



使用者指南

AWS Clean Rooms



AWS Clean Rooms: 使用者指南

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

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

Table of Contents

什麼是 AWS Clean Rooms ?	1
您是第一次 AWS Clean Rooms 使用嗎 ?	1
AWS Clean Rooms 運作方式	2
相關服務	2
AWS 服務	2
第三方服務	3
存取 AWS Clean Rooms	4
的定價 AWS Clean Rooms	4
的帳單 AWS Clean Rooms	4
分析規則	5
分析規則類型	5
彙總分析規則	7
列出分析規則	23
自訂分析規則	31
ID 映射表分析規則	37
AWS Clean Rooms 差異隱私權	46
差異隱私權	47
中的差異隱私權如何 AWS Clean Rooms 運作	47
差異性隱私權政策	48
SQL 功能	49
SQL 查詢秘訣和範例	90
限制	91
AWS 無塵室 ML	93
AWS Clean Rooms ML 術語	93
AWS Clean Rooms ML 如何與 AWS 模型搭配使用	94
AWS Clean Rooms ML 如何與自訂模型搭配使用	95
AWS Clean Rooms ML 中的模型	96
Clean Rooms ML 中的自訂模型	101
密碼編譯運算	108
考量事項	109
支援的檔案和資料類型	111
欄位名稱	116
資料欄類型	116
Parameters	118

選用旗標	123
使用 C3R 的查詢	125
指導方針	126
設定 AWS Clean Rooms	150
註冊 AWS	150
設定 的服務角色 AWS Clean Rooms	150
建立管理員使用者	151
為協同合作成員建立 IAM 角色	151
建立服務角色以從 Amazon S3 讀取資料	152
建立服務角色以從 Amazon Athena 讀取資料	156
建立服務角色以從 Snowflake 讀取資料	159
建立服務角色以從 S3 儲存貯體讀取程式碼 (PySpark 分析範本角色)	163
建立服務角色以寫入 PySpark 任務的結果	165
建立服務角色以接收結果	168
設定 AWS Clean Rooms ML 的服務角色	172
設定類似模型的服務角色	172
設定自訂建模的服務角色	186
協作和成員資格	200
建立協同合作	200
建立查詢的協同合作	201
為查詢和任務建立協同合作	209
為 ML 建模建立協作	218
建立成員資格並加入協同合作	225
編輯協同合作	230
編輯協同合作名稱和描述	231
更新協同合作分析引擎	231
關閉日誌儲存	231
編輯協同合作日誌設定	232
編輯協同合作標籤	233
編輯成員資格標籤	234
新增成員	234
編輯現有的成員功能	234
編輯協同合作自動核准設定	234
編輯相關聯的資料表標籤	235
編輯分析範本標籤	235
編輯差異隱私權政策標籤	236

變更請求	236
將新成員新增至協同合作	237
更新現有的成員能力	238
編輯協同合作自動核准設定	238
刪除協同合作	239
檢視協同合作	240
邀請成員	240
監控成員	241
新增成員	241
移除成員	242
離開協同合作	243
資料表	245
資料格式	246
PySpark 任務支援的資料格式	246
SQL 查詢支援的資料格式	246
支援的資料類型	247
的檔案壓縮類型 AWS Clean Rooms	247
的伺服器端加密 AWS Clean Rooms	248
Apache Iceberg 資料表	248
Iceberg 資料表支援的資料類型	249
準備資料表	250
在 Amazon S3 中準備資料表	250
在 Amazon Athena 中準備資料表	253
在 Snowflake 中準備資料表	255
準備加密資料表	257
步驟 1：完成先決條件	258
步驟 2：下載 C3R 加密用戶端	258
步驟 3：（選用）在 C3R 加密用戶端中檢視可用的命令	259
步驟 4：產生表格式檔案的加密結構描述	259
步驟 5：建立共用私密金鑰	265
步驟 6：將共用私密金鑰存放在 環境變數中	266
步驟 7：加密資料	267
步驟 8：驗證資料加密	268
（選用）建立結構描述（進階使用者）	269
解密資料表	278
設定的資料表	280

建立設定的資料表	281
Amazon S3 資料來源	281
Amazon Athena 資料來源	284
Snowflake 資料來源	287
將分析規則新增至設定的資料表	290
將彙總分析規則新增至資料表（引導流程）	291
將清單分析規則新增至資料表（引導流程）	295
將自訂分析規則新增至資料表（引導流程）	296
將分析規則新增至資料表（JSON 編輯器）	299
後續步驟	300
將已設定的資料表與協作建立關聯	301
資料存取預算	301
關聯已設定的資料表	301
設定資料存取預算	306
檢視資料存取預算	307
將資料存取預算新增至現有的關聯資料表	307
編輯資料存取預算	308
刪除資料存取預算	310
將協同合作分析規則新增至設定的資料表	311
設定差異隱私權政策（選用）	312
檢視差異隱私權用量日誌	313
編輯差異隱私權政策	313
刪除差異隱私權政策	314
檢視計算的差異隱私權參數	315
檢視資料表和分析規則	316
編輯設定的資料表	316
編輯設定的資料表標籤	317
編輯設定的資料表分析規則	317
刪除設定的資料表分析規則	318
設定的資料表不允許的資料欄	318
編輯設定的資料表關聯	322
取消已設定資料表的關聯	322
AWS Entity Resolution 在 中 AWS Clean Rooms	324
ID 命名空間	325
建立和關聯新的 ID 命名空間	325
關聯現有的 ID 命名空間	327

編輯 ID 命名空間關聯	329
取消 ID 命名空間關聯的關聯	330
ID 映射表	331
建立和填入新的 ID 映射表	332
填入現有的 ID 映射表	338
編輯 ID 映射表	339
刪除 ID 映射表	339
分析範本	341
SQL 分析範本	341
建立 SQL 分析範本	342
檢閱 SQL 分析範本	344
PySpark 分析範本	345
安全	346
限制	346
最佳實務	347
建立使用者指令碼	348
在 PySpark 分析範本中使用參數	352
建立虛擬環境（選用）	356
在 S3 中存放使用者指令碼和虛擬環境	357
建立 PySpark 分析範本	358
檢閱 PySpark 分析範本	362
故障診斷 PySpark 分析範本	364
對程式碼進行故障診斷	365
分析範本任務未啟動	365
分析範本任務開始，但在處理期間失敗	367
虛擬環境設定失敗	368
分析	370
執行 SQL 查詢	370
先決條件	371
SQL 查詢的 Spark 屬性組態	372
查詢設定的資料表	372
查詢 ID 映射表	387
使用分析範本查詢設定的資料表	389
使用分析建置器查詢	401
檢視差異隱私權的影響	406
檢視近期查詢	406

檢視查詢詳細資訊	407
執行 PySpark 任務	407
使用分析範本執行任務	408
檢視最近的任務	410
檢視任務詳細資訊	411
分析結果	412
接收查詢結果	412
接收任務結果	414
編輯查詢結果設定的預設值	414
編輯任務結果設定的預設值	416
在其他 中使用查詢輸出 AWS 服務	417
自訂建模	418
隱私權增強的合成資料集產生	419
合成資料產生的考量事項	420
建立和加入協同合作	422
建立機器學習的協同合作	423
加入協同合作	426
貢獻訓練資料	427
設定模型演算法	432
關聯設定的模型演算法	434
建立 ML 輸入通道	439
建立訓練過的模型	453
使用增量訓練	456
使用分散式訓練	460
匯出模型成品	463
在訓練過的模型上執行推論	465
訓練資料提供者的 ML 建模	467
匯入訓練資料	467
建立外觀模型	469
設定外觀模型	469
關聯已設定的外觀模型	470
更新已設定的類似模型	471
種子資料提供者的 ML 建模	473
建立類似樣貌的客群	473
匯出類似樣貌的客群	485
疑難排解	486

查詢參考的一或多個資料表無法由其相關聯的服務角色存取。資料表/角色擁有者必須授予 資料表的服務角色存取權。	486
其中一個基礎資料集具有不支援的檔案格式。	486
針對 使用密碼編譯運算時，查詢結果不如預期Clean Rooms。	487
AWS Clean Rooms Spark SQL：缺少分割區資料	487
安全	488
資料保護	488
靜態加密	489
傳輸中加密	490
加密基礎資料	490
金鑰政策	490
使用服務連結角色	496
的服務連結角色許可 AWS Clean Rooms	497
為 建立服務連結角色 AWS Clean Rooms	497
編輯 的服務連結角色 AWS Clean Rooms	497
刪除 的服務連結角色 AWS Clean Rooms	498
AWS Clean Rooms 服務連結角色支援的 區域	498
資料保留	498
最佳實務	499
的最佳實務 AWS Clean Rooms	499
在 中使用分析規則的最佳實務 AWS Clean Rooms	499
身分和存取權管理	500
目標對象	501
使用身分驗證	501
使用政策管理存取權	503
AWS Clean Rooms 如何使用 IAM	504
身分型政策範例	509
AWS 受管政策	511
疑難排解	518
預防跨服務混淆代理人	519
AWS Clean Rooms ML 的 IAM 行為	521
Clean Rooms ML 自訂模型的 IAM 行為	524
法規遵循驗證	526
恢復能力	526
基礎設施安全性	527
網路安全	527

AWS PrivateLink	527
考量事項	528
建立介面端點	528
監控	529
分析登入 AWS Clean Rooms	529
接收查詢和任務日誌	530
查詢和任務日誌的建議動作	531
在 中使用 CloudWatch 進行詳細監控AWS Clean Rooms	531
指標類型	532
指標維度	533
誰可以存取指標	534
定價	534
CloudTrail 日誌	534
AWS Clean Rooms CloudTrail 中的資訊	534
了解 AWS Clean Rooms 日誌檔案項目	535
Example AWS Clean Rooms CloudTrail 事件	535
整合 AWS Clean Rooms 到 EDAs	539
AWS Clean Rooms 事件	540
路由 AWS Clean Rooms 事件	543
事件詳細參照	544
成本分配標記	553
成本分配的可標記資源	553
AWS CloudFormation 資源	555
AWS Clean Rooms 和 CloudFormation 範本	555
進一步了解 CloudFormation	555
配額	556
AWS Clean Rooms 配額	556
AWS Clean Rooms 資源參數限制	567
AWS Clean Rooms ML 配額	568
Clean Rooms ML API 限流配額	587
文件歷史紀錄	592
詞彙表	604
彙總分析規則	604
分析規則	604
分析範本	604
C3R 加密用戶端	604

純文字欄	605
協作	605
協作建立者	605
設定的資料表	605
自訂分析規則	606
解密	606
差異隱私權	606
加密	606
指紋資料欄	606
ID 映射工作流程方法	606
ID 映射表	607
ID 映射表分析規則	607
ID 映射工作流程	607
ID 命名空間	607
ID 命名空間關聯	607
任務	607
列出分析規則	608
Lookalike 模型	608
Lookalike 區段	608
成員	608
可查詢的成員	608
可執行查詢和任務的成員	609
可接收結果的成員	609
成員支付查詢運算成本	609
成員支付查詢和任務運算成本	609
成員資格	610
密封資料欄	610
種子資料	610
Spark 分析引擎	610
Query	610
.....	dcxi

什麼是 AWS Clean Rooms ？

AWS Clean Rooms 可協助您和合作夥伴分析和協作您的集合資料集，以獲得新的洞見，而無需向彼此揭露基礎資料。AWS Clean Rooms 是一個安全的協作工作區，您可以在幾分鐘內建立自己的無塵室，只需幾個步驟即可分析集合資料集。您可以選擇要與其合作的合作夥伴、選取其資料集，並為這些合作夥伴設定隱私權增強控制。

透過 AWS Clean Rooms，您可以與數千家已使用 的公司協作 AWS。協同合作不需要將資料移出 AWS 或載入至其他雲端服務供應商。當您執行查詢或任務時，會從該資料的原始位置 AWS Clean Rooms 讀取資料，並套用內建分析規則，以協助您維持對該資料的控制。

AWS Clean Rooms 提供您可以設定的內建資料存取控制和稽核支援控制。這些控制項包括：

- 限制 SQL 查詢並提供輸出限制的[分析規則](#)。
- [密碼編譯運算 Clean Rooms](#)可保持資料加密，即使處理查詢也一樣，以符合嚴格的資料處理政策。
- [分析日誌](#)以檢閱 中的查詢和任務 AWS Clean Rooms，並協助支援稽核。
- 防止使用者識別嘗試的不同[隱私權](#)。AWS Clean Rooms 差異隱私權是一種全受管功能，可透過數學支援的技術和直覺式控制來保護使用者的隱私權，只需幾個步驟即可套用。
- [AWS Clean Rooms ML](#) 允許雙方在其資料中識別類似的使用者，而無需彼此共用其資料。第一方會從訓練資料建立和設定類似模型。然後，種子資料會帶入協同合作，以建立類似訓練資料的外觀客群。

以下影片會詳細說明 AWS Clean Rooms。

[AWS Clean Rooms](#)

您是第一次 AWS Clean Rooms 使用嗎？

如果您是第一次使用 AWS Clean Rooms，建議您先閱讀以下章節：

- [AWS Clean Rooms 運作方式](#)
- [存取 AWS Clean Rooms](#)
- [設定 AWS Clean Rooms](#)
- [AWS Clean Rooms 詞彙表](#)

AWS Clean Rooms 運作方式

在 中 AWS Clean Rooms，您可以建立協同合作並新增要邀請 AWS 帳戶 的，或建立成員資格來加入您獲邀請的協同合作。然後，您可以連結使用案例所需的資料資源：為事件資料設定的資料表、為 ML 建模設定的模型，或為實體解析設定的 ID 命名空間。您可以選擇建立或核准分析範本，以預先同意您希望在協同合作中允許的確切查詢和任務。最後，您可以透過在設定的資料表上執行 SQL 查詢或 PySpark 任務、在 ID 映射資料表中執行實體解析，或使用 ML 建模來產生外觀相似的受眾客群來分析關節資料。

下圖顯示 AWS Clean Rooms 的運作方式。

相關服務

AWS 服務

下列 AWS 服務 與 相關 AWS Clean Rooms：

- Amazon Athena

協作成員可以將他們帶入的資料儲存 AWS Clean Rooms 為 Amazon Athena 中的 AWS Glue Data Catalog 檢視。如需詳細資訊，請參閱下列主題：

如需詳細資訊，請參閱下列主題：

[為 中的查詢準備資料表 AWS Clean Rooms](#)

[建立設定的資料表 – Amazon Athena 資料來源](#)

《[Amazon Athena 使用者指南](#)》中的[什麼是 Amazon Athena](#)？

- CloudFormation

在 中建立下列資源 CloudFormation：協同合作、設定的資料表、設定的資料表關聯和成員資格

如需詳細資訊，請參閱[使用 建立 AWS Clean Rooms 資源 AWS CloudFormation](#)。

- AWS CloudTrail

使用 AWS Clean Rooms 搭配 CloudTrail 日誌來增強活動分析 AWS 服務。

如需詳細資訊，請參閱[使用 記錄 AWS Clean Rooms API 呼叫 AWS CloudTrail](#)。

- AWS Entity Resolution

使用 AWS Clean Rooms 搭配 AWS Entity Resolution 來執行實體解析。

如需詳細資訊，請參閱[AWS Entity Resolution 在 AWS Clean Rooms](#)。

- AWS Glue

協作成員可以從 Amazon S3 中的資料建立 AWS Glue 資料表，以便在 AWS Clean Rooms 中使用。

如需詳細資訊，請參閱下列主題：

[為中的查詢準備資料表 AWS Clean Rooms](#)

《AWS Glue 開發人員指南》中的[什麼是 AWS Glue ?](#)

- Amazon Simple Storage Service (Amazon S3)

協同合作成員可以存放他們在 Amazon S3 AWS Clean Rooms 中帶入的資料。

如需詳細資訊，請參閱下列主題：

[為中的查詢準備資料表 AWS Clean Rooms](#)

[建立設定的資料表 – Amazon S3 資料來源](#)

《Amazon Simple Storage Service 使用者指南》中的[什麼是 Amazon S3 ?](#)

- AWS Secrets Manager

協同合作成員可以建立秘密，以存取和讀取存放在 Snowflake 中的資料。

如需詳細資訊，請參閱下列主題：

[建立服務角色以從 Snowflake 讀取資料](#)

[為中的查詢準備資料表 AWS Clean Rooms](#)

《AWS Secrets Manager 使用者指南》中的[什麼是 AWS Secrets Manager ?](#)

第三方服務

下列第三方服務與相關 AWS Clean Rooms：

- Snowflake

協同合作成員可以將他們帶入 AWS Clean Rooms Snowflake 倉儲的資料存放。

如需詳細資訊，請參閱下列主題：

[為 中的查詢準備資料表 AWS Clean Rooms](#)

[建立設定的資料表 – Snowflake 資料來源](#)

存取 AWS Clean Rooms

您可以透過 AWS Clean Rooms 下列選項存取：

- 直接透過 AWS Clean Rooms 主控台前往 <https://console.aws.amazon.com/cleanrooms/>。
- 透過 AWS Clean Rooms API 以程式設計方式進行。如需詳細資訊，請參閱 [AWS Clean Rooms API 參考](#)。

的定價 AWS Clean Rooms

如需定價資訊，請參閱 [AWS Clean Rooms 定價](#)。

Note

對於與存放在 Snowflake 中的資料相關聯的協同合作成員，每次執行使用存放在這些位置的資料的查詢時，都會由其各自的資料倉儲供應商或雲端供應商向您收取資料輸出和運算的費用。

的帳單 AWS Clean Rooms

AWS Clean Rooms 讓協同合作建立者能夠指定哪些成員在協同合作中支付查詢或任務運算成本。

在大多數情況下，[可以查詢的成員](#)和[支付查詢運算成本的成員](#)是相同的。不過，如果可以查詢的成員和支付查詢運算成本的成員不同，則當可以查詢的成員根據自己的成員資源執行查詢時，會向支付查詢運算成本的成員的成員成員資源收費。

支付查詢運算成本的成員不會在其 CloudTrail 事件歷史記錄中看到正在執行查詢的任何事件，因為付款人不是執行查詢的人員，也不是執行查詢的資源擁有者。不過，付款人會看到其成員資源針對可在協同合作中執行查詢之成員執行的所有查詢所產生的費用。

如需如何建立協同合作並設定成員支付查詢運算成本的詳細資訊，請參閱 [建立協同合作](#)。

中的分析規則 AWS Clean Rooms

協作成員必須設定分析規則，才能在 中使用資料表 AWS Clean Rooms 進行協作分析。

分析規則是每個資料擁有者在已設定的資料表上設定的隱私權增強控制。分析規則決定如何分析設定的資料表。

分析規則是已設定資料表（帳戶層級資源）上的帳戶層級控制項，並在已設定資料表相關聯的任何協同合作中強制執行。如果未設定分析規則，則設定的資料表可以與協同合作相關聯，但無法查詢。查詢只能參考具有相同分析規則類型的已設定資料表。

若要設定分析規則，請先選取分析類型，然後指定分析規則。對於這兩個步驟，您應該考慮要啟用的使用案例，以及如何保護基礎資料。

AWS Clean Rooms 在查詢中參考的所有已設定資料表之間強制執行更嚴格的控制。

下列範例說明限制性控制。

Example 限制性控制：輸出限制

- 協作者 A 的識別符資料欄具有 100 的輸出限制。
- 協作者 B 在識別符資料欄上具有 150 的輸出限制。

參考兩個已設定資料表的彙總查詢在輸出列中至少需要 150 個不同的識別符值，才能顯示在查詢輸出中。查詢輸出不會指出由於輸出限制而移除結果。

Example 限制性控制：分析範本未核准

- 協作者 A 已允許具有查詢的分析範本，該查詢參考其自訂分析規則中來自協作者 A 和協作者 B 的已設定資料表。
- 協作者 B 不允許分析範本。

由於協作者 B 不允許分析範本，因此可以查詢的成員無法執行該分析範本。

分析規則類型

分析規則有三種類型：[彙總](#)、[列出](#)和[自訂](#)。下表比較分析規則類型。每種類型都有單獨的區段，描述指定分析規則。

Note

有一個稱為 ID 映射表分析規則的分析規則類型。不過，此分析規則是由 管理 AWS Clean Rooms，無法修改。如需詳細資訊，請參閱[ID 映射表分析規則](#)。

下列各節說明每個分析規則類型的支援使用案例和控制項。

支援的使用案例

下表顯示每個分析規則類型的支援使用案例比較摘要。

使用案例	彙總	清單	Custom (自訂)
支援的分析	使用 COUNT、SUM 和 AVG 函數沿著選用維度彙總統計資料的查詢	輸出多個資料表之間重疊之資料列層級清單的查詢	只要分析範本或分析建立者已檢閱並允許，任何自訂分析
常見使用案例	區段分析、測量、屬性	擴充、區段建置	第一次接觸歸因、增量分析、受眾探索
SQL 建構	JOIN 陳述式：INNER JOIN 彙總函數：COUNT/COUNT DISTINCT、SUM/SUM DISTINCT 和 AVG 純量函數：有限子集	JOIN 陳述式：INNER JOIN - 純量函數：無	大多數 SQL 函數和 SQL 建構可透過 SELECT 命令使用
子查詢和通用資料表表達式 (CTEs)	否	否	是
分析範本	否	否	是

支援的控制項

下表顯示每個分析規則類型如何保護基礎資料的比較摘要。

控制項	彙總	清單	Custom (自訂)
控制機制	控制如何在查詢中使用資料表中的資料 (例如，允許資料欄 hashed_email 的 COUNT 和 SUM。)	控制如何在查詢中使用資料表中的資料 (例如，僅允許將資料欄 hashed_email 用於聯結。)	控制允許在資料表上執行的查詢 (例如，僅允許分析範本「自訂查詢 1」中定義的查詢。)
內建隱私權增強技術	- 盲法比對 - 需要彙總 - 最小彙總閾值 $\geq$ 2 預先定義的查詢結構	- 盲法比對 - 需要重疊 - 預先定義的查詢結構 - 允許的額外分析	- 差異隱私權 - 不允許的輸出資料欄
檢閱查詢，然後才能執行	否	否	是，使用分析範本

如需 中可用分析規則的詳細資訊 AWS Clean Rooms，請參閱下列主題。

- [彙總分析規則](#)
- [列出分析規則](#)
- [中的自訂分析規則 AWS Clean Rooms](#)

彙總分析規則

在 中 AWS Clean Rooms，彙總分析規則會沿著選用維度使用 COUNT、SUM 和/或 AVG 函數產生彙總統計資料。當彙總分析規則新增至設定的資料表時，可讓可查詢的成員在設定的資料表上執行查詢。

彙總分析規則支援行銷活動規劃、媒體觸及、頻率測量和歸因等使用案例。

支援的查詢結構和語法在 中定義[彙總查詢結構和語法](#)。

中定義的分析規則參數[彙總分析規則 - 查詢控制項](#)包括查詢控制項和查詢結果控制項。其查詢控制項包含要求已設定資料表加入至少一個由成員擁有且可直接或暫時查詢的已設定資料表的能力。此要求可讓您確保查詢在資料表及其交集 (INNERJOIN) 上執行。

彙總查詢結構和語法

具有彙總分析規則之資料表的查詢必須遵循下列語法。

```
--select_aggregate_function_expression
SELECT
aggregation_function(column_name) [[AS] column_alias ] [, ...]

--select_grouping_column_expression
[, {column_name|scalar_function(arguments)} [[AS] column_alias ]][, ...]

--table_expression
FROM table_name [[AS] table_alias ]
    [[INNER] JOIN table_name [[AS] table_alias] ON join_condition] [...]

--where_expression
[WHERE where_condition]

--group_by_expression
[GROUP BY {column_name|scalar_function(arguments)}, ...]]

--having_expression
[HAVING having_condition]

--order_by_expression
[ORDER BY {column_name|scalar_function(arguments)} [{ASC|DESC}]] [,...]]
```

下表說明上述語法中列出的每個表達式。

表達式	定義	範例
<i>select_aggregate_function_expression</i>	逗號分隔清單，其中包含下列表達式： select_aggregation_function_expression	SELECT SUM(PRICE), user_segment

表達式	定義	範例
	<code>select_aggregate_expression</code>  Note <code>select_aggregation_function_expression</code> 中必須至少有一個 <code>select_aggregate_expression</code>。	
<i><code>select_aggregation_function_expression</code></i>	套用到一或多個資料欄的一或多個支援彙總函數。僅允許資料欄做為彙總函數的引數。  Note <code>select_aggregation_function_expression</code> 中必須至少有一個 <code>select_aggregate_expression</code>。	<code>AVG(PRICE)</code> <code>COUNT(DISTINCT user_id)</code>

表達式	定義	範例
<i>select_grouping_column_expression</i>	可使用下列項目包含任何表達式的表達式： - 資料表欄位名稱 - 支援的純量函數 - 字串常值 - 數值常值	TRUNC(timestampColumn) UPPER(campaignName)
	Note select_aggregate_expression 可以使用或不使用 AS 參數來別名資料欄。如需詳細資訊，請參閱 AWS Clean Rooms SQL 參考。	
<i>table_expression</i>	資料表或資料表聯結，以 連接聯結條件式表達式join_condition 。 join_condition 傳回布林值。 table_expression 支援： - 特定JOIN類型 (INNERJOIN) - join_condition (=) 內的等式比較條件 - 邏輯運算子 (AND、OR)。	<pre>FROM consumer_table INNER JOIN provider_table ON consumer_table.identifier1 = provider_table.identifier1 AND consumer_table.identifier2 = provider_table.identifier2</pre>

表達式	定義	範例
<i>where_expression</i>	傳回布林值的條件式表達式。它可能包含下列項目： - 資料表欄位名稱 - 支援的純量函數 - 數學運算子 - 字串常值 - 數值常值 支援的比較條件為 (=, >, <, <=, >=, <>, !=, NOT, IN, NOT IN, LIKE, IS NULL, IS NOT NULL)。 支援的邏輯運算子為 (AND, OR)。 where_expression 為選用。	<pre>WHERE where_condition WHERE price > 100 WHERE TRUNC(timestampColumn) = '1/1/2022' WHERE timestampColumn = timestampColumn2 - 14</pre>
<i>group_by_expression</i>	符合 需求的逗號分隔表達式清單select_grouping_column_expression 。	<pre>GROUP BY TRUNC(timestampColumn), UPPER(campaignName), segment</pre>

表達式	定義	範例
<i>having_expression</i>	傳回布林值的條件式表達式。它們具有套用至單一資料欄（例如，SUM(price)）的支援彙總函數，並與數值常值進行比較。 支援的條件為 (=, >, <, <=, >=, <>, !=)。 支援的邏輯運算子為 (AND, OR)。 having_expression 為選用。	HAVING SUM(SALES) > 500
<i>order_by_expression</i>	以逗號分隔的表達式清單，與select_aggregate_expression 先前定義的相同需求相容。 order_by_expression 為選用。  Note order_by_expression 允許 ASC 和 DESC 參數。如需詳細資訊，請參閱 AWS Clean Rooms SQL 參考 中的 ASC DESC 參數。	ORDER BY SUM(SALES), UPPER(campaignName)

對於彙總查詢結構和語法，請注意下列事項：

- 不支援 以外的 SQL SELECT 命令。
- 不支援子查詢和常用資料表表達式（例如 WITH）。
- 不支援合併多個查詢的運算子（例如 UNION）。
- TOP不支援 LIMIT、和 OFFSET 參數。

彙總分析規則 - 查詢控制項

透過彙總查詢控制項，您可以控制資料表中的資料欄用於查詢資料表的方式。例如，您可以控制哪些資料欄用於聯結、哪些資料欄可以計數，或哪些資料欄可用於 WHERE 陳述式。

以下各節說明每個控制項。

主題

- [彙總控制項](#)
- [聯結控制項](#)
- [維度控制](#)
- [純量函數](#)

彙總控制項

透過使用彙總控制項，您可以定義要允許哪些彙總函數，以及必須套用哪些資料欄。彙總函數可用於 SELECT、HAVING和 ORDERBY表達式。

控制項	定義	Usage
aggregateColumns	您允許在彙總函數中使用的已設定資料表資料欄資料欄。	aggregateColumns 可在 SELECT、HAVING和 ORDERBY表達式中的彙總函數內使用。 有些 aggregateColumns 也可以分類為 joinColumn（稍後定義）。 指定的也aggregate Column 無法分類為

控制項	定義	Usage
		<code>dimensionColumn</code> (稍後定義)。
<code>function</code>	您允許在 上使用的 <code>COUNT</code> 、 <code>SUM</code> 和 <code>AVG</code> 函數 <code>aggregateColumns</code> 。	<code>function</code> 可以套用至 <code>aggregateColumns</code> 與其相關聯的 。

聯結控制項

子 `JOIN` 句用於根據兩個或多個資料表之間的相關資料欄來結合資料列。

您可以使用聯結控制項來控制資料表如何聯結至 中的其他資料表 `table_expression`。AWS Clean Rooms 僅支援 `INNER JOIN`。`INNER JOIN` 陳述式只能使用在分析規則 `joinColumn` 中明確分類為 的資料欄，但需遵守您定義的控制項。

`INNER JOIN` 必須在 `joinColumn` 您設定的資料表和協同合作中 `joinColumn` 另一個設定的資料表的上操作。您決定資料表中的哪些資料欄可以用作 `joinColumn`。

子 `ON` 句中的每個相符條件都需要在兩個資料欄之間使用等式比較條件 (`=`)。

`ON` 子句中的多個相符條件可以是：

- 使用 `AND` 邏輯運算子合併
- 使用 `OR` 邏輯運算子分隔

Note

所有 `JOIN` 相符條件都必須符合 每一端的一列 `JOIN`。由 `OR` 或 `AND` 邏輯運算子連接的所有條件也必須遵守此要求。

以下是具有 `AND` 邏輯運算子的查詢範例。

```
SELECT some_col, other_col
FROM table1
  JOIN table2
  ON table1.id = table2.id AND table1.name = table2.name
```

以下是具有OR邏輯運算子的查詢範例。

```
SELECT some_col, other_col
FROM table1
  JOIN table2
  ON table1.id = table2.id OR table1.name = table2.name
```

控制項	定義	Usage
joinColumns	您想要允許可在 INNERJOIN 陳述式中使用之成員的資料欄 (如果有的話)。	特定 joinColumn 也可以分類為 aggregate Column (請參閱 彙總控制項)。 相同的資料欄不能同時使用為 joinColumn 和 dimension Columns (請參閱下文)。 除非它也被分類為 aggregate Column , 否則 joinColumn 無法在查詢的任何其他部分中使用INNERJOIN。
joinRequired	控制是否需要INNERJOIN具有可查詢之成員所設定資料表的。	如果您啟用此參數, INNERJOIN則需要。如果您未啟用此參數, INNERJOIN是選用的。 假設您啟用此參數, 可以查詢的成員需要在 INNER 中包含他們擁有的資料表JOIN。他們必須JOIN直接或暫時使用您的資料表 (也就是, 將他們的資料表加入另一個資料表, 而資料表本身會加入您的資料表)。

