中將角色 AWS CLI 和執行個體描述檔建立為個別動作。

```
$ aws iam create-instance-profile --instance-profile-name ecsInstanceRole
```

以下是回應範例。

```
{
    "InstanceProfile": {
        "Path": "/",
        "InstanceProfileName": "ecsInstanceRole",
        "InstanceProfileId": "AIPAT46P5RDITREXAMPLE",
        "Arn": "arn:aws:iam::123456789012:instance-profile/ecsInstanceRole",
        "CreateDate": "2022-06-30T23:53:34.093Z",
        "Roles": [],    }
```

```
}
```

4. 使用 add-role-to-instance-profile 命令，將ecsInstanceRole角色新增至ecsInstanceRole執行個體描述檔。

```
aws iam add-role-to-instance-profile \
    --role-name ecsInstanceRole --instance-profile-name ecsInstanceRole
```

5. 使用 attach-role-policy 命令將AmazonEC2ContainerServiceforEC2Role AWS 受管政策連接至ecsInstanceRole角色。

```
$ aws iam attach-role-policy \
    --policy-arn arn:aws:iam::aws:policy/service-role/
AmazonEC2ContainerServiceforEC2Role \
    --role-name ecsInstanceRole
```

Amazon EC2 Spot 機群角色

如果您建立使用 Amazon EC2 Spot Fleet 執行個體的受管運算環境，則必須建立AmazonEC2SpotFleetTaggingRole政策。此政策授予 Spot Fleet 代表您啟動、標記和終止執行個體的許可。在您的 Spot Fleet 請求中指定角色。您還必須擁有 Amazon EC2 Spot 和 Spot Fleet 的 AWSServiceRoleForEC2Spot 和 AWSServiceRoleForEC2SpotFleet 服務連結角色。 Amazon EC2 使用下列指示來建立所有這些角色。如需詳細資訊，請參閱《IAM 使用者指南》中的使用服務連結角色和建立角色以委派許可給 AWS 服務。

主題

- 在 中建立 Amazon EC2 Spot 機群角色 AWS 管理主控台
- 使用 建立 Amazon EC2 Spot 機群角色 AWS CLI

在 中建立 Amazon EC2 Spot 機群角色 AWS 管理主控台

為 **AmazonEC2SpotFleetTaggingRole** Amazon EC2 Spot Fleet 建立 IAM 服務連結角色

1. 前往 https://console.aws.amazon.com/iam/ 開啟 IAM 主控台。
2. 針對存取管理，選擇角色、
3. 針對角色，選擇建立角色。
4. 從選取信任實體類型的信任實體中，選擇 AWS 服務。

5. 對於其他的使用案例 AWS 服務，請選擇 EC2，然後選擇 EC2 - Spot 機群標記。
6. 選擇下一步。
7. 從政策名稱的許可政策中，驗證 AmazonEC2SpotFleetTaggingRole。
8. 選擇下一步。
9. 針對名稱、檢閱和建立：

    a. 針對角色名稱，輸入名稱以識別角色。

    b. 針對描述，輸入政策的簡短說明。

    c. （選用）對於步驟 1：選取信任的實體，選擇編輯以修改程式碼。

    d. （選用）對於步驟 2：新增許可，選擇編輯以修改程式碼。

    e. （選用）針對新增標籤，選擇新增標籤以將標籤新增至資源。

    f. 選擇建立角色。

> **Note**
>
> 過去，Amazon EC2 Spot Fleet 角色有兩個受管政策。
>
> - AmazonEC2SpotFleetRole：這是 Spot Fleet 角色的原始受管政策。不過，我們不再建議您將其與 搭配使用 AWS Batch。此政策不支援在運算環境中使用AWSServiceRoleForBatch服務連結角色所需的 Spot 執行個體標記。如果您先前已使用此政策建立 Spot Fleet 角色，請將新的建議政策套用至該角色。如需詳細資訊，請參閱建立時未標記 Spot 執行個體。
> - AmazonEC2SpotFleetTaggingRole：此角色提供標記 Amazon EC2 Spot 執行個體的所有必要許可。使用此角色允許在 AWS Batch 運算環境中標記 Spot 執行個體。

使用 建立 Amazon EC2 Spot 機群角色 AWS CLI

為您的 Spot Fleet 運算環境建立 AmazonEC2SpotFleetTaggingRole IAM 角色

1. 使用 執行下列命令 AWS CLI。

```
$ aws iam create-role --role-name AmazonEC2SpotFleetTaggingRole \
    --assume-role-policy-document '{
  "Version":"2012-10-17",
  "Statement": [
```

```
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "spotfleet.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}'
```

2. 若要將 AmazonEC2SpotFleetTaggingRole 受管 IAM 政策連接至您的 AmazonEC2SpotFleetTaggingRole 角色，請使用 執行下列命令 AWS CLI。

```
$ aws iam attach-role-policy \
  --policy-arn \
    arn:aws:iam::aws:policy/service-role/AmazonEC2SpotFleetTaggingRole \
  --role-name \
    AmazonEC2SpotFleetTaggingRole
```

為 **AWSServiceRoleForEC2Spot** Amazon EC2 Spot 建立 IAM 服務連結角色

> **Note**
>
> 如果 IAM AWSServiceRoleForEC2Spot 服務連結角色已存在，您會看到類似以下的錯誤訊息。
>
> ```
> An error occurred (InvalidInput) when calling the CreateServiceLinkedRole
>  operation:
> Service role name AWSServiceRoleForEC2Spot has been taken in this account,
>  please try a different suffix.
> ```

- 使用 執行下列命令 AWS CLI。

```
$ aws iam create-service-linked-role --aws-service-name spot.amazonaws.com
```

為 **AWSServiceRoleForEC2SpotFleet** Amazon EC2 Spot Fleet 建立 IAM 服務連結角色

> **Note**
>
> 如果 AWSServiceRoleForEC2SpotFleet IAM 服務連結角色已存在，您會看到類似以下的錯誤訊息。
>
> ```
> An error occurred (InvalidInput) when calling the CreateServiceLinkedRole
>  operation:
> Service role name AWSServiceRoleForEC2SpotFleet has been taken in this account,
>  please try a different suffix.
> ```

- 使用 執行下列命令 AWS CLI。

```
$ aws iam create-service-linked-role --aws-service-name spotfleet.amazonaws.com
```

EventBridge IAM 角色

Amazon EventBridge 提供近乎即時的系統事件串流，描述 AWS 資源的變更。 AWS Batch 任務可作為 EventBridge 目標使用。利用可快速設定的簡單規則，匹配事件並提交 AWS Batch 任務以回應事件。在使用 EventBridge 規則和目標提交 AWS Batch 任務之前，EventBridge 必須具有代表您執行 AWS Batch 任務的許可。

> **Note**
>
> 當您在 EventBridge 主控台中建立將 AWS Batch 佇列指定為目標的規則時，您可以建立此角色。如需範例演練，請參閱 AWS Batch 任務做為 EventBridge 目標。您可以使用 IAM 主控台手動建立 EventBridge 角色。如需說明，請參閱《IAM 使用者指南》中的使用自訂信任政策（主控台）建立角色。

EventBridge IAM 角色的信任關係必須為events.amazonaws.com服務主體提供擔任該角色的能力。

JSON

```
{
```

```
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "events.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

請確定連接至 EventBridge IAM 角色的政策允許 資源的`batch:SubmitJob`許可。在下列範例中，AWS Batch 提供 `AWSBatchServiceEventTargetRole`受管政策來提供這些許可。

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "batch:SubmitJob"
       ],
      "Resource": "*"
    }
  ]
}
```

建立 Virtual Private Cloud

運算環境中的運算資源需要外部網路存取，才能與 AWS Batch 和 Amazon ECS 服務端點通訊。不過，您可能有想要在私有子網路中執行的任務。若要彈性地在公有或私有子網路中執行任務，請建立同時具有公有和私有子網路的 VPC。

您可以使用 Amazon Virtual Private Cloud (Amazon VPC) 在您定義的虛擬網路中啟動 AWS 資源。本主題提供 Amazon VPC 精靈的連結，以及要選取的選項清單。

建立 VPC

如需有關如何建立 Amazon VPC 的資訊，請參閱《Amazon VPC 使用者指南》中的僅建立 VPC，並使用下表來決定要選取的選項。

選項	Value
要建立的資源	僅 VPC
名稱	可以選擇為 VPC 提供名稱。
IPv4 CIDR 區塊	IPv4 CIDR 手動輸入 CIDR 區塊大小必須為介於 /16 和 /28 之間的大小。
IPv6 CIDR 區塊	無 IPv6 CIDR 區塊
租用	預設

如需有關 Amazon VPC 的詳細資訊，請參閱《Amazon VPC 使用者指南》中的什麼是 Amazon VPC？。

後續步驟

建立 VPC 之後，請考慮下列後續步驟：

- 如果您的公有和私有資源需要入站網路存取，則為其建立安全群組。如需詳細資訊，請參閱《Amazon VPC 使用者指南》中的使用安全群組。
- 建立 AWS Batch 受管運算環境，將運算資源啟動到您的新 VPC。如需詳細資訊，請參閱建立運算環境。如果您在 AWS Batch 主控台中使用運算環境建立精靈，您可以指定剛建立的 VPC，以及要啟動執行個體的公有或私有子網路。
- 建立映射至新運算環境 AWS Batch 的任務佇列。如需詳細資訊，請參閱建立任務佇列。
- 建立任務定義來執行您的任務。如需詳細資訊，請參閱建立單一節點任務定義。
- 將任務和任務定義提交到新的任務佇列。此任務會登陸您使用新 VPC 和子網路建立的運算環境。如需詳細資訊，請參閱教學課程：提交任務。

使用界面端點存取 AWS Batch

您可以使用 在 VPC 與 之間 AWS PrivateLink 建立私有連線 AWS Batch。您可以 AWS Batch 像在 VPC 中一樣存取 ，無需使用網際網路閘道、NAT 裝置、VPN 連接或 Direct Connect 連線。VPC 中的執行個體不需要公有 IP 地址即可存取 AWS Batch。

您可以建立由 AWS PrivateLink提供支援的介面端點來建立此私有連線。我們會在您為介面端點啟用的每個子網中建立端點網路介面。這些是請求者管理的網路介面，可作為目的地為 AWS Batch之流量的進入點。

如需詳細資訊，請參閱《 AWS PrivateLink 指南》中的界面 VPC 端點。

的考量事項 AWS Batch

在您設定介面端點之前 AWS Batch，請檢閱《 AWS PrivateLink 指南》中的介面端點屬性和限制。

AWS Batch 支援透過介面端點呼叫其所有 API 動作。

設定介面 VPC 端點之前 AWS Batch，請注意下列考量：

- 使用 Fargate 資源啟動類型的任務不需要 Amazon ECS 的介面 VPC 端點，但您可能需要介面 VPC 端點 AWS Batch、Amazon ECR、Secrets Manager 或 Amazon CloudWatch Logs，如下所述。
  - 若要執行任務，您必須為 Amazon ECS 建立介面 VPC 端點。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的界面 VPC 端點 (AWS PrivateLink)。
  - 若要允許任務從 Amazon ECR 提取私有映像，您必須為 Amazon ECR 建立介面 VPC 端點。如需詳細資訊，請參閱 《Amazon Elastic Container Registry 使用者指南》中的介面 VPC 端點 (AWS PrivateLink)。
  - 若要允許任務從 Secrets Manager 提取敏感資料，您必須為 Secrets Manager 建立介面 VPC 端點。如需詳細資訊，請參閱 《AWS Secrets Manager 使用者指南》中的搭配使用 Secrets Manager 與 VPC 端點。
  - 如果您的 VPC 沒有網際網路閘道，且您的任務使用`awslogs`日誌驅動程式將日誌資訊傳送至 CloudWatch Logs，則必須為 CloudWatch Logs 建立介面 VPC 端點。如需詳細資訊，請參閱 《Amazon CloudWatch Logs 使用者指南》中的使用 CloudWatch Events 搭配介面 VPC 端點。
- 使用 EC2 資源的任務需要啟動它們的容器執行個體，才能執行版本 `1.25.1`或更新版本的 Amazon ECS 容器代理程式。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS Linux 容器代理程式版本。
- VPC 端點目前不支援跨區域請求。請確實在計劃發出 AWS Batch API 呼叫的相同區域中建立端點。

- 透過 Amazon Route 53，VPC 端點僅支援 Amazon 提供的 DNS。如果您想要使用自己的 DNS，您可以使用條件式 DNS 轉送。如需詳細資訊，請參閱《Amazon VPC 使用者指南》中的 DHCP 選項集。
- 連接到 VPC 端點的安全群組，必須允許從 VPC 的私有子網路，透過 443 埠傳入的連線。
- AWS Batch 不支援下列 VPC 介面端點 AWS 區域：
  - 亞太區域 (大阪) (ap-northeast-3)
  - 亞太區域 (雅加達) (ap-southeast-3)

建立 的介面端點 AWS Batch

您可以使用 Amazon VPC AWS Batch 主控台或 AWS Command Line Interface () 建立 的介面端點 AWS CLI。如需詳細資訊，請參閱《AWS PrivateLink 指南》中的「建立介面端點」。

AWS Batch 使用下列服務名稱建立 的介面端點：

- com.amazonaws.*region* .batch
- com.amazonaws.*region* .batch-fips （如需符合 FIPS 標準的端點，請參閱AWS Batch 端點和配額)

例如：

```
com.amazonaws.us-east-2.batch
```

```
com.amazonaws.us-east-2.batch-fips
```

在aws-cn分割區中，格式不同：

```
cn.com.amazonaws.region.batch
```

例如：

```
cn.com.amazonaws.cn-northwest-1.batch
```

AWS Batch 介面端點的私有 DNS 名稱

如果您為介面端點啟用私有 DNS，您可以使用特定 DNS 名稱來連線 AWS Batch，我們提供下列選項：

- batch.*region* .amazonaws.com
- batch.*region* .api.aws

對於 FIPS 相容端點：

- batch-fips.*region* .api.aws
- 不支援 fips.batch.*region* .amazonaws.com

如需詳細資訊，請參閱《 AWS PrivateLink 指南》中的透過介面端點存取服務。

為您的介面端點建立端點政策

端點政策為 IAM 資源，您可將其連接至介面端點。預設端點政策允許 AWS Batch 透過介面端點完整存取 。若需控制您的 VPC 對 AWS Batch 的存取範圍，請將自訂端點政策套用至介面型端點。

端點政策會指定以下資訊：

- 可執行動作的主體 (AWS 帳戶、使用者和 IAM 角色）。
- 可執行的動作。
- 可供執行動作的資源。

如需詳細資訊，請參閱《AWS PrivateLink 指南》中的「使用端點政策控制對服務的存取」。

範例： AWS Batch 動作的 VPC 端點政策

以下是自訂端點政策的範例。當您將此政策連接到介面端點時，它會授予所有資源上所有主體的所列 AWS Batch 動作的存取權。

```
{
   "Statement": [
      {
         "Principal": "*",
         "Effect": "Allow",
         "Action": [
            "batch:SubmitJob",
            "batch:ListJobs",
            "batch:DescribeJobs"
         ],
```

```
            "Resource":"*"
        }
    ]
}
```

的合規驗證 AWS Batch

若要了解 是否 AWS 服務 在特定合規計劃的範圍內，請參閱AWS 服務 合規計劃範圍內然後選擇您感興趣的合規計劃。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，請參閱下載報告 in AWS Artifact

您使用 時的合規責任 AWS 服務 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。如需使用 時合規責任的詳細資訊 AWS 服務，請參閱 AWS 安全文件。

中的基礎設施安全 AWS Batch

作為受管服務， AWS Batch 受到 AWS 全球網路安全的保護。如需 AWS 安全服務以及 如何 AWS 保護基礎設施的相關資訊，請參閱AWS 雲端安全。若要使用基礎設施安全的最佳實務來設計您的 AWS 環境，請參閱安全支柱 AWS Well-Architected Framework 中的基礎設施保護。

您可以使用 AWS 發佈的 API 呼叫， AWS Batch 透過網路存取 。使用者端必須支援下列專案：

- Transport Layer Security (TLS)。我們需要 TLS 1.2 並建議使用 TLS 1.3。
- 具備完美轉送私密(PFS)的密碼套件，例如 DHE (Ephemeral Diffie-Hellman)或 ECDHE (Elliptic Curve Ephemeral Diffie-Hellman)。現代系統(如 Java 7 和更新版本)大多會支援這些模式。

您可以從任何網路位置呼叫這些 API 操作，但 AWS Batch 支援以資源為基礎的存取政策，其中可能包括根據來源 IP 地址的限制。您也可以使用 AWS Batch 政策來控制來自特定 Amazon Virtual Private Cloud (Amazon VPC) 端點或特定 VPCs存取。實際上，這只會隔離網路內特定 VPC 對指定 AWS Batch 資源 AWS 的網路存取。

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權限的實體執行該動作。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全域條件內容索引鍵，以限制將另一個 服務 AWS Batch 提供給資源的許可。如果 aws:SourceArn 值不包含帳戶 ID (例如 Amazon S3 儲存貯體 ARN)，您必須使用這兩個全域條件內容金鑰來限制許可。如果同時使用這兩個全域條件內容金鑰，且 aws:SourceArn 值包含帳戶 ID，則在相同政策陳述式中使用 aws:SourceAccount 值和 aws:SourceArn 值中的帳戶時，必須使用相同的帳戶 ID。如果您想要僅允許一個資源與跨服務存取相關聯，則請使用 aws:SourceArn。如果您想要允許該帳戶中的任何資源與跨服務使用相關聯，請使用 aws:SourceAccount。

的值aws:SourceArn必須是 AWS Batch 存放的資源。

防範混淆代理人問題的最有效方法是使用 aws:SourceArn 全域條件內容索引鍵，以及資源的完整 ARN。如果不知道資源的完整 ARN，或者如果您指定了多個資源，請使用 aws:SourceArn 全域內容條件索引鍵搭配萬用字元 (*) 來表示 ARN 的未知部分。例如 arn:aws:*servicename*:*:*123456789012*:*。

下列範例示範如何在 中使用 aws:SourceArn和 aws:SourceAccount全域條件內容索引鍵 AWS Batch ，以防止混淆代理人問題。

範例：僅存取一個運算環境的角色

下列角色只能用於存取一個運算環境。任務名稱必須指定為 ，*因為任務佇列可以與多個運算環境相關聯。

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "batch.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "123456789012"
        },
```

```
        "ArnLike": {
          "aws:SourceArn": [
            "arn:aws:batch:us-east-1:123456789012:compute-environment/testCE",
            "arn:aws:batch:us-east-1:123456789012:job/*"
          ]
        }
      }
    }
  ]
}
```

範例：存取多個運算環境的角色

下列角色可用來存取多個運算環境。任務名稱必須指定為 ，*因為任務佇列可以與多個運算環境相關聯。

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "batch.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "123456789012"
        },
        "ArnLike": {
          "aws:SourceArn": [
            "arn:aws:batch:us-east-1:123456789012:compute-environment/*",
            "arn:aws:batch:us-east-1:123456789012:job/*"
          ]
        }
      }
    }
  ]
}
```

使用 記錄 AWS Batch API 呼叫 AWS CloudTrail

AWS Batch 已與 服務整合 AWS CloudTrail，此服務可提供使用者、角色或 AWS 服務在其中採取之動作的記錄 AWS Batch。CloudTrail 會將 的所有 API 呼叫擷取 AWS Batch 為事件。擷取的呼叫包括來自 AWS Batch 主控台的呼叫，以及對 AWS Batch API 操作的程式碼呼叫。如果您建立線索，您可以將 CloudTrail 事件持續交付至 Amazon S3 儲存貯體，包括 的事件 AWS Batch。即使您未設定追蹤，依然可以透過 CloudTrail 主控台中的 Event history (事件歷史記錄) 檢視最新事件。您可以利用 CloudTrail 所收集的資訊來判斷向 AWS Batch發出的請求，以及發出請求的 IP 地址、人員、時間和其他詳細資訊。

若要進一步了解 CloudTrail，請參閱AWS CloudTrail 《使用者指南》。

主題

- AWS Batch CloudTrail 中的資訊
- 參考：了解 AWS Batch 日誌檔案項目

AWS Batch CloudTrail 中的資訊

當您建立 AWS 帳戶時，會在您的帳戶上啟用 CloudTrail。當活動在 中發生時 AWS Batch，該活動會與事件歷史記錄中的其他服務 AWS 事件一起記錄在 CloudTrail 事件中。您可以在 AWS 帳戶中檢視、搜尋和下載最近的事件。如需詳細資訊，請參閱《使用 CloudTrail 事件歷史記錄檢視事件》https://docs.aws.amazon.com/awscloudtrail/latest/userguide/view-cloudtrail-events.html。