以下是可轉移性的範例。

```

ON
my_table.identifer = third_party_table.identifier
....
ON
third_party_table.identifier = member_who_can_query_table.id

```

Note

可以查詢的成員也可以使用 `joinRequired` 參數。在這種情況下，查詢必須與至少另一個資料表聯結其資料表。

維度控制

維度控制項控制可篩選、分組或彙總彙總資料欄的資料欄。

控制項	定義	Usage
<code>dimensionColumns</code>	您允許可在 SELECT、BY、和 中查詢使用之成員的資料欄 WHERE GROUP (如果有) ORDERBY。	<code>dimensionColumn</code> 可用於 SELECT(select_grouping_column_expression)BY、WHERE GROUP 和 ORDER BY。 相同的資料欄不能同時是 <code>dimensionColumn</code> 、 <code>joinColumn</code> 和/或 <code>aggregateColumn</code> 。

純量函數

純量函數控制哪些純量函數可用於維度資料欄。

控制項	定義	Usage
<code>scalarFunctions</code>	可在查詢 <code>dimensionColumns</code> 中使用的純量函數。	指定您允許 (例如,) 套用至的純量函數 (如果有的話 CAS T) <code>dimensionColumns</code> 。

控制項	定義	Usage
		純量函數無法用於其他函數之上或其他函數內。純量函數的引數可以是資料欄、字串常值或數值常值。

支援下列純量函數：

- 數學函數 – ABS、CEILING、FLOOR、LOG、LN、ROUND、SQRT
- 資料類型格式化函數 – CAST, CONVERT, TO_CHAR, TO_DATE, TO_NUMBER, TO_TIMESTAMP
- 字串函數 – LOWER、UPPER、TRIM、RTRIM、SUBSTRING
 - 對於 RTRIM，不允許要修剪的自訂字元集。
- 條件式表達式 – COALESCE
- 日期函數 – EXTRACT、GETDATE、CURRENT_DATE、DATEADD
- 其他 函數 – TRUNC

如需詳細資訊，請參閱 [AWS Clean Rooms SQL 參考](#)。

彙總分析規則 - 查詢結果控制

使用彙總查詢結果控制項，您可以透過指定每個輸出資料列必須符合的一或多個條件來控制傳回的結果。AWS Clean Rooms 支援形式的彙總限制 `COUNT (DISTINCT column) >= X`。此表單要求每一列至少從您設定的資料表彙總 X 個不同的值（例如，最低數量的不同 `user_id` 值）。即使提交的查詢本身未使用指定的資料欄，也會自動強制執行此最低閾值。它們會從協同合作中每個成員的已設定資料表，在查詢中的每個已設定資料表中集體強制執行。

每個設定的資料表在其分析規則中必須至少有一個彙總限制。設定的資料表擁有者可以新增多個 `columnName` 和相關聯的 `minimum`，而且它們會集體強制執行。

彙總限制

彙總限制會控制查詢結果中傳回的資料列。若要傳回，資料列必須符合彙總限制中指定之每個資料欄中指定的相異值下限。即使未在查詢或分析規則的其他部分中明確提及資料欄，此要求也適用。

控制項	定義	Usage
columnName	aggregateColumn 用於每個輸出資料列必須符合的條件的。	可以是已設定資料表中的任何資料欄。
minimum	aggregateColumn 輸出資料列必須擁有的關聯 (例如 COUNT DISTINCT) 之相異值的最小數量，才能在查詢結果中傳回。	值必須至少minimum為 2。

彙總分析規則結構

下列範例顯示彙總分析規則的預先定義結構。

在下列範例中，*MyTable*是指您的資料表。您可以將每個#####取代為您自己的資訊。

```
{
  "aggregateColumns": [
    {
      "columnNames": [MyTable column names], "function": [Allowed Agg Functions]
    },
  ],
  "joinRequired": ["QUERY_RUNNER"],
  "joinColumns": [MyTable column names],
  "dimensionColumns": [MyTable column names],
  "scalarFunctions": [Allowed Scalar functions],
  "outputConstraints": [
    {
      "columnName": [MyTable column names], "minimum": [Numeric value]
    },
  ]
}
```

彙總分析規則 - 範例

下列範例示範兩家公司如何使用彙總分析 AWS Clean Rooms 協作。

A 公司有客戶和銷售資料。A 公司有興趣了解產品退回活動。B 公司是 A 公司的零售商之一，並具有傳回資料。B 公司對客戶也有對 A 公司有用的客群屬性（例如，購買相關產品、使用零售商的客戶服務）。B 公司不想提供資料列層級的客戶傳回資料和屬性資訊。B 公司只想為 A 公司啟用一組查詢，以最低彙總閾值取得重疊客戶的彙總統計資料。

A 公司和 B 公司決定協同合作，讓 A 公司了解產品退回活動，並在 B 公司和其他管道提供更好的產品。

若要建立協同合作並執行彙總分析，公司會執行下列動作：

1. A 公司建立協同合作並建立成員資格。協同合作讓 B 公司成為協同合作中的另一個成員。公司 A 會在協同合作中啟用查詢記錄，並在其帳戶中啟用查詢記錄。
2. B 公司在協同合作中建立成員資格。它在其帳戶中啟用查詢記錄。
3. 公司 A 會建立銷售設定的資料表。
4. A 公司將下列彙總分析規則新增至銷售設定的資料表。

```
{
  "aggregateColumns": [
    {
      "columnNames": [
        "identifier"
      ],
      "function": "COUNT_DISTINCT"
    },
    {
      "columnNames": [
        "purchases"
      ],
      "function": "AVG"
    },
    {
      "columnNames": [
        "purchases"
      ],
      "function": "SUM"
    }
  ],
  "joinColumns": [
    "hashedemail"
  ],
  "dimensionColumns": [
```

```

    "demoseg",
    "purchasedate",
    "productline"
  ],
  "scalarFunctions": [
    "CAST",
    "COALESCE",
    "TRUNC"
  ],
  "outputConstraints": [
    {
      "columnName": "hashedemail",
      "minimum": 2,
      "type": "COUNT_DISTINCT"
    }
  ]
}

```

aggregateColumns – A 公司想要計算銷售資料和傳回資料之間重疊中的唯一客戶數量。A 公司也想要加總 **purchases** 要與 數目比較的 數目 **returns**。

joinColumns – 公司 A 想要使用 **identifier** 來比對客戶從銷售資料到客戶從傳回資料。這將有助於公司 A 配對傳回正確的購買。它還有助於公司 A 區段重疊的客戶。

dimensionColumns – A 公司使用 依特定產品 **dimensionColumns** 篩選、比較購買和在特定期間內傳回、確保傳回日期晚於產品日期，以及協助分割重疊的客戶。

scalarFunctions – A 公司會根據設定的資料表 A 與協同合作的關聯，視需要選取 **CAST** 純量函數以協助更新資料類型格式。它還會新增純量函數，以在需要時協助格式化資料欄。

outputConstraints – 公司 A 設定最小輸出限制。它不需要限制結果，因為允許分析師查看其銷售資料表中的資料列層級資料

Note

A 公司不包含在分析規則 **joinRequired** 中。它為分析師提供了單獨查詢銷售資料表的靈活性。

5. B 公司會建立傳回設定的資料表。
6. B 公司將下列彙總分析規則新增至傳回設定的資料表。

```
{
  "aggregateColumns": [
    {
      "columnNames": [
        "identifier"
      ],
      "function": "COUNT_DISTINCT"
    },
    {
      "columnNames": [
        "returns"
      ],
      "function": "AVG"
    },
    {
      "columnNames": [
        "returns"
      ],
      "function": "SUM"
    }
  ],
  "joinColumns": [
    "hashedemail"
  ],
  "joinRequired": [
    "QUERY_RUNNER"
  ],
  "dimensionColumns": [
    "state",
    "popularpurchases",
    "customerserviceuser",
    "productline",
    "returndate"
  ],
  "scalarFunctions": [
    "CAST",
    "LOWER",
    "UPPER",
    "TRUNC"
  ],
  "outputConstraints": [
    {
      "columnName": "hashedemail",
```



```

    "minimum": 100,
    "type": "COUNT_DISTINCT"
  },
  {
    "columnName": "producttype",
    "minimum": 2,
    "type": "COUNT_DISTINCT"
  }
]
}

```

aggregateColumns – 公司 B 可讓公司 A 的總和 **returns** 與購買次數進行比較。它們至少有一個彙總資料欄，因為它們正在啟用彙總查詢。

joinColumns – 公司 B 可讓公司 A 加入 **identifier**，以比對客戶從銷售資料傳回給客戶的資料。**identifier** 資料特別敏感，並將它做為 **joinColumn** 以確保資料永遠不會在查詢中輸出。

joinRequired – B 公司要求對傳回資料的查詢與銷售資料重疊。他們不希望讓 A 公司查詢其資料集中的所有個人。他們在協同合作協議中也同意該限制。

dimensionColumns – 公司 B 可讓公司 A 依 **state**、**popularpurchases** 和 **進行篩選和分組** **customerserviceuser**，這些是可協助進行公司 A 分析的唯一屬性。公司 B 可讓公司 A 使用 **returndate** 篩選 之後 **returndate** 發生的輸出 **purchasedate**。透過此篩選，輸出對於評估產品變更的影響更為準確。

scalarFunctions – B 公司啟用下列項目：

- 日期的 **TRUNC**
- **LOWER** 和 **UPPER**，以防在其資料中以不同的格式 **producttype** 輸入
- **CAST** 如果 A 公司需要將銷售中的資料類型轉換為與傳回中的資料類型相同

A 公司不會啟用其他純量函數，因為他們不認為查詢需要它們。

outputConstraints – B 公司設定 的最低輸出限制 **hashedemail**，以協助降低重新識別客戶的能力。它也會在 上新增最低輸出限制 **producttype**，以減少重新識別傳回之特定產品的能力。根據輸出的維度（例如，），某些產品類型可能更具優勢 **state**。無論 A 公司為其資料新增的輸出限制，一律都會強制執行其輸出限制。

7. A 公司建立與協同合作的銷售資料表關聯。

8. B 公司會建立與協同合作的傳回資料表關聯。

9. 公司 A 會執行查詢，例如下列範例，相較於 2022 年位置的總購買量，更了解公司 B 中的傳回數量。

```
SELECT
  companyB.state,
  SUM(companyB.returns),
  COUNT(DISTINCT companyA.hashemail)
FROM
  sales companyA
  INNER JOIN returns companyB ON companyA.identifier = companyB.identifier
WHERE
  companyA.purchasedate BETWEEN '2022-01-01' AND '2022-12-31' AND
  TRUNC(companyB.returndate) > companyA.purchasedate
GROUP BY
  companyB.state;
```

10. 公司 A 和公司 B 會檢閱查詢日誌。B 公司會驗證查詢是否符合協同合作協議中同意的內容。

對彙總分析規則問題進行故障診斷

當您使用彙總分析規則時，請使用此處的資訊來協助您診斷和修正常見問題。

問題

- [我的查詢未傳回任何結果](#)

我的查詢未傳回任何結果

當沒有相符的結果，或相符的結果不符合一或多個最小彙總閾值時，就會發生這種情況。

如需最低彙總閾值的詳細資訊，請參閱 [彙總分析規則 - 範例](#)。

列出分析規則

在中 AWS Clean Rooms，清單分析規則會輸出其新增至的已設定資料表與可查詢之成員的已設定資料表之間的重疊資料列層級清單。可查詢的成員會執行包含清單分析規則的查詢。

清單分析規則類型支援使用擴充和對象建置等案例。

如需此分析規則預先定義查詢結構和語法的詳細資訊，請參閱 [列出分析規則預先定義的結構](#)。

中定義的清單分析規則參數[列出分析規則 - 查詢控制項](#)具有查詢控制項。其查詢控制項包含選取可在輸出中列出的資料欄的功能。查詢需要至少有一個與成員所設定資料表的聯結，該成員可以直接或暫時進行查詢。

沒有像[彙總分析規則](#)一樣的查詢結果控制項。

清單查詢只能使用數學運算子。他們無法使用其他函數（例如彙總或純量）。

主題

- [列出查詢結構和語法](#)
- [列出分析規則 - 查詢控制項](#)
- [列出分析規則預先定義的結構](#)
- [列出分析規則 - 範例](#)

列出查詢結構和語法

具有清單分析規則之資料表的查詢必須遵循下列語法。

```
--select_list_expression
SELECT DISTINCT column_name [[AS] column_alias ] [, ...]

--table_expression
FROM table_name [[AS] table_alias ]
    [[INNER] JOIN table_name [[AS] table_alias] ON join_condition] [...]

--where_expression
[WHERE where_condition]

--limit_expression
[LIMIT number]
```

下表說明上述語法中列出的每個表達式。

表達式	定義	範例
<i>select_list_expression</i>	以逗號分隔的清單，包含至少一個資料表資料欄名稱。 需要 DISTINCT 參數。	SELECT DISTINCT segment

表達式	定義	範例
	 Note <code>select_list_expression</code> 可以使用或不使用 <code>AS</code> 參數來別名資料欄。如需詳細資訊，請參閱 AWS Clean Rooms SQL 參考。	
<i>table_expression</i>	資料表或資料表的聯結，<code>join_condition</code> 搭配 將其連接到 <code>join_condition</code>。 <code>join_condition</code> 傳回布林值。 <code>table_expression</code> 支援： • 特定 JOIN 類型 (INNER JOIN) • <code>join_condition</code> (=) 內的相等比較條件 • 邏輯運算子 (AND、OR)。	<pre>FROM consumer_table INNER JOIN provider_table ON consumer_table.identifier1 = provider_table.identifier1 AND consumer_table.identifier2 = provider_table.identifier2</pre>

表達式	定義	範例
<i>where_expression</i>	傳回布林值的條件式表達式。它可以包含下列項目： - 資料表欄位名稱 - 數學運算子 - 字串常值 - 數值常值 支援的比較條件為 (=, >, <, <=, >=, <>, !=, NOT, IN, NOT IN, LIKE, IS NULL, IS NOT NULL)。 支援的邏輯運算子為 (AND, OR)。 <i>where_expression</i> 為選用。	<pre>WHERE state + '_' + city = 'NY_NYC'</pre> <pre>WHERE timestampColumn = timestampColumn2 - 14</pre>
<i>limit_expression</i>	此表達式必須採用正整數。 <i>limit_expression</i> 為選用。	<pre>LIMIT 100</pre>

對於清單查詢結構和語法，請注意下列事項：

- 不支援 SELECT 以外的 SQL 命令。
- 不支援子查詢和常用資料表表達式（例如 WITH）
- 不支援 HAVINGBY、GROUP 和 ORDER BY子句
- 不支援 OFFSET 參數

列出分析規則 - 查詢控制項

透過清單查詢控制項，您可以控制資料表中的資料欄用於查詢資料表的方式。例如，您可以控制哪些資料欄用於聯結，或哪些資料欄可用於 SELECT 陳述式和WHERE子句。

以下各節說明每個控制項。

主題

- [聯結控制項](#)
- [列出控制項](#)

聯結控制項

使用聯結控制項，您可以控制資料表如何聯結至 table_expression 中的其他資料表。AWS Clean Rooms 僅支援 INNER JOIN。在清單分析規則中，至少需要一個 INNER JOIN，而且可以查詢的成員需要在 INNER JOIN 中包含他們擁有的資料表。這表示他們必須直接或暫時加入您的資料表。

以下是可轉移性的範例。

```
ON
my_table.identifer = third_party_table.identifier
....
ON
third_party_table.identifier = member_who_can_query_table.id
```

INNER JOIN 陳述式只能使用在您的分析規則joinColumn中明確歸類為 的資料欄。

INNER JOIN 必須在joinColumn您設定的資料表和協同合作中joinColumn另一個設定的資料表的上操作。您可以決定資料表中的哪些資料欄可以用作 joinColumn。

子ON句中的每個相符條件都需要使用兩個資料欄之間的相等比較條件 (=)。

ON 子句中的多個相符條件可以是：

- 使用AND邏輯運算子合併
- 使用OR邏輯運算子分隔

Note

所有JOIN相符條件都必須符合 每一端的一列JOIN。由 OR或 AND 邏輯運算子連接的所有條件也必須遵守此要求。

以下是具有AND邏輯運算子的查詢範例。

```
SELECT some_col, other_col
FROM table1
  JOIN table2
  ON table1.id = table2.id AND table1.name = table2.name
```

以下是具有OR邏輯運算子的查詢範例。

```
SELECT some_col, other_col
FROM table1
  JOIN table2
  ON table1.id = table2.id OR table1.name = table2.name
```

控制項	定義	Usage
joinColumns	您要允許可在 INNER JOIN 陳述式中查詢使用之成員的資料欄。	相同的資料欄無法同時歸類為 joinColumn 和 listColumn (請參閱 列出控制項)。 joinColumn 無法用於 INNER JOIN 以外的任何其他查詢部分。

列出控制項

清單控制項控制可在查詢輸出中列出的資料欄 (即 SELECT 陳述式中使用的資料欄) 或用於篩選結果 (即WHERE陳述式中使用的資料欄)。

控制項	定義	Usage
<code>listColumns</code>	您允許可在 SELECT 和 中查詢使用之成員的資料欄 WHERE	<code>listColumn</code> 可用於 SELECT 和 WHERE。 相同的資料欄無法同時做為 <code>listColumn</code> 和 使用 <code>joinColumn</code> 。

列出分析規則預先定義的結構

下列範例包含預先定義的結構，示範如何完成清單分析規則。

在下列範例中，*MyTable*是指您的資料表。您可以將每個#####取代為您自己的資訊。

```
{
  "joinColumns": [MyTable column name(s)],
  "listColumns": [MyTable column name(s)],
}
```

列出分析規則 - 範例

下列範例示範兩家公司如何使用 AWS Clean Rooms 清單分析協作。

A 公司有客戶關係管理 (CRM) 資料。A 公司想要取得客戶的其他客群資料，以進一步了解他們的客戶，並可能使用屬性做為其他分析的輸入。B 公司擁有區段資料，其中包含他們根據第一方資料建立的唯一區段屬性。公司 B 只希望在其資料與公司 A 資料之間重疊的客戶上，提供唯一的客群屬性給公司 A。

公司決定協同合作，讓 A 公司可以充實重疊的資料。A 公司是可以查詢的成員，而 B 公司是參與者。

若要建立協同合作並在協同合作中執行清單分析，公司會執行下列動作：

1. A 公司建立協同合作並建立成員資格。協同合作讓 B 公司成為協同合作的另一個成員。A 公司在協同合作中啟用查詢記錄，並在其帳戶中啟用查詢記錄。
2. B 公司在協同合作中建立成員資格。它在其帳戶中啟用查詢記錄。
3. 公司 A 會建立 CRM 設定的資料表
4. A 公司將分析規則新增至客戶設定的資料表，如下列範例所示。


```
{
  "joinColumns": [
    "identifier1",
    "identifier2"
  ],
  "listColumns": [
    "internalid",
    "segment1",
    "segment2",
    "customercategory"
  ]
}
```

joinColumns – A 公司想要使用 hashedemail 和/或 thirdpartyid (從身分廠商取得)，將客戶從 CRM 資料比對到客戶從客群資料。這將有助於確保 A 公司符合適當客戶的豐富資料。它們有兩個 joinColumns 可能改善分析的比對率。

listColumns – 公司 A 使用 在自己的系統中 internalid 使用 以外的 listColumns 來取得富集的資料欄。它們會新增 segment1、segment2 和 customercategory，以透過在篩選條件中使用特定客群來限制擴充。

5. B 公司會建立區段設定的資料表。
6. B 公司將分析規則新增至區段設定的資料表。

```
{
  "joinColumns": [
    "identifier2"
  ],
  "listColumns": [
    "segment3",
    "segment4"
  ]
}
```

joinColumns – B 公司可讓 A 公司加入 identifier2，將客戶從客群資料配對至 CRM 資料。A 公司和 B 公司與身分廠商合作，以取得 identifier2 符合此協同合作的項目。他們沒有新增其他 joinColumns，因為他們認為 identifier2 提供最高且最準確的配對率，而查詢不需要其他識別符。

listColumns – B 公司可讓 A 公司使用 segment3 和 segment4 屬性來擴充其資料，這些屬性是他們在（與客戶 A）上建立、收集和對齊的唯一屬性，成為資料擴充的一部分。他們希望 A 公司在資料列層級取得重疊的這些區段，因為這是資料擴充協作。

7. A 公司建立與協同合作的 CRM 資料表關聯。
8. B 公司建立與協同合作的區段資料表關聯。
9. A 公司執行查詢，例如下列查詢，以豐富重疊的客戶資料。

```
SELECT companyA.internalid, companyB.segment3, companyB.segment4
INNER JOIN returns companyB
  ON companyA.identifier2 = companyB.identifier2
WHERE companyA.customercategory > 'xxx'
```

10. 公司 A 和公司 B 會檢閱查詢日誌。B 公司會驗證查詢是否符合協同合作協議中同意的內容。

中的自訂分析規則 AWS Clean Rooms

在中 AWS Clean Rooms，自訂分析規則是一種新的分析規則類型，允許自訂查詢在設定的資料表上執行。自訂 SQL 查詢仍然限制為只有 SELECT 命令，但可以使用比[彙總](#)和[清單](#)查詢更多的 SQL 建構（例如，視窗函數、OUTER JOIN、CTEs 或子查詢；如需完整清單，請參閱[AWS Clean Rooms SQL 參考](#)）。自訂 SQL 查詢不需要遵循查詢結構，例如[彙總](#)和[列出](#)查詢。

自訂分析規則支援比彙總和列出分析規則所支援更進階的使用案例，例如自訂屬性分析、基準化、增量分析和對象探索。這是彙總和列出分析規則所支援之使用案例的超集。

自訂分析規則也支援差異隱私權。差異隱私權是資料隱私權保護的數學嚴格架構。如需詳細資訊，請參閱[AWS Clean Rooms 差異隱私權](#)。當您建立分析範本時，AWS Clean Rooms 差異隱私權會檢查範本，以判斷它是否與 AWS Clean Rooms 差異隱私權的一般用途查詢結構相容。此驗證可確保您不會建立不允許使用差異隱私權保護資料表的分析範本。

若要設定自訂分析規則，資料擁有者可以選擇允許儲存在[分析範本](#)中的特定自訂查詢在其設定的資料表上執行。資料擁有者會檢閱分析範本，再將其新增至自訂分析規則中允許的分析控制項。分析範本僅在建立它們的協同合作中可用和可見（即使資料表與其他協同合作相關聯），並且只能由可在該協同合作中查詢的成員執行。

或者，成員可以選擇允許其他成員（查詢提供者）建立查詢，而無需檢閱。成員在自訂分析規則中新增允許查詢提供者控制的查詢提供者帳戶。如果查詢提供者是可以查詢的成員，他們可以直接在設定的資料表上執行任何查詢。查詢提供者也可以透過建立[分析範本來建立](#)查詢。查詢提供者建立的任何查詢都會自動允許在 AWS 帳戶存在且與資料表相關聯的所有協同合作中，於資料表上執行。

資料擁有者只能允許分析範本或帳戶建立查詢，不能同時建立兩者。如果資料擁有者將其保留空白，可以查詢的成員就無法在設定的資料表上執行查詢。

主題

- [自訂分析規則預先定義的結構](#)
- [自訂分析規則範例](#)
- [具有差異隱私權的自訂分析規則](#)

自訂分析規則預先定義的結構

下列範例包含預先定義的結構，說明如何在開啟差異隱私權的情況下完成自訂分析規則。userIdentifier 值是可唯一識別使用者的欄，例如 user_id。當您在協同合作中開啟了兩個或多個具有差異隱私權的資料表時，AWS Clean Rooms 會要求您在兩個分析規則中設定與使用者識別符資料欄相同的資料欄，以維持資料表之間使用者的一致定義。

```
{
  "allowedAnalyses": ["ANY_QUERY"] | string[],
  "allowedAnalysisProviders": [],
  "differentialPrivacy": {
    "columns": [
      {
        "name": "userIdentifier"
      }
    ]
  }
}
```

您可擇一方法：

- 將分析範本 ARNs 新增至允許的分析控制項。在此情況下，不包含 allowedAnalysisProviders 控制項。

```
{
  allowedAnalyses: string[]
}
```

- 將 AWS 帳戶 IDs 新增至 allowedAnalysisProviders 控制項。在此情況下，您會將 ANY_QUERY 新增至 allowedAnalyses 控制項。

```
{
  allowedAnalyses: ["ANY_QUERY"],
  allowedAnalysisProviders: string[]
}
```

自訂分析規則範例

下列範例示範兩家公司如何使用自訂分析規則 AWS Clean Rooms 在 中協作。

A 公司有客戶和銷售資料。A 公司有興趣了解 B 公司網站上的廣告行銷活動銷售增量。B 公司具有對公司有用的瀏覽資料和客群屬性（例如，檢視廣告時所使用的裝置）。

A 公司有他們想要在協同合作中執行的特定增量查詢。

若要建立協同合作並在協同合作中執行自訂分析，公司會執行下列動作：

1. A 公司建立協同合作並建立成員資格。協同合作讓 B 公司成為協同合作的另一個成員。A 公司在協同合作中啟用查詢記錄，並在其帳戶中啟用查詢記錄。
2. B 公司在協同合作中建立成員資格。它在其帳戶中啟用查詢記錄。
3. 公司 A 會建立 CRM 設定的資料表
4. A 公司將空的自訂分析規則新增至銷售設定的資料表。
5. A 公司將銷售設定的資料表與協同合作建立關聯。
6. B 公司會建立檢視設定資料表。
7. B 公司將空的自訂分析規則新增至檢視器設定的資料表。
8. B 公司將檢視設定資料表與協同合作建立關聯。
9. 公司 A 檢視與協同合作相關聯的銷售資料表和檢視器資料表，並建立分析範本，新增行銷活動月份的增量查詢和參數。

```
{
  "analysisParameters": [
    {
      "defaultValue": ""
      "type": "DATE"
      "name": "campaign_month"
    }
  ],
  "description": "Monthly incrementality query using sales and viewership data"
```

```

"format": "SQL"
"name": "Incrementality analysis"
"source":
  "WITH labeleddata AS
  (
    SELECT hashedemail, deviceid, purchases, unitprice, purchasedate,
    CASE
      WHEN testvalue IN ('value1', 'value2', 'value3') THEN 0
      ELSE 1
    END AS testgroup
    FROM viewershipdata
  )
  SELECT labeleddata.purchases, provider.impressions
  FROM labeleddata
  INNER JOIN salesdata
    ON labeleddata.hashedemail = provider.hashedemail
  WHERE MONTH(labeleddata.purchasedate) > :campaignmonth
  AND testgroup = :group
  "
}

```

10A 公司將其帳戶（例如 444455556666）新增至自訂分析規則中允許的分析提供者控制項。他們使用允許的分析提供者控制項，因為他們想要允許他們建立的任何查詢在其銷售設定的資料表上執行。

```

{
  "allowedAnalyses": [
    "ANY_QUERY"
  ],
  "allowedAnalysisProviders": [
    "444455556666"
  ]
}

```

11B 公司會在協同合作中看到建立的分析範本，並檢閱其內容，包括查詢字串和參數。

12B 公司判斷分析範本達到增量使用案例，並符合其檢視設定資料表如何查詢的隱私權要求。

13B 公司將分析範本 ARN 新增至檢視器資料表自訂分析規則中允許的分析控制項。他們使用允許的分析控制項，因為他們只想要允許增量查詢在其檢視器設定的資料表上執行。

```

{
  "allowedAnalyses": [
    "arn:aws:cleanrooms:us-east-1:111122223333:membership/41327cc4-bbf0-43f1-b70c-a160dddceb08/analysistemplate/1ff1bf9d-781c-418d-a6ac-2b80c09d6292"
  ]
}

```

```
]
}
```

14A 公司執行分析範本，並使用參數值 05-01-2023。

具有差異隱私權的自訂分析規則

在中 AWS Clean Rooms，自訂分析規則支援差異隱私權。差異隱私權是一種數學上嚴格的資料隱私權保護架構，可協助您保護資料免於重新識別嘗試。

差異隱私權支援彙總分析，例如廣告行銷活動規劃、post-ad-campaign測量、金融機構協會中的基準測試，以及醫療保健研究的 A/B 測試。

支援的查詢結構和語法在 中定義[查詢結構和語法](#)。

具有差異隱私權的自訂分析規則範例

Note

AWS Clean Rooms 差異隱私權僅適用於資料存放在 Amazon S3 中的協同合作。

請考慮上一節中呈現的[自訂分析規則範例](#)。此範例示範如何使用差異隱私權來保護資料免於重新識別嘗試，同時讓合作夥伴從資料中學習關鍵業務洞見。假設擁有檢視器資料的 B 公司想要使用差異隱私權來保護其資料。若要完成差異隱私權設定，B 公司會完成下列步驟：

1. 公司 B 會開啟差異隱私權，同時將自訂分析規則新增至檢視器設定的資料表。B 公司會選取 `viewershipdata.hashemail` 做為使用者識別符欄。
2. B 公司在協同合作中[新增了差異性隱私權政策](#)，讓其瀏覽資料資料表可供查詢。B 公司會選取預設政策以快速完成設定。

A 公司想要了解 B 公司網站上的廣告行銷活動銷售增量，執行分析範本。由於查詢與 AWS Clean Rooms 差異隱私權的一般用途[查詢結構](#)相容，因此查詢會成功執行。

查詢結構和語法

包含至少一個已開啟差異隱私權的資料表的查詢必須遵循下列語法。

```
query_statement:
```

```

[cte, ...] final_select

cte:
  WITH sub_query AS (
    inner_select
    [ UNION | INTERSECT | UNION_ALL | EXCEPT/MINUS ]
    [ inner_select ]
  )

inner_select:
  SELECT [user_id_column, ] expression [, ...]
  FROM table_reference [, ...]
  [ WHERE condition ]
  [ GROUP BY user_id_column[, expression] [, ...] ]
  [ HAVING condition ]

final_select:
  SELECT [expression, ...] | COUNT | COUNT_DISTINCT | SUM | AVG | STDDEV
  FROM table_reference [, ...]
  [ WHERE condition ]
  [ GROUP BY expression [, ...] ]
  [ HAVING COUNT | COUNT_DISTINCT | SUM | AVG | STDDEV | condition ]
  [ ORDER BY column_list ASC | DESC ]
  [ OFFSET literal ]
  [ LIMIT literal ]

expression:
  column_name [, ...] | expression AS alias | aggregation_functions |
window_functions_on_user_id | scalar_function | CASE | column_name math_expression [,
expression]

window_functions_on_user_id:
  function () OVER (PARTITION BY user_id_column, [column_name] [ORDER BY column_list
ASC|DESC])

```

Note

對於差異隱私權查詢結構和語法，請注意下列事項：

- 不支援子查詢。
- 如果資料表或 CTEs 涉及受差異隱私權保護的資料，通用資料表表達式 (CTE) 應該發出使用者識別符資料欄。篩選、分組和彙總應在使用者層級完成。

- Final_select 允許 COUNT DISTINCT、COUNT、SUM、AVG 和 STDDEV 彙總函數。

如需差異隱私權支援哪些 SQL 關鍵字的詳細資訊，請參閱 [AWS Clean Rooms 差異隱私權的 SQL 功能](#)。

ID 映射表分析規則

在中 AWS Clean Rooms，ID 映射資料表分析規則不是獨立的分析規則。這種類型的分析規則由 管理 AWS Clean Rooms，用於聯結不同的身分資料，以方便查詢。它會自動新增至 ID 映射表，而且無法編輯。只要這些分析規則是同質的，就會繼承協同合作中其他分析規則的行為。

ID 映射表分析規則會對 ID 映射表強制執行安全性。它限制協作成員使用 ID 映射表直接選取或檢查兩個成員的資料集之間的非重疊人口。ID 映射表分析規則用於隱含在其他分析規則的查詢中使用時，保護 ID 映射表中的敏感資料。

使用 ID 映射表分析規則，會在擴展的 SQL 中 AWS Clean Rooms 強制執行 ID 映射表兩側的重疊。這可讓您執行下列任務：

- 在 JOIN 陳述式中使用 ID 映射表的重疊。

AWS Clean Rooms 允許 ID 映射資料表上的 INNER、LEFT 或 RIGHT 聯結，如果其遵守重疊。為了保護敏感映射資訊，ID 映射表必須一律位於任何 JOIN 操作的「inner」端。例如，下列 JOIN 操作有效：

- table LEFT JOIN id_mapping_table
- id_mapping_table RIGHT JOIN table
- table INNER JOIN id_mapping_table

下列 JOIN 操作無效：

- id_mapping_table LEFT JOIN table
- table RIGHT JOIN id_mapping_table

這可防止您的資料集中沒有對應相符項目的映射記錄暴露。允許此類操作可能會洩露有關其他協同合作成員資料映射的敏感資訊。

- 在 JOIN 陳述式中使用映射表資料欄。

您無法在下列陳述式中使用映射表欄：SELECT、GROUP BY、WHERE HAVING 或 ORDER BY（除非在來源 ID 命名空間關聯或目標 ID 命名空間關聯上修改保護）。

- 在擴展的 SQL 中，AWS Clean Rooms 也支援 OUTER JOIN、隱含 JOIN 和 CROSS JOIN。這些聯結無法滿足重疊要求。反之，AWS Clean Rooms 使用 `requireOverlap` 來指定必須加入哪些資料欄。

支援的查詢結構和語法在 [中定義 ID 映射表查詢結構和語法](#)。

中定義的分析規則參數 [ID 映射表分析規則查詢控制項](#) 包括查詢控制項和查詢結果控制項。其查詢控制項包含要求 JOIN 陳述式中 ID 映射資料表重疊的功能（即 `requireOverlap`）。

主題

- [ID 映射表查詢結構和語法](#)
- [ID 映射表分析規則查詢控制項](#)
- [ID 映射表分析規則預先定義的結構](#)
- [ID 映射表分析規則 – 範例](#)

ID 映射表查詢結構和語法

對具有 ID 映射資料表分析規則的資料表進行查詢必須遵循下列語法。

```
--select_list_expression
SELECT
provider.data_col, consumer.data_col

--table_expression
FROM provider

JOIN idMappingTable idmt ON provider.id = idmt.sourceId

JOIN consumer ON consumer.id = idmt.targetId
```

協同合作資料表

下表代表協同 AWS Clean Rooms 合作中存在的已設定資料表。`cr_drivers_license` 和 `cr_management` 資料表中的 ID 資料欄代表要與 ID 映射資料表相符的資料欄。

`cr_drivers_license`

id	driver_name	state_of_registration
1	Eduard	TX
2	Dana	MA
3	Gweneth	IL

cr_ 保險

id	policyholder_email	policy_number
a	eduardo@internal.company.com	17f9d04e-f5be-4426-bdc4-250ed59c6529
b	gwen@internal.company.com	3f0092db-2316-48a8-8d44-09cf8f6e6c64
c	rosa@internal.company.com	d7692e84-3d3c-47b8-b46d-a0d5345f0601

ID 映射表

下表代表與 cr_drivers_license 和 cr_management 資料表相符的現有 ID 映射資料表。並非所有項目都有兩個協同合作資料表IDs。

cr_drivers_license_id	cr_insurance_id
1	a
2	null
3	b
null	c

ID 映射資料表分析規則僅允許在一組重疊資料上執行查詢，如下所示：

cr_driver s_license_id	cr_insura nce_id	driver_name	state_of_ registration	policyhol der_email	policy_nu mber
1	a	Eduard	TX	eduardo@i nternal.c ompany.com	17f9d04e- f5be-4426 -bdc4-250 ed59c6529
3	b	Gweneth	IL	gwen@inte rnal.comp any.com	3f0092db- 2316-48a8 -8d44-09c f8f6e6c64

查詢範例

下列範例顯示 ID 映射資料表聯結的有效位置：

```
-- Single ID mapping table
SELECT
    [ select_items ]FROM
    cr_drivers_license cr_dl
    [ INNER | LEFT ] JOIN cr_identity_mapping_table idmt ON idmt.cr_drivers_license_id
= cr_dl.id
    [ INNER | RIGHT ] JOIN cr_insurance cr_in                ON idmt.cr_insurance_id
= cr_in.id
;
-- Single ID mapping table (Subquery)
SELECT
    [ select_items ]FROM (
    SELECT
        [ select_items ]
    FROM
        cr_drivers_license cr_dl
        [ INNER | LEFT ] JOIN cr_identity_mapping_table idmt ON
idmt.cr_drivers_license_id = cr_dl.id
        [ INNER | RIGHT ] JOIN cr_insurance cr_in                ON idmt.cr_insurance_id
= cr_in.id
    )
;
-- Single ID mapping table (CTE)
```

```

WITH
    matched_ids AS (
        SELECT
            [ select_items ]
        FROM
            cr_drivers_license cr_dl
            [ INNER | LEFT ] JOIN cr_identity_mapping_table idmt ON
            idmt.cr_drivers_license_id = cr_dl.id
            [ INNER | RIGHT ] JOIN cr_insurance cr_in ON
            idmt.cr_insurance_id = cr_in.id
        )SELECT
            [ select_items ]FROM
            matched_ids
;

```

考量事項

對於 ID 映射資料表查詢結構和語法，請注意下列事項：

- 您無法編輯它。
- 根據預設，它會套用至 ID 映射表。
- 它在協同合作中使用來源和目標 ID 命名空間關聯。
- 根據預設，ID 映射表會設定為為來自 ID 命名集的資料欄提供預設保護。您可以修改此組態，以便在查詢中的任何位置允許來自 ID 命名空間 (sourceID 或 targetID) 的資料欄。如需詳細資訊，請參閱 [中的 ID 命名空間 AWS Clean Rooms](#)。
- ID 映射資料表分析規則會繼承協同合作中其他分析規則的 SQL 限制。

ID 映射表分析規則查詢控制項

使用 ID 映射資料表查詢控制項，AWS Clean Rooms 控制資料表中的資料欄如何用於查詢資料表。例如，它會控制哪些資料欄用於聯結，以及哪些資料欄需要重疊。ID 映射資料表分析規則也包含的功能可讓您允許投影 sourceID、targetID 或兩者，而無需 JOIN。

下表說明每個控制項。

控制項	定義	Usage
<code>joinColumns</code>	可以查詢的成員可以在 INNER JOIN 陳述式中使用的資料欄。	除了 INNER JOIN 之外，您無法 <code>joinColumns</code> 在查詢的任何其他部分中使用。 如需詳細資訊，請參閱 聯結控制項 。
<code>dimensionColumns</code>	可在 SELECT 和 GROUP BY 陳述式中查詢的成員可以使用的資料欄（如果有）。	<code>dimensionColumn</code> 可用於 SELECT 和 GROUP BY。 <code>dimensionColumn</code> 可以顯示為 <code>joinKeys</code> 。 只有在括號 <code>dimensionColumns</code> 中指定 JOIN 子句時，才能在 JOIN 子句中使用。
<code>queryConstraints:RequireOverlap</code>	ID 映射表中必須聯結的資料欄，才能執行查詢。	這些資料欄必須用來加入 ID 映射資料表和協同合作資料表。

ID 映射表分析規則預先定義的結構

ID 映射表分析規則的預先定義結構隨附套用到 `sourceID` 和 `targetID` 的預設保護。這表示在查詢中必須使用套用保護的欄。

您可以透過下列方式設定 ID 映射表分析規則：

- `sourceID` 和 `targetID` 受保護

在此組態中，`sourceID` 和 `targetID` 無法同時投影。參考 ID 映射表時，`sourceID` 和 `targetID` 必須在 JOIN 中使用。

- 僅 `targetID` 受保護

在此組態中，`targetID` 無法投影。參考 ID 映射表時，`targetID` 必須在 JOIN 中使用。`sourceID` 可用於查詢。

- 僅sourceID受保護

在此組態中，sourceID 無法投影。參考 ID 映射表時，sourceID必須在 JOIN 中使用。targetID 可用於查詢。

- 既不sourceID受保護也不targetID受保護

在此組態中，ID 映射表不受可在查詢中使用的任何特定強制執行約束。

下列範例顯示 ID 映射資料表分析規則的預先定義結構，其預設保護會套用至 sourceID和 targetID。在此範例中，ID 映射資料表分析規則僅允許資料sourceID欄和targetID資料欄上的 INNER JOIN。

```
{
  "joinColumns": [
    "source_id",
    "target_id"
  ],
  "queryConstraints": [
    {
      "requireOverlap": {
        "columns": [
          "source_id",
          "target_id"
        ]
      }
    }
  ],
  "dimensionColumns": [] // columns that can be used in SELECT and JOIN
}
```

下列範例顯示 ID 映射資料表分析規則的預先定義結構，其中包含套用至 的保護targetID。在此範例中，ID 映射資料表分析規則僅允許資料sourceID欄上的 INNER JOIN。

```
{
  "joinColumns": [
    "source_id",
    "target_id"
  ],
  "queryConstraints": [
    {
      "requireOverlap": {
```

```

        "columns": [
            "target_id"
        ]
    }
},
"dimensionColumns": [
    "source_id"
]
}

```

下列範例顯示 ID 映射資料表分析規則的預先定義結構，其中包含套用至 的保護sourceID。在此範例中，ID 映射資料表分析規則僅允許資料targetID欄上的 INNER JOIN。

```

{
  "joinColumns": [
    "source_id",
    "target_id"
  ],
  "queryConstraints": [
    {
      "requireOverlap": {
        "columns": [
          "source_id"
        ]
      }
    }
  ],
  "dimensionColumns": [
    "target_id"
  ]
}

```

下列範例顯示未將保護套用至 sourceID或 的 ID 映射資料表分析規則的預先定義結構targetID。在此範例中，ID 映射資料表分析規則允許資料sourceID欄和targetID資料欄上的 INNER JOIN。

```

{
  "joinColumns": [
    "source_id",
    "target_id"
  ],
  "queryConstraints": [

```

```
{
  "requireOverlap": {
    "columns": []
  }
},
"dimensionColumns": [
  "source_id",
  "target_id"
]
}
```

ID 映射表分析規則 – 範例

例如，公司可以使用 ID 映射表分析規則來使用多方 LiveRamp 轉碼，而不是撰寫參考個人身分識別資訊 (PII) 的長瀑布陳述式。下列範例示範如何使用 ID AWS Clean Rooms 映射資料表分析規則在 中協作。

A 公司是擁有客戶和銷售資料的廣告商，將用作來源。A 公司也會代表合作中的各方執行轉碼，並帶來 LiveRamp 登入資料。

B 公司是具有事件資料的發佈者，將用作目標。

Note

A 公司或 B 公司都可以提供 LiveRamp 轉碼登入資料並執行轉碼。

若要建立協同合作以啟用協同合作 ID 映射表分析，公司會執行下列動作：

1. A 公司建立協同合作並建立成員資格。它新增了 B 公司，該公司也會在協同合作中建立成員資格。
2. 公司 A 關聯現有的 ID 命名空間來源，或使用 AWS Clean Rooms 主控台 AWS Entity Resolution 在中建立新的 ID 命名空間來源。

公司 A 會建立已設定的資料表及其銷售資料，並在 ID 映射資料表sourceId中建立索引鍵為 的資料欄。

ID 命名空間來源提供要轉碼的資料。

3. B 公司會關聯現有的 ID 命名空間目標，或使用 AWS Clean Rooms 主控台 AWS Entity Resolution 在中建立新的 ID 命名空間目標。

B 公司會建立已設定的資料表，其中包含其事件資料，以及 ID 映射資料表targetId中索引鍵為 的資料欄。

ID 命名空間目標不提供要轉碼的資料，僅提供 LiveRamp 組態周圍的中繼資料。

4. A 公司探索與協同合作相關聯的兩個 ID 命名空間，並建立和填入 ID 映射表。

5. A 公司透過加入 ID 映射表，跨兩個資料集執行查詢。

```
--- this would be valid for Custom or List
SELECT provider.data_col, consumer.data_col
FROM provider
  JOIN idMappingTable-123123123123-myMappingWFName idmt
    ON provider.id = idmt.sourceId
JOIN consumer
  ON consumer.id = idmt.targetId
```

AWS Clean Rooms 差異隱私權

AWS Clean Rooms 差異隱私權可協助您透過數學支援的技巧來保護使用者的隱私權，只需按幾下滑鼠，即可透過直覺式控制實作。作為全受管功能，不需要先前的差異隱私權體驗來協助您防止使用者重新識別。AWS Clean Rooms 會自動新增仔細校正的雜訊量，以在執行時間查詢結果，以協助保護您的個人層級資料。

AWS Clean Rooms 差異隱私權支援各種分析查詢，非常適合各種使用案例，其中查詢結果中的少量錯誤不會影響分析的實用性。有了它，您的合作夥伴就可以產生有關廣告行銷活動、投資決策、臨床研究等的業務關鍵洞見，而不需要您合作夥伴的任何額外設定。

AWS Clean Rooms 差異隱私權可防止以惡意方式使用純量函數或數學運算子符號的溢位或無效轉換錯誤。

如需 AWS Clean Rooms 差異隱私權的詳細資訊，請參閱下列主題。

主題

- [差異隱私權](#)
- [中的差異隱私權如何 AWS Clean Rooms 運作](#)
- [差異性隱私權政策](#)
- [AWS Clean Rooms 差異隱私權的 SQL 功能](#)

- [不同的隱私權查詢秘訣和範例](#)
- [AWS Clean Rooms 差異性隱私權的限制](#)

差異隱私權

差異隱私權僅允許彙總洞見，並混淆這些洞見中任何個人資料的貢獻。差異隱私權可保護協作資料，使其不會收到了解特定個人的結果。如果沒有差異隱私權，可以接收結果的成員可以嘗試透過新增或移除有關個人的記錄來推斷個別使用者資料，並觀察查詢結果的差異。

開啟差異隱私權時，指定的雜訊量會新增至查詢結果，以混淆個別使用者的貢獻。如果可接收結果的成員在從資料集中移除有關個人的記錄後嘗試觀察查詢結果的差異，查詢結果的變異性有助於防止識別個人的資料。AWS Clean Rooms 差異隱私權使用 [SampCert](#) 採樣器，這是一種經過驗證的正確採樣器實作，由 開發 AWS。

中的差異隱私權如何 AWS Clean Rooms 運作

在 中開啟差異隱私權[的工作流程](#)，在[完成工作流程 AWS Clean Rooms](#)時 AWS Clean Rooms 需要下列額外步驟：

1. 您可以在新增[自訂分析規則](#)時開啟差異隱私權。
2. [您可以為協同合作設定差異隱私權政策](#)，以使用差異隱私權來保護資料表以供查詢。

完成這些步驟後，可以查詢的成員可以開始對差異隱私權保護的資料執行查詢。AWS Clean Rooms 會傳回符合差異隱私權政策的結果。AWS Clean Rooms 差異隱私權會追蹤您可以執行的剩餘查詢預估數量，類似於顯示車輛目前油位的汽車氣位表。可以查詢的成員可以執行的查詢數量受限於 中設定的每個查詢參數的隱私權預算和新增的雜訊[差異性隱私權政策](#)。

考量事項

在 中使用差異隱私權時 AWS Clean Rooms，請考慮下列事項：

- 可以接收結果的成員無法使用差異隱私權。他們將針對其設定的資料表，設定關閉差異隱私權的自訂分析規則。
- 當兩個或多個資料提供者都開啟差異隱私權時，可以查詢的成員無法聯結來自兩個或多個資料提供者的資料表。

差異性隱私權政策

差異隱私權政策控制在協同合作中允許可查詢的成員執行多少彙總函數。隱私權預算定義了一個常見的有限資源，可套用協同合作中的所有資料表。每個查詢新增的雜訊會管理耗盡隱私權預算的速率。

需要差異隱私權政策，才能讓您的差異隱私權受保護資料表可供查詢。這是協作中的一次性步驟，包含兩個輸入：

- 隱私權預算 – 根據 epsilon 量化，隱私權預算控制隱私權保護的層級。這是常見的有限資源，適用於在協同合作中受到差異隱私權保護的所有資料表，因為目標是保留使用者隱私，其資訊可以存在於多個資料表中。

每次在資料表上執行查詢時，都會耗用隱私權預算。當隱私權預算用盡時，可以查詢的協同合作成員在增加或重新整理之前無法執行其他查詢。透過設定更大的隱私權預算，可以接收結果的成員可以減少他們對資料中個人的不確定性。選擇在諮詢業務決策者之後，平衡協同合作需求與隱私權需求的隱私權預算。

如果您打算定期將新資料帶入協同合作，您可以選取每月重新整理隱私權預算，以每月自動建立新的隱私權預算。選擇此選項可在跨重新整理重複查詢時，顯示有關資料列的任意資訊量。如果在隱私權預算重新整理之間重複查詢相同的資料列，請避免選擇此選項。

- 每個查詢新增的雜訊是根據您要隱藏貢獻的使用者數量來衡量。此值會管理耗盡隱私權預算的速率。較大的雜訊值可減少耗盡隱私權預算的速率，因此允許對資料執行更多查詢。不過，這應與發佈不準確的資料洞見相平衡。設定此值時，請考慮協作洞察所需的準確性。

您可以使用預設差異隱私權政策，根據您的使用案例快速完成設定或自訂差異隱私權政策。AWS Clean Rooms 差異隱私權提供直覺式控制來設定政策。AWS Clean Rooms 差異隱私權可讓您根據資料上所有查詢中可能的彙總數目預覽公用程式，並預估在資料協同合作中可以執行多少查詢。

您可以使用互動式範例，了解每個查詢新增的不同隱私權預算和雜訊值如何影響不同類型的 SQL 查詢的結果。一般而言，您需要平衡隱私權需求與要允許的查詢數量，以及這些查詢的準確性。每個查詢新增的較小隱私權預算或較大雜訊可以更好地保護使用者隱私權，但為您的協作合作夥伴提供較不有意義的洞見。

如果您增加隱私權預算，同時保持每個查詢參數新增的雜訊相同，則可以查詢的成員可以在協同合作中的資料表上執行更多彙總。您可以在協同合作期間隨時增加隱私權預算。如果您減少隱私權預算，同時保持每個查詢參數新增的雜訊相同，可以查詢的成員可以執行較少的彙總。在可以查詢的成員開始分析您的資料之後，您無法減少隱私權預算。

如果您增加每個查詢新增的雜訊，同時保持隱私權預算輸入相同，則可以查詢的成員可以在協同合作中對資料表執行更多彙總。如果您減少每個查詢新增的雜訊，同時保持隱私權預算輸入相同，可以查詢的成員可以執行較少的彙總。您可以在協同合作期間隨時增加或減少每個查詢新增的雜訊。

差異隱私權政策是由隱私權預算範本 API 動作管理。

AWS Clean Rooms 差異隱私權的 SQL 功能

AWS Clean Rooms 差異隱私權使用一般用途查詢結構來支援複雜的 SQL 查詢。根據此結構驗證自訂分析範本，以確保它們可以在受差異隱私權保護的資料表上執行。下表指出支援哪些 函數。如需詳細資訊，請參閱[查詢結構和語法](#)。

Category	類別
SELECT	檢取
表	表
SQLs)	SQLs)
建	建
構	構
SUMMED VALUE	加總值
with greatest	最大值
the:	的數
Average	平均數
APPROXIMATE COUNT,	估計數量，
COUNT PERCENTILE DISC,	COUNT PERCENTILE DISC,
STDDEV,	標準差，
and	和
summary AVG	摘要平均值
protected tables must result	受保護的表格必須產生結果

Category
表
欄
SELECT
群
數
表
權
成
的
\$(CLES)
建
構

inCOUNT
data和
withCOUNT
usedDISTINCT
levels
| 數
records.
You
數
should
write
• MEDIAN
the
函
SELECT
數
expressio
• MIN
n
函
in
數
those
• PERCENTIL
CTEs
using
`SELECT
數
userIdent
• STDDEV_S
inMP.'
和
format.
STDDEV_PO
P

Category

總

樞ELECT

研

數

表

權

的

(SCLs)

建

構

函

數

- SUM

和

SUM

DISTINCT

函

數

- VAR_SAMP

和

VAR_POP

函

數

構

write

format.

Category

類別

ELECT

群

表

表

權

的

(SCLs)

建

構

SELECT

HAVING

have

• JOIN

any

• JOIN

subquery

條

that

件

doesn't

• FROM

reference

WHERE

different

ial

privacy

relations

in

these

construct

s.

You

can

have

any

subquery

that

reference

s

Category
SELECT
FROM
JOIN
TABLES
SCHEMAs)
different
ial
privacy
relations
in
a
FROM
and
JOIN
clause
only.

Category	
查詢	
欄	
ELECT	
群	
數	
表	
權	
的	
(CQLs)	
建	
構	
內	
INNER	
權	
的	
LEFT	
條	
JOIN	
件	
• 左	
是	
半	
在	
聯	
使	
結	
用	
• 左	
者	
識	
ANTI	
加	
別	
符	
• 資	
RIGHT	
料	
JOIN	
欄	
完	
上	
整	
僅	
加	
支	
入	
援	
【JOIN】	
等	
OR	
同	
運	
聯	
算	
結	
子	

Category

終

樑
ELECT

研

數

表

權

的

SCLEs)

建

構

的CROSS

JOIN

函

數，

並

且

在

查

詢

開

啟

差

異

隱

私

權

的

兩

個

或

多

個

資

料

表

資

Category
SELECT
TABLES)
建
構
料
表
擁
有
者
已
在
所
有
資
料
表
中
設
定
相
同
的
使
用
者
識
別
符
欄 ,

Category

鑑

ELECT

研

製

表

權

的

SCIEs)

建

構

以便使用者的定義在資料表之間保持一致。

在開啟差異隱

Category
表
ELECT
聯
表
權
表
權
的
SCLEs)
建
構
私
權
的
情
況
下
結
合
兩
個
或
多
個
關
係
時，
不
支
援
CROSS
JOIN
函
數。

Category

類別

ELECT

群

表

表

權

的

(SCLs)

建

構

UNION,

UNIONed

ALL,

INTERSECT

”

EXCEPT

||

MINUS

(these

are

synonyms))

Category
窗格
ELECT
群
數
表
權
的
(SCLs)
建
構

Window
權限
Supported
Supported
數
the
• AVG
condition
節
that
圍
the
函
user
數
identifie
↑ COUNT
節
column
in 圍
the
函
數
window
function
COUNT_DIST
s 節
partition
class
is 數
required
• DENSE_RANK
when
you
節
query

Category

類別

ELECT

群

數

表

權

的

(SCLs)

建

構

a 函

relation

with FIRST_VAL

different

ial

privacy

turned

on

• LAG

範

圍

函

數

• LAST_VALU

E

範

圍

函

數

• LEAD

範

圍

函

數

Category

查詢
ELECT
聯
繫
表
撞
的
SCLEs)
建
構

- MAX
視
窗
函
數
- MEDIAN
視
窗
函
數
- MIN
視
窗
函
數
- NTH_VALUE
範
圍
函
數
- STDDEV_SAMP
和
STDDEV_PO

構

詞)

- 數

- 函

Category
數
據
庫
ELECT
研
究
數
據
表
權
限
的
SCLEs)
建
構

數
(VAR_SAMP
和
VARIANCE
是
同
義
詞)

排
名
函
數

- DENSE_RANK
範
圍
函
數
- NTILE
範
圍
函
數

Category
查詢
樹
ELECT
聯
繫
表
攪
的
SCLEs)
建
構

- PERCENT_R
ANK
範
圍
函
數
- RANK
範
圍
函
數
- ROW_NUMBE
R
範
圍
函
數

Category

查詢
ELECT
群
數
表
權
的
(SCLs)
建
構

AND/AS/En

and
supported
ns表
達
式

- COALESCE
表
達
式
- GREATEST
和
LEAST
函
數
- NVL
和
COALESCE
函
數
- NVL2
函
數

Category
終
極
ELECT
研
究
表
據
的
(SCLs)
建
構
• NULLIF
函
數

Category

類別
ELECT
群
羣
表
權
的
(SCLs)
建
構

Condition

條件
Supported
can't
be
used
because
they
require
subquerie
s.
All
others
are
supported
• BETWEEN
•
範圍
圍
條件
件
• Null
條件
件

Category	類別
ELECT	電
研	研
數	數
表	表
權	權
的	的
SCLEs)	SCLEs)
建	建
構	構
File-	File-
time	時間
supported	支援
日期	日期
期	期
與	與
時	時
間	間
函	函
數	數
• 串	• 串
連	連
運	運
算	算
子	子
• ADD_MONTH	• ADD_MONTH
S	S
函	函
數	數
• CONVERT_T	• CONVERT_T
IMEZONE	IMEZONE
函	函
數	數

Category

算

術

函

數

表

權

的

(SCLs)

建

構

- CURRENT_DATE
函
數
- DATEADD
函
數
- DATEDIFF
函
數
- DATE_PART
函
數
- DATE_TRUNC
函
數
- EXTRACT
函
數
- TO_TIMESTAMP
函
數

構

- 日期或時間戳記函數的日期部分

Category

查詢
ELECT
算
表
權
的
(SCLs)
建
構

SQL
functions
supported

運
算
子

- BTRIM
函
數
- CHAR_LENGTH
函
數
- CHARACTER_LENGTH
函
數
- CONCAT
函
數
- LEFT
和
RIGHT

Category

SQL
函數
ELECT
表
表
權
表
權
的
(SCLs)
建
構

函
數

- LEN

函
數

- LENGTH

函
數

- LOWER

函
數

- LPAD

和
RPAD
函
數

- LTRIM

函
數

- POSITION

函
數

構

-
- 77

Category

查詢
ELECT
算
表
權
的
(CIEs)
建
構

- REVERSE
函
數
- RTRIM
函
數
- SPLIT_PART
函
數
- SUBSTRING
函
數
- TRANSLATE
函
數
- TRIM
函
數
- UPPER
函
數

構

functions

- 79

Category

算
術
函
數
表
達
式
(SCLs)
建
構

AWSES_ENCRY

are

supported

AWSES_DECRY

PT

- ENCODE
- DECODE
- MD5

函
數

- SHA1

函
數

- SHA2

函
數

- XX_HASH64

Mathemati

are

supported

Symbols

and

@

Category

數據
庫
ELECT
算
術
表
達
式
(SCLs)
建
構

ABS
functions
supported

- ACOS
函
數
- ASIN
函
數
- ATAN
函
數
- ATAN2
函
數
- CBRT
函
數
- CEILING
(或
CEIL)
函
數

Category

算

算ELECT

算

算

表

據

的

(SCLs)

建

構

- COS
函
數
- COT
函
數
- DEGREES
函
數
- LTRIM
函
數
- EXP
函
數
- FLOOR
函
數
- LN
函
數

Category

算
術
函
數
ELECT
算
術
表
達
式
(SCLs)
建
構

- LOG
函
數
- MOD
函
數
- PI
函
數
- POWER
函
數
- RADIANS
函
數
- RANDOM
函
數
- ROUND
函
數

構

-
- 84

Category
Database
ELECT
Join
Query
Table
Join
Test
(SCTEs)
Join
Join

ARRAY,
functions
supported
• HEX

• HLL_SKETCH
H_AGG,

• HLL_SKETCH
H_ESTIMATE
E

• HLL_UNION

• HLL_UNION
_AGG

ADD_JSON
are
GET_JSON_
supported
OBJECT

Category
Database
ELECT
Join
Query
Table
Join
Test
(SCTEs)
Build
Structure

Not supported
Not supported

- ARRAY_DISTINCT
- ARRAY_EXCEPT
- ARRAY_INTERSECT
- ARRAY_JOIN
- ARRAY_REMOVE
- ARRAY_SORT
- ARRAY_UNION

Not supported
Not supported
ROLLUP,
CUBE

Category
支持
ELECT
群
表
表
權
的
(SCLs)
建
構

Supported
Operation
the
condition
that
an
ORDER
BY
clause
is
only
supported
in
a
window
function'
s
partition
clause
when
querying
tables
with
different

Category		
範圍函數		
ELECT		
算術		
表		
權		
的		
\$(CIEs)		
建		
構		
Supported functions on aggregate functions		
Supported functions within aggregate functions		

不支援 SQL 建構的常見替代方案

Category	SQL 建構	備用
範圍函數	• LISTAGG • PERCENTILE_CONT • PERCENTILE_DISC	You can use the equivalent aggregate function with GROUP BY.
Mathematical operator symbols	• \$column / 2 • \$column / 2 • \$column ^ 2	• CBRT • SQRT • POWER(\$column , 2)

Category	SQL 建構	備用
Scalar functions	• SYSDATE • \$column : integer • convert(type , \$column)	• CURRENT_DATE • CAST \$column AS 整數 • CAST \$column AS 類型
Literals	INTERVAL '1 SECOND'	INTERVAL '1' SECOND
Row limiting	TOP n	LIMIT n
Join	• USING • NATURAL	ON clause should explicitly contain a join criterion.