若要持續記錄您 AWS 帳戶中的事件，包括 的事件 AWS Batch，請建立追蹤。線索能讓 CloudTrail 將日誌檔案交付至 Amazon S3 儲存貯體。根據預設，當您在主控台中建立線索時，線索會套用至所有 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 日誌中所收集的事件資料。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
- 從多個區域接收 CloudTrail 日誌檔案，以及從多個帳戶接收 CloudTrail 日誌檔案

CloudTrail 會記錄所有 AWS Batch 動作，並記錄在 https：//https://docs.aws.amazon.com/batch/latest/APIReference/。例如，對 SubmitJob、ListJobs 和 DescribeJobs 區段的呼叫，會在 CloudTrail 日誌檔案中產生項目。

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：

- 該請求是否使用根或 IAM 使用者憑證提出。
- 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。

參考：了解 AWS Batch 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。

以下範例顯示的是展示 CreateComputeEnvironment 動作的 CloudTrail 日誌項目。

```
{
  "eventVersion": "1.05",
  "userIdentity": {
    "type": "AssumedRole",
    "principalId": "AIDACKCEVSQ6C2EXAMPLE:admin",
    "arn": "arn:aws:sts::012345678910:assumed-role/Admin/admin",
    "accountId": "012345678910",
    "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
    "sessionContext": {
      "attributes": {
        "mfaAuthenticated": "false",
        "creationDate": "2017-12-20T00:48:46Z"
      },
      "sessionIssuer": {
        "type": "Role",
        "principalId": "AIDACKCEVSQ6C2EXAMPLE",
        "arn": "arn:aws:iam::012345678910:role/Admin",
        "accountId": "012345678910",
        "userName": "Admin"
      }
```

```
    }
  },
  "eventTime": "2017-12-20T00:48:46Z",
  "eventSource": "batch.amazonaws.com",
  "eventName": "CreateComputeEnvironment",
  "awsRegion": "us-east-1",
  "sourceIPAddress": "203.0.113.1",
  "userAgent": "aws-cli/1.11.167 Python/2.7.10 Darwin/16.7.0 botocore/1.7.25",
  "requestParameters": {
    "computeResources": {
      "subnets": [
        "subnet-5eda8e04"
      ],
      "tags": {
        "testBatchTags": "CLI testing CE"
      },
      "desiredvCpus": 0,
      "minvCpus": 0,
      "instanceTypes": [
        "optimal"
      ],
      "securityGroupIds": [
        "sg-aba9e8db"
      ],
      "instanceRole": "ecsInstanceRole",
      "maxvCpus": 128,
      "type": "EC2"
    },
    "state": "ENABLED",
    "type": "MANAGED",
    "computeEnvironmentName": "Test"
  },
  "responseElements": {
    "computeEnvironmentName": "Test",
    "computeEnvironmentArn": "arn:aws:batch:us-east-1:012345678910:compute-environment/
Test"
  },
  "requestID": "890b8639-e51f-11e7-b038-EXAMPLE",
  "eventID": "874f89fa-70fc-4798-bc00-EXAMPLE",
  "readOnly": false,
  "eventType": "AwsApiCall",
  "recipientAccountId": "012345678910"
}
```

IAM AWS Batch 故障診斷

使用以下資訊來協助您診斷和修正使用 AWS Batch 和 IAM 時可能遇到的常見問題。

主題

- 我未獲授權，不得在 AWS Batch中執行動作
- 我未獲得執行 iam:PassRole 的授權
- 我想要允許 AWS 帳戶外的人員存取我的 AWS Batch 資源

我未獲授權，不得在 AWS Batch中執行動作

如果 AWS 管理主控台 告知您無權執行 動作，則必須聯絡您的管理員尋求協助。您的管理員是為您提供使用者名稱和密碼的人員。

下列範例錯誤會在 `mateojackson` 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 `batch:`*`GetWidget`* 許可時發生。

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
 batch:GetWidget on resource: my-example-widget
```

在此情況下，Mateo 會請求管理員更新他的政策，允許他使用 *`my-example-widget`* 動作存取 `batch:`*`GetWidget`* 資源。如需授予許可以傳遞角色的詳細資訊，請參閱授予使用者許可以傳遞角色至 AWS 服務。

我未獲得執行 iam:PassRole 的授權

如果您收到錯誤，告知您未獲授權執行 `iam:PassRole` 動作，您的政策必須更新，允許您將角色傳遞給 AWS Batch。

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。

名為 `marymajor` 的 IAM 使用者嘗試使用主控台在 AWS Batch中執行動作時，發生下列範例錯誤。但是，動作要求服務具備服務角色授予的許可。Mary 沒有將角色傳遞給服務的許可。

```
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 AWS 帳戶外的人員存取我的 AWS Batch 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 是否 AWS Batch 支援這些功能，請參閱 AWS Batch 如何使用 IAM。
- 若要了解如何提供您擁有 AWS 帳戶 的資源存取權，請參閱《IAM 使用者指南》中的在您 AWS 帳戶 擁有的另一個 中提供存取權給 IAM 使用者。
- 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。

AWS 步驟函數

您可以使用 AWS Batch 主控台來檢視 Step Functions 狀態機器的詳細資訊，以及它們所使用的函數。

章節

- 教學課程：檢視狀態機器詳細資訊
- 教學課程：編輯狀態機器
- 教學課程：執行狀態機器

教學課程：檢視狀態機器詳細資訊

AWS Batch 主控台會顯示目前 中狀態機器的清單 AWS 區域 ，其中包含至少一個提交 AWS Batch 任務的工作流程步驟。

選擇狀態機器以檢視代表工作流程的圖形。以藍色反白顯示的步驟代表 AWS Batch 任務。使用圖表控制項目來放大、縮小和置中圖表。

> **Note**
>
> 在狀態機器定義中使用 JsonPath 動態參考 AWS Batch 任務時，無法在 AWS Batch 主控台中顯示函數詳細資訊。相反地，任務名稱會列為動態參考，且圖形中的對應步驟會變成灰色。

檢視狀態機器詳細資料

1. 開啟 Step Functions 支援的 AWS Batch 主控台工作流程協調。 https://console.aws.amazon.com/batch/home#stepfunctions
2. 選擇狀態機器。
   <result>

   AWS Batch 主控台會開啟詳細資訊頁面。
   </result>

如需詳細資訊，請參閱 AWS Step Functions 開發人員指南中的 Step Functions。

教學課程：編輯狀態機器

當您想要編輯狀態機器時，請 AWS Batch 開啟 Step Functions 主控台的編輯定義頁面。

編輯狀態機器

1. 開啟 Step Functions 支援的 AWS Batch 主控台工作流程協調。 https://console.aws.amazon.com/batch/home#stepfunctions
2. 選擇狀態機器。
3. 選擇編輯。

   Step Functions 主控台會開啟 Edit definition (編輯定義) 頁面。
4. 編輯狀態機器，然後選擇儲存。

如需有關編輯狀態機器的詳細資訊，請參閱 AWS Step Functions 開發人員指南中的 Step Functions 狀態機器語言。

教學課程：執行狀態機器

當您想要執行狀態機器時，請 AWS Batch 開啟 Step Functions 主控台的新執行頁面。

執行狀態機器

1. 開啟 Step Functions 支援的 AWS Batch 主控台工作流程協調。 https://console.aws.amazon.com/batch/home#stepfunctions
2. 選擇狀態機器。
3. 選擇 Execute (執行)。

   Step Functions 主控台會開啟 New execution (新執行) 頁面。
4. (選用) 編輯狀態機器並選擇開始執行。

如需執行狀態機器的詳細資訊，請參閱 AWS Step Functions 開發人員指南中的 Step Functions 狀態機器執行概念。

AWS Batch Amazon EventBridge 的事件串流

您可以使用 Amazon EventBridge AWS Batch 的事件串流，接收有關任務佇列中任務目前狀態的近乎即時通知。

您可以使用 EventBridge 來深入了解您的 AWS Batch 服務。更具體地說，您可以使用它來檢查任務進度、建置 AWS Batch 自訂工作流程、產生用量報告或指標，或建置您自己的儀表板。使用 AWS Batch 和 EventBridge，您不需要排程和監控持續輪詢任務狀態變更 AWS Batch 的程式碼。反之，您可以使用各種 Amazon EventBridge 目標以非同步方式處理 AWS Batch 任務狀態變更。其中包括 AWS Lambda Amazon Simple Queue Service、Amazon Simple Notification Service 或 Amazon Kinesis Data Streams。

確保來自事件串流 AWS Batch 的事件至少交付一次。如果傳送重複的事件，則事件會提供足夠的資訊來識別重複項目。如此一來，您就可以比較事件的時間戳記和任務狀態。

AWS Batch 任務可作為 EventBridge 目標。使用簡單的規則，您可以比對事件並提交 AWS Batch 任務以回應它們。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的什麼是 EventBridge？。 EventBridge 您也可以使用 EventBridge，使用 cron或 速率表達式來排程在特定時間自動觸發的自動化動作。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的建立排程執行的 Amazon EventBridge 規則。如需範例演練，請參閱 AWS Batch 任務做為 EventBridge 目標。如需有關使用 EventBridge 排程器的資訊，請參閱《Amazon EventBridge 使用者指南》中的設定 Amazon EventBridge 排程器。 EventBridge

主題

- AWS Batch 事件
- 教學課程：搭配 使用 AWS 使用者通知 AWS Batch
- AWS Batch 任務做為 EventBridge 目標
- 教學課程：使用 EventBridge 接聽 AWS Batch 任務事件
- 教學課程：針對失敗的任務事件傳送 Amazon Simple Notification Service 提醒

AWS Batch 事件

AWS Batch 會將任務狀態變更事件傳送至 EventBridge。 會 AWS Batch 追蹤任務的狀態。如果先前提交的任務狀態變更，則會叫用 事件。例如，如果 `RUNNING` 狀態中的任務移至 `FAILED` 狀態。這些事件便歸類為任務狀態變更事件。

> **Note**
>
> AWS Batch 未來可能會新增其他事件類型、來源和詳細資訊。如果您以程式設計方式還原序列化事件 JSON 資料，請確定您的應用程式已準備好處理未知屬性。這是為了避免新增這些額外屬性時發生問題。

任務狀態變更事件

每當現有（先前提交的）任務變更狀態時，就會建立事件。如需 AWS Batch 任務狀態的詳細資訊，請參閱 任務狀態。

> **Note**
>
> 事件不會為初始任務提交建立。

Example任務狀態變更事件

任務狀態變更事件會以下列格式交付。`detail` 區段類似於 API AWS Batch 參考中從 DescribeJobs API 操作傳回的 JobDetail 物件。如需有關 EventBridge 參數的詳細資訊，請參閱 Amazon EventBridge 使用者指南中的事件和事件模式。

```
{
    "version": "0",
    "id": "c8f9c4b5-76e5-d76a-f980-7011e206042b",
    "detail-type": "Batch Job State Change",
    "source": "aws.batch",
    "account": "123456789012",
    "time": "2022-01-11T23:36:40Z",
    "region": "us-east-1",
    "resources": [
        "arn:aws:batch:us-east-1:123456789012:job/4c7599ae-0a82-49aa-ba5a-4727fcce14a8"
    ],
    "detail": {
        "jobArn": "arn:aws:batch:us-east-1:123456789012:job/4c7599ae-0a82-49aa-
ba5a-4727fcce14a8",
        "jobName": "event-test",
        "jobId": "4c7599ae-0a82-49aa-ba5a-4727fcce14a8",
        "jobQueue": "arn:aws:batch:us-east-1:123456789012:job-queue/
PexjEHappyPathCanary2JobQueue",
```

```
        "status": "RUNNABLE",
        "attempts": [],
        "createdAt": 1641944200058,
        "retryStrategy": {
            "attempts": 2,
            "evaluateOnExit": []
        },
        "dependsOn": [],
        "jobDefinition": "arn:aws:batch:us-east-1:123456789012:job-definition/first-
run-job-definition:1",
        "parameters": {},
        "container": {
            "image": "137112412989.dkr.ecr.us-east-1.amazonaws.com/amazonlinux:latest",
            "command": [
                "sleep",
                "600"
            ],
            "volumes": [],
            "environment": [],
            "mountPoints": [],
            "ulimits": [],
            "networkInterfaces": [],
            "resourceRequirements": [
                {
                    "value": "2",
                    "type": "VCPU"
                }, {
                    "value": "256",
                    "type": "MEMORY"
                }
            ],
            "secrets": []
        },
        "propagateTags": false,
        "platformCapabilities": []
    }
}
```

任務佇列封鎖事件

每當 AWS Batch 偵測到任務處於 RUNNABLE 狀態，因而封鎖佇列時，就會在 Amazon CloudWatch Events 中建立事件。如需支援的封鎖佇列原因的詳細資訊，請參閱 任務停滯在 RUNNABLE 狀態。API DescribeJobs動作中的 statusReason 欄位也提供相同的原因。

Example任務佇列封鎖事件

任務佇列封鎖事件會以下列格式交付。`detail` 區段類似 AWS Batch API 參考中 DescribeJobs API 操作傳回的 JobDetail 物件。如需有關 EventBridge 參數的詳細資訊，請參閱 Amazon EventBridge 使用者指南中的事件和事件模式。

```
{
    "version": "0",
    "id": "c8f9c4b5-76e5-d76a-f980-7011e206042b",
    "detail-type": "Batch Job Queue Blocked",
    "source": "aws.batch",
    "account": "123456789012",
    "time": "2022-01-11T23:36:40Z",
    "region": "us-east-1",
    "resources": [
        "arn:aws:batch:us-east-1:123456789012:job/4c7599ae-0a82-49aa-
ba5a-4727fcce14a8",
        "arn:aws:batch:us-east-1:123456789012:job-queue/PexjEHappyPathCanary2JobQueue"
    ],
    "detail": {
        "jobArn": "arn:aws:batch:us-east-1:123456789012:job/4c7599ae-0a82-49aa-
ba5a-4727fcce14a8",
        "jobName": "event-test",
        "jobId": "4c7599ae-0a82-49aa-ba5a-4727fcce14a8",
        "jobQueue": "arn:aws:batch:us-east-1:123456789012:job-queue/
PexjEHappyPathCanary2JobQueue",
        "status": "RUNNABLE",
        "statusReason": "blocked-reason",
        "attempts": [],
        "createdAt": 1641944200058,
        "retryStrategy": {
            "attempts": 2,
            "evaluateOnExit": []
        },
        "dependsOn": [],
        "jobDefinition": "arn:aws:batch:us-east-1:123456789012:job-definition/first-
run-job-definition:1",
        "parameters": {},
        "container": {
            "image": "137112412989.dkr.ecr.us-east-1.amazonaws.com/amazonlinux:latest",
            "command": [
                "sleep",
                "600"
```

```
            ],
            "volumes": [],
            "environment": [],
            "mountPoints": [],
            "ulimits": [],
            "networkInterfaces": [],
            "resourceRequirements": [
                {
                    "value": "2",
                    "type": "VCPU"
                }, {
                    "value": "256",
                    "type": "MEMORY"
                }
            ],
            "secrets": []
        },
        "propagateTags": false,
        "platformCapabilities": []
    }
}
```

服務任務狀態變更事件

只要現有服務任務變更狀態，就會建立事件。如需服務任務狀態的詳細資訊，請參閱 將 AWS Batch 服務任務狀態映射至 SageMaker AI 狀態。

> **Note**
>
> 事件不會為初始任務提交建立。

Example服務任務狀態變更事件

服務任務狀態變更事件會以下列格式交付。`detail` 區段類似 AWS Batch API 參考中 DescribeServiceJob API 操作傳回的回應。如需有關 EventBridge 參數的詳細資訊，請參閱 Amazon EventBridge 使用者指南中的事件和事件模式。

> **Note**
>
> 事件 中不包含 `tags`和 `serviceRequestPayload` 欄位`detail`。

```
{
  "version": "0",
  "id": "c8f9c4b5-76e5-d76a-f980-7011e206042b",
  "detail-type": "Batch Service Job State Change",
  "source": "aws.batch",
  "account": "123456789012",
  "time": "2022-01-11T23:36:40Z",
  "region": "us-east-1",
  "resources": [
    "arn:aws:batch:us-east-1:123456789012:service-job/4c7599ae-0a82-49aa-
ba5a-4727fcce14a8"
  ],
  "detail": {
    "attempts": [
      {
        "serviceResourceId": {
          "name": "TrainingJobArn",
          "value": "arn:aws:sagemaker:us-east-1:123456789012:training-job/AWSBatchmy-
training-job88b610a69aa8380ca5b0a7aba3f81cb8"
        },
        "startedAt": 1641944300058,
        "stoppedAt": 1641944400058,
        "statusReason": "Received status from SageMaker: Training job completed"
      }
    ],
    "createdAt": 1641944200058,
    "jobArn": "arn:aws:batch:us-east-1:123456789012:service-job/4c7599ae-0a82-49aa-
ba5a-4727fcce14a8",
    "jobId": "0bb17543-ece6-4480-b1a7-a556d344746b",
    "jobName": "event-test",
    "jobQueue": "arn:aws:batch:us-east-1:123456789012:job-queue/HappyPathJobQueue",
    "latestAttempt": {
      "serviceResourceId": {
        "name": "TrainingJobArn",
        "value": "arn:aws:sagemaker:us-east-1:123456789012:training-job/AWSBatchmy-
training-job88b610a69aa8380ca5b0a7aba3f81cb8"
      }
    },
    "serviceJobType": "SAGEMAKER_TRAINING",
    "startedAt": 1641944300058,
    "status": "SUCCEEDED",
    "statusReason": "Received status from SageMaker: Training job completed",
    "stoppedAt": 1641944400058,
```

```
    "timeoutConfig": {
      "attemptDurationSeconds": 60
    }
  }
}
```

服務任務佇列封鎖事件

每當 AWS Batch 偵測到封鎖佇列時，都會在 Amazon CloudWatch Events 中建立事件。封鎖佇列的原因可在 DescribeServiceJob API 動作的 statusReason 欄位中取得。

Example服務任務佇列封鎖事件

服務任務佇列封鎖事件會以下列格式交付。detail 區段類似 AWS Batch API 參考中 DescribeServiceJob API 操作傳回的回應。如需有關 EventBridge 參數的詳細資訊，請參閱 Amazon EventBridge 使用者指南中的事件和事件模式。

> **Note**
>
> 事件 中不包含 tags和 serviceRequestPayload 欄位detail。

```
{
  "version": "0",
  "id": "c8f9c4b5-76e5-d76a-f980-7011e206042b",
  "detail-type": "Batch Service Job Queue Blocked",
  "source": "aws.batch",
  "account": "123456789012",
  "time": "2022-01-11T23:36:40Z",
  "region": "us-east-1",
  "resources": [
    "arn:aws:batch:us-east-1:123456789012:service-job/4c7599ae-0a82-49aa-
ba5a-4727fcce14a8"
  ],
  "detail": {
    "attempts": [],
    "createdAt": 1641944200058,
    "jobArn": "arn:aws:batch:us-east-1:123456789012:service-job/4c7599ae-0a82-49aa-
ba5a-4727fcce14a8",
    "jobId": "6271dfdf-d8a7-41b1-a4d2-55a2224f5375",
    "jobName": "event-test",
    "jobQueue": "arn:aws:batch:us-east-1:123456789012:job-queue/HappyPathJobQueue",
```

```
    "serviceJobType": "SAGEMAKER_TRAINING",
    "status": "RUNNABLE",
    "statusReason": "blocked-reason",
    "timeoutConfig": {
      "attemptDurationSeconds": 60
    }
  }
}
```

教學課程：搭配 使用 AWS 使用者通知 AWS Batch

您可以使用AWS 使用者通知來設定交付管道，以取得 AWS Batch 事件的通知。當事件符合您指定的規則時，便會收到通知。您可以透過多個管道接收事件的通知，包括電子郵件、聊天應用程式中的 Amazon Q Developer 聊天通知或AWS Console Mobile Application推送通知。您也可以在主控台通知中心查看通知。使用者通知支援彙總，可減少您在特定事件期間收到的通知數目。

若要在 中設定使用者通知 AWS Batch：

1. 開啟 AWS Batch 主控台。
2. 選擇 Dashboard (儀表板)。
3. 選擇設定通知。
4. In AWS User Notifications，選擇建立通知組態。