不同的隱私權查詢秘訣和範例

AWS Clean Rooms 差異隱私權使用 [一般用途查詢結構](#) 來支援各種 SQL 建構，例如用於資料準備的通用資料表表達式 (CTEs)，以及常用的彙總函數，例如 COUNT、或 SUM。為了透過新增雜訊以在執行時間彙總查詢結果來混淆資料中任何可能使用者的貢獻，AWS Clean Rooms 差異隱私權要求最終的彙總函數 SELECT statement 在使用者層級資料上執行。

下列範例使用兩個名為 socialco_impressions 和 socialco_users 的資料表，來自媒體發佈者，他們想要使用差異隱私權保護資料，同時與運動品牌 athletic_brand_sales 資料協作。媒體發佈者已將資料 user_id 欄設定為使用者識別符資料欄，同時啟用差異隱私權 AWS Clean Rooms。廣告商不需要差異隱私權保護，並想要在合併資料上使用 CTEs 執行查詢。由於其 CTE 使用差異隱私權保護資料表，廣告商會在 CTE 資料欄清單中包含那些受保護資料表的使用者識別符資料欄，並聯結使用者識別符資料欄上的受保護資料表。

```
WITH matches_table AS(
  SELECT si.user_id, si.campaign_id, s.sale_id, s.sale_price
  FROM socialco_impressions si
  JOIN socialco_users su
    ON su.user_id = si.user_id
  JOIN athletic_brand_sales s
    ON s.emailsha256 = su.emailsha256
  WHERE s.timestamp > si.timestamp

UNION ALL

  SELECT si.user_id, si.campaign_id, s.sale_id, s.sale_price
```

```

FROM socialco_impressions si
JOIN socialco_users su
    ON su.user_id = si.user_id
JOIN athletic_brand_sales s
    ON s.phonesha256 = su.phonesha256
WHERE s.timestamp > si.timestamp
)

SELECT COUNT (DISTINCT user_id) as unique_users
FROM matches_table
GROUP BY campaign_id
ORDER BY COUNT (DISTINCT user_id) DESC
LIMIT 5

```

同樣地，如果您想要在差異隱私權保護資料表上執行視窗函數，則必須在 `PARTITION BY` 子句中包含使用者識別符欄。

```

ROW_NUMBER() OVER (PARTITION BY conversion_id, user_id ORDER BY match_type, match_age)
AS row

```

AWS Clean Rooms 差異性隱私權的限制

AWS Clean Rooms 差異隱私權不會處理下列情況：

1. AWS Clean Rooms 差異隱私權僅支援使用 Amazon S3 後端 AWS Glue 資料表的查詢。它不支援使用 Snowflake 或 Amazon Athena 資料表進行查詢。
2. AWS Clean Rooms 差異隱私權不會處理計時攻擊。例如，在個別使用者貢獻大量資料列，並新增或移除此使用者會大幅變更查詢運算時間的情況下，這些攻擊是可能的。
3. AWS Clean Rooms 當 SQL 查詢可能因使用特定 SQL 建構而導致執行時間溢位或無效的轉換錯誤時，差異隱私權並不保證差異隱私權。

下表列出一些但並非全部的 SQL 建構，這些建構可能會產生執行時間錯誤，並且應在分析範本中驗證。我們建議您核准分析範本，以將此類執行時間錯誤的機率降至最低，並定期檢閱查詢日誌，以判斷查詢是否符合協同合作協議。

下列 SQL 建構容易發生溢位錯誤：

Category	SQL 建構易受 Spark SQL 分析引擎中的溢位錯誤影響
Aggregate functions	• AVG • SUM/SUM_DISTINCT
Data type formatting functions	• TO_TIMESTAMP • TO_DATE
Date and time functions	• ADD_MONTHS • DATEADD • DATEDIFF
Math functions	• +, -, *, / • POWER
String functions	• • CONCAT • REPEAT
Window functions	• AVG • SUM

4. CAST 資料類型格式化函數容易發生無效的轉換錯誤。

您可以設定 [CloudWatch 為日誌群組建立指標篩選條件](#)，然後在該指標篩選條件上[建立 CloudWatch 警示](#)，以便在遇到潛在溢位或轉換錯誤時接收提醒。

具體而言，您應該監控錯誤代碼 `CastError`、`OverflowError`、`ConversionError`。這些錯誤代碼的存在表示潛在的側通道攻擊，但可能表示錯誤的 SQL 查詢。

如需詳細資訊，請參閱[分析登入 AWS Clean Rooms](#)。

AWS 無塵室 ML

AWS Clean Rooms ML 允許兩方或多方對其資料執行機器學習模型，而不需要彼此共用其資料。此服務提供隱私權增強控制，允許資料擁有者保護其資料及其模型 IP。您可以使用 AWS 撰寫的模型或自帶自訂模型。

如需如何運作的詳細說明，請參閱 [跨帳戶任務](#)。

如需 Clean Rooms ML 模型功能的詳細資訊，請參閱下列主題。

主題

- [AWS Clean Rooms ML 術語](#)
- [AWS Clean Rooms ML 如何與 AWS 模型搭配使用](#)
- [AWS Clean Rooms ML 如何與自訂模型搭配使用](#)
- [AWS Clean Rooms ML 中的模型](#)
- [Clean Rooms ML 中的自訂模型](#)

AWS Clean Rooms ML 術語

使用 Clean Rooms ML 時，請務必了解下列術語：

- 訓練資料提供者 – 提供訓練資料的一方，建立和設定類似模型，然後將類似模型與協同合作建立關聯。
- 種子資料提供者 – 貢獻種子資料、產生外觀客群，以及匯出其外觀客群的一方。
- 訓練資料 – 訓練資料提供者的資料，用於產生類似模型。訓練資料用於測量使用者行為中的相似性。

訓練資料必須包含使用者 ID、項目 ID 和時間戳記欄。或者，訓練資料可以包含數值或分類特徵的其他互動。互動的範例包括觀看的影片、購買的項目或文章讀取的清單。

- 種子資料 – 種子資料提供者的資料，用於建立外觀相似的區段。種子資料可以直接提供，也可以來自 AWS Clean Rooms 查詢的結果。類客群輸出是一組來自訓練資料的使用者，最接近種子使用者。
- Lookalike 模型 – 訓練資料的機器學習模型，用於在其他資料集中找到類似的使用者。

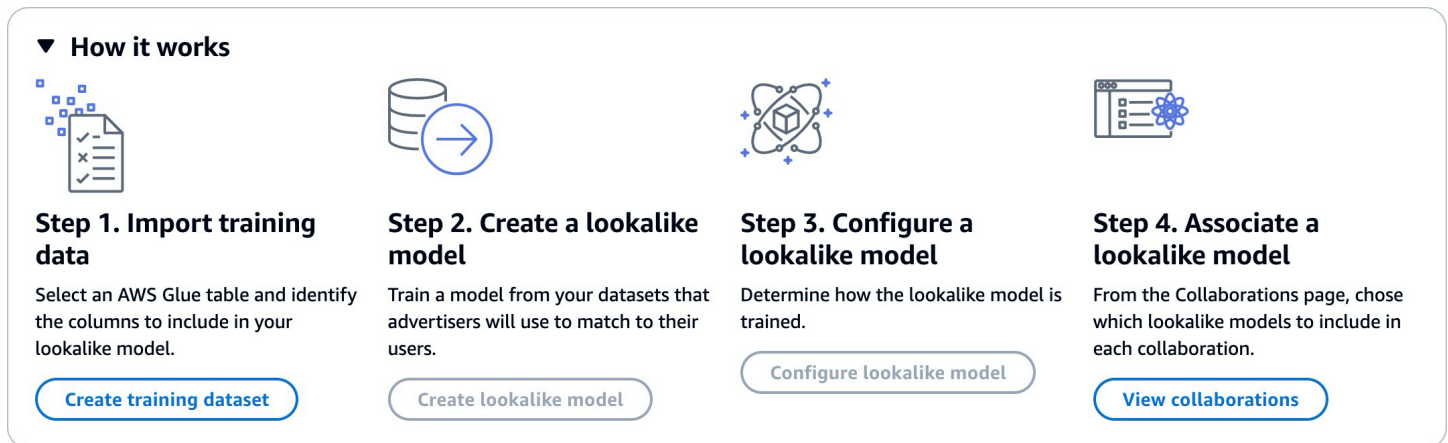
使用 API 時，對象模型一詞會同等用於看起來像模型。例如，您可以使用 [CreateAudienceModel](#) API 來建立類似模型。

- Lookalike 區段 – 訓練資料的子集，最接近種子資料。

使用 API 時，您可以使用 [StartAudienceGenerationJob](#) API 建立類似樣的客群。

訓練資料提供者的資料絕不會與種子資料提供者共用，而種子資料提供者的資料也絕不會與訓練資料提供者共用。看起來像區段的輸出會與訓練資料提供者共用，但永遠不會與種子資料提供者共用。

AWS Clean Rooms ML 如何與 AWS 模型搭配使用



使用類似模型需要兩個方，訓練資料提供者和種子資料提供者，依序在 中工作 AWS Clean Rooms ，將他們的資料帶入協同合作。這是訓練資料提供者必須先完成的工作流程：

1. 訓練資料提供者的資料必須存放在使用者項目互動 AWS Glue 的資料目錄資料表中。訓練資料至少必須包含使用者 ID 欄、互動 ID 欄和時間戳記欄。
2. 訓練資料提供者向 註冊訓練資料 AWS Clean Rooms。
3. 訓練資料提供者會建立外觀相似的模型，可與多個種子資料提供者共用。外觀模型是一種深度神經網路，最多可能需要 24 小時才能訓練。它不會自動重新訓練，我們建議您每週重新訓練模型。
4. 訓練資料提供者會設定類似模型，包括是否共用相關性指標和輸出區段的 Amazon S3 位置。訓練資料提供者可以從單一外觀模型建立多個已設定的外觀模型。
5. 訓練資料提供者會將設定的對象模型與與種子資料提供者共用的協同合作建立關聯。

這是種子資料提供者接下來必須完成的工作流程：

1. 種子資料提供者的資料可以存放在 Amazon S3 儲存貯體中，也可以來自查詢結果。
2. 種子資料提供者會開啟他們與訓練資料提供者共用的協同合作。
3. 種子資料提供者會從協作頁面的 Clean Rooms ML 索引標籤建立外觀相似的區段。

4. 如果已共用關聯性指標，則種子資料提供者可以評估它們，並匯出外觀區段以供外部使用 AWS Clean Rooms。

AWS Clean Rooms ML 如何與自訂模型搭配使用

透過 Clean Rooms ML，協同合作的成員可以使用存放在 Amazon ECR 中的停駐自訂模型演算法來共同分析其資料。若要這樣做，模型提供者必須建立映像並將其存放在 Amazon ECR 中。請遵循 [Amazon Elastic Container Registry 使用者指南](#) 中的步驟，建立將包含自訂 ML 模型的私有儲存庫。

協同合作的任何成員都可以是模型提供者，只要他們具有正確的許可。協同合作的所有成員都可以為模型提供訓練資料、推論資料或兩者。基於本指南的目的，貢獻資料的成員稱為資料提供者。建立協同合作的成員是協同合作建立者，而此成員可以是模型提供者、其中一個資料提供者，或兩者。

在最高層級，以下是執行自訂 ML 建模時必須完成的步驟：

1. 協同合作建立者會建立協同合作，並為每個成員指派適當的成員能力和付款組態。協同合作建立者必須指派成員接收模型輸出或接收推論結果的能力給此步驟中的適當成員，因為它無法在協同合作建立後更新。如需詳細資訊，請參閱 [在 AWS Clean Rooms ML 中建立和加入協同合作](#)。
2. 模型提供者會設定其容器化 ML 模型並將其與協同合作建立關聯，並確保針對匯出的資料設定隱私權限制。如需詳細資訊，請參閱 [在 AWS Clean Rooms ML 中設定模型演算法](#)。
3. 資料提供者將其資料提供給協同合作，並確保已指定其隱私權需求。資料提供者必須允許模型存取其資料。如需詳細資訊，請參閱 [在 AWS Clean Rooms ML 中貢獻訓練資料](#) 及 [在 AWS Clean Rooms ML 中關聯設定的模型演算法](#)。
4. 協同合作成員會建立 ML 組態，定義模型成品或推論結果的匯出位置。
5. 協同合作成員會建立 ML 輸入通道，提供訓練容器或推論容器的輸入。ML 輸入通道是一種查詢，可定義要在模型演算法內容中使用的資料。
6. 協作成員使用 ML 輸入通道和設定的模型演算法叫用模型訓練。如需詳細資訊，請參閱 [在 AWS Clean Rooms ML 中建立訓練過的模型](#)。
7. (選用) 模型訓練器會叫用模型匯出任務，並將模型成品傳送至模型結果接收者。只有具有有效 ML 組態和能夠接收模型輸出的成員才能接收模型成品。如需詳細資訊，請參閱 [從 AWS Clean Rooms ML 匯出模型成品](#)。
8. (選用) 協同合作成員使用 ML 輸入通道、訓練模型 ARN 和推論設定的模型演算法叫用模型推論。推論結果會傳送至推論輸出接收器。只有具有有效 ML 組態和能夠接收推論輸出的成員才能接收推論結果。

以下是模型提供者必須完成的步驟：

1. 建立與 SageMaker AI 相容的 Amazon ECR Docker 映像。Clean Rooms ML 僅支援 SageMaker AI 相容 Docker 映像。
2. 建立 SageMaker AI 相容 Docker 映像之後，請將映像推送至 Amazon ECR。遵循 [Amazon Elastic Container Registry 使用者指南](#) 中的指示來建立容器訓練映像。
3. 設定模型演算法以用於 Clean Rooms ML。
 - a. 提供設定模型演算法所需的 Amazon ECR 儲存庫連結和任何引數。
 - b. 提供服務存取角色，允許 Clean Rooms ML 存取 Amazon ECR 儲存庫。
 - c. 將設定的模型演算法與協同合作建立關聯。這包括提供隱私權政策，定義容器日誌、故障日誌、CloudWatch 指標的控制，以及可從容器結果匯出多少資料的限制。

以下是資料提供者必須完成的步驟，才能與自訂 ML 模型協作：

1. 使用自訂分析規則設定現有 AWS Glue 資料表。這允許一組特定的預先核准查詢或預先核准的帳戶使用您的資料。
2. 將設定的資料表與協同合作建立關聯，並提供可存取 AWS Glue 資料表的服務存取角色。
3. [將協同合作分析規則](#) 新增至資料表，以允許設定的模型演算法關聯存取設定的資料表。
4. 在 Clean Rooms ML 中關聯和設定模型和資料之後，具有執行查詢功能的成員會提供 SQL 查詢，並選取要使用的模型演算法。

模型訓練完成後，該成員會啟動模型訓練成品或推論結果的匯出。這些成品或結果會傳送給能夠接收訓練模型輸出的成員。結果接收者必須先設定其 `MachineLearningConfiguration` 才能接收模型輸出。

AWS Clean Rooms ML 中的模型

AWS Clean Rooms ML 為雙方提供隱私權保護方法，以識別其資料中的類似使用者，而無需彼此共用其資料。第一方將訓練資料帶到，AWS Clean Rooms 以便他們可以建立和設定類似模型，並將其與協同合作建立關聯。然後，種子資料會帶入協同合作，以建立類似訓練資料的外觀客群。

如需如何運作的詳細說明，請參閱 [跨帳戶任務](#)。

下列主題提供如何在 Clean Rooms ML 中建立和設定 AWS 模型的相關資訊。

主題

- [AWS Clean Rooms ML 的隱私權保護](#)

- [Clean Rooms ML 的訓練資料需求](#)
- [Clean Rooms ML 的種子資料需求](#)
- [AWS Clean Rooms ML 模型評估指標](#)

AWS Clean Rooms ML 的隱私權保護

Clean Rooms ML 旨在降低成員資格推論攻擊的風險，訓練資料提供者可以了解種子資料中的人物，而種子資料提供者可以了解訓練資料中的人物。採取幾個步驟來防止此攻擊。

首先，種子資料提供者不會直接觀察 Clean Rooms ML 輸出，而訓練資料提供者永遠無法觀察種子資料。種子資料提供者可以選擇在輸出區段中包含種子資料。

接著，從訓練資料的隨機樣本建立外觀模型。此範例包含大量不符合種子受眾的使用者。此程序可讓您更難判斷使用者是否不在資料中，這是成員資格推論的另一個途徑。

此外，多個種子客戶可用於種子特定樣模型訓練的每個參數。這限制了模型可過度擬合的程度，以及可推斷有關使用者的程度。因此，我們建議種子資料的大小下限為 500 個使用者。

最後，使用者層級指標永遠不會提供給訓練資料提供者，這會消除成員資格推論攻擊的另一個途徑。

Clean Rooms ML 的訓練資料需求

若要成功建立類似模型，您的訓練資料必須符合下列要求：

- 訓練資料必須是 Parquet、CSV 或 JSON 格式。

Note

不支援 Zstandard (ZSTD) 壓縮 Parquet 資料。

- 您的訓練資料必須編製目錄 AWS Glue。如需詳細資訊，請參閱《AWS Glue 開發人員指南》中的 [AWS Glue Data Catalog 入門](#)。我們建議您使用 AWS Glue 爬蟲程式來建立資料表，因為系統會自動推斷結構描述。
- 包含訓練資料和種子資料的 Amazon S3 儲存貯體位於與其他 Clean Rooms ML 資源相同的 AWS 區域。
- 訓練資料必須至少包含 100,000 個唯一使用者 IDs，每個 ID 至少有兩個項目互動。
- 訓練資料必須至少包含 100 萬筆記錄。

- [CreateTrainingDataset](#) 動作中指定的結構描述必須與建立 AWS Glue 資料表時定義的結構描述相符。
- 在 [CreateTrainingDataset](#) 動作中定義了所提供資料表中定義的必要欄位。

欄位類型	支援的資料類型	必要	Description
USER_ID	string , int , bigint	是	資料集中每個使用者的唯一識別符。它應該是非個人身分識別資訊 (PII) 值。這可能是雜湊識別符或客戶 ID。
ITEM_ID	string , int , bigint	是	使用者與之互動的每個項目的唯一識別符。
TIMESTAMP	bigint、int、時間戳記	是	使用者與項目互動的時間。值必須採用秒格式的 Unix epoch 時間。
CATEGORICAL_FEATURE	string , int , float , bigint , double , 布林值 , array	否	擷取與使用者或項目相關的分類資料。這可能包括事件類型 (例

欄位類型	支援的資料類型	必要	Description
			如點選或購買)、使用者人口統計資料 (年齡群組、性別 - 匿名)、使用者位置 (城市、國家 - 匿名)、項目類別 (例如衣服或電子) 或項目品牌。
NUMERICAL_FEATURE	雙、浮點數、int、biint	否	擷取與使用者或項目相關的數值資料。這可能包括使用者購買歷史記錄 (總花費金額)、項目價格、項目造訪次數，或項目的使用者評分。

- 或者，您可以提供總共最多 10 個分類或數值功能。

以下是 CSV 格式的有效訓練資料集範例

```
USER_ID,ITEM_ID,TIMESTAMP,EVENT_TYPE(CATEGORICAL FEATURE),EVENT_VALUE (NUMERICAL FEATURE)
```

```
196,242,881250949,click,15
186,302,891717742,click,13
22,377,878887116,click,10
244,51,880606923,click,20
166,346,886397596,click,10
```

Clean Rooms ML 的種子資料需求

類似模型的種子資料可以直接來自 Amazon S3 儲存貯體或 SQL 查詢的結果。

直接提供的種子資料必須符合下列要求：

- 種子資料必須是具有使用者 IDs 清單的 JSON 行格式。
- 種子大小應介於 25 到 500,000 個唯一的使用者 IDs 之間。
- 種子使用者的最小數量必須符合您建立設定的對象模型時指定的最小相符種子大小值。

以下是 CSV 格式的有效訓練資料集範例

```
{"user_id": "abc"}
{"user_id": "def"}
{"user_id": "ghijkl"}
{"user_id": "123"}
{"user_id": "456"}
{"user_id": "7890"}
```

AWS Clean Rooms ML 模型評估指標

Clean Rooms ML 會計算召回和相關性分數，以判斷模型的效能。Recall 會比較類似資料和訓練資料之間的相似性。相關性分數用於決定受眾應達到多大，而不是模型是否表現良好。

Recall 是衡量類似區段與訓練資料相似度的無偏差指標。Recall 是來自訓練資料範例中最相似的使用者百分比（預設為最相似的 20%），由受眾產生任務包含在種子受眾中。值範圍介於 0-1 之間，較大的值表示更好的對象。大約等於最大儲存貯體百分比的回收值表示對象模型等同於隨機選取。

我們認為這是比準確性、精確度和 F1 分數更好的評估指標，因為在建置模型時，Clean Rooms ML 沒有準確標記真正的負面使用者。

區段層級相關性分數是相似度的指標，其值範圍從 -1（最不相似）到 1（最相似）。Clean Rooms ML 會針對各種客群大小計算一組相關性分數，協助您判斷資料的最佳客群大小。關聯性分數會隨著區

段大小的增加而單調減少，因此隨著區段大小的增加，它可能會與種子資料較不相似。當區段層級相關性分數達到 0 時，模型會預測外觀區段中的所有使用者都來自與種子資料相同的分佈。增加輸出大小可能會包含在看起來像區段中的使用者，這些使用者不是來自與種子資料相同的分佈。

關聯性分數會在單一行銷活動中標準化，不應用於跨行銷活動進行比較。關聯性分數不應用作任何業務成果的單一來源證據，因為除了相關性之外，這些分數還受到多個複雜因素的影響，例如庫存品質、庫存類型、廣告時間等。

關聯性分數不應用來判斷種子的品質，而是是否可以增加或減少。請考量下列範例：

- 所有正分數 – 這表示預測為類似的輸出使用者比包含在類似區段中更多。這對於屬於大型市場一部分的種子資料很常見，例如過去一個月內購買過含括物的每個人。我們建議查看較小的種子資料，例如在過去一個月中購買一次以上的每個消費者。
- 您所需類似區段大小的所有負數分數或負數 – 這表示 Clean Rooms ML 預測所需類似區段大小中沒有足夠的類似使用者。這可能是因為種子資料太具體或市場太小。我們建議將較少的篩選條件套用到種子資料或擴大市場。例如，如果原始種子資料是購買助行器和汽車座位的客戶，您可以將市場擴展到購買多個嬰兒產品的客戶。

訓練資料提供者會判斷是否公開相關性分數，以及計算相關性分數的儲存貯體。

Clean Rooms ML 中的自訂模型

透過 Clean Rooms ML，協同合作的成員可以使用存放在 Amazon ECR 中的停駐自訂模型演算法來共同分析其資料。若要這樣做，模型提供者必須建立映像並將其存放在 Amazon ECR 中。請遵循 [Amazon Elastic Container Registry 使用者指南](#) 中的步驟，建立將包含自訂 ML 模型的私有儲存庫。

協同合作的任何成員都可以是模型提供者，只要他們具有正確的許可。協同合作的所有成員都可以為模型提供資料。基於本指南的目的，貢獻資料的成員稱為資料提供者。建立協同合作的成員是協同合作建立者，而此成員可以是模型提供者、其中一個資料提供者，或兩者。

下列主題說明建立自訂 ML 模型所需的資訊

主題

- [自訂 ML 建模先決條件](#)
- [訓練容器的模型撰寫準則](#)
- [推論容器的模型撰寫準則](#)
- [接收模型日誌和指標](#)

自訂 ML 建模先決條件

在執行自訂 ML 建模之前，您應該考慮下列事項：

- 判斷是否將在協同合作中對訓練模型執行模型訓練和推論。
 - 決定每個協同合作成員將執行的角色，並將適當的能力指派給他們。
 - 將CAN_QUERY功能指派給將訓練模型並在訓練模型上執行推論的成員。
 - CAN_RECEIVE_RESULTS 將指派給至少一位協同合作的成員。
 - 將 CAN_RECEIVE_MODEL_OUTPUT或 CAN_RECEIVE_INFERENCE_OUTPUT 功能指派給將分別接收訓練模型匯出或推論輸出的成員。如果您的使用案例需要這兩種功能，您可以選擇使用這兩種功能。
 - 決定您將允許匯出的訓練模型成品或推論結果的大小上限。
 - 我們建議所有使用者將 CleanroomsFullAccess和 CleanroomsMLFullAccess政策連接到其角色。使用自訂 ML 模型需要使用 AWS Clean Rooms 和 AWS Clean Rooms ML SDKs。
 - 請考慮下列有關 IAM 角色的資訊。
 - 所有資料提供者都必須具有服務存取角色，AWS Clean Rooms 允許 從其 AWS Glue 目錄和資料表以及基礎 Amazon S3 位置讀取資料。這些角色類似於 SQL 查詢所需的角色。這可讓您使用 CreateConfiguredTableAssociation動作。如需詳細資訊，請參閱[建立服務角色以建立設定的資料表關聯](#)。
 - 所有想要接收指標的成員都必須具有服務存取角色，允許他們撰寫 CloudWatch 指標和日誌。Clean Rooms ML 使用此角色，在模型訓練和推論 AWS 帳戶 期間將所有模型指標和日誌寫入成員的。我們也提供隱私權控制，以判斷哪些成員可以存取指標和日誌。這可讓您使用 CreateMLConfiguration動作。如需詳細資訊，請參閱 [建立自訂 ML 建模的服務角色 - ML 組態](#)。
- 接收結果的成員必須為服務存取角色提供寫入其 Amazon S3 儲存貯體的許可。此角色允許 Clean Rooms ML 將結果（訓練過的模型成品或推論結果）匯出至 Amazon S3 儲存貯體。這可讓您使用 CreateMLConfiguration動作。如需詳細資訊，請參閱[建立自訂 ML 建模的服務角色 - ML 組態](#)。
- 模型提供者必須為服務存取角色提供讀取其 Amazon ECR 儲存庫和映像的許可。這可讓您使用 CreateConfigureModelAlgorithm動作。如需詳細資訊，請參閱[建立服務角色以提供自訂 ML 模型](#)。
 - 建立 MLInputChannel 以產生資料集以進行訓練或推論的成員必須提供服務存取角色，允許 Clean Rooms ML 在其中執行 SQL 查詢 AWS Clean Rooms。這可讓您使用

CreateTrainedModel 和 StartTrainedModelInferenceJob 動作。如需詳細資訊，請參閱 [建立服務角色以查詢資料集](#)。

- 模型作者應遵循 [訓練容器的模型撰寫準則](#) 和 [推論容器的模型撰寫準則](#)，以確保模型輸入和輸出依預期進行設定 AWS Clean Rooms。

訓練容器的模型撰寫準則

本節詳細說明模型提供者在為 Clean Rooms ML 建立自訂 ML 模型演算法時應遵循的準則。

- 使用適當的 SageMaker AI 訓練支援的容器基礎映像，如 [SageMaker AI 開發人員指南](#) 中所述。下列程式碼可讓您從公有 SageMaker AI 端點提取支援的容器基礎映像。

```
ecr_registry_endpoint='763104351884.dkr.ecr.$REGION.amazonaws.com'
base_image='pytorch-training:2.3.0-cpu-py311-ubuntu20.04-sagemaker'
aws ecr get-login-password --region $REGION | docker login --username AWS --password-
stdin $ecr_registry_endpoint
docker pull $ecr_registry_endpoint/$base_image
```

- 在本機編寫模型時，請確保下列事項，以便您可以在本機、開發執行個體、中的 SageMaker AI Training AWS 帳戶和 Clean Rooms ML 上測試模型。
- 我們建議您撰寫訓練指令碼，透過各種環境變數存取有關訓練環境的實用屬性。Clean Rooms ML 使用以下引數來調用模型程式碼的訓練：SM_MODEL_DIR、SM_CHANNEL_TRAIN、SM_OUTPUT_DIR 和 FILE_FORMAT。Clean Rooms ML 使用這些預設值，在自己的執行環境中使用來自各方的資料來訓練 ML 模型。
- Clean Rooms ML 可讓您透過 Docker 容器中的 /opt/ml/input/data/*channel-name* 目錄使用訓練輸入通道。每個 ML 輸入通道都會根據其在 CreateTrainedModel 請求中 channel_name 提供的對應進行映射。

```
parser = argparse.ArgumentParser()# Data, model, and output directories

parser.add_argument('--model_dir', type=str, default=os.environ.get('SM_MODEL_DIR',
"/opt/ml/model"))
parser.add_argument('--output_dir', type=str,
default=os.environ.get('SM_OUTPUT_DIR', "/opt/ml/output/data"))
parser.add_argument('--train_dir', type=str,
default=os.environ.get('SM_CHANNEL_TRAIN', "/opt/ml/input/data/train"))
parser.add_argument('--train_file_format', type=str,
default=os.environ.get('FILE_FORMAT', "csv"))
```

- 確保您能夠根據模型程式碼中使用的協作器結構描述產生合成或測試資料集。
- 在建立模型演算法與協同合作的 AWS Clean Rooms 關聯 AWS 帳戶 之前，請確定您可以自行執行 SageMaker AI 訓練任務。

下列程式碼包含與本機測試、SageMaker AI Training 環境測試和 Clean Rooms ML 相容的範例 Docker 檔案

```
FROM 763104351884.dkr.ecr.us-west-2.amazonaws.com/pytorch-training:2.3.0-cpu-
py311-ubuntu20.04-sagemaker
MAINTAINER $author_name

ENV PYTHONDONTWRITEBYTECODE=1 \
    PYTHONUNBUFFERED=1 \
    LD_LIBRARY_PATH="${LD_LIBRARY_PATH}:/usr/local/lib"

ENV PATH="/opt/ml/code:${PATH}"

# this environment variable is used by the SageMaker PyTorch container to determine
our user code directory
ENV SAGEMAKER_SUBMIT_DIRECTORY /opt/ml/code

# copy the training script inside the container
COPY train.py /opt/ml/code/train.py
# define train.py as the script entry point
ENV SAGEMAKER_PROGRAM train.py
ENTRYPOINT ["python", "/opt/ml/code/train.py"]
```