如需如何設定和檢視使用者通知的詳細資訊，請參閱AWS 使用者通知入門。

AWS Batch 任務做為 EventBridge 目標

Amazon EventBridge 提供近乎即時的系統事件串流，描述 Amazon Web Services 資源中的變更。一般而言， AWS Batch 在 Amazon Elastic Container Service、Amazon Elastic Kubernetes Service 和 AWS Fargate 任務上可作為 EventBridge 目標。使用簡單的規則，您可以比對事件並提交 AWS Batch 任務以回應它們。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的什麼是 EventBridge？。 EventBridge

您也可以使用 EventBridge 來排程在特定時間使用 cron或 速率表達式調用的自動化動作。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的建立排程執行的 Amazon EventBridge 規則。

如需有關如何在事件符合事件模式時建立執行的規則的資訊，請參閱《Amazon EventBridge 使用者指南》中的建立對事件做出反應的 Amazon EventBridge 規則。 EventBridge

EventBridge 目標 AWS Batch 任務的常見使用案例包括下列使用案例：

- 排程任務會定期進行。例如，當 Amazon EC2 Spot 執行個體較便宜時，cron任務只會在低使用時間發生。
- AWS Batch 任務會執行以回應記錄在 CloudTrail 中的 API 操作。例如，每當物件上傳到指定的 Amazon S3 儲存貯體時，就會提交任務。每次發生這種情況時，EventBridge 輸入轉換器都會將物件的儲存貯體和金鑰名稱傳遞給 AWS Batch 參數。

> **Note**
>
> 在此案例中，所有相關 AWS 資源都必須位於相同的 區域。這包括 Amazon S3 儲存貯體、EventBridge 規則和 CloudTrail 日誌等資源。

在使用 EventBridge 規則和目標提交 AWS Batch 任務之前，EventBridge 服務需要多個執行 AWS Batch 任務的許可。當您在 EventBridge 主控台中建立將 AWS Batch 任務指定為目標的規則時，您也可以建立此角色。如需有關此角色必要的服務主體和 IAM 權限的詳細資訊，請參閱 EventBridge IAM 角色。

主題

- 教學課程：建立排程 AWS Batch 任務
- 教學課程：建立具有事件模式的規則
- 教學課程：使用 EventBridge 輸入轉換器，依排程將事件資訊傳遞至 AWS Batch 目標

教學課程：建立排程 AWS Batch 任務

下列程序說明如何建立排程 AWS Batch 任務和所需的 EventBridge IAM 角色。

使用 EventBridge 建立排程 AWS Batch 任務

> **Note**
>
> 此程序適用於所有 AWS Batch Amazon ECS、Amazon EKS 和 AWS Fargate 任務。

1. 前往 https://console.aws.amazon.com/events/ 開啟 Amazon EventBridge 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。

3. 在導覽窗格中，選擇規則。
4. 選擇建立規則。
5. 針對名稱，指定運算環境的唯一名稱。名稱最多可包含 64 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。

   > **Note**
   >
   > 在同一個區域和同一個事件匯流排上，規則不能與另一個規則同名。

6. （選用）針對描述，輸入規則的描述。
7. 針對事件匯流排，選擇要與此規則建立關聯的事件匯流排。如果您想要此規則匹配來自您的帳戶的事件，請選取預設值。當您帳戶中 AWS 服務 的 發出事件時，一律會前往您帳戶的預設事件匯流排。
8. （選用）如果您不想立即執行規則，請關閉所選匯流排上的規則。
9. 針對規則類型，選擇排程。
10. 選擇繼續以建立規則或下一步。
11. 針對 Schedule pattern (排程模式)，執行下列其中一項動作：

    - 選擇在特定時間執行的精細排程，例如上午 8:00 每月第一個星期一的 PST，然後輸入 Cron 表達式。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的 Cron Expressions。
    - 選擇以一般速率執行的排程，例如每 10 分鐘。 然後輸入速率表達式。

12. 選擇下一步。
13. 對於 Target types (目標類型)，選擇 AWS 服務。
14. 針對選取目標，選擇批次任務佇列。然後，設定下列項目：

    - Job queue (任務佇列)：輸入任務佇列的 Amazon Resource Name (ARN) 以排程任務。
    - Job definition (任務定義)：輸入用於任務之任務定義的名稱，及其修訂版或完整 ARN。
    - Job name (任務名稱)：輸入任務的名稱。
    - Array size (陣列大小)：(選擇性) 輸入任務要執行多個副本的陣列大小。如需詳細資訊，請參閱陣列任務。
    - Job attempts (任務嘗試)：(選擇性) 輸入任務失敗時的重試次數。如需詳細資訊，請參閱自動化任務重試。

15. 對於 Batch job queue (批次任務佇列) 目標類型而言，EventBridge 需要許可才能將事件傳送到目標。EventBridge 可建立執行您的規則所需的 IAM 角色。執行以下任意一項：

- 若要自動建立 IAM 角色，請選擇為此特定資源建立新角色。
- 若要使用您已建立的 IAM 角色，請選擇使用現有角色。

16. (選用) 展開 Additional settings (其他設定)。
    a. 針對設定目標輸入，選擇在傳遞至目標之前如何處理來自事件的文字。
    b. 針對事件的最長存留期，請指定未處理事件保留多久的時間間隔。
    c. 針對重試嘗試，輸入事件重試的次數。
    d. 針對無效字母佇列，選擇未處理事件處理方式的選項。如有必要，請指定要用作無效字母佇列的 Amazon SQS 佇列。
17. (選用) 選擇新增其他目標，為此規則新增另一個目標。
18. 選擇下一步。
19. （選用）針對標籤，選擇新增標籤以新增規則的資源標籤。如需詳細資訊，請參閱《Amazon EventBridge 標籤》。
20. 選擇下一步。
21. 對於檢閱和建立，請檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成時，請選擇 Create rule (建立規則)。

如需建立規則的詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的建立排程執行的 Amazon EventBridge 規則。

教學課程：建立具有事件模式的規則

下列程序說明如何建立具有事件模式的規則。

建立規則，當事件符合定義的模式時，將事件傳送至目標

> **Note**
> 此程序適用於所有 AWS Batch Amazon ECS、Amazon EKS 和 AWS Fargate 任務。

1. 前往 https://console.aws.amazon.com/events/ 開啟 Amazon EventBridge 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇規則。
4. 選擇建立規則。

5. 針對名稱，指定運算環境的唯一名稱。名稱最多可包含 64 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。

   > **Note**
   >
   > 在同一個區域和同一個事件匯流排上，規則不能與另一個規則同名。

6. （選用）針對描述，輸入規則的描述。
7. 針對事件匯流排，選擇要與此規則建立關聯的事件匯流排。如果您想要此規則匹配來自您的帳戶的事件，請選取預設值。當您帳戶中 AWS 服務 的 發出事件時，一律會前往您帳戶的預設事件匯流排。
8. （選用）如果您不想立即執行規則，請關閉所選匯流排上的規則。
9. 針對規則類型，選擇具有事件模式的規則。
10. 選擇下一步。
11. 針對事件來源，選擇AWS 事件或 EventBridge 合作夥伴事件。
12. （選用）對於範例事件：

    a. 針對範例事件類型，選擇AWS 事件。

    b. 針對範例事件，選擇批次任務狀態變更。

13. 針對建立方法，選取使用模式表單。
14. 對於事件模式：

    a. 在 Event source (事件來源)，選擇 AWS 服務。

    b. 針對 AWS 服務，選擇批次。

    c. 針對事件類型，選擇批次任務狀態變更。

15. 選擇下一步。
16. 對於 Target types (目標類型)，選擇 AWS 服務。
17. 針對選取目標，選擇目標類型。例如，選擇批次任務佇列。然後指定下列項目：

    - Job queue (任務佇列)：輸入任務佇列的 Amazon Resource Name (ARN) 以排程任務。
    - Job definition (任務定義)：輸入用於任務之任務定義的名稱，及其修訂版或完整 ARN。
    - Job name (任務名稱)：輸入任務的名稱。
    - Array size (陣列大小)：(選擇性) 輸入任務要執行多個副本的陣列大小。如需詳細資訊，請參閱陣列任務。

- Job attempts (任務嘗試)：(選擇性) 輸入任務失敗時的重試次數。如需詳細資訊，請參閱自動化任務重試。

18. 對於 Batch job queue (批次任務佇列) 目標類型而言，EventBridge 需要許可才能將事件傳送到目標。EventBridge 可建立執行您的規則所需的 IAM 角色。執行以下任意一項：

    - 若要自動建立 IAM 角色，請選擇為此特定資源建立新角色。
    - 若要使用您之前建立的 IAM 角色，請選擇 Use existing role (使用現有角色)。

19. (選用) 展開 Additional settings (其他設定)。

    a. 針對設定目標輸入，選擇如何處理來自事件的文字。

    b. 針對事件的最長存留期，請指定未處理事件保留多久的時間間隔。

    c. 針對重試嘗試，輸入事件重試的次數。

    d. 針對無效字母佇列，選擇未處理事件處理方式的選項。如有必要，請指定要用作無效字母佇列的 Amazon SQS 佇列。

20. （選用）選擇新增另一個目標以新增其他目標。

21. 選擇下一步。

22. （選用）針對標籤，選擇新增標籤以新增資源標籤。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的 Amazon EventBridge 標籤。

23. 選擇下一步。

24. 對於檢閱和建立，請檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立規則。

    如需建立規則的詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的建立對事件做出反應的 Amazon EventBridge 規則。

教學課程：使用 EventBridge 輸入轉換器，依排程將事件資訊傳遞至 AWS Batch 目標

您可以使用 EventBridge 輸入轉換器，在任務提交 AWS Batch 中將事件資訊傳遞至 。如果您因為其他 AWS 事件資訊而叫用任務，這會特別重要。其中一個範例是物件上傳至 Amazon S3 儲存貯體。您也可以在容器命令中使用任務定義與參數替代值。EventBridge 輸入轉換器可以根據事件資料提供參數值。

然後，您建立 AWS Batch 事件目標，從啟動該目標的事件剖析資訊，並將其轉換為`parameters`物件。當任務執行時，來自觸發事件的參數會傳遞至任務容器的 命令。

> **Note**
>
> 在此案例中，所有 AWS 資源 （例如 Amazon S3 儲存貯體、EventBridge 規則和 CloudTrail 日誌） 都必須位於相同的區域。

建立使用輸入轉換器 AWS Batch 的目標

1. 前往 https://console.aws.amazon.com/events/ 開啟 Amazon EventBridge 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇規則。
4. 選擇建立規則。
5. 針對名稱，指定運算環境的唯一名稱。名稱最多可包含 64 個字元。可以包含大小寫字母、數字、連字號 (-) 和底線 (_)。

   > **Note**
   >
   > 規則不能與相同 AWS 區域 和相同事件匯流排上的另一個規則具有相同的名稱。

6. （選用） 針對描述，輸入規則的描述。
7. 針對事件匯流排，選擇要與此規則建立關聯的事件匯流排。如果您想要此規則匹配來自您的帳戶的事件，請選取預設值。當您帳戶中 AWS 服務 的 發出事件時，一律會前往您帳戶的預設事件匯流排。
8. （選用） 如果您不想立即執行規則，請關閉所選匯流排上的規則。
9. 針對規則類型，選擇排程。
10. 選擇繼續以建立規則或下一步。
11. 針對 Schedule pattern (排程模式)，執行下列其中一項動作：

    - 選擇在特定時間執行的精細排程，例如上午 8:00 每月第一個星期一的 PST，然後輸入 Cron 表達式。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的 Cron Expressions。
    - 選擇以一般速率執行的排程，例如每 10 分鐘。然後輸入速率表達式。

12. 選擇下一步。
13. 對於 Target types (目標類型)，選擇 AWS 服務。
14. 針對選取目標，選擇批次任務佇列。然後，設定下列項目：

- Job queue (任務佇列)：輸入任務佇列的 Amazon Resource Name (ARN) 以排程任務。
- Job definition (任務定義)：輸入用於任務之任務定義的名稱，及其修訂版或完整 ARN。
- Job name (任務名稱)：輸入任務的名稱。
- Array size (陣列大小)：(選擇性) 輸入任務要執行多個副本的陣列大小。如需詳細資訊，請參閱陣列任務。
- Job attempts (任務嘗試)：(選擇性) 輸入任務失敗時的重試次數。如需詳細資訊，請參閱自動化任務重試。

15. 對於 Batch job queue (批次任務佇列) 目標類型而言，EventBridge 需要許可才能將事件傳送到目標。EventBridge 可建立執行您的規則所需的 IAM 角色。執行以下任意一項：

    - 若要自動建立 IAM 角色，請選擇為此特定資源建立新角色。
    - 若要使用您已建立的 IAM 角色，請選擇使用現有角色。

16. (選用) 展開 Additional settings (其他設定)。
17. 在 Additional settings (其他設定) 區段中，針對 Configure target input (設定目標輸入)，選擇 Input Transformer (輸入轉換器)。
18. 選擇設定輸入轉換器。
19. （選用）對於範例事件：

    a. 針對範例事件類型，選擇AWS 事件。

    b. 針對範例事件，選擇批次任務狀態變更。

20. 在 Target input transformer (目標輸入轉換器) 區段中，針對 Input path (輸入路徑)，指定透過觸發事件剖析的值。例如，若要剖析批次任務狀態變更事件，請使用下列 JSON 格式。

```
{
    "instance": "$.detail.jobId",
    "state": "$.detail.status"
}
```

21. 在範本中，輸入以下內容。

```
{
    "instance": <jobId> ,
    "status": <status>
}
```

22. 選擇確認。

23. 針對事件的最長存留期，請指定未處理事件保留多久的時間間隔。
24. 針對重試嘗試，輸入事件重試的次數。
25. 針對無效字母佇列，選擇未處理事件處理方式的選項。如有必要，請指定要用作無效字母佇列的 Amazon SQS 佇列。
26. （選用）選擇新增另一個目標以新增其他目標。
27. 選擇下一步。
28. （選用）針對標籤，選擇新增標籤以新增資源標籤。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的 Amazon EventBridge 標籤。
29. 選擇下一步。
30. 對於檢閱和建立，請檢閱組態步驟。如需變更，請選擇 Edit (編輯)。完成後，請選擇建立規則。

教學課程：使用 EventBridge 接聽 AWS Batch 任務事件

在本教學課程中，您會設定簡單的 AWS Lambda 函數來接聽 AWS Batch 任務事件，並將其寫入 CloudWatch Logs 日誌串流。

先決條件

此教學課程假設您有一個運作中的運算環境，和已準備好要接受任務的任務佇列。如果您沒有執行中的運算環境和任務佇列可從中擷取事件，請依照中的步驟AWS Batch 教學課程入門建立事件。在本教學課程結束時，您可以選擇將任務提交至此任務佇列，以測試您已正確設定 Lambda 函數。

主題

- 教學課程：建立 Lambda 函數
- 教學課程：註冊事件規則
- 教學課程：測試您的組態

教學課程：建立 Lambda 函數

在此程序中，您會建立簡單的 Lambda 函數，做為 AWS Batch 事件串流訊息的目標。

若要建立目標 Lambda 函數

1. 在 https：//https://console.aws.amazon.com/lambda/ 開啟 AWS Lambda 主控台。
2. 依序選擇 Create function (建立函數)、Author from scratch (從頭開始撰寫)。

3. 針對函數名稱，輸入 batch-event-stream-handler。
4. 針對執行階段，選擇 Python 3.8。
5. 選擇 Create function (建立函數)。
6. 在程式碼來源區段中，編輯範例程式碼以符合下列範例：

```
import json


def lambda_handler(event, _context):
    # _context is not used
    del _context
    if event["source"] != "aws.batch":
        raise ValueError("Function only supports input from events with a source
 type of: aws.batch")

    print(json.dumps(event))
```

這是簡單的 Python 3.8 函數，可列印 傳送的事件。 AWS Batch如果一切設定正確，則在此教學課程的最後，事件詳細資訊出現在與此 Lambda 函數建立關聯的 CloudWatch Logs 日誌串流中。

7. 選擇部署。

教學課程：註冊事件規則

在本節中，您會建立 EventBridge 事件規則，以擷取來自 AWS Batch 資源的任務事件。此規則會擷取來自定義帳戶 AWS Batch 內的所有事件。任務訊息本身包含事件來源的相關資訊，包括提交該來源的任務佇列。您可以使用此資訊以程式設計方式篩選和排序事件。

> **Note**
>
> 如果您使用 AWS 管理主控台 建立事件規則，主控台會自動新增 EventBridge 的 IAM 許可，以呼叫 Lambda 函數。不過，如果您使用 建立事件規則 AWS CLI，則必須明確授予許可。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的事件和事件模式。

建立 EventBridge 規則

1. 前往 https://console.aws.amazon.com/events/ 開啟 Amazon EventBridge 主控台。
2. 在導覽窗格中，選擇規則。

3. 選擇建立規則。
4. 輸入規則的名稱和描述。

   在同一個區域和同一個事件匯流排上，規則不能與另一個規則同名。
5. 針對事件匯流排，選擇要與此規則建立關聯的事件匯流排。如果您想要此規則匹配來自您的帳戶的事件，請選取 AWS 預設事件匯流排。當您帳戶中的 AWS 服務發出事件時，一律會前往您帳戶的預設事件匯流排。
6. 針對規則類型，選擇具有事件模式的規則。
7. 選擇下一步。
8. 在事件來源中，選擇其他。
9. 針對事件模式，選取自訂模式 (JSON 編輯器）。
10. 將下列的事件模式貼到文字區域。

```
{
  "source": [
    "aws.batch"
  ]
}
```

   此規則適用於所有 AWS Batch 群組和每個 AWS Batch 事件。或者，您可以建立一個更針對性的規則，來篩選掉一些結果。
11. 選擇下一步。
12. 在目標類型欄位中，選擇 AWS 服務。
13. 針對選取目標，選擇 Lambda 函數，然後選取您的 Lambda 函數。
14. (選用) 針對其他設定，請執行下列動作：

    a. 針對 Maximum age of event (事件的最長存留期)，輸入介於一分鐘 (00:01) 到 24 小時 (24:00) 之間的某個值。

    b. 針對重試嘗試，輸入介於 0 到 185 之間的某個數。

    c. 針對無效字母佇列，選擇是否使用標準 Amazon SQS 佇列做為無效字母佇列。若與此規則匹配的事件未成功傳送到目標，則 EventBridge 會將其傳送至無效字母佇列。執行以下任意一項：

       - 選擇無，即不使用無效字母佇列。
       - 選擇目前 AWS 帳戶中的選取 Amazon SQS 佇列以用作無效字母佇列，然後從下拉式清單中選取要使用的佇列。

- 選擇在其他 AWS 帳戶中選取 Amazon SQS 佇列做為無效字母佇列，然後輸入要使用的佇列 ARN。您必須將以資源為基礎政策連接到佇列，而且該佇列授與 EventBridge 向其傳送簡訊的許可。如需詳細資訊，請參閱 Amazon EventBridge 使用者指南中的授與無效字母佇列的許可。

15. 選擇下一步。
16. (選用) 為規則輸入一或多個標籤。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的 Amazon EventBridge 標籤。
17. 選擇下一步。
18. 檢閱規則的詳細資訊，然後選擇建立規則。

教學課程：測試您的組態

您現在可以透過將任務提交至任務佇列來測試 EventBridge 組態。如果一切設定正確，則會觸發 Lambda 函數，並將事件資料寫入函數的 CloudWatch Logs 日誌串流。

若要測試組態

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 提交新 AWS Batch 任務。如需詳細資訊，請參閱教學課程：提交任務。
3. 透過 https://console.aws.amazon.com/cloudwatch/ 開啟 CloudWatch 主控台。
4. 在導覽窗格上，選擇 Logs (日誌)，然後選取 Lambda 函數的日誌群組 (例如，/aws/lambda/*my-function*)。
5. 選取日誌串流，以檢視事件資料。

教學課程：針對失敗的任務事件傳送 Amazon Simple Notification Service 提醒

在本教學課程中，您會設定 Amazon EventBridge 事件規則，該規則只會擷取任務已移至 `FAILED` 狀態的任務事件。在本教學課程結束時，您也可以選擇性地將任務提交至此任務佇列。這是為了測試您已正確設定 Amazon SNS 提醒。

先決條件

此教學課程假設您有一個運作中的運算環境，和已準備好要接受任務的任務佇列。如果您沒有執行中的運算環境和任務佇列可從中擷取事件，請依照中的步驟AWS Batch 教學課程入門建立事件。