- 為了最佳監控容器故障，基於故障原因，我們建議您匯出日誌和偵錯。在GetTrainedModel回應中，Clean Rooms ML 會在下傳回此檔案的前 1024 個字元StatusDetails。
- 完成任何模型變更，並準備好在 SageMaker AI 環境中進行測試後，請依提供的順序執行下列命令。

```
export ACCOUNT_ID=xxx
export REPO_NAME=xxx
export REPO_TAG=xxx
export REGION=xxx

docker build -t $ACCOUNT_ID.dkr.ecr.us-west-2.amazonaws.com/$REPO_NAME:$REPO_TAG

# Sign into AWS $ACCOUNT_ID/ Run aws configure
# Check the account and make sure it is the correct role/credentials
aws sts get-caller-identity
```

```
aws ecr create-repository --repository-name $REPO_NAME --region $REGION
aws ecr describe-repositories --repository-name $REPO_NAME --region $REGION

# Authenticate Docker
aws ecr get-login-password --region $REGION | docker login --username AWS --password-
stdin $ACCOUNT_ID.dkr.ecr.$REGION.amazonaws.com

# Push To ECR Images
docker push $ACCOUNT_ID.dkr.ecr.$REGION.amazonaws.com$REPO_NAME:$REPO_TAG

# Create Sagemaker Training job
# Configure the training_job.json with
# 1. TrainingImage
# 2. Input DataConfig
# 3. Output DataConfig
aws sagemaker create-training-job --cli-input-json file://training_job.json --region
$REGION
```

SageMaker AI 任務完成且您對模型演算法感到滿意後，您就可以向 AWS Clean Rooms ML 註冊 Amazon ECR 登錄檔。使用 `CreateConfiguredModelAlgorithm` 動作來註冊模型演算法，並使用 `CreateConfiguredModelAlgorithmAssociation` 將其與協同合作建立關聯。

推論容器的模型撰寫準則

本節詳細說明模型提供者在為 Clean Rooms ML 建立推論演算法時應遵循的準則。

- 使用適當的 SageMaker AI 推論支援的容器基礎映像，如 [SageMaker AI 開發人員指南](#) 中所述。下列程式碼可讓您從公有 SageMaker AI 端點提取支援的容器基礎映像。

```
ecr_registry_endpoint='763104351884.dkr.ecr.$REGION.amazonaws.com'
base_image='pytorch-inference:2.3.0-cpu-py311-ubuntu20.04-sagemaker'
aws ecr get-login-password --region $REGION | docker login --username AWS --password-
stdin $ecr_registry_endpoint
docker pull $ecr_registry_endpoint/$base_image
```

- 在本機編寫模型時，請確保下列事項，以便您可以在本機、開發執行個體、中的 SageMaker AI Batch Transform AWS 帳戶和 Clean Rooms ML 上測試模型。
- Clean Rooms ML 透過 Docker 容器中的 `/opt/ml/model` 目錄，讓您的推論模型成品可供推論程式碼使用。

- Clean Rooms ML 會依行分割輸入、使用MultiRecord批次策略，並在每個轉換的記錄結尾新增換行字元。
- 確保您能夠根據模型程式碼中使用的協作者結構描述產生合成或測試推論資料集。
- 在建立模型演算法與協同合作的 AWS Clean Rooms 關聯 AWS 帳戶 之前，請確定您可以自行執行 SageMaker AI 批次轉換任務。

下列程式碼包含與本機測試、SageMaker AI 轉換環境測試和 Clean Rooms ML 相容的範例 Docker 檔案

```
FROM 763104351884.dkr.ecr.us-east-1.amazonaws.com/pytorch-inference:1.12.1-cpu-py38-ubuntu20.04-sagemaker

ENV PYTHONUNBUFFERED=1

COPY serve.py /opt/ml/code/serve.py
COPY inference_handler.py /opt/ml/code/inference_handler.py
COPY handler_service.py /opt/ml/code/handler_service.py
COPY model.py /opt/ml/code/model.py

RUN chmod +x /opt/ml/code/serve.py

ENTRYPOINT ["/opt/ml/code/serve.py"]
```

- 完成任何模型變更，並準備好在 SageMaker AI 環境中進行測試後，請依提供的順序執行下列命令。

```
export ACCOUNT_ID=xxx
export REPO_NAME=xxx
export REPO_TAG=xxx
export REGION=xxx

docker build -t $ACCOUNT_ID.dkr.ecr.us-west-2.amazonaws.com/$REPO_NAME:$REPO_TAG

# Sign into AWS $ACCOUNT_ID/ Run aws configure
# Check the account and make sure it is the correct role/credentials
aws sts get-caller-identity
aws ecr create-repository --repository-name $REPO_NAME --region $REGION
aws ecr describe-repositories --repository-name $REPO_NAME --region $REGION

# Authenticate Docker
aws ecr get-login-password --region $REGION | docker login --username AWS --password-stdin $ACCOUNT_ID.dkr.ecr.$REGION.amazonaws.com
```

```
# Push To ECR Repository
docker push $ACCOUNT_ID.dkr.ecr.$REGION.amazonaws.com$REPO_NAME:$REPO_TAG

# Create Sagemaker Model
# Configure the create_model.json with
# 1. Primary container -
#   a. ModelDataUrl - S3 Uri of the model.tar from your training job
aws sagemaker create-model --cli-input-json file://create_model.json --region $REGION

# Create Sagemaker Transform Job
# Configure the transform_job.json with
# 1. Model created in the step above
# 2. MultiRecord batch strategy
# 3. Line SplitType for TransformInput
# 4. AssembleWith Line for TransformOutput
aws sagemaker create-transform-job --cli-input-json file://transform_job.json --
region $REGION
```

在 SageMaker AI 任務完成且您對批次轉換感到滿意之後，您就可以向 AWS Clean Rooms ML 註冊 Amazon ECR 登錄檔。使用 `CreateConfiguredModelAlgorithm` 動作來註冊模型演算法，並使用 `CreateConfiguredModelAlgorithmAssociation` 將其與協同合作建立關聯。

接收模型日誌和指標

若要從自訂模型訓練或推論接收日誌和指標，成員必須使用提供必要 CloudWatch 許可的有效角色[來建立 ML 組態](#)（請參閱[建立自訂 ML 模型的服務角色 - ML 組態](#)）。

系統指標

訓練和推論的系統指標，例如 CPU 和記憶體使用率，都會發佈給與有效 ML 組態協作的所有成員。這些指標可以隨著任務進度，分別透過 `/aws/cleanroomsm1/TrainedModels` 或 `/aws/cleanroomsm1/TrainedModelInferenceJobs` 命名空間中的 CloudWatch 指標進行檢視。

模型日誌

每個設定模型演算法的隱私權組態政策都會提供對模型日誌的存取。模型作者在將設定的模型演算法（透過主控台或 `CreateConfiguredModelAlgorithmAssociation` API）與協同合作建立關聯時，會設定隱私權組態政策。設定隱私權組態政策可控制哪些成員可以接收模型日誌。

此外，模型作者可以在隱私權組態政策中設定篩選模式，以篩選日誌事件。模型容器傳送至 `stdout` 或 `stderr` 以及符合篩選條件模式（如果設定）的所有日誌都會傳送至 Amazon CloudWatch Logs。模

型日誌/aws/cleanroomsm1/TrainedModelInferenceJobs分別在 CloudWatch 日誌群組/aws/cleanroomsm1/TrainedModels或 中提供。

自訂定義的指標

當您設定模型演算法（透過主控台或 CreateConfiguredModelAlgorithm API）時，模型作者可以提供特定的指標名稱和 regex 陳述式，以便在輸出日誌中搜尋。當任務透過 /aws/cleanroomsm1/TrainedModels 命名空間中的 CloudWatch Metrics 進行時，可以檢視這些項目。關聯設定的模型演算法時，模型作者可以在指標隱私權組態中設定選用的雜訊層級，以避免輸出原始資料，同時仍提供自訂指標趨勢的可見性。如果設定雜訊層級，則指標會在任務結束時發佈，而不是即時發佈。

的加密運算 Clean Rooms

Clean Rooms (C3R) 密碼編譯運算是 中的功能 AWS Clean Rooms，除了[分析規則](#)之外，還可以使用此功能。透過 C3R，組織可以將敏感資料結合在一起，從資料分析中衍生新洞見，同時以密碼編譯方式限制流程中的任一方可以學習的內容。C3R 可供想要與其敏感資料協作，但只需要在雲端中使用加密資料的兩個或多個方使用。

C3R 加密用戶端是一種用戶端加密工具，可用來[加密](#)資料以搭配使用 AWS Clean Rooms。當您使用 C3R 加密用戶端時，資料在 AWS Clean Rooms 協同合作中使用時仍會受到密碼編譯保護。如同定期 AWS Clean Rooms 協同合作，輸入資料是關聯式資料庫資料表，而運算會以 SQL 查詢表示。不過，C3R 僅支援加密資料上有限的 SQL 查詢子集。

具體而言，C3R 支援密碼編譯受保護資料的 SQL JOIN 和 SELECT 陳述式。輸入資料表中的每個資料欄只能用於下列其中一個 SQL 陳述式類型：

- 受密碼編譯保護以用於 JOIN 陳述式的資料欄稱為 資料fingerprint欄。
- 受密碼編譯保護以用於 SELECT 陳述式的資料欄稱為 資料sealed欄。
- 未受密碼編譯保護的資料欄可用於 JOIN 或 SELECT 陳述式，稱為資料cleartext欄。

在某些情況下，資料fingerprint欄支援 GROUP BY 陳述式。如需詳細資訊，請參閱[Fingerprint 資料欄](#)。目前，C3R 不支援在加密資料上使用其他 SQL 建構，例如子WHERE句或彙總函數，例如 SUM 和 AVERAGE，即使相關分析規則允許這些函數。

C3R 旨在保護資料表個別儲存格中的資料。使用 C3R 的預設組態，客戶透過協同合作提供給第三方的基礎資料會在內使用內容時保持加密 AWS Clean Rooms。C3R 對所有sealed資料欄使用業界標準的 AES-GCM 加密，以及稱為雜湊型訊息驗證碼 (HMAC) 的業界標準虛擬隨機函數來保護資料fingerprint欄。

雖然 C3R 會加密資料表中的資料，但仍可能推論下列資訊：

- 資料表本身的相關資訊，包括資料表中的資料欄數、資料欄名稱和資料列數。
- 如同大多數標準形式的加密，C3R 不會嘗試隱藏加密值的長度。C3R 確實提供填補加密值的功能，以隱藏純文字的確切長度。不過，每個資料欄中純文字長度的上限仍然可以向另一方公開。
- 記錄層級資訊，例如將特定資料列新增至加密 C3R 資料表的時間。

如需 C3R 的詳細資訊，請參閱下列主題。

主題

- [針對 使用密碼編譯運算時的考量事項 Clean Rooms](#)
- [適用於 的密碼編譯運算中支援的檔案和資料類型 Clean Rooms](#)
- [的密碼編譯運算中的資料欄名稱 Clean Rooms](#)
- [適用於 的密碼編譯運算中的資料欄類型 Clean Rooms](#)
- [密碼編譯運算參數](#)
- [適用於 的密碼編譯運算中的選用旗標 Clean Rooms](#)
- [適用於 的加密運算查詢 Clean Rooms](#)
- [C3R 加密用戶端的準則](#)

針對 使用密碼編譯運算時的考量事項 Clean Rooms

適用於 Clean Rooms(C3R) 的加密運算力力求將資料保護最大化。不過，某些使用案例可能受益於較低層級的資料保護，以換取額外的功能。您可以從最安全的組態修改 C3R，藉此做出這些特定的權衡。身為客戶，您應該了解這些權衡，並判斷它們是否適合您的使用案例。要考慮的權衡包括下列項目：

主題

- [允許資料表中的混合cleartext和加密資料](#)
- [允許資料fingerprint欄中的重複值](#)
- [解除資料fingerprint欄命名方式的限制](#)
- [決定NULL值的呈現方式](#)

如需如何設定這些案例參數的詳細資訊，請參閱 [密碼編譯運算參數](#)。

允許資料表中的混合cleartext和加密資料

所有資料都經過用戶端加密，可提供最大的資料保護。不過，這會限制特定類型的查詢（例如SUM彙總函數）。允許cleartext資料的風險是，任何有權存取加密資料表的人都可以推斷一些有關加密值的資訊。這可以透過對 cleartext和相關聯的資料執行統計分析來完成。

例如，假設您有 City和 的資料欄State。資料City欄為 cleartext，且資料State欄已加密。當您Chicago在 City欄中看到 值時，這可協助您以高機率判斷 State為 Illinois。相反地，如果一個資料欄是 City，另一個資料欄是 EmailAddress，則 cleartext City 不太可能顯示有關加密 的任何內容EmailAddress。

如需此案例參數的詳細資訊，請參閱 [允許cleartext資料欄參數](#)。

允許資料fingerprint欄中的重複值

為了最安全的方法，我們假設任何fingerprint資料欄只包含一個變數的執行個體。fingerprint 欄中不可重複任何項目。C3R 加密用戶端會將這些cleartext值映射到與隨機值不區分的唯一值。因此，無法 cleartext從這些隨機值推斷有關 的資訊。

資料fingerprint欄中重複值的風險是重複值將導致重複的隨機顯示值。因此，任何有權存取加密資料表的人，理論上都可以對可能顯示cleartext值資訊的fingerprint欄執行統計分析。

同樣地，假設資料fingerprint欄是 State，且資料表的每一列都對應至美國家庭。透過執行頻率分析，可以推斷哪個狀態為 California，以及哪個狀態Wyoming為高機率。這種推論是可能的，因為 California 的居民比 多Wyoming。相反地，假設資料fingerprint欄位於家庭識別符上，而且每個家庭在數百萬個項目的資料庫中出現 1 到 4 次。頻率分析不太可能顯示任何有用的資訊。

如需此案例參數的詳細資訊，請參閱 [允許重複參數](#)。

解除資料fingerprint欄命名方式的限制

根據預設，我們假設當兩個資料表使用加密資料fingerprint欄聯結時，這些資料欄在每個資料表中具有相同的名稱。此結果的技術原因是，根據預設，我們會衍生不同的密碼編譯金鑰來加密每個fingerprint資料欄。該金鑰衍生自協同合作的共用私密金鑰和資料欄名稱的組合。如果我們嘗試使用不同的資料欄名稱聯結兩個資料欄，我們會衍生不同的金鑰，而且無法計算有效的聯結。

若要解決此問題，您可以關閉從每個資料欄名稱衍生索引鍵的功能。然後，C3R 加密用戶端對所有 fingerprint資料欄使用單一衍生金鑰。風險是可以進行另一種可能顯示資訊的頻率分析。

讓我們再次使用 City和 State 範例。如果我們為每個fingerprint資料欄衍生相同的隨機值（透過不合併資料欄名稱）。在 City和 資料State欄中New York具有相同的隨機值。紐約是美國少數幾個

城市之一，其City名稱與State名稱相同。相反地，如果您的資料集在每個資料欄中具有完全不同的值，則不會洩漏任何資訊。

如需此案例參數的詳細資訊，請參閱 [允許具有不同名稱參數JOIN的資料欄](#)。

決定NULL值的呈現方式

您可用的選項是是否像任何其他NULL值一樣，以密碼編譯方式處理（加密和 HMAC）值。如果您不像任何其他NULL值一樣處理值，可能會顯示資訊。

例如，假設 NULL 的 Middle Name欄中的 cleartext表示沒有中間名的人員。如果您不加密這些值，則會洩漏加密資料表中哪些資料列用於沒有中間名的人員。該資訊可能是某些人口中某些人的識別訊號。但是，如果您以密碼編譯方式處理NULL值，某些 SQL 查詢的運作方式會有所不同。例如，GROUP BY子句不會將fingerprint資料欄中fingerprintNULL的值分組在一起。

如需此案例參數的詳細資訊，請參閱 [保留NULL值參數](#)。

適用於 的密碼編譯運算中支援的檔案和資料類型 Clean Rooms

C3R 加密用戶端會辨識下列檔案類型：

- CSV 檔案
- Parquet 檔案

您可以使用 C3R 加密用戶端中的 `--fileFormat` 旗標明確指定檔案格式。明確指定時，檔案格式不會由副檔名決定。

主題

- [CSV 檔案](#)
- [Parquet 檔案](#)
- [加密非字串值](#)

CSV 檔案

假設副檔名為 .csv 的檔案為 CSV 格式，並包含 UTF-8 編碼的文字。C3R 加密用戶端會將所有值視為字串。

.csv 檔案中支援的屬性

C3R 加密用戶端需要 .csv 檔案具有下列屬性：

- 可能或可能不包含可唯一命名每個資料欄的初始標頭列。
- 逗號分隔。(目前不支援自訂分隔符號。)
- UTF-8 編碼的文字。

從 .csv 項目裁剪的空格

前置和後置空格都會從 .csv 項目修剪。

.csv 檔案的自訂NULL編碼

.csv 檔案可以使用自訂NULL編碼。

使用 C3R 加密用戶端，您可以使用 `--csvInputNULLValue=<csv-input-null>` 旗標為輸入資料中的NULL項目指定自訂編碼。C3R 加密用戶端可以使用 `--csvOutputNULLValue=<csv-output-null>` 旗標，在產生的輸出檔案中針對 NULL 項目使用自訂編碼。

Note

NULL 項目被視為缺少內容，特別是在 SQL 資料表等更豐富的表格格式環境中。雖然 .csv 基於歷史原因未明確支援此特性描述，但一般慣例是將僅包含空格的空白項目視為 NULL。因此，這是 C3R 加密用戶端的預設行為，可以視需要自訂。

C3R 如何解譯 .csv 項目

下表提供 .csv 項目如何根據為 `--csvInputNULLValue=<csv-input-null>` 和 `--csvOutputNULLValue=<csv-output-null>` 旗標提供的值 (cleartext如果有的話) 進行封送 (到 cleartext 以獲得明確性) 的範例。在 C3R 解譯任何值的意義之前，會修剪引號外部的前後空格。

<code><csv-input-null></code>	<code><csv-output-null></code>	輸入項目	輸出項目
None	None	<code>#AnyProduct#</code>	<code>#AnyProduct#</code>
None	None	<code>#AnyProduct#</code>	<code>#AnyProduct#</code>
None	None	<code>"AnyProduct"</code>	<code>#AnyProduct#</code>
None	None	<code>##AnyProduct##</code>	<code>#AnyProduct#</code>

<csv-input-null>	<csv-output-null>	輸入項目	輸出項目
None	None	,,	,,
None	None	, ,	,,
None	None	, "",	,,
None	None	, " ",	, " ",
None	None	, " " ,	, " ",
#AnyProduct#	#NULL#	#AnyProduct#	#NULL#
#AnyProduct#	#NULL#	#AnyProduct#	#NULL#
#AnyProduct#	#NULL#	#"AnyProduct"#	#NULL#
#AnyProduct#	#NULL#	##AnyProduct##	#NULL#
None	#NULL#	,,	#NULL#
None	#NULL#	, ,	#NULL#
None	#NULL#	, "",	#NULL#
None	#NULL#	, " ",	, " ",
None	#NULL#	, " " ,	, " ",
""	#NULL#	,,	#NULL#
""	#NULL#	, ,	#NULL#
""	#NULL#	, "",	, "",
""	#NULL#	, " ",	, " ",
""	#NULL#	, " " ,	, " ",
"\""	#NULL#	,,	,,

<csv-input-null>	<csv-output-null>	輸入項目	輸出項目
"\"\""	#NULL#	, ,	, ,
"\"\""	#NULL#	, "",	#NULL#
"\"\""	#NULL#	, " ",	, " ",
"\"\""	#NULL#	, " " ,	, " " ,

不含標頭的 CSV 檔案

來源 .csv 檔案不需要在第一列中具有唯一命名每個資料欄的標頭。不過，沒有標頭列的 .csv 檔案需要位置加密結構描述。位置加密結構描述是必要的，而不是用於具有標頭列和檔案的 .csv Parquet 檔案的一般映射結構描述。

位置加密結構描述會依位置而非名稱指定輸出資料欄。映射的加密結構描述會將來源資料欄名稱映射至目標資料欄名稱。如需詳細資訊，包括這兩種結構描述格式的詳細討論和範例，請參閱 [映射和位置資料表結構描述](#)。

Parquet 檔案

假設副.parquet檔名為 的檔案Apache Parquet格式。

支援的Parquet資料類型

C3R 加密用戶端可以處理檔案中Parquet代表 支援之資料類型的任何非複雜（即基本類型）資料 AWS Clean Rooms。

不過，只有字串資料欄可用於sealed資料欄。

支援下列 Parquet 資料類型：

- Binary 具有下列邏輯註釋的基本類型：
 - 如果--parquetBinaryAsString已設定 則無 (STRING 資料類型)
 - Decimal(scale, precision) (DECIMAL 資料類型)
 - String (STRING 資料類型)
- Boolean 沒有邏輯註釋的基本資料類型 (BOOLEAN 資料類型)

- Double 沒有邏輯註釋的基本資料類型 (DOUBLE 資料類型)
- Fixed_Len_Binary_Array 具有邏輯註釋的基本類型 `Decimal(scale, precision)` (DECIMAL 資料類型)
- Float 沒有邏輯註釋的基本資料類型 (FLOAT 資料類型)
- Int32 具有下列邏輯註釋的基本類型：
 - 無 (INT 資料類型)
 - Date (DATE 資料類型)
 - `Decimal(scale, precision)` (DECIMAL 資料類型)
 - `Int(16, true)` (SMALLINT 資料類型)
 - `Int(32, true)` (INT 資料類型)
- Int64 具有下列邏輯註釋的基本資料類型：
 - 無 (BIGINT 資料類型)
 - `Decimal(scale, precision)` (DECIMAL 資料類型)
 - `Int(64, true)` (BIGINT 資料類型)
 - `Timestamp(isUTCAdjusted, TimeUnit.MILLIS)` (TIMESTAMP 資料類型)
 - `Timestamp(isUTCAdjusted, TimeUnit.MICROS)` (TIMESTAMP 資料類型)
 - `Timestamp(isUTCAdjusted, TimeUnit.NANOS)` (TIMESTAMP 資料類型)

加密非字串值

目前，sealed欄僅支援字串值。

對於 .csv 檔案，C3R 加密用戶端會將所有值視為 UTF-8 編碼文字，並且在加密之前不會嘗試以不同的方式解譯這些值。

對於指紋資料欄，類型會分組為等效類別。等效類別是一組資料類型，可透過代表性資料類型明確比較相等性。

對等類別允許將相同的指紋指派給相同的語意值，無論原始表示法為何。不過，兩個等效類別中的相同值不會產生相同的指紋資料欄。

例如，無論該INTEGRAL值原本是 SMALLINT、或 INT，42都會指派相同的指紋BIGINT。此外，該INTEGRAL值永遠0不會符合該BOOLEAN值 FALSE (以值 表示0)。

指紋資料欄支援下列等效類別和對應的 AWS Clean Rooms 資料類型：

對等類別	支援的 AWS Clean Rooms 資料類型
BOOLEAN	BOOLEAN
DATE	DATE
INTEGRAL	BIGINT, INT, SMALLINT
STRING	CHAR, STRING, VARCHAR

的密碼編譯運算中的資料欄名稱 Clean Rooms

根據預設，資料欄的名稱在 的加密運算中很重要Clean Rooms。

如果允許具有不同名稱參數的資料JOIN欄的值為 false，則會在資料欄加密期間使用fingerprint資料欄名稱。因此，根據預設，協作者必須事先協調，並為將在查詢中使用JOIN陳述式的資料使用相同的目標資料欄名稱。根據預設，JOIN以不同名稱加密的 資料欄無法在任何值JOIN上成功。

如果允許具有不同名稱參數的資料JOIN欄的值為 true，則資料欄之間的JOIN陳述式會加密為資料fingerprint欄成功。使用此參數加密資料可能會允許對cleartext值進行一些推論。例如，如果資料列在City資料欄和State資料欄中具有相同的雜湊型訊息驗證碼 (HMAC) 值，則該值可能是 New York。

資料欄標頭名稱的標準化

資料欄標頭名稱由 C3R 加密用戶端標準化。任何開頭和結尾的空格都會移除，且轉換輸出的資料欄名稱會變成小寫。

標準化會在所有其他運算、計算或可能受到資料欄名稱影響的其他操作之前套用。發出的輸出檔案只包含標準化名稱。

適用於 的密碼編譯運算中的資料欄類型 Clean Rooms

本主題提供 密碼編譯運算中資料欄類型的相關資訊Clean Rooms。

主題

- [Fingerprint 資料欄](#)

- [已密封的資料欄](#)
- [Cleartext 資料欄](#)

Fingerprint 資料欄

Fingerprint 資料欄是以密碼編譯方式保護的資料欄，可用於 JOIN 陳述式。

資料 fingerprint 欄的資料無法解密。只有來自密封資料欄的資料才能解密。

Fingerprint 欄只能用於下列 SQL 子句和函數：

- JOIN (INNER, OUTER, LEFT, RIGHT, or FULL) 針對其他 fingerprint 資料欄：
 - 如果 `allowJoinsOnColumnsWithDifferentNames` 參數的值設定為 `false`，則的兩個 fingerprint 資料欄 JOIN 也必須具有相同的名稱。
- SELECT COUNT()
- SELECT COUNT(DISTINCT)
- GROUP BY (只有在協同合作已將 `preserveNulls` 參數的值設定為 `true` 時，才能使用 `true`。)

違反這些限制條件的查詢可能會導致不正確的結果。

已密封的資料欄

密封資料欄是以密碼編譯方式保護的資料欄，可用於 SELECT 陳述式。

密封資料欄只能用於下列 SQL 子句和函數：

- SELECT
- SELECT ... AS
- SELECT COUNT()

Note

不支援 SELECT COUNT(DISTINCT)。

違反這些限制條件的查詢可能會導致不正確的結果。

在加密之前填補資料sealed欄的資料

當您指定資料欄應為資料sealed欄時，C3R 會要求您選擇哪種填補。在加密之前填補資料是選用的。不使用填補 (的鍵盤類型none)，加密資料的長度表示 的大小cleartext。在某些情況下， 的大小cleartext可能會公開純文字。使用填補 (鍵盤類型為 fixed或 max) 時，所有值會先填補至通用大小，然後加密。使用填補時，加密資料的長度不會提供有關原始cleartext長度的資訊，除了在大小上加上上限。

如果您想要填補資料欄，且已知該資料欄中的資料位元組長度上限，請使用fixed填補。使用至少與該資料欄中最長length值的位元組長度相同的值。

Note

如果值超過提供的 ，則發生錯誤且加密失敗length。

如果您想要填補資料欄，而且該資料欄中的資料位元組長度上限不明，請使用max填補。此填補模式會將所有資料填入最長值加上額外length位元組的長度。

Note

您可能想要批次加密資料，或定期使用新資料更新資料表。請注意，max填補會將項目填補至指定批次中最長純文字項目的長度 length (加位元組)。這表示加密文字長度可能因批次而異。因此，如果您知道資料欄的最大位元組長度，則應該使用 fixed而非 max。

Cleartext 資料欄

Cleartext 資料欄是未受密碼編譯保護的資料欄，可用於 JOIN或 SELECT陳述式。

Cleartext 資料欄可用於 SQL 查詢的任何部分。

密碼編譯運算參數

密碼編譯運算參數可在建立協作時使用 Cryptographic Computing for Clean Rooms(C3R) 進行協作。

[???](#)您可以使用 AWS Clean Rooms 主控台或 CreateCollaboration API 操作建立協同合作。在主控台中，您可以在開啟支援密碼編譯運算選項之後，在密碼編譯運算參數中設定參數的值。如需詳細資訊，請參閱下列主題。

主題

- [允許cleartext資料欄參數](#)
- [允許重複參數](#)
- [允許具有不同名稱參數JOIN的資料欄](#)
- [保留NULL值參數](#)

允許cleartext資料欄參數

在主控台中，您可以在[建立協同合作](#)時設定允許cleartext資料欄參數，以指定具有加密cleartext資料的資料表中是否允許資料。

下表說明允許cleartext資料欄參數的值。

參數值	Description
否	Cleartext 加密資料表中不允許 欄。所有資料都會受到密碼編譯保護。
是	Cleartext 加密資料表中允許 欄。 Cleartext 資料欄不受密碼編譯保護，並會包含為 cleartext。您應該記下資料列cleartext的資料在資料表中可能顯示的其他資料。 若要在特定資料欄AVG上執行 SUM或 ，資料欄必須位於 中 cleartext。

使用 CreateCollaboration API 操作，對於 dataEncryptionMetadata 參數，您可以將 的值設定為 allowCleartexttrue或 false。如需 API 操作的詳細資訊，請參閱 [AWS Clean Rooms API 參考](#)。

Cleartext 資料欄對應至資料表特定結構描述cleartext中分類為 的資料欄。這些資料欄中的資料不會加密，而且可以以任何方式使用。如果資料不敏感和/或需要比加密Cleartext資料欄或fingerprint資料欄允許的更多彈性，資料sealed欄可能很有用。

允許重複參數

在主控台中，您可以在[建立協同合作](#)時設定允許重複參數，以指定為JOIN查詢加密的資料欄是否可以包含重複的非NULL值。

⚠ Important

允許重複、[允許具有不同名稱的資料JOIN欄](#)，以及[保留NULL值](#)參數具有個別但相關的效果。

下表說明 Allow duplicates 參數的值。

參數值	Description
否	資料fingerprint欄中不允許重複的值。單一資料fingerprint欄中的所有值都必須是唯一的。
是	資料fingerprint欄中允許重複的值。 如果您需要使用重複值聯結資料欄，請將此值設定為是。當設定為是時，C3R 資料表或結果中的資料fingerprint欄中出現的頻率模式可能表示有關cleartext資料結構的一些額外資訊。

使用 CreateCollaboration API 操作，對於 dataEncryptionMetadata 參數，您可以將 的值設定為 allowDuplicatestrue或 false。如需 API 操作的詳細資訊，請參閱 [AWS Clean Rooms API 參考](#)。

根據預設，如果必須在JOIN查詢中使用加密的資料，C3R 加密用戶端會要求這些資料欄沒有重複的值。這項要求是為了提高資料保護。此行為有助於確保資料中重複的模式無法觀察。不過，如果您想要在JOIN查詢中使用加密的資料，並且不擔心重複值，則允許重複參數可以停用此保守檢查。

允許具有不同名稱參數JOIN的資料欄

在 主控台中，您可以在[建立協同](#)合作時設定允許具有不同名稱參數JOIN的資料欄，以指定是否支援不同名稱的資料欄之間的JOIN陳述式。

如需詳細資訊，請參閱[資料欄標頭名稱的標準化](#)

下表說明使用不同名稱參數JOIN允許資料欄的值。

參數值	Description
否	不支援聯結具有不同名稱的資料fingerprint欄。 JOIN陳述式只會在具有相同名稱的資料欄上提供準確的結果。

參數值	Description
	 Important 無值可提高資訊安全，但要求協同合作參與者先前同意資料欄名稱。如果兩個資料欄在加密為fingerprint資料欄時具有不同的名稱，且允許具有不同名稱JOIN的資料欄設定為否，則這些資料欄上的JOIN陳述式不會產生任何結果。這是因為加密後不會在它們之間共用任何值。
是	支援聯結具有不同名稱的資料fingerprint欄。為了提高靈活性，使用者可以將此值設定為是，這允許資料欄上的JOIN陳述式，無論其資料欄名稱為何。 如果設定為是，C3R 加密用戶端在保護資料欄時不會考慮資料fingerprint欄名稱。因此，可在 C3R 資料表中觀察不同fingerprint資料欄之間共同值。 例如，如果資料列在City資料欄和State資料欄中具有相同的加密JOIN值，則推斷該值為 可能是合理的New York。

使用 CreateCollaboration API 操作，對於 dataEncryptionMetadata 參數，您可以將 的值設定為 allowJoinsOnColumnsWithDifferentNamestrue 或 false。如需 API 操作的詳細資訊，請參閱 [AWS Clean Rooms API 參考](#)。

根據預設，資料fingerprint欄加密會受到該資料欄targetHeader的影響，在 中設定[步驟 4：產生表格式檔案的加密結構描述](#)。因此，相同的cleartext值在加密的每個不同資料fingerprint欄中都有不同的加密表示法。

在某些情況下，此參數有助於防止cleartext值推論。例如，在資料fingerprint欄中看到相同的加密值City，並State可能用於合理推斷該值為 New York。不過，此參數的使用需要事先進行額外的協調，因此要加入查詢的所有資料欄都有共用的名稱。

您可以使用允許具有不同名稱參數JOIN的資料欄來鬆散此限制。當參數值設定為 時Yes，它允許將 加密的任何資料欄一起使用JOIN，無論名稱為何。

保留NULL值參數

在主控台中，您可以在[建立協同](#)合作時設定保留NULL值參數，以表示該資料欄不存在任何值。

下表說明保留值參數NULL的值。

參數值	Description
否	NULL 值不會保留。 NULL 值不會在加密資料表NULL中顯示為 。 NULL值會在 C3R 資料表中顯示為唯一的隨機值。
是	NULL 值會保留。 NULL 值會在加密的資料表NULL中顯示為 。如果您需要NULL值的 SQL 語意，您可以將此值設定為是。因此，無論資料欄是否已加密，無論允許重複項目的參數設定為何，NULL項目都會在 C3R 資料表NULL中顯示為 。

使用 CreateCollaboration API 操作，對於 dataEncryptionMetadata 參數，您可以將 的值設定為 preserveNullstrue或 false。如需 API 操作的詳細資訊，請參閱 [AWS Clean Rooms API 參考](#)。

協同合作的保留NULL值參數設定為否時：

1. NULL cleartext資料欄中的項目保持不變。
2. NULL 加密資料fingerprint欄中的項目會加密為隨機值，以隱藏其內容。在具有資料欄中NULL項目的加密cleartext資料欄上聯結不會產生任何項目的任何相符NULL項目。不會進行任何比對，因為它們都會收到自己的唯一隨機內容。
3. NULL 加密sealed欄中的項目會加密。

當保留NULL值參數的值針對協同合作設為是時，NULL無論資料欄是否加密，所有資料欄NULL的項目都會保持為 。

保留NULL值參數在資料擴充等案例中非常有用，其中您想要分享的資訊不足，以 表示NULL。如果您在想要 fingerprint或 的欄中有NULL值，則保留NULL值參數在 JOIN或 HMAC 格式中也很有用GROUP BY。

如果允許重複和保留NULL值參數的值設定為否，則fingerprint欄中有一個以上的NULL項目會產生錯誤並停止加密。如果任一參數的值設定為是，則不會發生此類錯誤。

適用於 的密碼編譯運算中的選用旗標 Clean Rooms

下列各節說明當您使用 C3R 加密用戶端加密[資料](#)以進行表格式檔案自訂和測試時，可以設定的選用旗標。

主題

- [--csvInputNULLValue 標記](#)
- [--csvOutputNULLValue 標記](#)
- [--enableStackTraces 標記](#)
- [--dryRun 標記](#)
- [--tempDir 標記](#)

--csvInputNULLValue 標記

當您使用 C3R 加密用戶端[加密資料](#)時，您可以使用 `--csvInputNULLValue` 旗標為輸入資料中的 NULL 項目指定自訂編碼。

下表摘要說明此旗標的用量和參數。

Usage	Parameters
選用。使用者可以為輸入資料中的 NULL 項目指定自訂編碼。	輸入 CSV 檔案中 NULL 值的使用者指定編碼

NULL 項目是視為缺少內容的項目，特別是在 SQL 資料表等更豐富的表格格式環境中。雖然 .csv 基於歷史原因未明確支援此特性描述，但一般慣例是將僅包含空格的空白項目視為 NULL。因此，這是 C3R 加密用戶端的預設行為，可以視需要自訂。

--csvOutputNULLValue 標記

當您使用 C3R 加密用戶端[加密資料](#)時，您可以使用 `--csvOutputNULLValue` 旗標為輸出資料中的 NULL 項目指定自訂編碼。

下表摘要說明此旗標的用量和參數。

Usage	Parameters
選用。使用者可以在產生的輸出檔案中為 NULL 項目指定自訂編碼。	使用者指定的輸出 CSV 檔案中 NULL 值編碼

NULL 項目是視為缺少內容的項目，特別是在 SQL 資料表等更豐富的表格格式環境中。雖然 .csv 基於歷史原因未明確支援此特性描述，但一般慣例是將僅包含空格的空白項目視為 NULL。因此，這是 C3R 加密用戶端的預設行為，可以視需要自訂。

--enableStackTraces 標記

當您使用 C3R 加密用戶端加密[資料](#)時，請使用 --enableStackTraces 旗標提供其他內容資訊，以便在 C3R 遇到錯誤時回報錯誤。

AWS 不會收集錯誤。如果您遇到錯誤，請使用堆疊追蹤自行疑難排解錯誤，或將堆疊追蹤傳送至支援以取得協助。

下表摘要說明此旗標的用量和參數。

Usage	Parameters
選用。用於在 C3R 加密用戶端遇到錯誤時提供錯誤報告的其他內容資訊。	無

--dryRun 標記

[加密](#)和[解密](#) C3R 加密用戶端命令包含選用的 --dryRun 旗標。旗標會取得所有使用者提供的引數，並檢查它們的有效性和一致性。

您可以使用 --dryRun 旗標來檢查您的結構描述檔案是否有效，並與對應的輸入檔案一致。

下表摘要說明此旗標的用量和參數。

Usage	Parameters
選用。導致 C3R 加密用戶端剖析參數和檢查檔案，但不會執行加密或解密。	無

--tempDir 標記

您可能想要使用暫時目錄，因為加密的檔案有時可能會大於未加密的檔案，取決於其設定。資料集也必須在每次協同合作時加密，才能正常運作。

當您使用 C3R [加密資料](#)時，請使用 --tempDir 旗標指定在處理輸入時可以建立暫存檔案的位置。

下表摘要說明此旗標的用量和參數。

Usage	Parameters
使用者可以指定在處理輸入時可以建立暫存檔案的位置。	預設為系統暫時目錄。

適用於 的加密運算查詢 Clean Rooms

本主題提供使用 Cryptographic Computing for 加密之資料表撰寫查詢的相關資訊Clean Rooms。

主題

- [在上分支的查詢 NULL](#)
- [將一個來源資料欄映射至多個目標資料欄](#)
- [對 JOIN 和 SELECT 查詢使用相同的資料](#)

在上分支的查詢 NULL

在 NULL 陳述式上具有查詢分支表示使用語法，例如 `IF x IS NULL THEN 0 ELSE 1`。

查詢一律可以對 cleartext 資料欄中的 NULL 陳述式進行分支。

只有在保留 NULL 值參數 (preserveNulls) 的值設定為 `true` 時，查詢才能對 sealed 資料欄和資料 fingerprint 欄中的 NULL 陳述式進行分支。

違反這些限制條件的查詢可能會導致不正確的結果。

將一個來源資料欄映射至多個目標資料欄

一個來源資料欄可以映射到多個目標資料欄。例如，您可能想要在資料欄 SELECT 上同時使用 JOIN 和 `UNION`。

如需詳細資訊，請參閱[對 JOIN 和 SELECT 查詢使用相同的資料](#)。

對 JOIN 和 SELECT 查詢使用相同的資料

如果資料欄中的資料不敏感，它可以出現在cleartext目標資料欄中，這允許將其用於任何用途。

如果資料欄中的資料很敏感，且必須同時用於 JOIN 和 SELECT 查詢，請將該來源資料欄映射到輸出檔案中的兩個目標資料欄。一個資料欄是以 type 做為資料fingerprint欄加密，一個資料欄是以 type 做為密封資料欄加密。C3R 加密用戶端的互動式結構描述產生會建議 `_fingerprint` 和 的標頭尾碼 `_sealed`。這些標頭尾碼可以是快速區分此類資料欄的實用慣例。

C3R 加密用戶端的準則

C3R 加密用戶端是一種工具，可讓組織將敏感資料結合在一起，從資料分析中衍生新的洞見。工具以密碼編譯方式限制任何一方和 AWS 程序中可學習的內容。雖然這至關重要，但以密碼編譯方式保護資料的程序可能會在運算和儲存資源方面增加大量額外負荷。因此，請務必了解使用每個設定的權衡，以及如何最佳化設定，同時仍維持所需的密碼編譯保證。本主題著重於 C3R 加密用戶端和結構描述中不同設定的效能影響。

所有 C3R 加密用戶端加密設定都提供不同的密碼編譯保證。協作層級設定預設為最安全。在建立協同合作時啟用其他功能會降低隱私權保證，允許在加密文字上執行頻率分析等活動。如需如何使用這些設定及其影響的詳細資訊，請參閱 [the section called “密碼編譯運算”](#)。

主題

- [資料欄類型的效能影響](#)
- [對加密文字大小的非預期增加進行故障診斷](#)

資料欄類型的效能影響

C3R 使用三種資料欄類型：cleartext、fingerprint 和 sealed。每個資料欄類型都提供不同的密碼編譯保證，並有不同的預期用途。在下列各節中，會討論資料欄類型的效能影響，以及每個設定的效能影響。

主題

- [Cleartext 資料欄](#)
- [Fingerprint 資料欄](#)
- [Sealed 資料欄](#)

Cleartext 資料欄

Cleartext 資料欄不會從原始格式變更，也不會以任何方式以密碼編譯方式處理。無法設定此資料欄類型，也不會影響儲存或運算效能。

Fingerprint 資料欄

Fingerprint 資料欄旨在用於跨多個資料表聯結資料。為此，產生的加密文字大小必須一律相同。不過，這些資料欄會受到協同合作層級設定的影響。資料Fingerprint欄可能會對輸出檔案大小產生不同程度的影響，取決於輸入中cleartext包含的。

主題

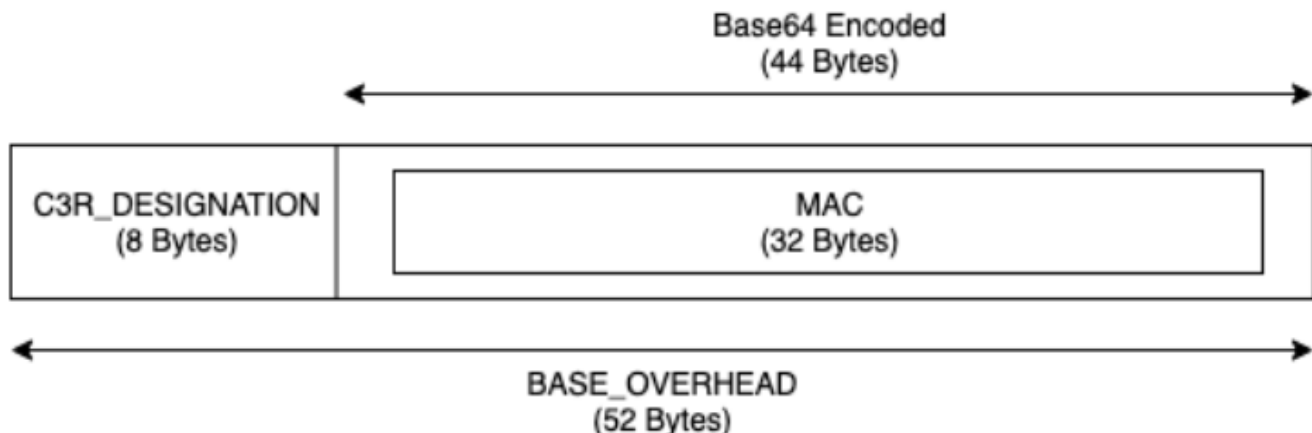
- [fingerprint 資料欄的基本額外負荷](#)
- [fingerprint 資料欄的協同作業設定](#)
- [資料fingerprint欄的範例資料](#)
- [對fingerprint資料欄進行故障診斷](#)

fingerprint 資料欄的基本額外負荷

資料fingerprint欄有基本額外負荷。此額外負荷是固定的，並取代cleartext位元組的大小。

資料fingerprint欄中的資料會透過雜湊型訊息驗證碼 (HMAC) 函數以密碼編譯方式處理，將資料轉換為 32 位元組訊息驗證碼 (MAC)。然後，這些資料會透過 base64 編碼器處理，將大約 33% 新增至位元組大小。它預先附加 8 位元組 C3R 指定，以指定資料所屬的資料欄類型，以及產生它的用戶端版本。最終結果為 52 個位元組。然後，此結果會乘以資料列計數，以取得總基本額外負荷（如果 preserveNulls 設為 true，請使用非null 值的總數）。

下圖顯示 如何 $BASE\_OVERHEAD = C3R\_DESIGNATION + (MAC * 1.33)$



fingerprint 欄中的輸出加密文字一律為 52 個位元組。如果輸入cleartext資料平均超過 52 個位元組（例如完整的街道地址），這可能會大幅減少儲存空間。如果輸入cleartext資料平均小於 52 個位元組（例如，客戶年齡），這可能會大幅增加儲存空間。

fingerprint 資料欄的協同作業設定

preserveNulls 設定

當協同合作層級設定為 preserveNulls false（預設）時，每個null值都會以唯一的隨機 32 個位元組取代，並如同不是一樣進行處理null。結果是每個null值現在都是 52 個位元組。相較於此設定為 true且null值以傳遞時，這可能會為包含非常稀疏資料的資料表新增重要的儲存需求null。

如果您不需要此設定的隱私權保證，並偏好在資料集內保留null值，請在建立協同合作時啟用preserveNulls設定。建立協同合作之後，就無法變更preserveNulls設定。

資料fingerprint欄的範例資料

以下是資料fingerprint欄的輸入和輸出資料集範例，其中包含要重現的設定。其他協作層級設定，例如allowCleartext和allowDuplicates不會影響結果，false而且如果嘗試在本機重現，則可以設定為true或。

共用秘密範例：wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY

協作 ID 範例：a1b2c3d4-5678-90ab-cdef-EXAMPLE11111

allowJoinsOnColumnsWithDifferentNames：True此設定不會影響效能或儲存需求。不過，此設定會在重新產生下表中顯示的值時，使資料欄名稱選擇無關。

範例 1

輸入	null
preserveNulls	TRUE
Output	null
確定性	Yes
輸入位元組	0
輸出位元組	0

範例 2

輸入	null
preserveNulls	FALSE
Output	01:hmac:3lkFjthvV3IUu6mMvFc1a +XAHwgw/Elm0q4p3Yg25kk=
確定性	No
輸入位元組	0
輸出位元組	52

範例 3

輸入	empty string
preserveNulls	-
Output	01:hmac:oKTgi3Gba+eUb3JteSz 2EMgXUkF1WgM77UP0Ydw5kPQ=
確定性	Yes
輸入位元組	0
輸出位元組	52

範例 4

輸入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:hmac:kU/IqwG7FMmzzshr0B9 scomE0UJUEE7j9keTctp1Gww=
確定性	Yes

輸入位元組	26
輸出位元組	52

範例 5

輸入	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:hmac:ks3htnQbw2vdhCRFF6J NzW5LMndJaHG57uvE26mBtSs=
確定性	Yes
輸入位元組	62
輸出位元組	52

對fingerprint資料欄進行故障診斷

為什麼我fingerprint欄中的加密文字比cleartext進去的 大小大上數倍？

資料fingerprint欄中的加密文字長度一律為 52 個位元組。如果您的輸入資料很小（例如，客戶的年齡），其大小將會大幅增加。如果preserveNulls設定設為 `false`，也可能發生這種情況。

為什麼我fingerprint欄中的加密文字比cleartext進去的 大小小數倍？

資料fingerprint欄中的加密文字長度一律為 52 個位元組。如果您的輸入資料很大（例如，客戶的完整街道地址），其大小將明顯減少。

如何知道是否需要 提供的密碼編譯保證 **preserveNulls**？

不幸的是，答案是它取決於 [the section called “Parameters”](#) 應檢閱，了解preserveNulls設定如何保護您的資料。不過，我們建議您參考組織的資料處理需求，以及適用於個別協同合作的任何合約。

為什麼我必須產生 base64 的額外負荷？

若要允許 與 CSV 等表格式檔案格式相容，需要 base64 編碼。雖然某些 檔案格式Parquet可能支援資料的二進位表示法，但協同合作中的所有參與者都以相同的方式代表資料，以確保適當的查詢結果。

Sealed 資料欄

Sealed 欄旨在用於在協同合作的成員之間傳輸資料。這些資料欄中的加密文字是非確定性的，並根據資料欄的設定方式對效能和儲存體產生重大影響。這些欄可以個別設定，通常對 C3R 加密用戶端的效能和產生的輸出檔案大小有最大的影響。

主題

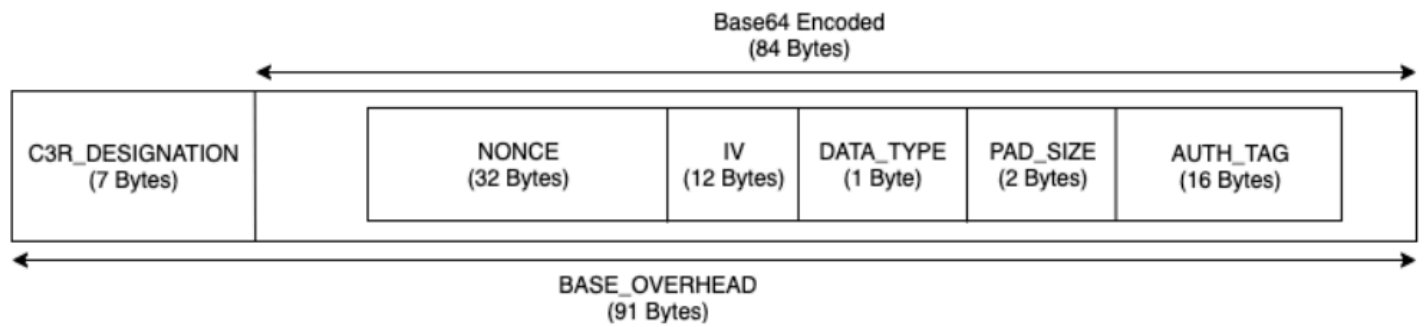
- [sealed 資料欄的基本額外負荷](#)
- [sealed 資料欄的協同作業設定](#)
- [結構描述設定sealed欄：填補類型](#)
- [資料sealed欄的範例資料](#)
- [對sealed資料欄進行故障診斷](#)

sealed 資料欄的基本額外負荷

資料sealed欄有基本額外負荷。此額外負荷是固定的，除了 cleartext和填補（如果有的話）位元組的大小之外。

在任何加密之前，資料sealed欄中的資料都會預先附加 1 位元組字元，以指定包含的資料類型。如果選取填補，則資料會填入並附加 2 個位元組，說明填補大小。新增這些位元組之後，會使用 AES-GCM 以密碼編譯方式處理資料，並以 IV(12 個位元組)、nonce(32 個位元組) 和 Auth Tag(16 個位元組) 存放。然後，這些資料會透過 base64 編碼器處理，將大約 33% 新增至位元組大小。資料會加上 7 位元組 C3R 指定，以指定資料所屬的欄類型，以及用於產生它的用戶端版本。最終基本額外負荷為 91 個位元組。然後，此結果可以乘以資料列計數，以取得總基本額外負荷（如果 preserveNulls 設為 true，請使用非空值總數）。

下圖顯示如何 $BASE\_OVERHEAD = C3R\_DESIGNATION + ((NONCE + IV + DATA\_TYPE + PAD\_SIZE + AUTH\_TAG) * 1.33)$



sealed 資料欄的協同作業設定

preserveNulls 設定

當協同合作層級設定為 `preserveNulls false` (預設) 時，每個 `null` 值都是唯一的隨機 32 個位元組，並如同不是一樣進行處理 `null`。結果是每個 `null` 值現在都是 91 個位元組 (如果有填充則更多)。相較於此設定為 `true` 且 `null` 值以傳遞時，這可能會為包含非常稀疏資料的資料表新增重要的儲存需求 `null`。

如果您不需要此設定的隱私權保證，並偏好在資料集內保留 `null` 值，請在建立協同合作時啟用 `preserveNulls` 設定。建立協同合作之後，就無法變更 `preserveNulls` 設定。

結構描述設定 sealed 欄：填補類型

主題

- [的填補類型 none](#)
- [的填補類型 fixed](#)
- [的填補類型 max](#)

的填補類型 **none**

選取 的護墊類型 `none` 不會將任何填補新增至 `cleartext`，也不會為先前所述的基本額外額外負荷新增額外負荷。沒有填補會產生最節省空間的輸出大小。不過，它不提供與 `fixed` 和 `max` 填補類型相同的隱私權保證。這是因為基礎的大小與加密文字的大小 `cleartext` 不同。

的填補類型 **fixed**

選取 的鍵盤類型 `fixed` 是一種隱私權保留措施，可隱藏資料欄中包含的資料長度。這是在加密 `pad_length` 之前，將所有填補 `cleartext` 到提供的。超過該大小的任何資料都會導致 C3R 加密用戶端失敗。

由於填補會在加密cleartext之前新增至 `padding`，AES-GCM 具有 1 對 1 映射cleartext至加密文字位元組。base64 編碼將增加 33%。填補的額外儲存體額外負荷的計算方式是將 `padding` 的平均長度cleartext從 `padding` 的值減去pad_length，再乘以 1.33。結果是每筆記錄填補的平均額外負荷。然後，此結果可以乘以資料列數，以取得總填補開銷（如果 preserveNulls 設定為 `true`，請使用非null 值的總數true）。

$$PADDING\_OVERHEAD = (PAD\_LENGTH - AVG\_CLEARTEXT\_LENGTH) * 1.33 * ROW\_COUNT$$

建議您選取包含資料欄中最大值pad_length的最小值。例如，如果最大值是 50 個位元組，則 50 pad_length的 就已足夠。大於 `padding` 的值只會增加額外的儲存體額外負荷。

固定填補不會增加任何顯著的運算額外負荷。

的填補類型 **max**

選取 `padding` 的鍵盤類型max是一種隱私權保留措施，可隱藏資料欄中包含的資料長度。方法是將所有 填入資料欄中cleartext的最大值，加上其他 `padding`，pad_length再加密。一般而言，max填補提供與單一資料集fixed填補相同的保證，同時允許 不知道欄中的最大cleartext值。不過，max填補可能無法提供與fixed填補跨更新相同的隱私權保證，因為個別資料集中的最大值可能不同。

建議您在使用max填補時選取額外的 `padding` 0。此長度會將所有值調整為與資料欄中最大值相同的大小。大於 `padding` 的值只會增加額外的儲存體額外負荷。

如果指定資料欄已知cleartext最大值，建議您改用 `padding` 鍵盤類型。使用fixed填補可在更新後的資料集之間建立一致性。使用max填補會導致每個資料子集填充到子集中最大的值。

資料sealed欄的範例資料

以下是資料sealed欄的輸入和輸出資料集範例，其中包含要重現的設定。allowCleartext、allowJoinsOnColumnsWithDifferentNames和 等其他協同合作層級設定allowDuplicates不會影響結果，false而且如果嘗試在本機重現，可以設定為 true或 `true`。雖然這些是要重現的基本設定，但資料sealed欄是非確定性的，且值每次都會變更。目標是顯示 `padding` 中的位元組與傳出的位元組比較。範例pad_length值是刻意選擇的。它們顯示fixed填補會產生與建議的最小pad_length設定max或需要額外填補時相同的值。

共用秘密範例：wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY

協作 ID 範例：a1b2c3d4-5678-90ab-cdef-EXAMPLE11111

主題

- [的填補類型 none](#)

- [的填補類型 fixed \(範例 1\)](#)
- [的填補類型 fixed \(範例 2\)](#)
- [的填補類型 max \(範例 1\)](#)
- [的填補類型 max \(範例 2\)](#)

的填補類型 **none**

範例 1

輸入	null
preserveNulls	TRUE
Output	null
確定性	Yes
輸入位元組	0
輸出位元組	0

範例 2

輸入	null
preserveNulls	FALSE
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbWlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfssGSPbNIJfG3iXmu6cbCUrizuV
確定性	No
輸入位元組	0
輸出位元組	91

範例 3

輸入	empty string
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc40TBqfRY Z98t5KU6aWfstGSPeM6qR8DWC2P B2GM1X41YK
確定性	No
輸入位元組	0
輸出位元組	91

範例 4

輸入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc40TBqfRY Z98t5KU6aWfsteEE1GKEPiRzyh0 h7t60mWMLTWCv02ckr6pkx9sGL5 VLDQeHzh6DmPpyWNuI=
確定性	No
輸入位元組	26
輸出位元組	127

範例 5

輸入	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqdP4/ Q0Q3cXb/pbvPcnnohrHIGSX54ua+1/ JfcVjc=
確定性	No
輸入位元組	62
輸出位元組	175

的填補類型 **fixed** (範例 1)

在此範例中，pad_length是 62，最大輸入是 62 個位元組。

範例 1

輸入	null
preserveNulls	TRUE
Output	null
確定性	Yes
輸入位元組	0
輸出位元組	0

範例 2

輸入	null
preserveNulls	FALSE
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc40TBqfRY Z98t5KU6aWfssGSNWfMRp7nSb7S MX2s3JKL0hK1+7r75Tk+Mx9jy48 Fcgl1y0PvBqRSZ7oqy1V3UKfYTLEZb/ hCz7oaIneVsrcoNpATs0GzbnLkor4L+/ aSuA=
確定性	No
輸入位元組	0
輸出位元組	175

範例 3

輸入	empty string
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc40TBqfRY Z98t5KU6aWfstGSNWfMRp7nSb7S MX2s3JKL0hK1+7r75Tk+Mx9jy48 Fcgl1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcoLB53l07VZp A60wkuXu29CA=
確定性	No
輸入位元組	0
輸出位元組	175

範例 4

輸入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstEE1GKEPiRzyh0h7t60mWMLTWCv02ckr6pkx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcutBAc0+Mb9tuU2KIIHH31AWg=
確定性	No
輸入位元組	26
輸出位元組	175

範例 5

輸入	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstEE1GKEPiRzyh0h7t60mWMLTWCv02ckr6plwtH/8tRFnn2rF91bcB9G4+n8GiRfJNmqdP4/QOQ3cXb/pbvPcnnohrHIGSX54ua+1/JfcVjc=
確定性	No
輸入位元組	62

輸出位元組	175

的填補類型 **fixed** (範例 2)

在此範例中，`pad_length`是 162，最大輸入是 62 個位元組。

範例 1

輸入	null
<code>preserveNulls</code>	TRUE
Output	null
確定性	Yes
輸入位元組	0
輸出位元組	0

範例 2

輸入	null
<code>preserveNulls</code>	FALSE
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfssGSNWfMRp7nSb7SMX2s3JKLOhK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcnkB0xbLWD7zNdAqQGR0rXoSESdW0I0vpNoGcBfv4cJbG0A3h1DvtkSSVc2B8000GppzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn+8o4WtG/ClipNcjDXvXVtK4vfCohcCA6uwrmwv/xAySX+xcntotL703aBTBb

確定性	No
輸入位元組	0
輸出位元組	307

範例 3

輸入	empty string
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstGSNWfMRp7nSb7SMX2s3JKL0hK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcnkB0xbLWD7zNdAqQGR0rXoSESdW0I0vpNoGcBfv4cJbG0A3h1DvtkSSVc2B8000GppzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn+8o4WtG/ClipNcjDXvXVtK4vfCohcCA6uwrmwv84lVaT9Yd+6oQx65/+gdVT
確定性	No
輸入位元組	0
輸出位元組	307

範例 4

輸入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRY

	Z98t5KU6aWfsteEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6pkx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwT5Hn1+Wyf06ks3QMaRDGSf
確定性	No
輸入位元組	26
輸出位元組	307

範例 5

輸入	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfsteEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqd P4/Q0Q3cXb/pbvPcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwjkJXQZ0gPdeFX9Yr/8a1V5i
確定性	No

輸入位元組	62
輸出位元組	307

的填補類型 **max** (範例 1)

在此範例中， `pad_length` 為 0，最大輸入為 62 個位元組。

範例 1

輸入	<code>null</code>
<code>preserveNulls</code>	<code>TRUE</code>
Output	<code>null</code>
確定性	<code>Yes</code>
輸入位元組	0
輸出位元組	0

範例 2

輸入	<code>null</code>
<code>preserveNulls</code>	<code>FALSE</code>
Output	<code>01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfssGSNWfMRp7nSb7SMX2s3JKLOhK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTL EZb/hCz7oaIneVsrcoNpATs0GzbnLkor4L+/aSuA=</code>
確定性	<code>No</code>
輸入位元組	0

輸出位元組	175

範例 3

輸入	empty string
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstGSNWfMRp7nSb7SMX2s3JKL0hK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsarcoLB53l07VZpA60wkuXu29CA=
確定性	No
輸入位元組	0
輸出位元組	175

範例 4

輸入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstEE1GKEPiRzyh0h7t60mWMLTWCv02ckr6pkx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsircutBAc0+Mb9tuU2KIIHH31AWg=
確定性	No

輸入位元組	26
輸出位元組	175

範例 5

輸入	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqdP4/ QQQ3cXb/pbvPcnnohrHIGSX54ua+1/ JfcVjc=
確定性	No
輸入位元組	62
輸出位元組	175

的填補類型 **max** (範例 2)

在此範例中，pad_length是 100，最大輸入是 62 個位元組。

範例 1

輸入	null
preserveNulls	TRUE
Output	null
確定性	Yes

輸入位元組	0
輸出位元組	0

範例 2

輸入	null
preserveNulls	FALSE
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfssGSNWfMRp7nSb7SMX2s3JKLOhK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcnkB0xbLWD7zNdAqQGR0rXoSESdW0I0vpNoGcBfv4cJbG0A3h1DvtkSSVc2B8000GppzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn+8o4WtG/ClipNcjDXvXVtK4vfCohcCA6uwrmwv/xAySX+xcntotL703aBTBb
確定性	No
輸入位元組	0
輸出位元組	307