主題

- 教學課程：建立和訂閱 Amazon SNS 主題
- 教學課程：註冊事件規則
- 教學課程：測試您的規則
- 替代規則：批次任務佇列已封鎖

教學課程：建立和訂閱 Amazon SNS 主題

在此教學課程中，您會設定 Amazon SNS 主題，做為新事件規則的事件目標。

建立 Amazon SNS 主題

1. 在 https://console.aws.amazon.com/sns/v3/home 開啟 Amazon SNS 主控台。
2. 選擇 Topics (主題)、Create topic (建立主題)。
3. 針對類型，選擇標準。
4. 針對名稱，輸入 **`JobFailedAlert`** 並選擇建立主題。
5. 在 JobFailedAlert 畫面上，選擇建立訂閱。
6. 對於通訊協定，選擇電子郵件。
7. 對於 Endpoint (端點)，輸入您目前能存取的電子郵件地址，並選擇 Create subscription (建立訂閱)。
8. 檢查您的電子郵件帳戶，並等待收到訂閱確認的電子郵件訊息。您收到訊息時，請選擇 Confirm subscription (確認訂閱)。

教學課程：註冊事件規則

接著，註冊事件規則，使其只擷取任務失敗的事件。

註冊您的 EventBridge 規則

1. 前往 https://console.aws.amazon.com/events/ 開啟 Amazon EventBridge 主控台。
2. 在導覽窗格中，選擇規則。
3. 選擇建立規則。
4. 輸入規則的名稱和描述。

   在同一個區域和同一個事件匯流排上，規則不能與另一個規則同名。

5. 針對事件匯流排，選擇要與此規則建立關聯的事件匯流排。如果您想要此規則匹配來自您的帳戶的事件，請選取 AWS 預設事件匯流排。當您帳戶中的 AWS 服務發出事件時，一律會前往您帳戶的預設事件匯流排。
6. 針對規則類型，選擇具有事件模式的規則。
7. 選擇下一步。
8. 在事件來源中，選擇其他。
9. 針對事件模式，選取自訂模式 (JSON 編輯器）。
10. 將下列的事件模式貼到文字區域。

```
{
  "detail-type": [
    "Batch Job State Change"
  ],
  "source": [
    "aws.batch"
  ],
  "detail": {
    "status": [
      "FAILED"
    ]
  }
}
```

此程式碼會定義符合任務狀態為 之任何事件的 EventBridge 規則FAILED。如需事件模式的詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的事件和事件模式。

11. 選擇下一步。
12. 在目標類型欄位中，選擇 AWS 服務。
13. 對於選取目標，選擇 SNS 主題，對於主題，選擇 JobFailedAlert。
14. (選用) 針對其他設定，請執行下列動作：

    a. 針對 Maximum age of event (事件的最長存留期)，輸入介於一分鐘 (00:01) 到 24 小時 (24:00) 之間的某個值。

    b. 針對重試嘗試，輸入介於 0 到 185 之間的某個數。

    c. 針對無效字母佇列，選擇是否使用標準 Amazon SQS 佇列做為無效字母佇列。若與此規則匹配的事件未成功傳送到目標，則 EventBridge 會將其傳送至無效字母佇列。執行以下任意一項：

- 選擇無，即不使用無效字母佇列。
- 選擇目前 AWS 帳戶中的選取 Amazon SQS 佇列以用作無效字母佇列，然後從下拉式清單中選取要使用的佇列。
- 選擇在其他 AWS 帳戶中選取 Amazon SQS 佇列做為無效字母佇列，然後輸入要使用的佇列 ARN。您必須將以資源為基礎政策連接到佇列，而且該佇列授與 EventBridge 向其傳送簡訊的許可。如需詳細資訊，請參閱 Amazon EventBridge 使用者指南中的授與無效字母佇列的許可。

15. 選擇下一步。
16. (選用) 為規則輸入一或多個標籤。如需詳細資訊，請參閱《Amazon EventBridge 使用者指南》中的 Amazon EventBridge 標籤。
17. 選擇下一步。
18. 檢閱規則的詳細資訊，然後選擇建立規則。

教學課程：測試您的規則

為了測試您的規則，提交在使用非零結束代碼啟動後不久結束的任務。如果您的事件規則設定正確，您應該會在幾分鐘內收到包含事件文字的電子郵件訊息。

測試規則

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 提交新 AWS Batch 任務。如需詳細資訊，請參閱教學課程：提交任務。針對任務的命令，將結束容器的此命令換成結束代碼 1。

```
/bin/sh, -c, 'exit 1'
```

3. 檢查您的電子郵件以確認您收到任務通知失敗的電子郵件提醒。

替代規則：批次任務佇列已封鎖

若要建立監控批次任務佇列已封鎖的事件規則，請使用下列變更重複這些教學課程：

1. 在 教學課程：建立和訂閱 Amazon SNS 主題 中，使用 *BlockedJobQueue* 做為主題名稱。
2. 在 教學課程：註冊事件規則 中，在 JSON 編輯器中使用下列模式：

```
{
```

```
    "detail-type": [
      "Batch Job Queue Blocked"
    ],
    "source": [
      "aws.batch"
    ]
}
```

Elastic Fabric Adapter

Elastic Fabric Adapter (EFA) 是一種用於加速高效能運算 (HPC) 應用程式的裝置。如果符合以下條件， AWS Batch 支援使用 EFA 的應用程式。

- 如需支援 EFAs執行個體類型清單，請參閱《Amazon EC2 使用者指南》中的支援的執行個體類型。

  > **Tip**
  >
  > 若要查看 中支援 EFAs執行個體類型清單 AWS 區域，請執行下列命令。然後，交叉參考 AWS Batch 主控台中與可用執行個體類型清單一起傳回的清單。
  >
  > ```
  > $ aws ec2 describe-instance-types  --region us-east-1  --filters Name=network-info.efa-supported,Values=true  --query "InstanceTypes[*].[InstanceType]"  --output text | sort
  > ```

- 如需支援 EFA 的作業系統清單，請參閱支援的作業系統。
- AMI 已載入 EFA 驅動程式。
- EFA 的安全群組必須允許往返於其本身的所有傳入和傳出流量。
- 使用 EFA 的所有執行個體都必須位於相同的叢集置放群組中。
- 任務定義必須包含 `hostPath` 設定為 `/dev/infiniband/uverbs0` 的 `devices` 成員，以允許 EFA 裝置傳遞到容器。如果指定 `containerPath` ，則它也必須設定為 `/dev/infiniband/uverbs0`。如果已設定 `permissions`，它必須設定為 `READ` | `WRITE` | `MKNOD`。

  LinuxParameters 成員的位置對於多節點平行任務和單節點容器任務不同。下列範例顯示差異，但缺少必要的值。

  Example多節點平行任務範例

  ```
  {
    "jobDefinitionName": "EFA-MNP-JobDef",
    "type": "multinode",
    "nodeProperties": {
      ...
      "nodeRangeProperties": [
        {
          ...
          "container": {
  ```

```
          ...
          "linuxParameters": {
            "devices": [
              {
                "hostPath": "/dev/infiniband/uverbs0",
                "containerPath": "/dev/infiniband/uverbs0",
                "permissions": [
                    "READ", "WRITE", "MKNOD"
                ]
              },
            ],
          },
        },
      },
    ],
  },
}
```

Example單一節點容器任務範例

```
{
  "jobDefinitionName": "EFA-Container-JobDef",
  "type": "container",
  ...
  "containerProperties": {
    ...
    "linuxParameters": {
      "devices": [
        {
          "hostPath": "/dev/infiniband/uverbs0",
        },
      ],
    },
  },
}
```

如需 EFA 的詳細資訊，請參閱《Amazon EC2 使用者指南》中的彈性布料轉接器。

監控 AWS Batch

監控是維護 和 AWS 解決方案的可靠性、可用性 AWS Batch 和效能的重要部分。

我們強烈建議您從 AWS 解決方案的所有部分收集監控資料，以便在發生多點故障時更輕鬆地偵錯。請在一開始先建立可回答下列問題的監視計劃。如果您不確定如何回答這些問題，仍然可以使用 Amazon CloudWatch Logs 來建立效能基準。

- 監控目標是什麼？
- 監控哪些資源？
- 監控這些資源的頻率為何？
- 將使用哪些監控工具？
- 誰將執行監控任務？
- 發生問題時應該通知誰？

您的下一個步驟是在不同的時間及不同的負載條件下測量 AWS Batch 效能，以建立環境中正常效能的基準。進行監控時 AWS Batch，請保留歷史監控資料，以便與目前的效能資料進行比較。這可協助您識別正常效能模式和效能異常情況，並策劃解決這些情況的方法。

本節中的主題可協助您開始記錄和監視 AWS Batch。

主題

- 搭配 使用 CloudWatch Logs AWS Batch
- AWS Batch CloudWatch Container Insights
- 使用 CloudWatch Logs 監控 AWS Batch Amazon EKS 任務

搭配 使用 CloudWatch Logs AWS Batch

您可以在 EC2 資源上設定 AWS Batch 任務，將詳細的日誌資訊和指標傳送至 CloudWatch Logs。這樣做，您可以在一個方便的位置檢視與任務不同的日誌。如需 CloudWatch Logs 的詳細資訊，請參閱《Amazon CloudWatch 使用者指南》中的什麼是 Amazon CloudWatch Logs？。 Amazon CloudWatch

> **Note**
>
> 根據預設， AWS Fargate 容器的 CloudWatch Logs 已開啟。

若要開啟和自訂 CloudWatch Logs 記錄，請檢閱下列一次性組態任務：

- 對於以 EC2 資源為基礎的 AWS Batch 運算環境，請將 IAM 政策新增至`ecsInstanceRole`角色。如需詳細資訊，請參閱the section called “教學課程：新增 CloudWatch Logs IAM 政策”。
- 建立包含詳細 CloudWatch 監控的 Amazon EC2 啟動範本，然後在建立 AWS Batch 運算環境時指定範本。您也可以在現有映像上安裝 CloudWatch 代理程式，然後在 AWS Batch 初次執行精靈中指定映像。
- （選用）設定 awslogs 驅動程式。您可以新增參數，以變更 EC2 和 Fargate 資源的預設行為。如需詳細資訊，請參閱the section called “使用 awslogs 日誌驅動程式”。

主題

- 教學課程：新增 CloudWatch Logs IAM 政策
- 安裝及設定 CloudWatch 代理程式
- 教學課程：檢視 CloudWatch Logs

教學課程：新增 CloudWatch Logs IAM 政策

您必須先建立使用 CloudWatch Logs APIs IAM 政策，才能將日誌資料和詳細指標傳送至 CloudWatch Logs。建立 IAM 政策後，將其連接到`ecsInstanceRole`角色。

> **Note**
>
> 如果`ECS-CloudWatchLogs`政策未連接至`ecsInstanceRole`角色，仍可將基本指標傳送至 CloudWatch Logs。不過，基本指標不包含日誌資料或詳細指標，例如可用磁碟空間。

AWS Batch 運算環境使用 Amazon EC2 資源。當您使用 AWS Batch 初次執行精靈建立運算環境時，會 AWS Batch 建立角色並使用該`ecsInstanceRole`角色設定環境。

如果您未使用初次執行精靈，您可以在 AWS Command Line Interface 或 AWS Batch API 中建立運算環境時指定 `ecsInstanceRole`角色。如需詳細資訊，請參閱 AWS CLI 命令參考或 AWS Batch API 參考。

若要建立 **ECS-CloudWatchLogs** IAM 政策

1. 前往 https://console.aws.amazon.com/iam/ 開啟 IAM 主控台。
2. 在導覽窗格中，選擇政策。
3. 選擇 Create policy (建立政策)。
4. 選擇 JSON，然後輸入下列政策：

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:PutLogEvents",
                "logs:DescribeLogStreams"
            ],
            "Resource": [
                "arn:aws:logs:*:*:*"
            ]
        }
    ]
}
```

5. 選擇下一步：標籤。
6. （選用）對於新增標籤，選擇新增標籤以將標籤新增至政策。
7. 選擇下一步：檢閱。
8. 在檢閱政策頁面上，針對名稱輸入 **ECS-CloudWatchLogs**，然後輸入選用的描述。
9. 選擇建立政策。

將 **ECS-CloudWatchLogs** 政策連接至 **ecsInstanceRole**

1. 在以下網址開啟 IAM 主控台：https://console.aws.amazon.com/iam/。
2. 在導覽窗格中，選擇角色。
3. 選擇 ecsInstanceRole。如果角色不存在，請遵循 中的程序Amazon ECS 執行個體角色來建立角色。
4. 選擇新增許可，然後選擇連接政策。
5. 選擇 ECS-CloudWatchLogs 政策，然後選擇連接政策。

安裝及設定 CloudWatch 代理程式

您可以建立包含 CloudWatch 監控的 Amazon EC2 啟動範本。如需詳細資訊，請參閱《Amazon EC2 使用者指南》中的從啟動範本啟動執行個體和進階詳細資訊。

您也可以在現有的 Amazon EC2 AMI 上安裝 CloudWatch 代理程式，然後在 AWS Batch 初次執行精靈中指定映像。如需詳細資訊，請參閱安裝 CloudWatch 代理程式和 AWS Batch 教學課程入門。

> **Note**
>
> AWS Fargate 資源不支援啟動範本。

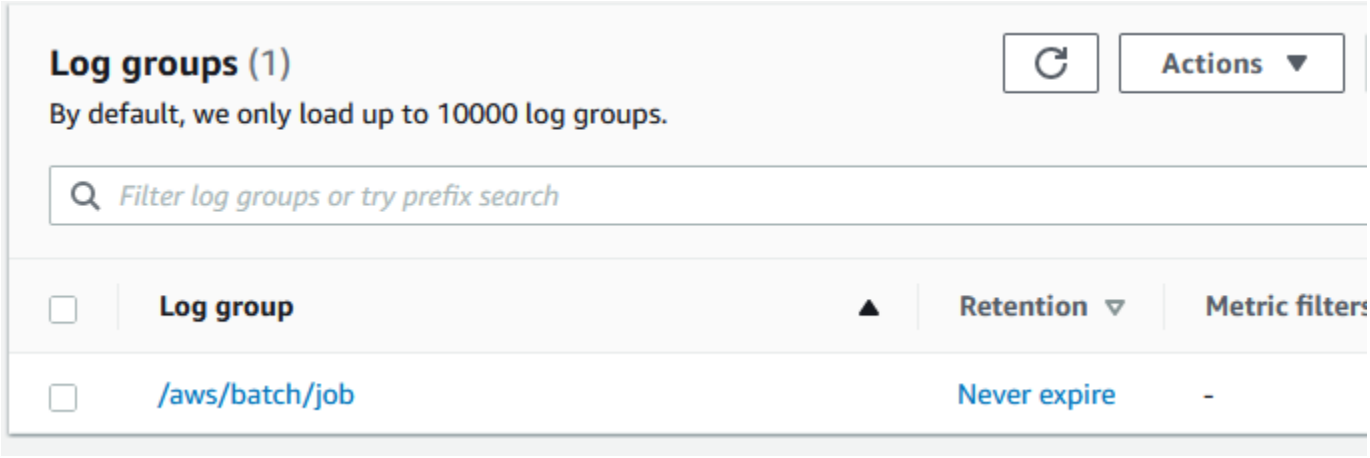
教學課程：檢視 CloudWatch Logs

您可以在 中檢視和搜尋 CloudWatch Logs 日誌 AWS 管理主控台。

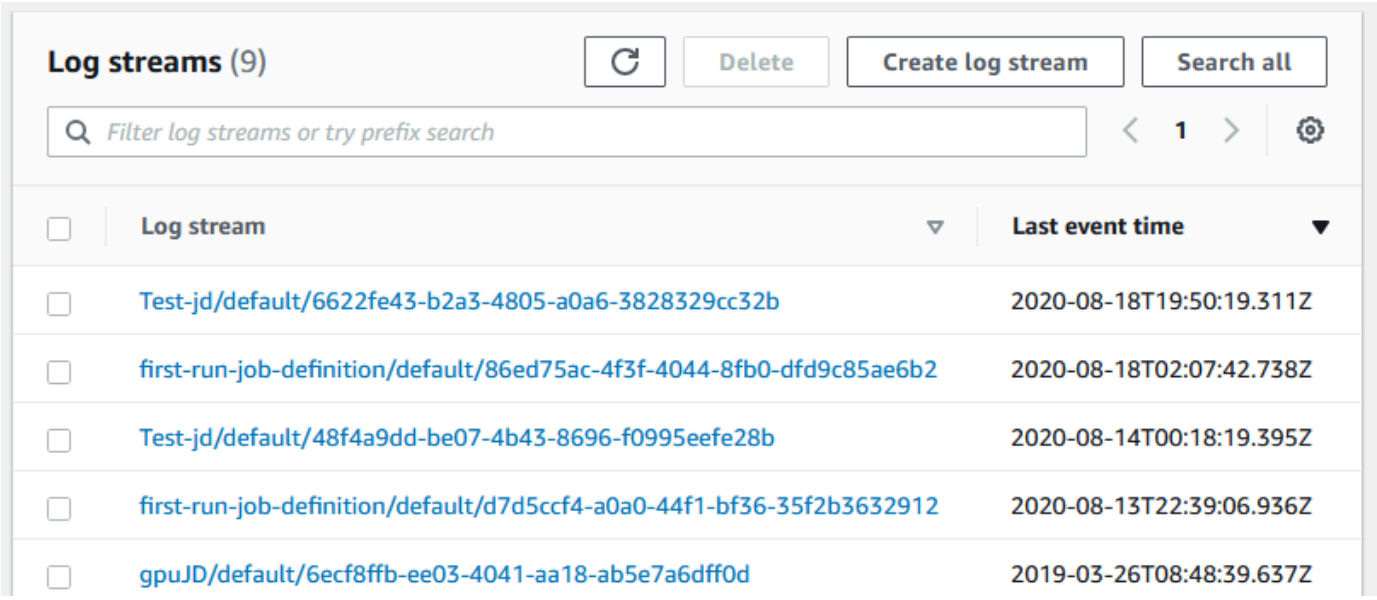
> **Note**
>
> 資料可能需要幾分鐘的時間才會顯示在 CloudWatch Logs 中。

若要檢視您的 CloudWatch Logs 資料

1. 透過 https://console.aws.amazon.com/cloudwatch/ 開啟 CloudWatch 主控台。
2. 在左側導覽窗格中，選擇日誌，然後選擇日誌群組。

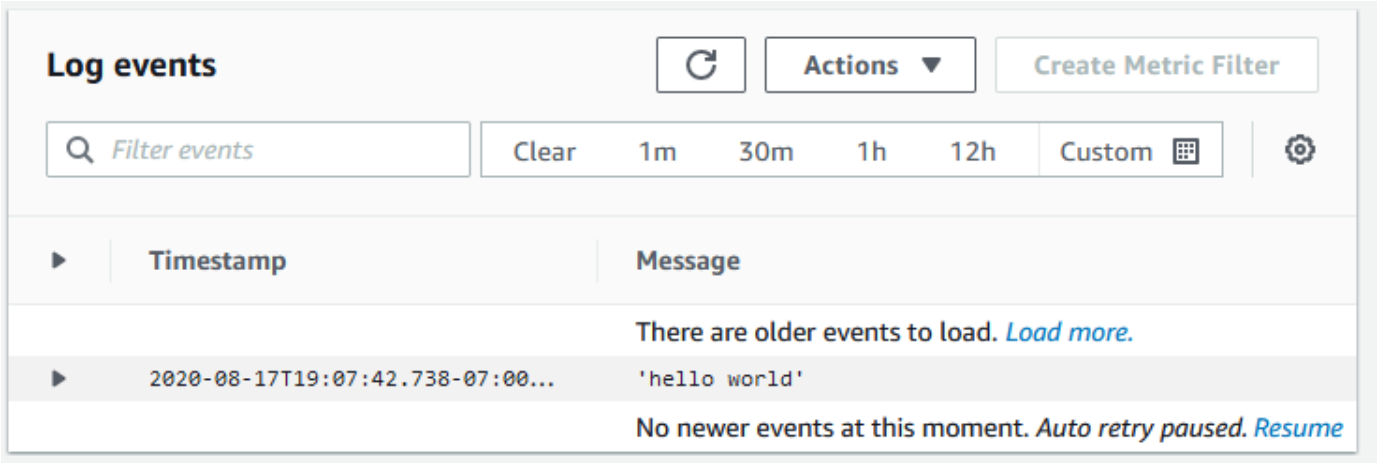


3. 選擇要檢視的日誌群組。



4. 選擇要檢視的日誌串流。根據預設，串流會以任務名稱的前 200 個字元和 Amazon ECS 任務 ID 來識別。

> **Tip**
>
> 若要下載日誌串流資料，請選擇動作。



AWS Batch CloudWatch Container Insights

CloudWatch Container Insights 會從 AWS Batch 運算環境和任務收集、彙總和摘要指標和日誌。這些指標包括 CPU、記憶體、磁碟和網路使用率。您可以將這些指標新增至 CloudWatch 儀表板。

營運資料的收集形式為「效能日誌事件」。這些項目使用的是結構化的 JSON 結構描述，可擷取高基數資料並大量地進行存放。從這些資料中，CloudWatch 會在運算環境和任務層級建立更高層級的彙總指標，做為 CloudWatch 指標。如需詳細資訊，請參閱 Amazon CloudWatch 使用者指南中的 Amazon ECS 的 Container Insights 結構式日誌。

> **Important**
>
> CloudWatch Container Insights 會以自訂指標的形式由 CloudWatch 收費。如需詳細資訊，請參閱 Amazon CloudWatch Events 定價

主題

- 開啟容器洞見

開啟容器洞見

完成下列步驟以開啟 AWS Batch 運算環境的 Container Insights。

1. 開啟 AWS Batch 主控台。

2. 選擇 Environments (環境)。
3. 選擇您想要的運算環境。
4. 在容器洞見索引標籤上，開啟運算環境的容器洞見。

> **Tip**
>
> 您可以選取預設間隔來彙總指標或建立自訂間隔。

根據預設，會顯示下列指標。要檢視 Amazon ECS Container Insights 指標的完整清單，請參閱 Amazon CloudWatch 使用者指南中的 Amazon ECS Container Insights 指標。

- **JobCount** – 在運算環境中執行的任務數量。
- **ContainerInstanceCount** – 執行 Amazon ECS 代理程式並在運算環境中註冊的 Amazon Elastic Compute Cloud 執行個體數目。
- **MemoryReserved** – 運算環境任務預留的記憶體。此指標只會針對在其任務定義中具有已定義記憶體保留的任務收集。
- **MemoryUtilized** – 運算環境任務正在使用的記憶體。此指標只會針對在其任務定義中具有已定義記憶體保留的任務收集。
- **CpuReserved** – 運算環境任務預留的 CPU 單位。此指標只會針對在其任務定義中具有已定義 CPU 保留的任務收集。
- **CpuUtilized** – 運算環境中任務使用的 CPU 單位。此指標只會針對在其任務定義中具有已定義 CPU 保留的任務收集。
- **NetworkRxBytes** - 收到的位元組數。此指標僅適用於使用 awsvpc或橋接網路模式的任務中的容器。
- **NetworkTxBytes** – 傳輸的位元組數。此指標僅適用於使用 awsvpc或橋接網路模式的任務中的容器。
- **StorageReadBytes** – 從儲存體讀取的位元組數。
- **StorageWriteBytes** – 寫入儲存體的位元組數。

使用 CloudWatch Logs 監控 AWS Batch Amazon EKS 任務

您可以使用 Amazon CloudWatch Logs 在一個位置監控、存放和檢視所有日誌檔案。使用 CloudWatch Logs，您可以搜尋、篩選和分析來自多個來源的日誌資料。

您可以下載 AWS Fluent Bit映像的 ，其中包含在 CloudWatch Logs 中監控 Amazon EKS 任務 AWS Batch 的外掛程式。 Fluent Bit 是 Docker 和Kubernetes相容的開放原始碼日誌處理器和轉送器。我們建議您使用 Fluent Bit做為日誌路由器，因為它的資源密集度低於 Fluentd。如需詳細資訊，請參閱使用 Amazon CloudWatch 可觀測性 EKS 附加元件或 Helm Chart 安裝 CloudWatch 代理 Amazon CloudWatch程式。

先決條件

- 將CloudWatchAgentServerPolicy政策連接至工作者節點 AWS Identity and Access Management 的政策。如需詳細資訊，請參閱驗證先決條件。

安裝 附加元件

如需如何 AWS 安裝 Fluent Bit 和建立 CloudWatch 群組的指示，請參閱使用 Amazon CloudWatch 可觀測性 EKS 附加元件或 Helm Chart 安裝 CloudWatch 代理 Amazon CloudWatch程式。

安裝附加元件時，您必須提供下列其他組態資料：

- 如果您使用 安裝附加元件 AWS 管理主控台 ，則需要在組態值中提供下列容錯：