範例 3

輸入	empty string
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstGSNWfMRp7nSb7SMX2s3JKLOhK1+7r75Tk+Mx9jy48

	Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv84lVaT9Yd+6oQx65/+gdVT
確定性	No
輸入位元組	0
輸出位元組	307

範例 4

輸入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6pkx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwX5Hn1+Wyf06ks3QMaRDGSf
確定性	No
輸入位元組	26
輸出位元組	307

範例 5

輸入	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqd P4/Q0Q3cXb/pbvPcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwjkJXQZ0gPdeFX9Yr/8a1V5i
確定性	No
輸入位元組	62
輸出位元組	307

對sealed資料欄進行故障診斷

為什麼我sealed欄中的加密文字比cleartext進去的 大小大上數倍？

這取決於幾個因素。其中一個資料Cleartext欄中的加密文字長度一律至少為 91 個位元組。如果您的輸入資料很小（例如，客戶的年齡），其大小將會大幅增加。其次，如果 preserveNulls 設定為 false 且您的輸入資料包含許多 null 值，則每個 null 值都會轉換為 91 個位元組的加密文字。最後，如果您使用填補，則在加密資料之前，依定義，系統會將位元組新增至 cleartext 資料。

資料sealed欄中的大部分資料都很小，我需要使用填補。我是否可以移除大值並分別處理它們以節省空間？

不建議您移除大型值並分別處理。這樣做會變更 C3R 加密用戶端提供的隱私權保證。作為威脅模型，假設觀察者可以看到兩個加密的資料集。如果觀察者發現一個資料子集的資料欄填充量明顯大於或小於另一個子集，他們可以對每個子集中的資料大小進行推論。例如，假設資料 `fullName` 欄在一個檔案中填入總計 40 個位元組，並在另一個檔案中填入 800 個位元組。觀察者可以假設一個資料集包含世界最長的名稱 747 位元組）。

使用填補類型時，是否需要提供額外的 `max` 填補？

否。使用 `max` 填補時，我們建議您將 `pad_length` 設定為 `0`，也稱為超過資料欄中最大值的額外填補。

使用 `fixed` 填補 `pad_length` 時，我可以直接挑選大型物件，以避免擔心最大值是否合適？

是，但大型鍵盤長度效率不佳，使用的儲存空間比所需更多。建議您檢查最大值的大小，並將 `pad_length` 設定為該值。

如何知道是否需要 提供的密碼編譯保證 `preserveNulls`？

不幸的是，答案是它取決於 [您的加密運算 Clean Rooms](#)。應檢閱 `preserveNulls` 的設定如何保護您的資料。不過，我們建議您參考組織的資料處理需求，以及適用於個別協同合作的任何合約。

為什麼我必須產生 `base64` 的額外負荷？

若要允許與 CSV 等表格式檔案格式相容，需要 `base64` 編碼。雖然某些 檔案格式 Parquet 可能支援資料的二進位表示法，但協同合作中的所有參與者都以相同的方式代表資料，以確保適當的查詢結果。

對加密文字大小的非預期增加進行故障診斷

假設您已加密資料，而產生的資料大小非常大。下列步驟可協助您識別大小增加的位置，以及您可以採取哪些動作。

識別大小增加的位置

您必須先識別大小增加的位置，才能對加密的資料明顯大於 `cleartext` 資料的原因進行故障診斷。資料 `Cleartext` 欄可以放心地被忽略，因為它們保持不變。查看其餘 `fingerprint` 和 `sealed` 欄，然後選擇顯示顯著的項目。

識別大小增加的原因

資料 `fingerprint` 欄或資料 `sealed` 欄可能會導致大小增加。

主題

- [大小增加是否來自資料 `fingerprint` 欄？](#)

- [大小增加是否來自資料sealed欄？](#)

大小增加是否來自資料fingerprint欄？

如果對增加儲存體貢獻最大的資料欄是資料fingerprint欄，這可能是因為cleartext資料很小（例如，客戶年齡）。每個產生的fingerprint加密文字長度為 52 個位元組。遺憾的是，無法逐column-by-column處理此問題。如需詳細資訊，請參閱 [fingerprint 資料欄的基本額外負荷](#) 以取得此欄的詳細資訊，包括其如何影響儲存需求。

資料fingerprint欄大小增加的另一個可能原因是協同合作設定 preserveNulls。如果停用 preserveNulls 的協同合作設定（預設設定），資料fingerprint欄中的所有null值都會變成 52 個位元組的加密文字。在目前的協同合作中，無法對此執行任何操作。此preserveNulls設定是在建立協同合作時設定，且所有協同合作者必須使用相同的設定，以確保正確的查詢結果。如需preserveNulls設定及其如何影響資料隱私權保證的詳細資訊，請參閱 [the section called “密碼編譯運算”](#)。

大小增加是否來自資料sealed欄？

如果對儲存體增加貢獻最大的資料欄是資料sealed欄，則有一些詳細資訊可能會導致大小增加。

如果cleartext資料很小（例如，客戶年齡），則每個產生的sealed加密文字長度至少為 91 個位元組。遺憾的是，此問題無法解決。如需詳細資訊，請參閱 [sealed 資料欄的基本額外負荷](#) 以取得此欄的詳細資訊，包括其如何影響儲存需求。

資料sealed欄儲存增加的第二個主要原因是填補。在加密cleartext之前，填補會將額外的位元組新增至，以隱藏資料集中個別值的大小。建議您將填補設定為資料集的最小可能值。至少，pad_length對於fixed填補，必須設定為包含資料欄中最大的可能值。任何高於的設定都不會新增額外的隱私權保證。例如，如果您知道欄中最大的可能值可以是 50 個位元組，建議您pad_length將設定為 50 個位元組。不過，如果資料sealed欄使用max填補，建議您pad_length將設定為 0 位元組。這是因為max填補是指超出資料欄中最大值的額外填補。

資料sealed欄大小增加的最終可能原因是協同合作設定 preserveNulls。如果停用 preserveNulls 的協同合作設定（預設設定），資料sealed欄中的所有null值都會變成 91 個位元組的加密文字。在目前的協同合作中，無法對此執行任何操作。preserveNulls 設定是在建立協同合作時設定，且所有協同合作者必須使用相同的設定，以確保正確的查詢結果。如需此設定的詳細資訊，以及啟用它如何影響資料的隱私權保證，請參閱 [the section called “密碼編譯運算”](#)。

設定 AWS Clean Rooms

下列主題說明如何設定 AWS Clean Rooms。

主題

- [註冊 AWS](#)
- [設定 的服務角色 AWS Clean Rooms](#)
- [設定 AWS Clean Rooms ML 的服務角色](#)

註冊 AWS

您必須 AWS 使用 AWS 服務註冊 AWS Clean Rooms，才能使用 或任何 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。

在註冊程序期間，您會收到一通電話，其中包含您將在電話鍵盤中輸入的驗證碼。

3. 當您註冊 時 AWS 帳戶，會建立AWS 帳戶 根使用者。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，[將管理存取權指派給管理使用者](#)，並且僅使用根使用者來執行 [需要根使用者存取權的任務](#)。

設定 的服務角色 AWS Clean Rooms

下列各節說明執行每個任務所需的角色。

主題

- [建立管理員使用者](#)
- [為協同合作成員建立 IAM 角色](#)
- [建立服務角色以從 Amazon S3 讀取資料](#)
- [建立服務角色以從 Amazon Athena 讀取資料](#)
- [建立服務角色以從 Snowflake 讀取資料](#)

- [建立服務角色以從 S3 儲存貯體讀取程式碼 \(PySpark 分析範本角色 \)](#)
- [建立服務角色以寫入 PySpark 任務的結果](#)
- [建立服務角色以接收結果](#)

建立管理員使用者

若要使用 AWS Clean Rooms，您需要為自己建立管理員使用者，並將管理員使用者新增至管理員群組。

若要建立管理員使用者，請選擇下列其中一個選項。

選擇一種管理管理員的方式	到	根據	您也可以
在 IAM Identity Center (建議)	使用短期憑證存取 AWS。 這與安全性最佳實務一致。有關最佳實務的資訊，請參閱 IAM 使用者指南中的 IAM 安全最佳實務 。	請遵循 AWS IAM Identity Center 使用者指南的 入門 中的說明。	在 AWS Command Line Interface 使用者指南中設定 AWS CLI 以使用來設定 AWS IAM Identity Center 程式設計存取。
在 IAM 中 (不建議使用)	使用長期憑證存取 AWS。	請遵循《IAM 使用者指南》中 建立 IAM 使用者以進行緊急存取 的指示。	請依照《IAM 使用者指南》中的 管理 IAM 使用者的存取金鑰 設定以程式設計方式存取。

為協同合作成員建立 IAM 角色

成員是參與協作 AWS 的客戶。

為協同合作成員建立 IAM 角色

1. 遵循《AWS Identity and Access Management 使用者指南》中的[建立角色將許可委派給 IAM 使用者](#)程序。
2. 在建立政策步驟中，選取政策編輯器中的 JSON 索引標籤，然後根據授予協同合作成員的能力來新增政策。

AWS Clean Rooms 根據常見的使用案例提供下列 受管政策。

如果您想要...	然後使用 ...
檢視資源和中繼資料	AWS 受管政策：AWSCleanRoomsReadOnlyAccess
Query	AWS 受管政策：AWSCleanRoomsFullAccess
查詢和執行任務	AWS 受管政策：AWSCleanRoomsFullAccess
查詢和接收結果	AWS 受管政策：AWSCleanRoomsFullAccess
管理協同合作資源，但不查詢	AWS 受管政策：AWSCleanRoomsFullAccessNoQuerying

如需 所提供不同受管政策的相關資訊 AWS Clean Rooms，請參閱 [AWS 的 受管政策 AWS Clean Rooms](#)、

建立服務角色以從 Amazon S3 讀取資料

AWS Clean Rooms 使用服務角色從 Amazon S3 讀取資料。

有兩種方法可以建立此服務角色。

- 如果您有建立服務角色所需的 IAM 許可，請使用 AWS Clean Rooms 主控台建立服務角色。
- 如果您沒有 `iam:CreateRole`、`iam:CreatePolicy`和 `iam:AttachRolePolicy`許可，或想要手動建立 IAM 角色，請執行下列其中一項操作：

- 使用下列程序，使用自訂信任政策建立服務角色。
- 請您的管理員使用下列程序建立服務角色。

Note

只有在您沒有使用 AWS Clean Rooms 主控台建立服務角色的必要許可時，您或您的 IAM 管理員才應該遵循此程序。

建立服務角色以使用自訂信任政策從 Amazon S3 讀取資料

1. 使用自訂信任政策建立角色。如需詳細資訊，請參閱AWS Identity and Access Management 《使用者指南》中的[使用自訂信任政策（主控台）建立角色](#)程序。
2. 根據使用自訂信任政策 [（主控台）程序建立角色，使用下列自訂信任政策](#)。

Note

如果您想要協助確保角色僅用於特定協同合作成員資格的內容，您可以進一步縮小信任政策的範圍。如需詳細資訊，請參閱[預防跨服務混淆代理人](#)。

JSON

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

3. 根據[使用自訂信任政策（主控台）](#)程序建立角色，使用下列許可政策。

Note

下列範例政策支援讀取 AWS Glue 中繼資料及其對應 Amazon S3 資料所需的許可。不過，您可能需要根據設定 Amazon S3 資料的方式修改此政策。例如，如果您已為 Amazon S3 資料設定自訂 KMS 金鑰，您可能需要使用其他 AWS Key Management Service (AWS KMS) 許可來修改此政策。

您的 AWS Glue 資源和基礎 Amazon S3 資源必須與 AWS Clean Rooms 協同合作 AWS 區域 位於相同的 中。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "NecessaryGluePermissions",
      "Effect": "Allow",
      "Action": [
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetTable",
        "glue:GetTables",
        "glue:GetPartition",
        "glue:GetPartitions",
        "glue:BatchGetPartition"
      ],
      "Resource": [
        "arn:aws:glue:us-east-1:111122223333:database/databaseName",
        "arn:aws:glue:us-east-1:111122223333:table/databaseName/tableName",
        "arn:aws:glue:us-east-1:111122223333:catalog"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "glue:GetSchema",
        "glue:GetSchemaVersion"
      ]
    }
  ]
}
```

```

    ],
    "Resource": [
        "*"
    ]
},
{
    "Sid": "NecessaryS3BucketPermissions",
    "Effect": "Allow",
    "Action": [
        "s3:GetBucketLocation",
        "s3:ListBucket"
    ],
    "Resource": [
        "arn:aws:s3:::bucket"
    ],
    "Condition": {
        "StringEquals": {
            "s3:ResourceAccount": [
                "444455556666"
            ]
        }
    }
},
{
    "Sid": "NecessaryS3ObjectPermissions",
    "Effect": "Allow",
    "Action": [
        "s3:GetObject"
    ],
    "Resource": [
        "arn:aws:s3:::bucket/prefix/*"
    ],
    "Condition": {
        "StringEquals": {
            "s3:ResourceAccount": [
                "444455556666"
            ]
        }
    }
}
]
}

```

Note

此政策參考兩個不同的 AWS 帳戶 IDs 以支援由 AWS Clean Rooms 不同方管理資料目錄中繼資料和實際資料儲存的協同合作：

- 111122223333 - 這是擁有 AWS Glue Data Catalog 資源（資料庫、資料表和目錄）的帳戶。第一個陳述式授予許可，以從此帳戶的 AWS Glue 目錄中存取資料表結構描述、分割區資訊和中繼資料。
- 444455556666 - 這是擁有包含實際資料檔案之 Amazon S3 儲存貯體的帳戶。Amazon S3 許可（陳述式 3 和 4）僅限於此帳戶透過 `s3:ResourceAccount` 條件擁有的儲存貯體。

此組態支援常見的企業資料架構，其中一個團隊管理資料目錄和結構描述定義，而另一個團隊擁有基礎資料儲存基礎設施。`s3:ResourceAccount` 條件透過確保 Amazon S3 操作僅適用於指定帳戶擁有的儲存貯體，提供額外的安全層。

4. 將每個####取代為您自己的資訊。
5. 繼續遵循[使用自訂信任政策（主控台）建立角色](#)程序來建立角色。

建立服務角色以從 Amazon Athena 讀取資料

AWS Clean Rooms 使用服務角色從 Amazon Athena 讀取資料。

建立服務角色以使用自訂信任政策從 Athena 讀取資料

1. 使用自訂信任政策建立角色。如需詳細資訊，請參閱AWS Identity and Access Management 《使用者指南》中的[使用自訂信任政策（主控台）建立角色](#)程序。
2. 根據使用自訂信任政策 [（主控台）程序建立角色，使用下列自訂信任政策](#)。

Note

如果您想要協助確保角色僅用於特定協同合作成員資格的內容，您可以進一步縮小信任政策的範圍。如需詳細資訊，請參閱[預防跨服務混淆代理人](#)。

JSON

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

3. 根據[使用自訂信任政策（主控台）](#)程序建立角色，使用下列許可政策。

 Note

下列範例政策支援讀取 AWS Glue 中繼資料及其對應 Athena 資料所需的許可。不過，您可能需要根據設定 Amazon S3 資料的方式修改此政策。例如，如果您已為 Amazon S3 資料設定自訂 KMS 金鑰，您可能需要使用其他 AWS KMS 許可來修改此政策。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "athena:GetWorkGroup",
        "athena:GetTableMetadata",
        "athena:GetDataCatalog",
        "athena:StartQueryExecution",
        "athena:GetQueryExecution",
        "athena:GetQueryResults"
      ]
    }
  ],
}
```



```

        "Resource": [
            "arn:aws:athena:region:accountId:workgroup/workgroup",
            "arn:aws:athena:region:accountId:datacatalog/federatedCatalogName"
        ],
    },
    {
        "Effect": "Allow",
        "Action": [
            "glue:GetDatabase",
            "glue:GetTable",
            "glue:GetCatalog"
        ],
        "Resource": [
            "arn:aws:glue:region:accountId:catalog",
            "arn:aws:glue:region:accountId:catalog/federatedCatalogName",
            "arn:aws:glue:region:accountId:database/federatedCatalogName/databaseName",
            "arn:aws:glue:region:accountId:table/federatedCatalogName/databaseName/tableName"
        ],
    },
    {
        "Effect": "Allow",
        "Action": [
            "s3:GetObject",
            "s3:GetBucketLocation",
            "s3:AbortMultipartUpload",
            "s3:ListBucket",
            "s3:PutObject",
            "s3:ListMultipartUploadParts"
        ],
        "Resource": [
            "arn:aws:s3:::athenaResultsBucket",
            "arn:aws:s3:::athenaResultsBucket/*"
        ],
        "Condition": {
            "StringEquals": {
                "aws:ResourceAccount": "accountId"
            }
        }
    },
    {
        "Effect": "Allow",
        "Action": "lakeformation:GetDataAccess",
    }

```

```
        "Resource": "*"
      }
    ]
  }
}
```

4. 將每個####取代為您自己的資訊。
5. 繼續遵循[使用自訂信任政策（主控台）建立角色](#)程序來建立角色。

設定 Lake Formation 許可

如果您查詢受 Lake Formation 許可保護的資源，服務角色必須具有table/view/catalog的選取和描述存取許可，以及 AWS Glue 資料庫的描述許可。

如需詳細資訊，請參閱：

- 《Amazon Athena [Athena 使用者指南](#)》中的[使用 Athena 查詢向 註冊的資料 AWS Lake Formation](#)
- 《AWS Lake Formation 開發人員指南》中的[加入 Lake Formation 許可](#)

建立服務角色以從 Snowflake 讀取資料

AWS Clean Rooms 使用服務角色擷取 Snowflake 的登入資料，以從此來源讀取您的資料。

有兩種方法可以建立此服務角色：

- 如果您有建立服務角色所需的 IAM 許可，請使用 AWS Clean Rooms 主控台建立服務角色。
- 如果您沒有 iam:CreateRole、iam:CreatePolicy和 iam:AttachRolePolicy許可，或想要手動建立 IAM 角色，請執行下列其中一項操作：
 - 使用下列程序，使用自訂信任政策建立服務角色。
 - 請您的管理員使用下列程序建立服務角色。

Note

只有在您沒有使用 AWS Clean Rooms 主控台建立服務角色的必要許可時，您或您的 IAM 管理員才應該遵循此程序。

建立服務角色以使用自訂信任政策從 Snowflake 讀取資料

1. 使用自訂信任政策建立角色。如需詳細資訊，請參閱AWS Identity and Access Management 《使用者指南》中的[使用自訂信任政策（主控台）程序建立角色](#)。
2. 根據使用自訂信任政策 [（主控台）程序建立角色](#)，使用下列自訂信任政策。

Note

如果您想要協助確保角色僅用於特定協同合作成員資格的内容，您可以進一步縮小信任政策的範圍。如需詳細資訊，請參閱[預防跨服務混淆代理人](#)。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowIfSourceArnMatches",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "ForAnyValue:ArnEquals": {
          "aws:SourceArn": [
            "arn:aws:cleanrooms:us-east-1:111122223333:membership/membershipId",
            "arn:aws:cleanrooms:us-east-1:444455556666:membership/queryRunnerMembershipId"
          ]
        }
      }
    }
  ]
}
```

Note

此信任政策參考兩個不同的 AWS 帳戶 IDs，以支援在 AWS Clean Rooms 多方之間分配查詢執行責任的協同合作：

- 111122223333 - 這是包含參與協作之成員資格的帳戶。此成員資格可能擁有資料表、分析規則或其他需要角色存取的協同合作資源。
- 444455556666 - 這是帳戶，其中包含負責執行查詢的成員資格（「查詢執行器」）。此成員資格會執行受保護的查詢，並需要擔任此角色才能存取必要的運算和資料資源。

此組態可讓一方在另一方執行實際查詢時提供資料或分析範本的情況。這兩個角色都需要透過相同執行角色的不同但互補的許可。aws:SourceArn 條件可確保只有來自這兩個特定成員資格 AWS Clean Rooms 的操作才能擔任該角色，同時維護安全性，同時支援分散式任務執行和結果管理工作流程。

3. 根據[使用自訂信任政策（主控台）](#)程序建立角色，使用下列其中一個許可政策。

使用客戶擁有的 KMS 金鑰加密之秘密的許可政策

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": "secretsmanager:GetSecretValue",
      "Resource": "arn:aws:secretsmanager:us-east-1:111122223333:secret:secretIdentifier",
      "Effect": "Allow"
    },
    {
      "Sid": "AllowDecryptViaSecretsManagerForKey",
      "Action": "kms:Decrypt",
      "Resource": "arn:aws:kms:us-east-1:444455556666:key/keyIdentifier",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
```

```

        "kms:ViaService": "secretsmanager.us-  
east-1.amazonaws.com",  
        "kms:EncryptionContext:SecretARN":  
        "arn:aws:secretsmanager:us-east-1:111122223333:secret:secretIdentifier"  
    }  
}  
}  
]  
}

```

Note

此政策參考兩個不同的 AWS 帳戶 IDs 以支援跨帳戶秘密管理案例：

- 111122223333 - 這是在 中擁有和存放秘密的帳戶。第一個陳述式授予從此帳戶擷取秘密值的許可。
- 444455556666 - 這是擁有用於加密秘密之 AWS KMS 金鑰的帳戶。第二個陳述式授予許可，以使用此帳戶中的 AWS KMS 金鑰解密秘密。

此組態在企業環境中很常見，其中：

- 秘密集中管理在一個帳戶中（帳戶 1）
- 加密金鑰由單獨的安全或共享服務帳戶管理（帳戶 2）
- 帳戶 2 中的 AWS KMS 金鑰政策也必須允許帳戶 1 中的服務使用金鑰進行加密/解密操作

kms:EncryptionContext:SecretARN 條件可確保 AWS KMS 金鑰只能用於解密此特定秘密，為跨帳戶存取提供額外的安全層。

使用 加密之秘密的許可政策 AWS 受管金鑰

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [

```

```

    {
      "Action": "secretsmanager:GetSecretValue",
      "Resource": "arn:aws:secretsmanager:us-
east-1:111122223333:secret:secretIdentifier",
      "Effect": "Allow"
    }
  ]
}

```

4. 將每個####取代為您自己的資訊。
5. 繼續遵循[使用自訂信任政策（主控台）建立角色](#)程序來建立角色。

建立服務角色以從 S3 儲存貯體讀取程式碼 (PySpark 分析範本角色)

AWS Clean Rooms 使用 PySpark 分析範本時，會使用服務角色從協同合作成員指定的 S3 儲存貯體讀取程式碼。

建立服務角色以從 S3 儲存貯體讀取程式碼

1. 使用自訂信任政策建立角色。如需詳細資訊，請參閱AWS Identity and Access Management 《使用者指南》中的[使用自訂信任政策（主控台）程序建立角色](#)。
2. 根據使用自訂信任政策 [（主控台）程序建立角色](#)，使用下列自訂信任政策。

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "ForAnyValue:ArnEquals": {
          "aws:SourceArn": [
            "arn:aws:cleanrooms:us-
east-1:111122223333:membership/jobRunnerMembershipId",
            "arn:aws:cleanrooms:us-
east-1:444455556666:membership/analysisTemplateOwnerMembershipId"
          ]
        }
      }
    }
  ]
}

```

```

    ]
  }
}
]
}

```

Note

此信任政策參考兩個不同的 AWS 帳戶 IDs 以支援多方 AWS Clean Rooms 協同合作案例：

- 111122223333 - 這是帳戶，其中包含負責執行查詢的成員資格 (「任務執行器」)。此成員資格會執行分析任務，並需要擔任此角色才能存取必要的資源。
- 444455556666 - 這是擁有分析範本及其相關聯成員資格的帳戶 (「分析範本擁有者」)。此成員資格定義可以執行哪些查詢，也需要擔任此角色才能管理和執行分析。

此組態在多方參與相同協作的 AWS Clean Rooms 協作中很常見，每個協作都有自己的 AWS 帳戶 和 成員資格。查詢執行器和分析範本擁有者都需要存取共用資源。aws:SourceArn 條件可確保只有來自這兩個特定成員資格 AWS Clean Rooms 的操作可以擔任 角色，為多方協同合作提供精確的存取控制。

3. 根據[使用自訂信任政策 \(主控台\)](#) 程序建立角色，使用下列許可政策。

Note

下列範例政策支援從 Amazon S3 讀取程式碼所需的許可。不過，您可能需要根據設定 S3 資料的方式修改此政策。

您的 Amazon S3 資源必須與協同合作 AWS 區域 位於 AWS Clean Rooms 相同的 中。

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {

```

```

        "Effect": "Allow",
        "Action": [
            "s3:GetObject",
            "s3:GetObjectVersion"
        ],
        "Resource": ["arn:aws:s3:::s3Path"],
        "Condition": {
            "StringEquals": {
                "s3:ResourceAccount": [
                    "s3BucketOwnerAccountId"
                ]
            }
        }
    }
}

```

4. 以您自己的資訊取代每個####：

- *s3Path* – 程式碼的 S3 儲存貯體位置。
- *s3BucketOwnerAccountId* – S3 儲存貯體擁有者的 AWS 帳戶 ID。
- *region* – 的名稱 AWS 區域。例如 **us-east-1**。
- *jobRunnerAccountId* – 可執行查詢和任務的成員 AWS 帳戶 ID。
- *jobRunnerMembershipId* – 可查詢和執行任務之成員的成員 ID。您可以在協同合作的詳細資訊索引標籤中找到成員 ID。這可確保 AWS Clean Rooms 只有在此成員在此協同合作中執行分析時，才會擔任該角色。
- *analysisTemplateAccountId* – 分析範本的 AWS 帳戶 ID。
- *analysisTemplateOwnerMembershipId* – 擁有分析範本之成員的成員 ID。您可以在協同合作的詳細資訊索引標籤中找到成員 ID。

5. 繼續遵循[使用自訂信任政策（主控台）建立角色](#)程序來建立角色。

建立服務角色以寫入 PySpark 任務的結果

AWS Clean Rooms 使用服務角色將 PySpark 任務的結果寫入指定的 S3 儲存貯體。

建立服務角色以寫入 PySpark 任務的結果

1. 使用自訂信任政策建立角色。如需詳細資訊，請參閱AWS Identity and Access Management 《使用者指南》中的[使用自訂信任政策（主控台）程序建立角色](#)。

2. 根據使用自訂信任政策 (主控台) 程序建立角色，使用下列自訂信任政策。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "ForAnyValue:ArnEquals": {
          "aws:SourceArn": [
            "arn:aws:cleanrooms:us-east-1:111122223333:membership/jobRunnerMembershipId",
            "arn:aws:cleanrooms:us-east-1:444455556666:membership/rrMembershipId"
          ]
        }
      }
    }
  ]
}
```

Note

此信任政策參考兩個不同的 AWS 帳戶 IDs，以支援與不同操作角色的 AWS Clean Rooms 協同合作：

- 111122223333 - 這是帳戶，其中包含負責執行分析任務的成員資格 (「任務執行器」)。此成員資格會執行運算工作負載，並需要擔任此角色才能存取處理資源。
- 444455556666 - 這是包含具有結果接收者 (RR) 責任成員資格的帳戶。此成員資格獲授權接收和存取分析任務的輸出，並需要角色存取權才能將結果寫入指定的位置。

此組態可讓一方在另一方接收和管理結果時執行運算分析的 AWS Clean Rooms 案例。這兩個角色都需要透過相同執行角色的不同但互補的許可。aws:SourceArn 條件可確保

只有來自這兩個特定成員資格 AWS Clean Rooms 的操作可以擔任該角色，同時維護安全性，同時支援分散式任務執行和結果管理工作流程。

3. 根據[使用自訂信任政策（主控台）](#)程序建立角色，使用下列許可政策。

Note

下列範例政策支援寫入 Amazon S3 所需的許可。不過，您可能需要根據設定 S3 的方式修改此政策。

您的 Amazon S3 資源必須與協同合作 AWS 區域 位於 AWS Clean Rooms 相同的 中。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:PutObject"
      ],
      "Resource": "arn:aws:s3::bucket/optionalPrefix/*",
      "Condition": {
        "StringEquals": {
          "s3:ResourceAccount": [
            "s3BucketOwnerAccountId"
          ]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetBucketLocation",
        "s3:ListBucket"
      ],
      "Resource": "arn:aws:s3::bucket",
      "Condition": {
        "StringEquals": {
          "s3:ResourceAccount": [
```

```

}
[
  }
  }
  }
  ]
  "s3BucketOwnerAccountId"
]

```

- 將每個####取代為您自己的資訊：
 - region* – 的名稱 AWS 區域。例如 **us-east-1**。
 - jobRunnerAccountId* – S3 儲存貯體所在的 AWS 帳戶 ID。
 - jobRunnerMembershipId* – 可查詢和執行任務之成員的成員 ID。您可以在協同合作的詳細資訊索引標籤中找到成員 ID。這可確保 AWS Clean Rooms 只有在此成員在此協同合作中執行分析時，才會擔任該角色。
 - rrAccountId* – S3 AWS 帳戶 儲存貯體所在的 ID。
 - rrMembershipId* – 可接收結果之成員的成員 ID。您可以在協同合作的詳細資訊索引標籤中找到成員 ID。這可確保 AWS Clean Rooms 只有在此成員在此協同合作中執行分析時，才會擔任該角色。
 - ###體* – S3 儲存貯體的名稱和位置。
 - optionalPrefix* – 如果您想要將結果儲存在特定 S3 字首下，則為選用字首。
 - s3BucketOwnerAccountId* – S3 儲存貯體擁有者的 AWS 帳戶 ID。
- 繼續遵循使用自訂信任政策（主控台）建立角色程序來建立角色。

建立服務角色以接收結果

 Note

如果您是只能接收結果的成員（在主控台中，您的成員能力只是接收結果），請遵循此程序。
如果您是可以同時查詢和接收結果的成員（在主控台中，您的成員能力同時是查詢和接收結果），您可以略過此程序。

對於只能接收結果的協同合作成員，AWS Clean Rooms 會使用服務角色將協同合作中查詢資料的結果寫入指定的 S3 儲存貯體。

有兩種方法可以建立此服務角色：

- 如果您有建立服務角色所需的 IAM 許可，請使用 AWS Clean Rooms 主控台建立服務角色。
- 如果您沒有 `iam:CreateRole`、`iam:CreatePolicy` 和 `iam:AttachRolePolicy` 許可，或想要手動建立 IAM 角色，請執行下列其中一項操作：
 - 使用下列程序，使用自訂信任政策建立服務角色。
 - 請您的管理員使用下列程序建立服務角色。

Note

只有在您沒有使用 AWS Clean Rooms 主控台建立服務角色的必要許可時，您或您的 IAM 管理員才應該遵循此程序。

使用自訂信任政策建立服務角色以接收結果

1. 使用自訂信任政策建立角色。如需詳細資訊，請參閱 AWS Identity and Access Management 《使用者指南》中的 [使用自訂信任政策（主控台）程序建立角色](#)。
2. 根據使用自訂信任政策 [（主控台）程序建立角色](#)，使用下列自訂信任政策。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowIfExternalIdMatches",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "ArnLike": {
          "sts:ExternalId":
            "arn:aws:*:region:*:dbuser:*/a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaa*"
        }
      }
    },
    {
      "Sid": "AllowIfSourceArnMatches",
```

```

        "Effect": "Allow",
        "Principal": {
            "Service": "cleanrooms.amazonaws.com"
        },
        "Action": "sts:AssumeRole",
        "Condition": {
            "ForAnyValue:ArnEquals": {
                "aws:SourceArn": [
                    "arn:aws:cleanrooms:us-  
east-1:555555555555:membership/a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaaa"
                ]
            }
        }
    }
}