```
{
  "tolerations": [
    {
      "key": "batch.amazonaws.com/batch-node",
      "operator": "Exists"
    }
  ]
}
```

- 如果您使用 安裝附加元件 AWS CLI ，請新增下列引數：

```
--configuration-values '{"tolerations":[{"key":"batch.amazonaws.com/batch-
node","operator":"Exists"}]}'
```

> **Tip**
>
> 請記住， Fluent Bit會在 AWS Batch 節點上使用 .5 個 CPU 和 100 MB 的記憶體。這可減少 AWS Batch 任務的總可用容量。當您調整任務大小時，請考慮這一點。

標記您的 AWS Batch 資源

為了協助您管理 AWS Batch 資源，您可以以標籤的形式將自己的中繼資料指派給每個資源。本主題說明標籤並示範如何建立它們。

主題

- 標籤基本概念
- 標記您的 資源
- 標籤限制
- 教學課程：使用主控台管理標籤
- 使用 CLI 或 API 管理標籤

標籤基本概念

標籤是您指派給 AWS 資源的標籤。每個標籤皆包含由您定義的一個金鑰與一個選用值。

標籤可讓您依用途、擁有者或環境等方式將 AWS 資源分類。當您有許多相同類型的資源時，您可以依據先前指派的標籤，快速識別特定的資源。例如，您可以為您的 AWS Batch 服務定義一組標籤，以協助您追蹤每個服務的擁有者和堆疊層級。建議您為每個資源類型設計一組一致的標籤金鑰。

標籤不會自動指派給您的資源。新增標籤後，您可以隨時編輯標籤索引鍵和值，或從資源移除標籤。如果您刪除資源，也會刪除任何該資源的標籤。

標籤對 沒有任何語意意義， AWS Batch 並嚴格解譯為字元字串。您可以將標籤的值設為空白字串，但您無法將標籤的值設為 Null。若您將與現有標籤具有相同鍵的標籤新增到該資源，則新值會覆寫舊值。

您可以使用 AWS 管理主控台、 AWS CLI和 AWS Batch API 來使用標籤。

如果您使用的是 AWS Identity and Access Management (IAM)，您可以控制 AWS 帳戶中哪些使用者具有建立、編輯或刪除標籤的許可。

標記您的 資源

您可以標記新的或現有的 AWS Batch 運算環境、任務、任務定義、任務佇列和排程政策。

如果您使用的是 AWS Batch 主控台，您可以隨時使用相關資源頁面上的標籤索引標籤，將標籤套用至新資源或現有資源。

如果您使用的是 AWS Batch API AWS CLI、 或 AWS SDK，您可以使用相關 API 動作上的 `tags` 參數將標籤套用至新資源，或使用 `TagResource` API 動作將標籤套用至現有資源。如需詳細資訊，請參閱 TagResource。

有些資源建立動作可讓您在建立資源時指定資源的標籤。如果無法在資源建立時套用標籤，則資源建立程序會失敗。這可確保您要在建立時標記的資源是以指定的標籤建立，不然就根本不會建立。如果您在建立時標記資源，則不需要在建立資源之後執行自訂標記指令碼。

下表說明可標記 AWS Batch 的資源，以及在建立時可標記的資源。

資源的標籤支援 AWS Batch

資源	支援標籤	支援標籤傳播	支援建立時標記 (AWS Batch API、 AWS CLI AWS SDK)
AWS Batch 運算環境	是	否。運算環境標籤不會傳播到任何其他資源。資源的標籤是在 CreateComputeEnvironment API 操作中傳遞之 computeResources 物件的標籤成員中指定。	是
AWS Batch 任務	是	是	是
AWS Batch 任務定義	是	否	是
AWS Batch 任務佇列	是	否	是
AWS Batch 排程政策	是	否	是

標籤限制

以下基本限制適用於標籤：

- 每一資源最多標籤數 – 50
- 對於每一個資源，每個標籤金鑰必須是唯一的，且每個標籤金鑰只能有一個值。

- 索引鍵長度上限 - 128 個 UTF-8 Unicode 字元
- 值的長度上限 - 256 個 UTF-8 Unicode 字元
- 如果您的標記結構描述用於多個 AWS 服務和資源，請記住，其他服務可能有允許的字元限制。通常允許的字元包括：可用 UTF-8 表示的英文字母、數字和空格，還有以下字元：+ - = . _ : / @。
- 標籤鍵與值皆區分大小寫。
- 請勿使用 `aws:`、 `AWS:`或任何大寫或小寫的組合，例如索引鍵或值的字首，因為其保留供 AWS 使用。您不可編輯或刪除具此字首的標籤金鑰或值。具此字首的標籤不算在每一資源的標籤數限制內。

教學課程：使用主控台管理標籤

使用 AWS Batch 主控台，您可以管理與新的或現有的運算環境、任務、任務定義和任務佇列相關聯的標籤。

建立時在個別資源上新增標籤

您可以在建立時將標籤新增至 AWS Batch 運算環境、任務、任務定義、任務佇列和排程政策。

新增和刪除個別資源上的標籤

AWS Batch 可讓您直接從資源的頁面新增或刪除與叢集相關聯的標籤。

在個別資源上新增或刪除標籤

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選擇要使用的 區域。
3. 在導覽窗格中，選擇資源類型 （ 例如，任務佇列)。
4. 選擇特定資源，然後選擇編輯標籤。
5. 視需要新增或刪除標籤。
   - 若要新增標籤 — 在清單結尾的空白文字方塊中指定索引鍵和值。
   - 若要刪除標籤 — 選擇標籤旁的

     Delete icon
     按鈕。
6. 針對您要新增或刪除的每個標籤重複此程序，然後選擇編輯標籤以完成。

使用 CLI 或 API 管理標籤

使用下列 AWS CLI 命令或 AWS Batch API 操作來新增、更新、列出和刪除 資源的標籤。

資源的標籤支援 AWS Batch

任務	API 動作	AWS CLI	AWS Tools for Windows PowerShell
新增或覆寫一或多個標籤。	TagResource	tag-resource	Add-BATResourceTag
刪除一或多個標籤。	UntagResource	untag-resource	Remove-BATResourceTag
列出資源的標籤	ListTagsForResource	list-tags-for-resource	Get-BATResourceTag

下列範例示範如何使用 AWS CLI來標記或取消標記資源。

範例 1：標記現有資源

以下命令會標記現有的資源。

```
aws batch tag-resource --resource-arn resource_ARN --tags team=devs
```

範例 2：取消標記現有的資源

以下命令會從現有的資源刪除標籤。

```
aws batch untag-resource --resource-arn resource_ARN --tag-keys tag_key
```

範例 3：列出資源的標籤

以下命令列出與現有資源相關聯的標籤。

```
aws batch list-tags-for-resource --resource-arn resource_ARN
```

有些資源建立動作可讓您在建立資源時指定標籤。下列動作支援在建立時新增標籤。

任務	API 動作	AWS CLI	AWS Tools for Windows PowerShell
建立運算環境	CreateComputeEnvironment	create-compute-environment	New-BATComputeEnvironment
建立任務佇列	CreateJobQueue	create-job-queue	New-BATJobQueue
建立排程政策	CreateSchedulingPolicy	create-scheduling-policy	New-BATSchedulingPolicy
註冊任務定義	RegisterJobDefinition	register-job-definition	Register-BATJobDefinition
提交工作	SubmitJob	submit-job	Submit-BATJob
建立消耗性資源	CreateConsumableResource	create-consumable-resource	Create-BATConsumableResource

的最佳實務 AWS Batch

您可以使用 大規模 AWS Batch 執行各種嚴苛的運算工作負載，而無需管理複雜的架構。 AWS Batch 任務可用於流行病學、遊戲和機器學習等領域的各種使用案例。

本主題涵蓋使用 時要考慮的最佳實務， AWS Batch 以及使用 時如何執行和最佳化工作負載的指引 AWS Batch。

主題

- 使用時機 AWS Batch
- 大規模執行的檢查清單
- 最佳化容器和 AMIs
- 選擇正確的運算環境資源
- Amazon EC2 隨需或 Amazon EC2 Spot
- 使用適用於 的 Amazon EC2 Spot 最佳實務 AWS Batch
- 常見錯誤和故障診斷

使用時機 AWS Batch

AWS Batch 會以低成本大規模執行任務，並提供佇列服務和成本最佳化擴展。不過，並非所有工作負載都適合使用 執行 AWS Batch。

- 短期任務 – 如果任務只執行幾秒鐘，則排程批次任務的額外負荷可能需要比任務本身的執行時間更長的時間。作為解決方法，binpack您的任務會在您提交之前一起進行 AWS Batch。然後，設定您的 AWS Batch 任務以反覆執行任務。例如，將個別任務引數暫存到 Amazon DynamoDB 資料表或 Amazon S3 儲存貯體中的檔案。考慮將任務分組，讓任務執行 3-5 分鐘。在您binpack執行任務之後，請在 AWS Batch 任務中循環執行任務群組。
- 必須立即執行的任務 - AWS Batch 可以快速處理任務。不過， AWS Batch 是排程器，並針對成本效能、任務優先順序和輸送量進行最佳化。 AWS Batch 可能需要時間來處理請求。如果您在幾秒鐘內需要回應，則使用 Amazon ECS 或 Amazon EKS 的服務型方法更合適。

大規模執行的檢查清單

在 50，000 個或更多 vCPUs 上執行大型工作負載之前，請考慮下列檢查清單。

> **Note**
>
> 如果您計劃在數百萬個或更多vCPUs 上執行大型工作負載，或需要大規模執行的指引，請聯絡您的 AWS 團隊。

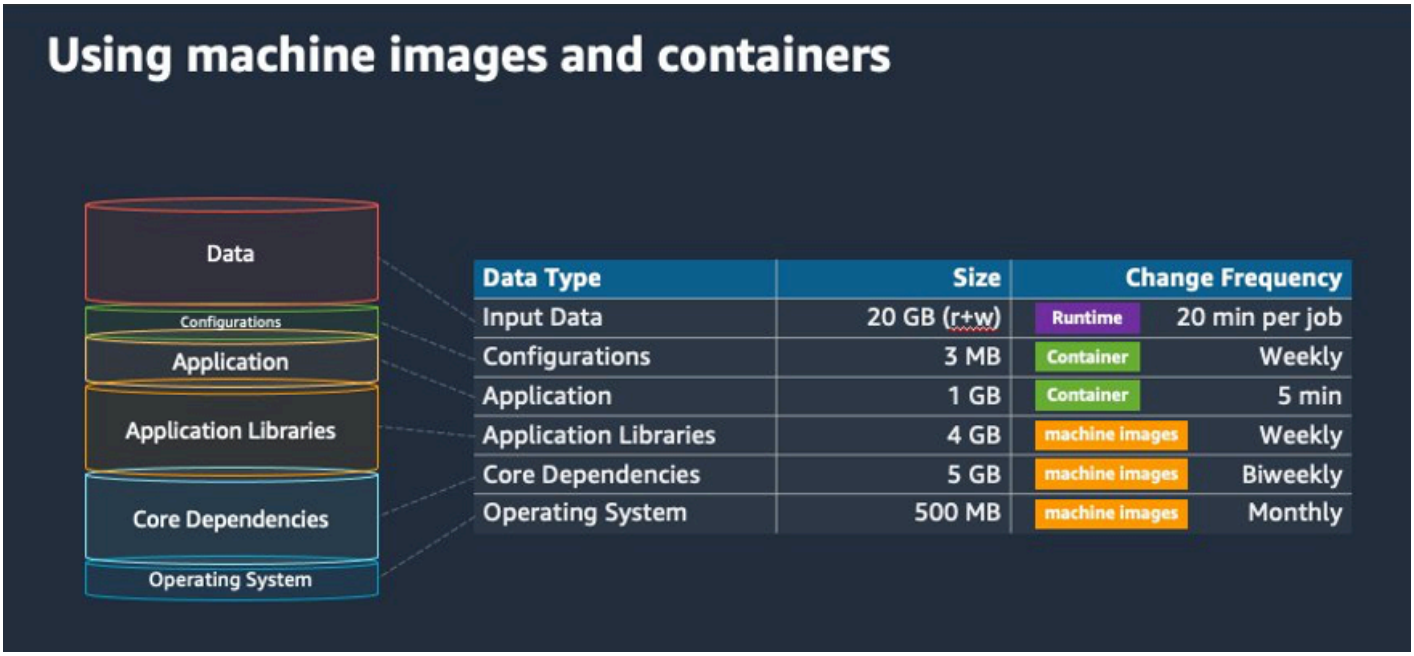
- 檢查您的 Amazon EC2 配額 – 在 Service Quotas面板中檢查您的 Amazon EC2 配額 （ 也稱為限制 ） AWS 管理主控台。如有必要，請請求提高 Amazon EC2 執行個體尖峰數量的配額。請記住，Amazon EC2 Spot 和 Amazon 隨需執行個體有不同的配額。如需詳細資訊，請參閱 Service Quotas 入門。
- 驗證每個區域的 Amazon Elastic Block Store 配額 – 每個執行個體都會使用作業系統的 GP2 或 GP3 磁碟區。根據預設，每個 的配額 AWS 區域 為 300 TiB。不過，每個執行個體都會使用計數做為此配額的一部分。因此，當您驗證每個區域的 Amazon Elastic Block Store 配額時，請務必將其納入考量。如果達到配額，則無法建立更多執行個體。如需詳細資訊，請參閱 Amazon Elastic Block Store 端點和配額
- 使用 Amazon S3 進行儲存 – Amazon S3 提供高輸送量，並有助於根據每個可用區域中的任務和執行個體數量，消除對要佈建多少儲存體的猜測。如需詳細資訊，請參閱最佳實務設計模式：最佳化 Amazon S3 效能。
- 逐步擴展以提早識別瓶頸 – 對於在數百萬個或更多 vCPUs上執行的任務，請開始降低並逐步增加，以便您可以提早識別瓶頸。例如，從在 50，000 個 vCPUs 上執行開始。然後，將計數增加到 20 萬 vCPUs，然後增加到 50 萬vCPUs，以此類推。換句話說，繼續逐步增加 vCPU 計數，直到您達到所需的 vCPUs 數量。
- 及早監控以識別潛在問題 – 為了避免大規模執行時的潛在中斷和問題，請務必同時監控您的應用程式和架構。即使從 1，000 擴展到 5，000 vCPUs，也可能發生中斷。您可以使用 Amazon CloudWatch Logs 檢閱日誌資料，或使用用戶端程式庫的 CloudWatch Embedded Metrics。如需詳細資訊，請參閱 CloudWatch Logs 代理程式參考和 `aws-embedded-metrics`

最佳化容器和 AMIs

容器大小和結構對於您執行的第一組任務很重要。如果容器大於 4 GB，則尤其如此。容器映像內建於 layer 中。Docker 會使用三個並行執行緒平行擷取圖層。您可以使用 `max-concurrent-downloads` 參數增加並行執行緒的數量。如需詳細資訊，請參閱 Dockerd 文件。

雖然您可以使用較大的容器，但我們建議您最佳化容器結構和大小，以縮短啟動時間。

- 較小型的容器擷取速度較快 – 較小型的容器可能會導致較快的應用程式啟動時間。若要減少容器大小，請將不常更新的程式庫或檔案卸載至 Amazon Machine Image (AMI)。您也可以使用綁定掛載來提供容器的存取權。如需詳細資訊，請參閱繫結掛載。
- 建立大小相同的圖層並分解大型圖層 – 每個圖層都會由一個執行緒擷取。因此，大型 layer 可能會大幅影響您的任務啟動時間。我們建議最大層大小為 2 GB，以便在較大的容器大小和更快的啟動時間之間取得良好的權衡。您可以執行 `docker history your_image_id`命令來檢查您的容器映像結構和圖層大小。如需詳細資訊，請參閱 Docker 文件。
- 使用 Amazon Elastic Container Registry 做為容器儲存庫 – 當您平行執行數千個任務時，自我管理的儲存庫可能會失敗或調節輸送量。Amazon ECR 可大規模運作，並且可以使用超過一百萬vCPUs處理工作負載。



選擇正確的運算環境資源

AWS Fargate 需要的初始設定和組態比 Amazon EC2 更少，而且可能更容易使用，特別是第一次使用時。搭配使用 Fargate，您無需管理伺服器、處理容量規劃，或出於安全性而隔離容器工作負載。

如果您有下列需求，建議您使用 Fargate 執行個體：

- 您的任務必須快速啟動，特別是不到 30 秒。
- 任務的需求為 16 vCPUs或更少、沒有 GPUs，以及 120 GiB 或更少的記憶體。

如需詳細資訊，請參閱何時使用 Fargate。

如果您有下列需求，建議您使用 Amazon EC2 執行個體：

- 您需要加強對執行個體選擇的控制，或使用特定的執行個體類型。
- 您的任務需要 AWS Fargate 無法提供的資源，例如 GPUs、更多記憶體、自訂 AMI 或 Amazon Elastic Fabric Adapter。
- 您需要高水準的輸送量或並行。
- 您需要自訂您的 AMI、Amazon EC2 啟動範本，或存取特殊 Linux 參數。

使用 Amazon EC2，您可以根據特定需求更精細地調整工作負載，並視需要大規模執行。

Amazon EC2 隨需或 Amazon EC2 Spot

大多數 AWS Batch 客戶使用 Amazon EC2 Spot 執行個體，因為比隨需執行個體節省成本。不過，如果您的工作負載執行數小時且無法中斷，則隨需執行個體可能更適合您。您一律可以先嘗試 Spot 執行個體，並視需要切換到隨需。

如果您有下列需求和期望，請使用 Amazon EC2 隨需執行個體：

- 任務的執行時間超過一小時，您無法容忍工作負載中斷。
- 您的整體工作負載具有嚴格的 SLO（服務層級目標），無法增加運算時間。
- 您需要的執行個體更有可能看到中斷。

如果您有下列需求和期望，請使用 Amazon EC2 Spot 執行個體：

- 任務的執行時間通常為 30 分鐘或更短。
- 您可以在工作負載中容忍潛在的中斷和任務重新排程。如需詳細資訊，請參閱 Spot 執行個體建議程式。
- 如果中斷，可以從檢查點重新啟動長時間執行的任務。

您可以先在 Spot 執行個體上提交，然後使用隨需執行個體做為備用選項，來混合這兩種購買模式。例如，在連線至在 Amazon EC2 Spot 執行個體上執行之運算環境的佇列上提交您的任務。如果任務中斷，請從 Amazon EventBridge 擷取事件，並將其與 Spot 執行個體回收相關聯。然後，使用 AWS Lambda 函數或 將任務重新提交至隨需佇列 AWS Step Functions。如需詳細資訊，請參閱 教學課程：針對失敗的任務事件傳送 Amazon Simple Notification Service 提醒、處理 Amazon EC2 Spot 執行個體中斷的最佳實務，以及AWS Batch 使用 Step Functions 管理。

> **Important**
>
> 針對隨需運算環境使用不同的執行個體類型、大小和可用區域，以維持 Amazon EC2 Spot 執行個體集區可用性並降低中斷率。

使用適用於 的 Amazon EC2 Spot 最佳實務 AWS Batch

當您選擇 Amazon Elastic Compute Cloud (EC2) Spot 執行個體時，您可以最佳化工作流程以節省成本，有時可大幅節省成本。如需詳細資訊，請參閱 Amazon EC2 Spot 的最佳實務。

若要最佳化工作流程以節省成本，請考慮下列 Amazon EC2 Spot 最佳實務 AWS Batch：

- 選擇**SPOT_CAPACITY_OPTIMIZED**配置策略 – 從最深層的 Amazon EC2 Spot 容量集區 AWS Batch 中選擇 Amazon EC2 執行個體。如果您擔心中斷，這是適當的選擇。如需詳細資訊，請參閱的執行個體類型配置策略 AWS Batch。
- 多樣化執行個體類型 – 若要多樣化執行個體類型，請考慮相容的大小和系列，然後根據價格或可用性來 AWS Batch 選擇。例如，將 `c5.24xlarge`視為 `c5.12xlarge`或 `c5a`、`c5n`、`m5`、 `c5d`和 `m5d` 系列的替代方案。如需詳細資訊，請參閱靈活了解執行個體類型和可用區域。
- 減少任務執行時間或檢查點 – 我們建議不要在使用 Amazon EC2 Spot 執行個體時執行需要一小時或更長時間以避免中斷的任務。如果您將任務分割或檢查點為包含 30 分鐘或更短的較小部分，則可以大幅降低中斷的可能性。
- 使用自動重試 – 為了避免 AWS Batch 任務中斷，請設定任務的自動重試。批次任務可能因下列任何原因中斷：傳回非零結束碼、發生服務錯誤，或發生執行個體回收。您最多可以設定 10 次自動重試。首先，我們建議您設定至少 1-3 次自動重試。如需追蹤 Amazon EC2 Spot 中斷的詳細資訊，請參閱 Spot 中斷儀表板。

  對於 AWS Batch，如果您設定重試參數，任務會放置在任務佇列的前面。也就是說，任務會獲得優先順序。當您建立任務定義或在 中提交任務時 AWS CLI，您可以設定重試策略。如需詳細資訊，請參閱 submit-job。

```
$ aws batch submit-job --job-name MyJob \
    --job-queue MyJQ \
    --job-definition MyJD \
    --retry-strategy attempts=2
```

- 使用自訂重試 – 您可以將任務重試策略設定為特定應用程式結束程式碼或執行個體回收。在下列範例中，如果主機造成失敗，任務最多可重試五次。不過，如果任務因不同原因而失敗，則任務會結束，且狀態會設為 FAILED。

```
"retryStrategy": {
    "attempts": 5,
    "evaluateOnExit":
    [{
        "onStatusReason" :"Host EC2*",
        "action": "RETRY"
    },{
      "onReason" : "*",
        "action": "EXIT"
    }]
}
```

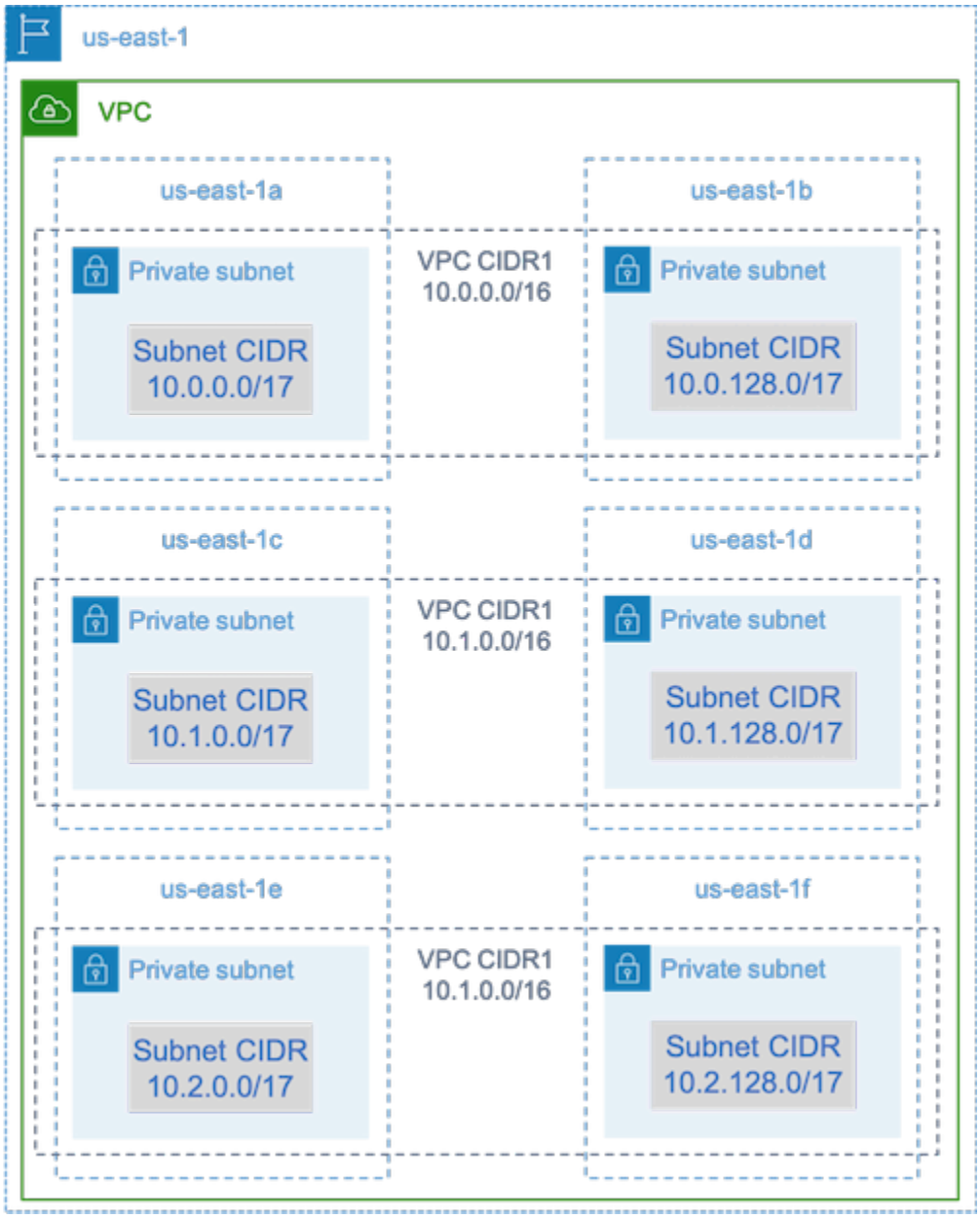
- 使用 Spot 中斷儀表板 – 您可以使用 Spot 中斷儀表板來追蹤 Spot 中斷。應用程式會在回收的 Amazon EC2 Spot 執行個體上提供指標，以及 Spot 執行個體所在的可用區域。如需詳細資訊，請參閱 Spot 中斷儀表板

常見錯誤和故障診斷

中的錯誤 AWS Batch 通常發生在應用程式層級，或是由不符合特定任務需求的執行個體組態所造成。其他問題包括任務卡在 RUNNABLE 狀態，或運算環境卡在 INVALID 狀態。如需有關故障診斷任務卡在 RUNNABLE 狀態的詳細資訊，請參閱 任務停滯在 RUNNABLE 狀態。如需 INVALID 狀態中運算環境故障診斷的資訊，請參閱 INVALID 運算環境。

- 檢查 Amazon EC2 Spot vCPU 配額 – 驗證您目前的服務配額是否符合任務要求。例如，假設您目前的服務配額為 256 個 vCPUs且任務需要 10，000 vCPUs。然後，服務配額不符合任務要求。如需詳細資訊和疑難排解說明，請參閱 Amazon EC2 服務配額和如何提高 Amazon EC2resources的服務配額？。
- 任務在應用程式執行之前失敗 – 有些任務可能因為DockerTimeoutError錯誤或CannotPullContainerError錯誤而失敗。如需疑難排解資訊，請參閱如何解決 中的「DockerTimeoutError」錯誤 AWS Batch？。
- IP 地址不足 – VPC 和子網路中的 IP 地址數目可以限制您可以建立的執行個體數目。使用無類別網域間路由 (CIDRs) 來提供比執行工作負載所需的更多 IP 地址。如有必要，您也可以建置具有大型地址空間的專用 VPC。例如，您可以在 中建立具有多個 CIDRs VPC，10.x.0.0/16並在每個可用區

域中建立具有 CIDR 為 的子網路`10.x.y.0/17`。在此範例中，x 介於 1-4 之間，y 為 0 或 128。此組態在每個子網路中提供 36，000 個 IP 地址。



- 確認執行個體已向 Amazon EC2 註冊 – 如果您在 Amazon EC2 主控台中看到執行個體，但 Amazon ECS 叢集中沒有 Amazon Elastic Container Service 容器執行個體，Amazon ECS 代理程式可能不會安裝在 Amazon Machine Image (AMI) 上。您的 AMI 中的 Amazon ECS 代理程式、Amazon EC2 資料或啟動範本可能也未正確設定。若要隔離根本原因，請建立個別的 Amazon EC2 執行個體，或使用 SSH 連線到現有的執行個體。如需詳細資訊，請參閱 Amazon ECS 容器代理程式組態、Amazon ECS 日誌檔案位置和 運算資源 AMI。
- 檢閱 AWS 儀表板 – 檢閱 AWS 儀表板，以確認預期的任務狀態和運算環境如預期擴展。您也可以在 CloudWatch 中檢閱任務日誌。

- 確認您的執行個體已建立 – 如果執行個體已建立，這表示您的運算環境會如預期擴展。如果未建立執行個體，請在運算環境中尋找要變更的關聯子網路。如需詳細資訊，請參閱驗證 Auto Scaling 群組的擴展活動。

  我們也建議您驗證執行個體是否可以滿足相關的任務需求。例如，任務可能需要 1 TiB 的記憶體，但運算環境使用的 C5 執行個體類型限制為 192 GB 的記憶體。
- 驗證您的執行個體是否正由 請求 AWS Batch – 檢查 Auto Scaling 群組歷史記錄，以確認您的執行個體正由 請求 AWS Batch。這是 Amazon EC2 如何嘗試取得執行個體的指示。如果您收到錯誤，指出 Amazon EC2 Spot 無法取得特定可用區域中的執行個體，這可能是因為可用區域不提供特定執行個體系列。
- 確認執行個體已向 Amazon ECS 註冊 – 如果您在 Amazon EC2 主控台中看到執行個體，但 Amazon ECS 叢集中沒有 Amazon ECS 容器執行個體，Amazon ECS 代理程式可能不會安裝在 Amazon Machine Image (AMI) 上。此外，Amazon ECS 代理程式、AMI 中的 Amazon EC2 資料或啟動範本可能未正確設定。若要隔離根本原因，請建立個別的 Amazon EC2 執行個體，或使用 SSH 連線到現有的執行個體。如需詳細資訊，請參閱 CloudWatch 代理程式組態檔案：日誌區段、Amazon ECS 日誌檔案位置和 運算資源 AMI。
- 開啟支援票證 – 如果您在進行故障診斷後仍遇到問題，並擁有支援計畫，請開啟支援票證。在支援票證中，請務必包含有關問題、工作負載詳細資訊、組態和測試結果的資訊。如需詳細資訊，請參閱比較 支援 計劃。
- 檢閱 AWS Batch 和 HPC 論壇 – 如需詳細資訊，請參閱 AWS Batch和 HPC 論壇。
- 檢閱 AWS Batch 執行期監控儀表板 – 此儀表板使用無伺服器架構從 Amazon ECS AWS Batch和 Amazon EC2 擷取事件，以提供任務和執行個體的洞見。如需詳細資訊，請參閱AWS Batch 執行期監控儀表板解決方案。

故障診斷 AWS Batch

您可能需要疑難排解與運算環境、任務佇列、任務定義或任務相關的問題。本章說明如何故障診斷和解決您 AWS Batch 環境中的此類問題。

AWS Batch 使用 IAM 政策、角色和許可，並在 Amazon EC2、Amazon ECS AWS Fargate和 Amazon Elastic Kubernetes Service 基礎設施上執行。若要對與這些服務相關的問題進行故障診斷，請參閱以下內容：

- 《IAM 使用者指南》中的 IAM 故障診斷
- 《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 疑難排解
- 《Amazon EKS 使用者指南》中的 Amazon EKS 疑難排解
- 《Amazon EC2 使用者指南》中的 EC2 執行個體故障診斷 Amazon EC2

內容

- AWS Batch
  - 接收自動執行個體系列更新的最佳執行個體類型組態
  - INVALID 運算環境
    - 不正確的角色名稱或 ARN
    - 修復INVALID運算環境
  - 任務停滯在 RUNNABLE 狀態
  - 建立時未標記 Spot 執行個體
  - Spot 執行個體未縮減規模
    - 將 AmazonEC2SpotFleetTaggingRole 受管政策連接至 中的 Spot Fleet 角色 AWS 管理主控台
    - 使用 將 AmazonEC2SpotFleetTaggingRole 受管政策連接至 Spot Fleet 角色 AWS CLI
  - 無法擷取 Secrets Manager 秘密
  - 無法覆寫任務定義資源需求
  - 更新desiredvCpus設定時的錯誤訊息
- AWS Batch 在 Amazon EKS 上
  - INVALID 運算環境
    - 不支援的Kubernetes版本

    - 執行個體描述檔不存在
    - 無效的Kubernetes命名空間
    - 已刪除的運算環境
    - 節點不會加入 Amazon EKS 叢集
- AWS Batch Amazon EKS 任務上的 卡在 RUNNABLE 狀態
- AWS Batch Amazon EKS 任務上的 卡在 STARTING 狀態
    - 案例：持續磁碟區宣告連接或掛載失敗
- 確認 aws-auth ConfigMap 已正確設定
- RBAC 許可或繫結未正確設定

AWS Batch

檢閱下列主題，尋找您使用 時可能遇到之常見問題的檢閱程序和潛在解決方案 AWS Batch。

主題

- 接收自動執行個體系列更新的最佳執行個體類型組態
- INVALID 運算環境
- 任務停滯在 RUNNABLE 狀態
- 建立時未標記 Spot 執行個體
- Spot 執行個體未縮減規模
- 無法擷取 Secrets Manager 秘密
- 無法覆寫任務定義資源需求
- 更新desiredvCpus設定時的錯誤訊息

接收自動執行個體系列更新的最佳執行個體類型組態

> **Note**
>
> 從 11/01/2025 開始， `optimal` 的行為將變更為符合 `default_x86_64`。在變更期間，您的執行個體系列可能會更新為較新一代。您不需要執行任何動作，即可進行升級。

AWS Batch 支援 的 instanceTypes 中的單一選項optimal，以符合任務佇列的需求。我們已推出兩個新的執行個體類型選項： default_x86_64和 default_arm64。default_x86_64 如果您未選擇執行個體類型，我們將使用 。這些新選項會根據您的任務佇列需求，自動選取不同系列和世代之間具成本效益的執行個體類型，讓您快速執行工作負載。

當 中有足夠容量的新執行個體類型可用時 AWS 區域，對應的預設集區會自動更新為新的執行個體類型。現有optimal選項將繼續受到支援，而且不會遭到取代，因為基礎預設集區會支援它，以提供後續更新的執行個體。如果您使用的是 'optimal，則不需要採取任何動作。

不過，請注意，系統只會使用新的執行個體類型自動更新 ENABLED和 VALID 運算環境 (CEs)。如果您有任何 DISABLED或 INVALID CEs，它們會在重新啟用並設為 VALID 狀態時收到更新。

INVALID 運算環境

您可能未正確設定受管運算環境。如果您這麼做，運算環境會進入 INVALID 狀態，且無法接受要放置的任務。下列各節說明可能的原因，以及如何根據原因進行故障診斷。

> **Important**
>
> AWS Batch 會代表您和在您的帳戶中建立和管理多個 AWS 資源，包括 Amazon EC2 啟動範本、Amazon EC2 Auto Scaling 群組、Amazon EC2 Spot 機群和 Amazon ECS 叢集。這些受管資源經過專門設定，以確保最佳 AWS Batch 操作。除非文件中明確說明 AWS Batch ，否則手動修改這些批次受管資源可能會導致意外行為，導致INVALID運算環境、次佳執行個體擴展行為、延遲工作負載處理或意外成本。 AWS Batch 服務無法確定是否支援這些手動修改。一律使用支援的 Batch APIs或 Batch 主控台來管理您的運算環境。

不正確的角色名稱或 ARN

運算環境進入 INVALID 狀態的最常見原因是 AWS Batch 服務角色或 Amazon EC2 Spot Fleet 角色的名稱或 Amazon Resource Name (ARN) 不正確。這在使用 AWS CLI 或 AWS SDKs 建立的運算環境中更為常見。當您在 中建立運算環境時 AWS 管理主控台， AWS Batch 可協助您選擇正確的服務或 Spot 機群角色。不過，假設您手動輸入名稱或 ARN，並不正確地輸入。然後，產生的運算環境也是 INVALID。

不過，假設您在 命令或 SDK 程式碼中 AWS CLI 手動輸入 IAM 資源的名稱或 ARN。在此情況下， AWS Batch 無法驗證字串。反之， AWS Batch 必須接受錯誤的值，並嘗試建立環境。如果 AWS Batch 無法建立環境，環境會移至 INVALID 狀態，而您會看到下列錯誤。

如為無效的服務角色：

```
CLIENT_ERROR - Not authorized to perform sts:AssumeRole (Service:
AWSSecurityTokenService; Status Code: 403; Error Code: AccessDenied;
Request ID: dc0e2d28-2e99-11e7-b372-7fcc6fb65fe7)
```

如為無效的 Spot Fleet 角色：

```
CLIENT_ERROR - Parameter: SpotFleetRequestConfig.IamFleetRole
is invalid. (Service: AmazonEC2; Status Code: 400; Error Code:
InvalidSpotFleetRequestConfig; Request ID: 331205f0-5ae3-4cea-
bac4-897769639f8d) Parameter: SpotFleetRequestConfig.IamFleetRole is
invalid
```

此問題的一個常見原因如下。您只能在使用 AWS CLI 或 AWS SDKs 時指定 IAM 角色的名稱，而不是完整的 Amazon Resource Name (ARN)。根據您建立角色的方式，ARN 可能包含`aws-service-role`路徑字首。例如，如果您使用 中的程序手動建立 AWS Batch 服務角色使用 AWS Batch的服務連結角色，您的服務角色 ARN 可能如下所示。

```
arn:aws:iam::123456789012:role/AWSBatchServiceRole
```

不過，如果您今天在主控台初次執行精靈中建立了服務角色，您的服務角色 ARN 可能如下所示。

```
arn:aws:iam::123456789012:role/aws-service-role/AWSBatchServiceRole
```

如果您將 AWS Batch 服務層級政策 (`AWSBatchServiceRole`) 連接至非服務角色，也會發生此問題。例如，在此案例中，您可能會收到類似下列內容的錯誤訊息：

```
CLIENT_ERROR - User: arn:aws:sts::account_number:assumed-role/batch-replacement-role/
aws-batch is not
   authorized to perform: action on resource ...
```

若要解決此問題，請執行下列其中一項操作。

- 當您建立 AWS Batch 運算環境時，請使用服務角色的空字串。
- 以下列格式指定服務角色：`arn:aws:iam::account_number:role/aws-service-role/batch.amazonaws.com/AWSServiceRoleForBatch`。

當您只在使用 AWS CLI 或 AWS SDKs時指定 IAM 角色的名稱時， 會 AWS Batch 假設您的 ARN 不使用`aws-service-role`路徑字首。因此，我們建議您在建立運算環境時，為您的 IAM 角色指定完整的 ARN。

若要修復以此方式設定錯誤的運算環境，請參閱 修復INVALID運算環境。

修復INVALID運算環境

當您的運算環境處於 INVALID 狀態時，請更新它以修復無效的參數。對於 不正確的角色名稱或 ARN，請使用正確的服務角色更新運算環境。

修復設定錯誤的運算環境

1. 在 https：//https://console.aws.amazon.com/batch/ 開啟 AWS Batch 主控台。
2. 從導覽列中，選取要 AWS 區域 使用的 。
3. 在導覽窗格中，選擇 Compute environments (運算環境)。
4. 在 Compute environments (運算環境) 頁面，選擇要編輯的運算環境旁的選項按鈕，然後選擇 Edit (編輯)。
5. 在更新運算環境頁面上，針對服務角色，選擇要與運算環境搭配使用的 IAM 角色。 AWS Batch 主控台只會顯示與運算環境有正確信任關係的角色。

   > **Tip**
   >
   > 如需如何建立服務連結角色的指示，請參閱 使用 的角色 AWS Batch。

6. 選擇 Save (儲存)，更新運算環境。

任務停滯在 RUNNABLE 狀態

假設您的運算環境包含運算資源，但您的任務不會超過 RUNNABLE 狀態。然後，某件事可能會阻止任務放置在運算資源上，並導致您的任務佇列遭到封鎖。以下是如何知道您的任務正在等待輪換或卡住並封鎖佇列的方法。

如果 AWS Batch 偵測到您有一個前端RUNNABLE任務並封鎖佇列，您會收到來自 Amazon CloudWatch Events 任務佇列封鎖事件的事件，其中包含原因。相同的原因也會更新為 `statusReason` `ListJobs`和 `DescribeJobs` API 呼叫的一部分。

或者，您可以透過 `CreateJobQueue`和 `UpdateJobQueue` API 動作來設定 `jobStateTimeLimitActions` 參數。

> **Note**
>
> 目前，您可以搭配 使用的唯一動作`jobStateLimitActions.action`是取消任務。

`jobStateTimeLimitActions` 參數用於指定一組在特定狀態下對任務 AWS Batch 執行的動作。您可以透過 `maxTimeSeconds` 欄位以秒為單位設定時間閾值。

當任務處於具有定義 `RUNNABLE`的狀態時`statusReason`， 會 AWS Batch 執行 `maxTimeSeconds` 經過之後指定的動作。

例如，您可以設定 `jobStateTimeLimitActions` 參數，為 `RUNNABLE` 狀態中等待足夠容量的任何任務等待最多 4 小時。您可以在取消任務之前`statusReason`將 設定為 `CAPACITY:INSUFFICIENT_INSTANCE_CAPACITY`，並將 `maxTimeSeconds`設定為 144000，並允許下一個任務進入任務佇列的前端。

以下是當 偵測到任務佇列遭到封鎖時， AWS Batch 提供的原因。此清單提供從 `ListJobs`和 `DescribeJobs` API 動作傳回的訊息。這些也是您可以為 `jobStateLimitActions.statusReason` 參數定義的相同值。

1. 原因：所有連線的運算環境的容量錯誤不足。請求時， 會 AWS Batch 偵測遇到容量不足錯誤的 Amazon EC2 執行個體。手動取消任務將允許後續任務移至佇列前端，但不解決服務角色問題（也可能），下一個任務也會遭到封鎖。最好手動調查並解決此問題。
   - **`statusReason`** 任務停滯時的 訊息： `CAPACITY:INSUFFICIENT_INSTANCE_CAPACITY - Service cannot fulfill the capacity requested for instance type [instanceTypeName]`
   - **`reason`** 用於 **`jobStateTimeLimitActions`**： `CAPACITY:INSUFFICIENT_INSTANCE_CAPACITY`
   - **`statusReason`** 任務取消後的訊息： `Canceled by JobStateTimeLimit action due to reason: CAPACITY:INSUFFICIENT_INSTANCE_CAPACITY`

   請注意：

   a. AWS Batch 服務角色需要`autoscaling:DescribeScalingActivities`許可，此偵測才能運作。如果您使用[使用 AWS Batch的服務連結角色](#)服務連結角色 (SLR) 或 [AWS 受管政策：AWSBatchServiceRole 政策](#)受管政策，則不需要採取任何動作，因為其許可政策已更新。

   b. 如果您使用 SLR 或 受管政策，則必須新增 `autoscaling:DescribeScalingActivities`和 `ec2:DescribeSpotFleetRequestHistory`許可，以便在 中接收封鎖的任務佇列事件和更新的任務狀態`RUNNABLE`。此外， AWS Batch 需要這些許可，才能透過

jobStateTimeLimitActions 參數執行cancellation動作，即使它們是在任務佇列上設定。

c. 在多節點平行 (MNP) 任務中，如果連接的高優先順序 Amazon EC2 運算環境發生錯誤insufficient capacity，即使較低優先順序的運算環境確實遇到此錯誤，也會封鎖佇列。

2. 原因：所有運算環境的maxvCpus參數都小於任務需求。在 上手動或透過設定 jobStateTimeLimitActions 參數來取消任務statusReason，可讓後續任務移至佇列的前端。或者，您可以增加主要運算環境的 maxvCpus 參數，以滿足封鎖任務的需求。

   - **statusReason** 當任務停滯時的訊息：MISCONFIGURATION:COMPUTE_ENVIRONMENT_MAX_RESOURCE - CE(s) associated with the job queue cannot meet the CPU requirement of the job.
   - **reason** 用於 **jobStateTimeLimitActions**：MISCONFIGURATION:COMPUTE_ENVIRONMENT_MAX_RESOURCE
   - **statusReason** 任務取消後的訊息：Canceled by JobStateTimeLimit action due to reason: MISCONFIGURATION:COMPUTE_ENVIRONMENT_MAX_RESOURCE

3. 原因：沒有任何運算環境的執行個體符合任務需求。當任務請求資源時， 會 AWS Batch 偵測沒有連接的運算環境能夠容納傳入的任務。在 上手動或透過設定 jobStateTimeLimitActions 參數來取消任務statusReason，可讓後續任務移至佇列的前端。或者，您可以重新定義運算環境允許的執行個體類型，以新增必要的任務資源。

   - **statusReason** 當任務停滯時的訊息：MISCONFIGURATION:JOB_RESOURCE_REQUIREMENT - The job resource requirement (vCPU/memory/GPU) is higher than that can be met by the CE(s) attached to the job queue.
   - **reason** 用於 **jobStateTimeLimitActions**：MISCONFIGURATION:JOB_RESOURCE_REQUIREMENT
   - **statusReason** 任務取消後的訊息：Canceled by JobStateTimeLimit action due to reason: MISCONFIGURATION:JOB_RESOURCE_REQUIREMENT

4. 原因：所有運算環境都有服務角色問題。若要解決此問題，請將您的服務角色許可與 進行比較，AWS 的 受管政策 AWS Batch並解決任何差距。注意：您無法透過 jobStateTimeLimitActions 參數設定可程式化動作來解決此錯誤。

   最佳實務是使用 使用 AWS Batch的服務連結角色來避免類似的錯誤。

   在 上手動或透過設定 jobStateTimeLimitActions 參數來取消任務statusReason，可讓後續任務移至佇列的前端。如果沒有解決服務角色問題 （ 服務角色 ），則可能也會封鎖下一個任務。最好手動調查並解決此問題。

- **statusReason** 當任務停滯時的訊息： `MISCONFIGURATION:SERVICE_ROLE_PERMISSIONS – Batch service role has a permission issue.`

5. 原因：您的運算環境具有不支援的執行個體類型組態。當您選取的可用區域中無法使用執行個體類型，或您的啟動範本或啟動組態包含與指定執行個體類型不相容的設定時，就會發生這種情況。若要解決此問題，請確認您指定的 AWS 區域 和可用區域中支援執行個體類型，檢查您的啟動範本設定是否與您的執行個體類型相容，並考慮更新至較新一代的執行個體類型。如需尋找支援執行個體類型的詳細資訊，請參閱《Amazon EC2 使用者指南》中的尋找 Amazon EC2 執行個體類型。Amazon EC2
   - **statusReason** 當任務停滯時的訊息： `MISCONFIGURATION:EC2_INSTANCE_CONFIGURATION_UNSUPPORTED - Your compute environment associated with this job queue has an unsupported instance type configuration.`
6. 原因：所有運算環境都無效。如需詳細資訊，請參閱INVALID 運算環境。注意：您無法透過 `jobStateTimeLimitActions` 參數設定可程式化動作來解決此錯誤。
   - **statusReason** 當任務停滯時的訊息： `ACTION_REQUIRED - CE(s) associated with the job queue are invalid.`
7. 原因： AWS Batch 偵測到封鎖的佇列，但無法判斷原因。注意：您無法透過 `jobStateTimeLimitActions` 參數設定可程式化動作來解決此錯誤。如需疑難排解的詳細資訊，請參閱 re：Post 中的為什麼我的 AWS Batch 任務卡在 RUNNABLE AWS中。
   - **statusReason** 任務停滯時的 訊息： `UNDETERMINED - Batch job is blocked, root cause is undetermined.`

如果您未從 CloudWatch Events 收到事件，或收到不明原因事件，以下是一些常見的此問題原因。

未在運算資源上設定**awslogs**日誌驅動程式

AWS Batch 任務會將日誌資訊傳送至 CloudWatch Logs。若要啟用此功能，您必須設定運算資源使用 `awslogs` 日誌驅動程式。假設您從 Amazon ECS 最佳化 AMI （或 Amazon Linux) 基礎運算資源 AMI。然後，此驅動程式預設會向 `ecs-init`套件註冊。現在假設您使用不同的基本 AMI。然後，您必須驗證在 Amazon ECS 容器代理程式啟動時，`awslogs`日誌驅動程式已指定為具有 `ECS_AVAILABLE_LOGGING_DRIVERS`環境變數的可用日誌驅動程式。如需詳細資訊，請參閱運算資源 AMI 規格及教學課程：建立運算資源 AMI。

資源不足

如果您的任務定義指定比運算資源可配置更多的 CPU 或記憶體資源，則不會放置您的任務。例如，假設您的任務指定 4 GiB 的記憶體，而且您的運算資源少於可用的記憶體。然後，任務便無

法放置在這些運算資源上。在此情況下，您必須減少任務定義中所指定的記憶體，或在環境中加入更多運算資源。有些記憶體保留給 Amazon ECS 容器代理程式和其他關鍵系統程序。如需詳細資訊，請參閱運算資源記憶體管理。

運算資源無法存取網際網路

運算資源需要存取，才可以與 Amazon ECS 服務端點通訊。可透過介面 VPC 端點或透過具備公有 IP 地址的運算資源來實現。

如需介面 VPC 端點的詳細資訊，請參閱 Amazon Elastic Container Service 開發人員指南中的 Amazon ECS 介面 VPC 端點 (AWS PrivateLink)。

如果您沒有設定介面 VPC 端點，且運算資源沒有公有 IP 地址，則它們必須使用網路地址轉譯 (NAT) 來提供此存取。如需詳細資訊，請參閱 Amazon VPC 使用者指南中的 NAT 閘道。如需詳細資訊，請參閱the section called “建立 VPC”。

已達到 Amazon EC2 執行個體限制

您的帳戶可以在 中啟動的 Amazon EC2 執行個體數量 AWS 區域 取決於您的 EC2 執行個體配額。某些執行個體類型也有per-instance-type配額。如需帳戶 Amazon EC2 執行個體配額的詳細資訊，包括如何請求提高限制，請參閱《Amazon EC2 使用者指南》中的 Amazon EC2 服務限制。Amazon EC2

未安裝 Amazon ECS 容器代理程式

Amazon ECS 容器代理程式必須安裝在 Amazon Machine Image (AMI) 上，才能讓 AWS Batch 執行任務。根據預設，Amazon ECS 容器代理程式會安裝在 Amazon ECS 最佳化 AMIs上。如需 Amazon ECS 容器代理程式的詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 容器代理程式。

如需詳細資訊，請參閱 re：Post 中的為什麼我的 AWS Batch 任務卡在 RUNNABLE 狀態？。

建立時未標記 Spot 執行個體

自 2017 年 10 月 25 日起，支援 AWS Batch 運算資源的 Spot 執行個體標記。之前，Amazon EC2 Spot Fleet 角色的建議 IAM 受管政策 (`AmazonEC2SpotFleetRole`) 不包含在啟動時標記 Spot 執行個體的許可。新的建議 IAM 受管政策稱為 `AmazonEC2SpotFleetTaggingRole`。它支援在啟動時標記 Spot 執行個體。

若要修正建立時的 Spot 執行個體標記，請依照下列程序，將目前建議的 IAM 受管政策套用至 Amazon EC2 Spot Fleet 角色。如此一來，使用該角色建立的任何未來 Spot 執行個體都具有在建立執行個體標籤時套用它們的許可。

將目前的 IAM 受管政策套用至 Amazon EC2 Spot Fleet 角色

1. 前往 https://console.aws.amazon.com/iam/ 開啟 IAM 主控台。
2. 選擇角色，然後選擇您的 Amazon EC2 Spot 機群角色。
3. 選擇連接政策。
4. 選擇 AmazonEC2SpotFleetTaggingRole，然後選擇 Attach policy (連結政策)。
5. 再次選擇您的 Amazon EC2 Spot Fleet 角色，以移除先前的政策。
6. 選擇 AmazonEC2SpotFleetRole 政策右側的 x，然後選擇 Detach (分離)。

Spot 執行個體未縮減規模

AWS Batch 於 2021 年 3 月 10 日推出 AWSServiceRoleForBatch 服務連結角色。如果運算環境的 `serviceRole` 參數中未指定角色，則會使用此服務連結角色做為服務角色。不過，假設服務連結角色是在 EC2 Spot 運算環境中使用，但使用的 Spot 角色不包含 AmazonEC2SpotFleetTaggingRole 受管政策。然後，Spot 執行個體不會縮減規模。因此，您會收到錯誤訊息：「您無權執行此操作」。使用下列步驟來更新您在 `spotIamFleetRole` 參數中使用的 Spot 機群角色。如需詳細資訊，請參閱《IAM 使用者指南》中的使用服務連結角色和建立角色以委派許可給 AWS 服務。

主題

- 將 AmazonEC2SpotFleetTaggingRole 受管政策連接至 中的 Spot Fleet 角色 AWS 管理主控台
- 使用 將 AmazonEC2SpotFleetTaggingRole 受管政策連接至 Spot Fleet 角色 AWS CLI

將 AmazonEC2SpotFleetTaggingRole 受管政策連接至 中的 Spot Fleet 角色 AWS 管理主控台

將目前的 IAM 受管政策套用至您的 Amazon EC2 Spot Fleet 角色

1. 前往 https://console.aws.amazon.com/iam/ 開啟 IAM 主控台。
2. 選擇角色，然後選擇您的 Amazon EC2 Spot 機群角色。
3. 選擇連接政策。
4. 選擇 AmazonEC2SpotFleetTaggingRole，然後選擇 Attach policy (連結政策)。
5. 再次選擇您的 Amazon EC2 Spot Fleet 角色，以移除先前的政策。
6. 選擇 AmazonEC2SpotFleetRole 政策右側的 x，然後選擇 Detach (分離)。

使用 將 AmazonEC2SpotFleetTaggingRole 受管政策連接至 Spot Fleet 角色 AWS CLI

範例命令假設您的 Amazon EC2 Spot Fleet 角色名為 *AmazonEC2SpotFleetRole*。如果您的角色使用不同的名稱，請調整命令以符合。

將 AmazonEC2SpotFleetTaggingRole 受管政策連接至 Spot Fleet 角色

1. 若要將 AmazonEC2SpotFleetTaggingRole 受管 IAM 政策連接至您的 *AmazonEC2SpotFleetRole* 角色，請使用 執行下列命令 AWS CLI。

```
$ aws iam attach-role-policy \
    --policy-arn arn:aws:iam::aws:policy/service-role/AmazonEC2SpotFleetTaggingRole
 \
    --role-name AmazonEC2SpotFleetRole
```

2. 若要從 AmazonEC2SpotFleetRole 角色分離 *AmazonEC2SpotFleetRole* 受管 IAM 政策，請使用 執行下列命令 AWS CLI。

```
$ aws iam detach-role-policy \
    --policy-arn arn:aws:iam::aws:policy/service-role/AmazonEC2SpotFleetRole \
    --role-name AmazonEC2SpotFleetRole
```

無法擷取 Secrets Manager 秘密

如果您使用 AMI 搭配早於 1.16.0-1 版的 Amazon ECS 代理程式，則必須使用 Amazon ECS 代理程式組態變數ECS_ENABLE_AWSLOGS_EXECUTIONROLE_OVERRIDE=true來使用此功能。當您建立該執行個體時，您可以將其新增至./etc/ecs/ecs.config檔案到新的容器執行個體。或者，您可以將其新增至現有的執行個體。如果您將其新增至現有的執行個體，則必須在新增 ECS 代理程式之後重新啟動它。如需詳細資訊，請參閱《Amazon Elastic Container Service 開發人員指南》中的 Amazon ECS 容器代理程式組態。

無法覆寫任務定義資源需求

傳遞給 SubmitJob memoryvcpus的 containerOverrides 結構成員中指定的記憶體和 vCPU 覆寫無法覆寫任務定義中 resourceRequirements 結構中指定的記憶體和 vCPU 需求。

如果您嘗試覆寫這些資源需求，您可能會看到下列錯誤訊息：

「此值已在已棄用金鑰中提交，可能與任務定義的資源需求所提供的值衝突。」

若要修正此問題，請在 containerOverrides 的 resourceRequirements 成員中指定記憶體和 vCPU 需求。例如，如果您的記憶體和 vCPU 覆寫是在下列幾行中指定。

```
"containerOverrides": {
   "memory": 8192,
   "vcpus": 4
}
```

將它們變更為以下內容：

```
"containerOverrides": {
   "resourceRequirements": [
      {
         "type": "MEMORY",
         "value": "8192"
      },
      {
         "type": "VCPU",
         "value": "4"
      }
   ],
}
```

對任務定義中 containerProperties 物件中指定的記憶體和 vCPU 要求進行相同的變更。例如，如果您的記憶體和 vCPU 需求在下列幾行中指定。

```
{
   "containerProperties": {
      "memory": 4096,
      "vcpus": 2,
}
```

將它們變更為以下內容：

```
"containerProperties": {
   "resourceRequirements": [
      {
         "type": "MEMORY",
         "value": "4096"
      },
      {
```

```
            "type": "VCPU",
            "value": "2"
        }
    ],
}
```

更新desiredvCpus設定時的錯誤訊息

當您使用 AWS Batch API 更新所需的 vCPUs(desiredvCpus) 設定時，您會看到下列錯誤訊息。

```
Manually scaling down compute environment is not supported. Disconnecting job queues from compute environment will cause it to scale-down to minvCpus.
```

如果更新desiredvCpus的值小於目前desiredvCpus值，就會發生此問題。當您更新desiredvCpus值時，下列兩項都必須為 true：

- desiredvCpus 值必須介於 minvCpus和 maxvCpus值之間。
- 更新desiredvCpus的值必須大於或等於目前的desiredvCpus值。

AWS Batch 在 Amazon EKS 上

主題

- INVALID 運算環境
- AWS Batch Amazon EKS 任務上的 卡在 RUNNABLE 狀態
- AWS Batch Amazon EKS 任務上的 卡在 STARTING 狀態
- 確認 aws-auth ConfigMap 已正確設定
- RBAC 許可或繫結未正確設定

檢閱下列主題，尋找您在 Amazon Elastic Kubernetes Service AWS Batch 上使用 時可能遇到的常見問題的檢閱程序和潛在解決方案。

INVALID 運算環境

您可能未正確設定受管運算環境。如果您這麼做，運算環境會進入 INVALID 狀態，且無法接受要放置的任務。下列各節說明可能的原因，以及如何根據原因進行故障診斷。

不支援的Kubernetes版本

當您使用 CreateComputeEnvironment API 操作或 UpdateComputeEnvironmentAPI 操作來建立或更新運算環境時，您可能會看到類似以下內容的錯誤訊息。如果您在 中指定不支援的Kubernetes版本，就會發生此問題EC2Configuration。

```
At least one imageKubernetesVersion in EC2Configuration is not supported.
```

若要解決此問題，請刪除運算環境，然後使用支援的Kubernetes版本重新建立它。

您可以在 Amazon EKS 叢集上執行次要版本升級。例如，1.yy即使不支援次要版本，您也可以將叢集從 1.xx 升級至 。

不過，在主要版本更新INVALID之後，運算環境狀態可能會變更為 。例如，如果您執行從 1.xx升級至 的主要版本升級2.yy。如果 不支援主要版本 AWS Batch，您會看到類似以下的錯誤訊息。

```
reason=CLIENT_ERROR - ... EKS Cluster version [2.yy] is unsupported
```

若要解決此問題，請在使用 API 操作建立或更新運算環境時指定支援的Kubernetes版本。

AWS Batch Amazon EKS 上的 目前支援下列Kubernetes版本：

- 1.34
- 1.33
- 1.32
- 1.31
- 1.30
- 1.29

執行個體描述檔不存在

如果指定的執行個體描述檔不存在，Amazon EKS 運算環境 AWS Batch 上的 狀態會變更為 INVALID。您可以在類似以下內容的 statusReason 參數中看到錯誤集。

```
CLIENT_ERROR - Instance profile arn:aws:iam::...:instance-profile/<name> does not exist
```

若要解決此問題，請指定或建立運作中的執行個體描述檔。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS 節點 IAM 角色。

無效的Kubernetes命名空間

如果在 Amazon EKS AWS Batch 上無法驗證運算環境的命名空間，則運算環境狀態會變更為 INVALID。例如，如果命名空間不存在，可能會發生此問題。

您可以在類似以下內容的 statusReason 參數中看到錯誤訊息集。

```
CLIENT_ERROR - Unable to validate Kubernetes Namespace
```

如果以下任何一項成立，可能會發生此問題：

- CreateComputeEnvironment 呼叫中的Kubernetes命名空間字串不存在。如需詳細資訊，請參閱 CreateComputeEnvironment。
- 管理命名空間所需的角色型存取控制 (RBAC) 許可未正確設定。
- AWS Batch 無法存取 Amazon EKS Kubernetes API 伺服器端點。

若要解決此問題，請參閱 確認 aws-auth ConfigMap 已正確設定。如需詳細資訊，請參閱在 Amazon EKS AWS Batch 上開始使用。

已刪除的運算環境

假設您先刪除 Amazon EKS 叢集，再刪除連接到 Amazon EKS 運算環境 AWS Batch 的 。然後，運算環境狀態會變更為 INVALID。在此案例中，如果您以相同名稱重新建立 Amazon EKS 叢集，則運算環境無法正常運作。

若要解決此問題，請刪除並重新建立 Amazon EKS 運算環境 AWS Batch 上的 。

節點不會加入 Amazon EKS 叢集

AWS Batch 如果 Amazon EKS 上的 判斷並非所有節點都加入 Amazon EKS 叢集，則會縮減運算環境。在 AWS Batch Amazon EKS 上縮減運算環境時，運算環境狀態會變更為 INVALID。

> **Note**
>
> AWS Batch 不會立即變更運算環境狀態，因此您可以偵錯問題。

您可以在類似下列其中一項的 statusReason 參數中看到錯誤訊息集：

```
Your compute environment has been INVALIDATED and scaled down because
none of the instances joined the underlying ECS Cluster. Common issues
preventing instances joining are the following: VPC/Subnet configuration
preventing communication to ECS, incorrect Instance Profile policy
preventing authorization to ECS, or customized AMI or LaunchTemplate
configurations affecting ECS agent.

Your compute environment has been INVALIDATED and scaled down because
none of the nodes joined the underlying Amazon EKS Cluster. Common issues
preventing nodes joining are the following: networking configuration
preventing communication to Amazon EKS Cluster, incorrect Amazon EKS
Instance Profile or Kubernetes RBAC policy preventing authorization
to Amazon EKS Cluster, customized AMI or LaunchTemplate configurations
affecting Amazon EKS/Kubernetes node bootstrap.
```

使用預設 Amazon EKS AMI 時，此問題最常見的原因如下：

- 執行個體角色未正確設定。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS 節點 IAM 角色。
- 子網路未正確設定。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS VPC 和子網路需求和考量事項。
- 未正確設定安全群組。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS 安全群組需求和考量事項。

> **Note**
>
> 您也可以在個人運作狀態儀表板 (PHD) 中看到錯誤通知。

AWS Batch Amazon EKS 任務上的 卡在 **RUNNABLE** 狀態

當您使用 建立受管節點群組或節點群組時， `aws-authConfigMap`會自動建立並套用至您的叢集`eksctl`。一開始`aws-authConfigMap`會建立 ，以允許節點加入您的叢集。不過，您也可以使用`aws-authConfigMap`將角色型存取控制 (RBAC) 存取權新增至使用者和角色。

若要確認 `aws-auth ConfigMap` 已正確設定：

1. 在 中擷取映射的角色`aws-authConfigMap`：

```
$ kubectl get configmap -n kube-system aws-auth -o yaml
```

2. 確認 roleARN 已設定如下。

```
rolearn: arn:aws:iam::aws_account_number:role/AWSServiceRoleForBatch
```

> **Note**
>
> 您也可以檢閱 Amazon EKS 控制平面日誌。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS 控制平面記錄。

若要解決任務卡在 RUNNABLE 狀態的問題，建議您使用 kubectl重新套用資訊清單。如需詳細資訊，請參閱步驟 2：準備您的 Amazon EKS 叢集 AWS Batch。或者，您可以使用 kubectl手動編輯 aws-auth ConfigMap。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的啟用叢集的 IAM 使用者和角色存取權。

AWS Batch Amazon EKS 任務上的 卡在 **STARTING** 狀態

當 Pod 因來自 kubelet (pull、exec、 和 attach) ContainerCreating的任何長時間執行請求而卡在 Pod PENDING上時log，任務可能會保持 STARTING 狀態，直到 Pod 啟動問題解決或任務終止為止。在下列合格案例中， AWS Batch 會代表您終止任務，否則任務必須使用 TerminateJob API 手動終止。

若要驗證任務可能卡在 中的原因STARTING，請使用 教學課程：將執行中的任務映射至 Pod 和節點 尋找 podName，並描述 Pod：

```
% kubectl describe pod aws-batch.000c8190-87df-31e7-8819-176fe017a24a -n my-aws-batch-
namespace
Name:             aws-batch.000c8190-87df-31e7-8819-176fe017a24a
Namespace:        my-aws-batch-namespace
...
Containers:
  default:
...
    State:          Waiting
      Reason:       ContainerCreating
    Ready:          False
...
Conditions:
```

```
  Type                        Status
  PodReadyToStartContainers   False
  Initialized                 True
  Ready                       False
  ContainersReady             False
  PodScheduled                True
...
Events:
  Type     Reason       Age    From     Message
  ----     ------       ----   ----     -------
  Warning  FailedMount  2m32s  kubelet  Unable to attach or mount volumes: ...
```

請考慮將您的 EKS 叢集設定為將控制平面日誌傳送至 CloudWatch Logs，以獲得完整可見性。

案例：持續磁碟區宣告連接或掛載失敗

使用持久性磁碟區宣告的任務，其中磁碟區無法連接或掛載是終止的候選項目。這可能是未正確設定任務定義的結果。如需詳細資訊，請參閱在 Amazon EKS 資源上建立單一節點任務定義。

確認 aws-auth ConfigMap 已正確設定

若要確認 aws-auth ConfigMap 已正確設定：

1. 在 中擷取映射的角色aws-authConfigMap。

   ```
   $ kubectl get configmap -n kube-system aws-auth -o yaml
   ```

2. 確認 roleARN 已設定如下。

   ```
   rolearn: arn:aws:iam::aws_account_number:role/AWSServiceRoleForBatch
   ```

   > **Note**
   >
   > 路徑aws-service-role/batch.amazonaws.com/已從服務連結角色的 ARN 中移除。這是因為aws-auth組態映射發生問題。如需詳細資訊，請參閱 中的具有路徑的角色在路徑包含在其 ARN 中時無法運作aws-authconfigmap。

> **Note**
>
> 您也可以檢閱 Amazon EKS 控制平面日誌。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的 Amazon EKS 控制平面記錄。

若要解決任務卡在 RUNNABLE 狀態的問題，建議您使用 kubectl重新套用資訊清單。如需詳細資訊，請參閱步驟 2：準備您的 Amazon EKS 叢集 AWS Batch。或者，您可以使用 kubectl手動編輯 aws-auth ConfigMap。如需詳細資訊，請參閱《Amazon EKS 使用者指南》中的啟用叢集的 IAM 使用者和角色存取權。

RBAC 許可或繫結未正確設定

如果您遇到任何 RBAC 許可或繫結問題，請確認aws-batchKubernetes角色可以存取Kubernetes命名空間：

```
$ kubectl get namespace namespace --as=aws-batch
```

```
$ kubectl auth can-i get ns --as=aws-batch
```

您也可以使用 **kubectl describe**命令來檢視叢集角色或Kubernetes命名空間的授權。

```
$ kubectl describe clusterrole aws-batch-cluster-role
```

下列為範例輸出。

```
Name:         aws-batch-cluster-role
Labels:       <none>
Annotations:  <none>
PolicyRule:
  Resources                                  Non-Resource URLs  Resource Names
 Verbs
  ---------                                  -----------------  --------------
 -----
  configmaps                                 []                 []
 [get list watch]
  nodes                                      []                 []
 [get list watch]
```

```
  pods                                              []                 []
 [get list watch]
  daemonsets.apps                                   []                 []
 [get list watch]
  deployments.apps                                  []                 []
 [get list watch]
  replicasets.apps                                  []                 []
 [get list watch]
  statefulsets.apps                                 []                 []
 [get list watch]
  clusterrolebindings.rbac.authorization.k8s.io     []                 []
 [get list]
  clusterroles.rbac.authorization.k8s.io            []                 []
 [get list]
  namespaces                                        []                 []
 [get]
  events                                            []                 []
 [list]
```

```
$ kubectl describe role aws-batch-compute-environment-role -n my-aws-batch-namespace
```

下列為範例輸出。

```
Name:         aws-batch-compute-environment-role
Labels:       <none>
Annotations:  <none>
PolicyRule:
  Resources                                 Non-Resource URLs  Resource Names  Verbs
  ---------                                 -----------------  --------------  -----
  pods                                      []                 []              [create
 get list watch delete patch]
  serviceaccounts                           []                 []              [get list]
  rolebindings.rbac.authorization.k8s.io    []                 []              [get list]
  roles.rbac.authorization.k8s.io           []                 []              [get list]
```

若要解決此問題，請重新套用 RBAC 許可和`rolebinding`命令。如需詳細資訊，請參閱步驟 2：準備您的 Amazon EKS 叢集 AWS Batch。

資源： AWS Batch 服務配額

下表提供無法變更 AWS Batch 的 服務配額。每個配額都是區域特定的。

資源	配額
任務佇列的數量上限 如需詳細資訊，請參閱任務佇列。	50
跨 Amazon ECS 和 Amazon EKS 的運算環境數量上限。如需詳細資訊，請參閱的運算環境 AWS Batch。	50
每個 Amazon EKS 叢集的運算環境數目上限。	5
每個任務佇列的運算環境數目上限	3
任務相依性上限數量	20
最大任務定義大小 (適用於 RegisterJobDefinition API 操作)	24 KiB
最大任務承載大小 (適用於 SubmitJob API 操作)	30 KiB
陣列任務的最大陣列大小	10000
SUBMITTED 狀態任務的最大數量	1000000
SubmitJob 每個 帳戶操作的每秒交易數上限 (TPS)	50
消耗性資源的數量上限	5 萬
服務環境的最大數量。如需詳細資訊，請參閱服務環境。	50
每個任務佇列的服務環境數量上限	1
SubmitServiceJob 請求的大小上限	30 KiB
任務服務請求承載大小上限 （ 適用於 SubmitServiceJob API 操作 ）	10 KiB
SubmitServiceJob 每個 帳戶操作的每秒交易數上限 (TPS)	5
服務任務重試策略的嘗試次數上限	10

視您的使用方式而定 AWS Batch，可能會套用額外的配額。若要了解 Amazon EC2 配額，請參閱《》中的 Amazon EC2 Service QuotasAWS 一般參考。如需 Amazon ECS 配額的詳細資訊，請參閱《》中的 Amazon ECS Service QuotasAWS 一般參考。如需 Amazon EKS 配額的詳細資訊，請參閱《》中的 Amazon EKS Service QuotasAWS 一般參考。

文件歷史紀錄

下表說明 文件自初始發行以來的重要變更 AWS Batch。我們也會經常更新文件，以處理您傳送給我們的意見回饋。

變更	描述	日期
新增 default_x86_64 和 default_arm64	新增`default_arm64` 允許執行個體類型的 `default_x86_64` 和 。	2025 年 8 月 15 日
新增服務環境和服務任務	新增 AWS Batch 搭配 SageMaker AI 使用 的服務環境和服務任務。	2025 年 7 月 30 日
新增 AWSServiceRoleForAWSBatchWithSagemaker 和 AWSBatchServiceRolePolicyForSageMaker	新增了新的 AWS 服務連結角色`AWSServiceRoleForAWSBatchWithSagemaker` 和管理政策`AWSBatchServiceRolePolicyForSageMaker` ，AWS Batch 允許 代表您管理 SageMaker AI。	2025 年 7 月 30 日
新增對 EKS AL 2023 AMIs支援	如何從 EKS AL2 升級到 EKS AL2023	2025 年 6 月 24 日
新增 FireLens 和 ECS Exec 命令的支援	新增 FireLens 和 ECS Exec 命令的支援。	2025 年 4 月 15 日
新增 的資源感知排程支援 AWS Batch	新增支援 AWS Batch 適用於 Amazon Elastic Container Service、Amazon Elastic Kubernetes Service 和 的 資源感知排程 AWS Fargate。	2025 年 2 月 27 日

更新 AWS Batch 支援的 Amazon EKS 版本	已更新 AWS Batch 支援 的 Amazon EKS 版本，以移除 1.22 版。	2024 年 3 月 11 日
更新 AWS Batch 支援的 Amazon EKS 版本	已更新 AWS Batch 支援的 Amazon EKS 版本，以包含 1.29 版。	2024 年 2 月 29 日
自動化任務重試	更正程式碼範例。	2024 年 2 月 29 日
新增 的多容器任務支援 AWS Batch	新增對 AWS Batch for Amazon Elastic Container Service、Amazon Elastic Kubernetes Service 和 的多容器任務的支援 AWS Fargate。	2024 年 2 月 28 日
更新 AWS Batch 支援的 Amazon EKS 版本	已更新 AWS Batch 支援的 Amazon EKS 版本，以包含 1.28 版	2024 年 1 月 27 日
已更新 BatchServiceRolePolicy和 AWSBatchServiceRole	BatchServiceRolePolicy 更新以新增描述 Spot Fleet 請求歷史記錄和 Amazon EC2 Auto Scaling 活動的支援。 AWSBatchServiceRole 更新以新增陳述式 IDs，將 AWS Batch 許可授予 `ec2:DescribeSpotFleetRequestHistory` 和 `autoscaling:DescribeScalingActivities` 。	2023 年 12 月 5 日
AWS Batch 在 Amazon EKS 上	AWS Batch 新增對在 Amazon EKS 叢集上執行任務的支援。	2022 年 10 月 25 日

的跨服務混淆代理人預防 AWS Batch	AWS Batch 現在提供混淆代理人安全問題的解決方法，當不同的實體強制實體執行動作時，就會發生此問題。	2022 年 6 月 6 日
介面 VPC 端點 (AWS PrivateLink)	新增設定介面 VPC 端點的支援 AWS PrivateLink。這表示您可以在 VPC 與 之間建立私有連線， AWS Batch 而不需要透過 NAT 執行個體、VPN 連線或 進行存取 Direct Connect。	2022 年 4 月 15 日
增強型運算環境更新	AWS Batch 增強支援運算環境的更新。	2022 年 4 月 14 日
AWS 受管政策更新 - 更新現有政策	AWS Batch 已更新現有的 受管政策。	2021 年 12 月 6 日
公平共用排程	AWS Batch 新增將排程政策新增至任務佇列的支援。	2021 年 11 月 9 日
Amazon EFS	AWS Batch 新增將 Amazon EFS 檔案系統新增至任務定義的支援。	2021 年 4 月 1 日
新增服務連結角色	AWS Batch 新增 AWSServiceRoleForBatch 服務連結角色。	2021 年 3 月 10 日
AWS Fargate 支援	AWS Batch 新增對 Fargate 資源上執行任務的支援。	2020 年 12 月 3 日
資源標記	AWS Batch 新增支援將中繼資料標籤新增至您的運算環境、任務定義、任務佇列和任務。	2020 年 10 月 7 日
秘密	AWS Batch 新增支援將秘密傳遞至 任務。	2020 年 10 月 1 日

日誌	AWS Batch 新增了為任務指定其他日誌驅動程式的支援。	2020 年 10 月 1 日
配置策略	AWS Batch 新增對多個策略的支援，以選擇執行個體類型。	2019 年 10 月 16 日
EFA 支援	AWS Batch 新增對 Elastic Fabric Adapter (EFA) 裝置的支援。	2019 年 8 月 2 日
GPU 排程	AWS Batch 新增 GPU 排程。透過此功能，您可以指定每個任務所需的 GPUs 數量，並相應 AWS Batch 地擴展執行個體。	2019 年 4 月 4 日
多節點平行任務	AWS Batch 新增對多節點平行任務的支援。您可以使用此功能執行跨越多個 Amazon EC2 執行個體的單一任務。	2018 年 11 月 19 日
資源層級許可	AWS Batch 支援數個 API 操作的資源層級許可。	2018 年 11 月 12 日
Amazon EC2 啟動範本支援	AWS Batch 新增對搭配運算環境使用啟動範本的支援。	2018 年 11 月 12 日
AWS Batch 任務逾時	AWS Batch 新增對任務逾時的支援。透過此支援，您可以設定任務的特定逾時持續時間，以便在任務執行時間超過預期時， AWS Batch 終止任務。	2018 年 4 月 5 日
AWS Batch 任務做為 EventBridge 目標	AWS Batch 任務以 EventBridge 目標的形式提供。透過建立簡單的規則，您可以比對事件並提交 AWS Batch 任務以回應它們。	2018 年 3 月 1 日

的 CloudTrail 稽核 AWS Batch	CloudTrail 可以稽核對 AWS Batch API 動作發出的呼叫。	2018 年 1 月 10 日
陣列任務	AWS Batch 新增對陣列任務的支援。您可以使用陣列任務進行參數掃描和 Monte Carlo 工作負載。	2017 年 11 月 28 日
擴展 AWS Batch 標記	AWS Batch 展開對標記函數的支援。您可以使用此函數來指定在受管運算環境中啟動之 Amazon EC2 Spot 執行個體的標籤。	2017 年 10 月 26 日
AWS Batch EventBridge 的事件串流	AWS Batch 新增 EventBridge 的事件串流。您可以使用 AWS Batch 事件串流來接收有關提交至任務佇列之任務狀態的近乎即時通知。	2017 年 10 月 24 日
自動化任務重試	AWS Batch 新增對任務重試的支援。透過此更新，您可以將重試策略套用至您的任務和任務定義，以便在任務失敗時自動重試。	2017 年 3 月 28 日
AWS Batch 一般可用性	AWS Batch 已推出 ，設計為讓您在 上執行批次運算工作負載的方法 AWS 雲端。	2017 年 1 月 5 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。