```

3. 根據[使用自訂信任政策（主控台）](#)程序建立角色，使用下列許可政策。

Note

下列範例政策支援讀取 AWS Glue 中繼資料及其對應 Amazon S3 資料所需的許可。不過，您可能需要根據設定 S3 資料的方式修改此政策。

您的 AWS Glue 資源和基礎 Amazon S3 資源必須與 AWS Clean Rooms 協同合作 AWS 區域 位於相同的 中。

JSON

```

{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetBucketLocation",
                "s3:ListBucket"
            ],
            "Resource": [

```

```

        "arn:aws:s3:::bucket_name"
    ],
    "Condition": {
        "StringEquals": {
            "aws:ResourceAccount": "accountId"
        }
    }
},
{
    "Effect": "Allow",
    "Action": [
        "s3:PutObject"
    ],
    "Resource": [
        "arn:aws:s3:::bucket_name/optional_key_prefix/*"
    ],
    "Condition": {
        "StringEquals": {
            "aws:ResourceAccount": "accountId"
        }
    }
}
]
}

```

4. 將每個####取代為您自己的資訊：

- *region* – 的名稱 AWS 區域。例如 **us-east-1**。
- *a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaaa* – 可查詢之成員的成員 ID。您可以在協同合作的詳細資訊索引標籤中找到成員 ID。這可確保 AWS Clean Rooms 只有在此成員在此協同合作中執行分析時，才會擔任該角色。
- *arn#aws#cleanrooms#us-east-1#555555555555#membership/a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaaa* – 可查詢成員的單一成員 ARN。您可以在協同合作的詳細資訊索引標籤中找到成員資格 ARN。這可確保只有當此成員在此協同合作中執行分析時，才會 AWS Clean Rooms 擔任該角色。
- *bucket_name* – S3 儲存貯體的 Amazon Resource Name (ARN)。您可以在 Amazon S3 中儲存貯體的屬性索引標籤中找到 Amazon Resource Name (ARN)。 Amazon S3
- *accountId* – S3 儲存貯體所在的 AWS 帳戶 ID。

bucket_name/optional_key_prefix – Amazon S3 中結果目的地的 Amazon Resource Name (ARN)。Amazon S3 您可以在 Amazon S3 中儲存貯體的屬性索引標籤中找到 Amazon Resource Name (ARN)。Amazon S3

5. 繼續遵循[使用自訂信任政策（主控台）建立角色](#)程序來建立角色。

設定 AWS Clean Rooms ML 的服務角色

執行外觀建模所需的角色與使用自訂模型所需的角色不同。下列各節說明執行每個任務所需的角色。

主題

- [設定類似模型的服務角色](#)
- [設定自訂建模的服務角色](#)

設定類似模型的服務角色

主題

- [建立服務角色以讀取訓練資料](#)
- [建立服務角色以撰寫外觀相似的客群](#)
- [建立服務角色以讀取種子資料](#)

建立服務角色以讀取訓練資料

AWS Clean Rooms 使用服務角色來讀取訓練資料。如果您有必要的 IAM 許可，您可以使用 主控台 建立此角色。如果您沒有 CreateRole 許可，請要求您的管理員建立服務角色。

建立服務角色以訓練資料集

1. 使用您的管理員帳戶登入 IAM 主控台 (<https://console.aws.amazon.com/iam/> : //)。
2. 在 Access management (存取管理) 下，請選擇 Policies (政策)。
3. 選擇 Create policy (建立政策)。
4. 在政策編輯器中，選取 JSON 索引標籤，然後複製並貼上下列政策。

 Note

下列範例政策支援讀取 AWS Glue 中繼資料及其對應 Amazon S3 資料所需的許可。不過，您可能需要根據設定 S3 資料的方式修改此政策。此政策不包含用於解密資料的 KMS 金鑰。

您的 AWS Glue 資源和基礎 Amazon S3 資源必須與 AWS Clean Rooms 協同合作 AWS 區域 位於相同的 中。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetTable",
        "glue:GetTables",
        "glue:GetPartitions",
        "glue:GetPartition",
        "glue:BatchGetPartition",
        "glue:GetUserDefinedFunctions"
      ],
      "Resource": [
        "arn:aws:glue:us-east-1:111122223333:database/databases",
        "arn:aws:glue:us-east-1:111122223333:table/databases/tables",
        "arn:aws:glue:us-east-1:111122223333:catalog",
        "arn:aws:glue:us-east-1:111122223333:database/default"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "glue:CreateDatabase"
      ],
      "Resource": [
        "arn:aws:glue:us-east-1:111122223333:database/default"
      ]
    }
  ]
}
```



```

    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "s3:ListBucket",
      "s3:GetBucketLocation"
    ],
    "Resource": [
      "arn:aws:s3:::bucket"
    ],
    "Condition": {
      "StringEquals": {
        "s3:ResourceAccount": [
          "111122223333"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "s3:GetObject"
    ],
    "Resource": [
      "arn:aws:s3:::bucketFolders/*"
    ],
    "Condition": {
      "StringEquals": {
        "s3:ResourceAccount": [
          "111122223333"
        ]
      }
    }
  }
]
}

```

如果您需要使用 KMS 金鑰來解密資料，請將此 AWS KMS 陳述式新增至先前的範本：

```

{
    "Effect": "Allow",

```

```

    "Action": [
        "kms:Decrypt",
    ],
    "Resource": [
        "arn:aws:kms:region:accountId:key/keyId"
    ],
    "Condition": {
        "ArnLike": {
            "kms:EncryptionContext:aws:s3:arn":
                "arn:aws:s3:::bucketFolders*"
        }
    }
}

```

5. 將每個####取代為您自己的資訊：

- *region* – 的名稱 AWS 區域。例如 **us-east-1**。
- *accountId* – S3 儲存貯體所在的 AWS 帳戶 ID。
- *###/###*、*table/databases/tables*、*##*和*###/##* – AWS Clean Rooms 需要存取的訓練資料位置。
- 儲存*#*體 – S3 儲存貯體的 Amazon Resource Name (ARN)。您可以在 Amazon S3 中儲存貯體的屬性索引標籤中找到 Amazon Resource Name (ARN)。 Amazon S3
- *bucketFolders* – AWS Clean Rooms 需要存取的 S3 儲存貯體中特定資料夾的名稱。

6. 選擇下一步。

7. 針對檢閱和建立，輸入政策名稱和描述，然後檢閱摘要。

8. 選擇建立政策。

您已為 建立政策 AWS Clean Rooms。

9. 在 Access management (存取管理) 下，請選擇 Roles (角色)。

使用 角色，您可以建立短期登入資料，為提高安全性而建議這麼做。您也可以選擇使用者來建立長期登入資料。

10. 選擇建立角色。

11. 在建立角色精靈中，針對信任的實體類型，選擇自訂信任政策。

12. 將下列自訂信任政策複製並貼到 JSON 編輯器。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowAssumeRole",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEqualsIfExists": {
          "aws:SourceAccount": ["111122223333"]
        },
        "ArnLikeIfExists": {
          "aws:SourceArn": "arn:aws:cleanrooms-ml:us-east-1:111122223333:training-dataset/*"
        }
      }
    }
  ]
}
```

一律SourceAccount是您的 AWS 帳戶。SourceArn 可以限制為特定訓練資料集，但僅限於建立該資料集之後。由於您尚不知道訓練資料集 ARN，在此指定萬用字元。

accountId 是 的 ID AWS 帳戶，其中包含訓練資料。

13. 選擇下一步，然後在新增許可下，輸入您剛建立的政策名稱。（您可能需要重新載入頁面。）
14. 選取您建立的政策名稱旁的核取方塊，然後選擇下一步。
15. 針對名稱、檢閱和建立，輸入角色名稱和描述。

 Note

角色名稱必須符合授予成員passRole許可中的模式，該成員可以查詢和接收結果和成員角色。

- a. 檢閱選取信任的實體，並視需要編輯。
- b. 檢閱新增許可中的許可，並視需要編輯。
- c. 檢閱標籤，並視需要新增標籤。
- d. 選擇建立角色。

您已為 建立 服務角色 AWS Clean Rooms。

建立服務角色以撰寫外觀相似的客群

AWS Clean Rooms 使用服務角色將類似區段寫入儲存貯體。如果您有必要的 IAM 許可，您可以使用主控台建立此角色。如果您沒有 CreateRole 許可，請要求您的管理員建立服務角色。

建立服務角色以撰寫類似樣子的客群

1. 使用您的管理員帳戶登入 IAM 主控台 (<https://console.aws.amazon.com/iam/> : //)。
2. 在 Access management (存取管理) 下，請選擇 Policies (政策)。
3. 選擇 Create policy (建立政策)。
4. 在政策編輯器中，選取 JSON 索引標籤，然後複製並貼上下列政策。

Note

下列範例政策支援讀取 AWS Glue 中繼資料及其對應 Amazon S3 資料所需的許可。不過，您可能需要根據設定 Amazon S3 資料的方式修改此政策。此政策不包含用於解密資料的 KMS 金鑰。

您的 AWS Glue 資源和基礎 Amazon S3 資源必須與 AWS Clean Rooms 協同合作 AWS 區域 位於相同的 中。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
```

```

        "s3:ListBucket",
        "s3:GetBucketLocation"
    ],
    "Resource": [
        "arn:aws:s3:::buckets"
    ],
    "Condition": {
        "StringEquals": {
            "s3:ResourceAccount": [
                "accountId"
            ]
        }
    }
},
{
    "Effect": "Allow",
    "Action": [
        "s3:PutObject"
    ],
    "Resource": [
        "arn:aws:s3:::bucketFolders/*"
    ],
    "Condition": {
        "StringEquals": {
            "s3:ResourceAccount": [
                "accountId"
            ]
        }
    }
}
]
}

```

如果您需要使用 KMS 金鑰來加密資料，請將此 AWS KMS 陳述式新增至範本：

```

{
    "Effect": "Allow",
    "Action": [
        "kms:Encrypt",
        "kms:GenerateDataKey*",
        "kms:ReEncrypt*",
    ],

```

```

    "Resource": [
      "arn:aws:kms:region:accountId:key/keyId"
    ],
    "Condition": {
      "ArnLike": {
        "kms:EncryptionContext:aws:s3:arn":
          "arn:aws:s3:::bucketFolders*"
      }
    }
  }
]
}

```

5. 將每個####取代為您自己的資訊：

- ####體 – S3 儲存貯體的 Amazon Resource Name (ARN)。您可以在 Amazon S3 中儲存貯體的屬性索引標籤中找到 Amazon Resource Name (ARN)。 Amazon S3
- *accountId* – S3 儲存貯體所在的 AWS 帳戶 ID。
- *bucketFolders* – AWS Clean Rooms 需要存取的 S3 儲存貯體中特定資料夾的名稱。
- *region* – 的名稱 AWS 區域。例如 **us-east-1**。
- *keyId* – 加密資料所需的 KMS 金鑰。

6. 選擇下一步。

7. 針對檢閱和建立，輸入政策名稱和描述，然後檢閱摘要。

8. 選擇建立政策。

您已為 建立政策 AWS Clean Rooms。

9. 在 Access management (存取管理) 下，請選擇 Roles (角色)。

透過 角色，您可以建立短期登入資料，為提高安全性而建議這麼做。您也可以選擇使用者來建立長期登入資料。

10. 選擇建立角色。

11. 在建立角色精靈中，針對信任的實體類型，選擇自訂信任政策。

12. 將下列自訂信任政策複製並貼到 JSON 編輯器。

JSON

```
{
```

```

"Version": "2012-10-17",
"Statement": [
  {
    "Sid": "AllowAssumeRole",
    "Effect": "Allow",
    "Principal": {
      "Service": "cleanrooms-ml.amazonaws.com"
    },
    "Action": "sts:AssumeRole",
    "Condition": {
      "StringEqualsIfExists": {
        "aws:SourceAccount": ["111122223333"]
      },
      "ArnLikeIfExists": {
        "aws:SourceArn": "arn:aws:cleanrooms-ml:us-
east-1:111122223333:configured-audience-model/*"
      }
    }
  }
]
}

```

一律SourceAccount是您的 AWS 帳戶。SourceArn 可以限制為特定訓練資料集，但僅限於建立該資料集之後。由於您尚不知道訓練資料集 ARN，在此指定萬用字元。

13. 選擇下一步。
14. 選取您建立的政策名稱旁的核取方塊，然後選擇下一步。
15. 針對名稱、檢閱和建立，輸入角色名稱和描述。

Note

角色名稱必須符合授予成員passRole許可中的模式，該成員可以查詢和接收結果和成員角色。

- a. 檢閱選取信任的實體，並視需要編輯。
- b. 檢閱新增許可中的許可，並視需要編輯。
- c. 檢閱標籤，並視需要新增標籤。

d. 選擇建立角色。

您已為 建立 服務角色 AWS Clean Rooms。

建立服務角色以讀取種子資料

AWS Clean Rooms 使用服務角色讀取種子資料。如果您有必要的 IAM 許可，您可以使用 主控台 建立此角色。如果您沒有 CreateRole 許可，請要求您的管理員建立服務角色。

建立服務角色以讀取存放在 S3 儲存貯體中的種子資料。

1. 使用您的管理員帳戶登入 IAM 主控台 (<https://console.aws.amazon.com/iam/> : //)。
2. 在 Access management (存取管理) 下，請選擇 Policies (政策)。
3. 選擇 Create policy (建立政策)。
4. 在政策編輯器中，選取 JSON 索引標籤，然後複製並貼上下列其中一個政策。

Note

下列範例政策支援讀取 AWS Glue 中繼資料及其對應 Amazon S3 資料所需的許可。不過，您可能需要根據設定 Amazon S3 資料的方式修改此政策。此政策不包含用於解密資料的 KMS 金鑰。

您的 AWS Glue 資源和基礎 Amazon S3 資源必須與 AWS Clean Rooms 協同合作 AWS 區域 位於相同的 中。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:ListBucket"
      ],
      "Resource": [
        "arn:aws:s3:::buckets"
      ],
    }
  ]
}
```



```

        "Condition": {
            "StringEquals": {
                "s3:ResourceAccount": [
                    "accountId"
                ]
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucketFolders/*"
            ],
            "Condition": {
                "StringEquals": {
                    "s3:ResourceAccount": [
                        "accountId"
                    ]
                }
            }
        }
    ]
}

```

Note

下列範例政策支援讀取 SQL 查詢結果並使用它做為輸入資料所需的許可。不過，您可能需要根據查詢的結構來修改此政策。此政策不包含用於解密資料的 KMS 金鑰。

JSON

```

{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCleanRoomsStartQuery",

```

```

        "Effect": "Allow",
        "Action": [
            "cleanrooms:GetCollaborationAnalysisTemplate",
            "cleanrooms:GetSchema",
            "cleanrooms:StartProtectedQuery"
        ],
        "Resource": "*"
    },
    {
        "Sid": "AllowCleanRoomsGetAndUpdateQuery",
        "Effect": "Allow",
        "Action": [
            "cleanrooms:GetProtectedQuery",
            "cleanrooms:UpdateProtectedQuery"
        ],
        "Resource": [
            "arn:aws:cleanrooms:us-east-1:111122223333:membership/queryRunnerMembershipId"
        ]
    }
]
}

```

如果您需要使用 KMS 金鑰來解密資料，請將此 AWS KMS 陳述式新增至範本：

```

{
    "Effect": "Allow",
    "Action": [
        "kms:Decrypt",
        "kms:DescribeKey"
    ],
    "Resource": [
        "arn:aws:kms:region:accountId:key/keyId"
    ],
    "Condition": {
        "ArnLike": {
            "kms:EncryptionContext:aws:s3:arn":
                "arn:aws:s3::bucketFolders*"
        }
    }
}
]

```

```
}
```

5. 將每個####取代為您自己的資訊：

- ####體 – S3 儲存貯體的 Amazon Resource Name (ARN)。您可以在 Amazon S3 中儲存貯體的屬性索引標籤中找到 Amazon Resource Name (ARN)。 Amazon S3
- *accountId* – S3 儲存貯體所在的 AWS 帳戶 ID。
- *bucketFolders* – AWS Clean Rooms 需要存取的 S3 儲存貯體中特定資料夾的名稱。
- *region* – 的名稱 AWS 區域。例如 **us-east-1**。
- *queryRunnerAccountId* – 將執行查詢的帳戶 AWS 帳戶 ID。
- *queryRunnerMembershipId* – 可查詢之成員的成員 ID。您可以在協同合作的詳細資訊索引標籤中找到成員 ID。這可確保只有當此成員在此協同合作中執行分析時，AWS Clean Rooms 才會擔任該角色。
- *keyId* – 加密資料所需的 KMS 金鑰。

6. 選擇下一步。

7. 針對檢閱和建立，輸入政策名稱和描述，然後檢閱摘要。

8. 選擇建立政策。

您已為 建立政策 AWS Clean Rooms。

9. 在 Access management (存取管理) 下，請選擇 Roles (角色)。

透過 角色，您可以建立短期登入資料，為提高安全性而建議這麼做。您也可以選擇使用者來建立長期登入資料。

10. 選擇建立角色。

11. 在建立角色精靈中，針對信任的實體類型，選擇自訂信任政策。

12. 將下列自訂信任政策複製並貼到 JSON 編輯器。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowAssumeRole",
      "Effect": "Allow",
      "Principal": {
```

```

        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEqualsIfExists": {
          "aws:SourceAccount": ["111122223333"]
        },
        "ArnLikeIfExists": {
          "aws:SourceArn": "arn:aws:cleanrooms-ml:us-
east-1:111122223333:audience-generation-job/*"
        }
      }
    }
  ]
}

```

一律SourceAccount是您的 AWS 帳戶。SourceArn 可以限制為特定訓練資料集，但僅限於建立該資料集之後。由於您尚不知道訓練資料集 ARN，在此指定萬用字元。

13. 選擇下一步。
14. 選取您建立的政策名稱旁的核取方塊，然後選擇下一步。
15. 針對名稱、檢閱和建立，輸入角色名稱和描述。

Note

角色名稱必須符合授予成員passRole許可中的模式，該成員可以查詢和接收結果和成員角色。

- a. 檢閱選取信任的實體，並視需要編輯。
- b. 檢閱新增許可中的許可，並視需要編輯。
- c. 檢閱標籤，並視需要新增標籤。
- d. 選擇建立角色。

您已為 建立 服務角色 AWS Clean Rooms。

設定自訂建模的服務角色

主題

- [建立自訂 ML 建模的服務角色 - ML 組態](#)
- [建立服務角色以提供自訂 ML 模型](#)
- [建立服務角色以查詢資料集](#)
- [建立服務角色以建立設定的資料表關聯](#)

建立自訂 ML 建模的服務角色 - ML 組態

AWS Clean Rooms 使用服務角色來控制誰可以建立自訂 ML 組態。如果您有必要的 IAM 許可，您可以使用主控台建立此角色。如果您沒有 `CreateRole` 許可，請要求管理員建立服務角色。

此角色可讓您使用 [PutMLConfiguration](#) 動作。

建立服務角色以允許建立自訂 ML 組態

1. 使用您的管理員帳戶登入 IAM 主控台 (<https://console.aws.amazon.com/iam/> : //)。
2. 在 Access management (存取管理) 下，請選擇 Policies (政策)。
3. 選擇 Create policy (建立政策)。
4. 在政策編輯器中，選取 JSON 索引標籤，然後複製並貼上下列政策。

Note

下列範例政策支援存取和寫入資料至 S3 儲存貯體以及發佈 CloudWatch 指標所需的許可。不過，您可能需要根據設定 Amazon S3 資料的方式修改此政策。此政策不包含用於解密資料的 KMS 金鑰。

您的 Amazon S3 資源必須與 AWS Clean Rooms 協同合作 AWS 區域 位於相同的 中。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowS3ObjectWriteForExport",
```

```

        "Effect": "Allow",
        "Action": [
            "s3:PutObject"
        ],
        "Resource": [
            "arn:aws:s3:::bucket/*"
        ],
        "Condition": {
            "StringEquals": {
                "s3:ResourceAccount": [
                    "111122223333"
                ]
            }
        }
    },
    {
        "Sid": "AllowS3KMSEncryptForExport",
        "Effect": "Allow",
        "Action": [
            "kms:Encrypt",
            "kms:GenerateDataKey*"
        ],
        "Resource": [
            "arn:aws:kms:us-east-1:111122223333:key/keyId"
        ],
        "Condition": {
            "StringLike": {
                "kms:EncryptionContext:aws:s3:arn":
                    "arn:aws:s3:::bucket*"
            }
        }
    },
    {
        "Sid": "AllowCloudWatchMetricsPublishingForTrainingJobs",
        "Action": "cloudwatch:PutMetricData",
        "Resource": "*",
        "Effect": "Allow",
        "Condition": {
            "StringLike": {
                "cloudwatch:namespace": "/aws/cleanroomsml/*"
            }
        }
    },
    {

```

```

        "Sid": "AllowCloudWatchLogsPublishingForTrainingOrInferenceJobs",
        "Effect": "Allow",
        "Action": [
            "logs:CreateLogGroup",
            "logs:CreateLogStream",
            "logs:DescribeLogStreams",
            "logs:PutLogEvents"
        ],
        "Resource": [
            "arn:aws:logs:us-east-1:111122223333:log-group:/aws/cleanroomsml/*"
        ]
    }
}

```

5. 將每個####取代為您自己的資訊：

- 儲存#體 – S3 儲存貯體的 Amazon Resource Name (ARN)。您可以在 Amazon S3 中儲存貯體的屬性索引標籤中找到 Amazon Resource Name (ARN)。 Amazon S3
- *region* – 的名稱 AWS 區域。例如 **us-east-1**。
- *accountId* – S3 儲存貯體所在的 AWS 帳戶 ID。
- *keyId* – 加密資料所需的 KMS 金鑰。

6. 選擇下一步。

7. 針對檢閱和建立，輸入政策名稱和描述，然後檢閱摘要。

8. 選擇建立政策。

您已為 建立政策 AWS Clean Rooms。

9. 在 Access management (存取管理) 下，請選擇 Roles (角色)。

使用 角色，您可以建立短期登入資料，為提高安全性而建議這麼做。您也可以選擇使用者來建立長期登入資料。

10. 選擇建立角色。

11. 在建立角色精靈中，針對信任的實體類型，選擇自訂信任政策。

12. 將下列自訂信任政策複製並貼到 JSON 編輯器。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "111122223333"
        },
        "ArnLike": {
          "aws:SourceArn": "arn:aws:cleanrooms:us-east-1:111122223333:membership/membershipID"
        }
      }
    }
  ]
}
```

一律SourceAccount是您的 AWS 帳戶。SourceArn 可以限制為特定訓練資料集，但僅限於建立該資料集之後。由於您尚不知道訓練資料集 ARN，在此指定萬用字元。

13. 選擇下一步。
14. 選取您建立的政策名稱旁的核取方塊，然後選擇下一步。
15. 針對名稱、檢閱和建立，輸入角色名稱和描述。

 Note

角色名稱必須符合授予成員passRole許可中的模式，該成員可以查詢和接收結果和成員角色。

- a. 檢閱選取信任的實體，並視需要編輯。

- b. 檢閱新增許可中的許可，並視需要編輯。
- c. 檢閱標籤，並視需要新增標籤。
- d. 選擇建立角色。

您已為 建立 服務角色 AWS Clean Rooms。

建立服務角色以提供自訂 ML 模型

AWS Clean Rooms 使用服務角色來控制誰可以建立自訂 ML 模型演算法。如果您有必要的 IAM 許可，您可以使用 主控台 建立此角色。如果您沒有 `CreateRole` 許可，請要求管理員建立服務角色。

此角色可讓您使用 [CreateConfiguredModelAlgorithm](#) 動作。

建立服務角色以允許成員提供自訂 ML 模型

1. 使用您的管理員帳戶登入 IAM 主控台 (<https://console.aws.amazon.com/iam/> : //)。
2. 在 Access management (存取管理) 下，請選擇 Policies (政策)。
3. 選擇 Create policy (建立政策)。
4. 在政策編輯器中，選取 JSON 索引標籤，然後複製並貼上下列政策。

Note

下列範例政策支援擷取包含模型演算法之 Docker 映像所需的許可。不過，您可能需要根據設定 Amazon S3 資料的方式修改此政策。此政策不包含用於解密資料的 KMS 金鑰。您的 Amazon S3 資源必須與 AWS Clean Rooms 協同合作 AWS 區域 位於相同的 中。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowECRIImageDownloadForTrainingAndInferenceJobs",
      "Effect": "Allow",
      "Action": [
        "ecr:BatchGetImage",
        "ecr:BatchCheckLayerAvailability",
```

```

        "ecr:GetDownloadUrlForLayer"
      ],
      "Resource": "arn:aws:ecr:us-east-1:111122223333:repository/repoName"
    }
  ]
}

```

5. 將每個####取代為您自己的資訊：

- *region* – 的名稱 AWS 區域。例如 **us-east-1**。
- *accountId* – S3 儲存貯體所在的 AWS 帳戶 ID。
- *repoName* – 包含您資料的儲存庫名稱。

6. 選擇下一步。

7. 針對檢閱和建立，輸入政策名稱和描述，然後檢閱摘要。

8. 選擇建立政策。

您已為 建立政策 AWS Clean Rooms。

9. 在 Access management (存取管理) 下，請選擇 Roles (角色)。

透過 角色，您可以建立短期登入資料，為提高安全性而建議這麼做。您也可以選擇使用者來建立長期登入資料。

10. 選擇建立角色。

11. 在建立角色精靈中，針對信任的實體類型，選擇自訂信任政策。

12. 將下列自訂信任政策複製並貼到 JSON 編輯器。

JSON

```

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

```

```
}
```

一律SourceAccount是您的 AWS 帳戶 SourceArn可以限制在特定訓練資料集內，但僅限於建立該資料集之後。由於您尚不知道訓練資料集 ARN，在此指定萬用字元。

13. 選擇下一步。
14. 選取您建立的政策名稱旁的核取方塊，然後選擇下一步。
15. 針對名稱、檢閱和建立，輸入角色名稱和描述。

Note

角色名稱必須符合授予passRole成員許可中的模式，該成員可以查詢和接收結果和成員角色。

- a. 檢閱選取信任的實體，並視需要編輯。
- b. 檢閱新增許可中的許可，並視需要編輯。
- c. 檢閱標籤，並視需要新增標籤。
- d. 選擇建立角色。

您已為 建立 服務角色 AWS Clean Rooms。

建立服務角色以查詢資料集

AWS Clean Rooms 使用服務角色來控制誰可以查詢將用於自訂 ML 建模的資料集。如果您有必要的 IAM 許可，您可以使用 主控台 建立此角色。如果您沒有CreateRole許可，請要求管理員建立服務角色。

此角色可讓您使用 [CreateMLInputChannel](#) 動作。

建立服務角色以允許成員查詢資料集

1. 使用您的管理員帳戶登入 IAM 主控台 (<https://console.aws.amazon.com/iam/> : //)。
2. 在 Access management (存取管理) 下，請選擇 Policies (政策)。
3. 選擇 Create policy (建立政策)。
4. 在政策編輯器中，選取 JSON 索引標籤，然後複製並貼上下列政策。

Note

下列範例政策支援查詢將用於自訂 ML 建模的資料集所需的許可。不過，您可能需要根據設定 Amazon S3 資料的方式修改此政策。此政策不包含用於解密資料的 KMS 金鑰。您的 Amazon S3 資源必須與 AWS Clean Rooms 協同合作 AWS 區域 位於相同的 中。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowCleanRoomsStartQueryForMLInputChannel",
      "Effect": "Allow",
      "Action": "cleanrooms:StartProtectedQuery",
      "Resource": "*"
    },
    {
      "Sid": "AllowCleanroomsGetSchemaAndGetAnalysisTemplateForMLInputChannel",
      "Effect": "Allow",
      "Action": [
        "cleanrooms:GetSchema",
        "cleanrooms:GetCollaborationAnalysisTemplate"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AllowCleanRoomsGetAndUpdateQueryForMLInputChannel",
      "Effect": "Allow",
      "Action": [
        "cleanrooms:GetProtectedQuery",
        "cleanrooms:UpdateProtectedQuery"
      ],
      "Resource": [
        "arn:aws:cleanrooms:us-  
east-1:111122223333:membership/queryRunnerMembershipId"
      ]
    }
  ]
}
```

```
]
}
```

5. 將每個####取代為您自己的資訊：

- *region* – 的名稱 AWS 區域。例如 **us-east-1**。
- *queryRunnerAccountId* – 將執行查詢的帳戶 AWS 帳戶 ID。
- *queryRunnerMembershipId* – 可查詢之成員的成員 ID。您可以在協同合作的詳細資訊索引標籤中找到成員 ID。這可確保只有當此成員在此協同合作中執行分析時，AWS Clean Rooms 才會擔任該角色。

6. 選擇下一步。

7. 針對檢閱和建立，輸入政策名稱和描述，然後檢閱摘要。

8. 選擇建立政策。

您已為 建立政策 AWS Clean Rooms。

9. 在 Access management (存取管理) 下，請選擇 Roles (角色)。

透過 角色，您可以建立短期登入資料，為提高安全性而建議這麼做。您也可以選擇使用者來建立長期登入資料。

10. 選擇建立角色。

11. 在建立角色精靈中，針對信任的實體類型，選擇自訂信任政策。

12. 將下列自訂信任政策複製並貼到 JSON 編輯器。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-m1.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

一律SourceAccount是您的 AWS 帳戶 SourceArn可以限制在特定訓練資料集內，但僅限於建立該資料集之後。由於您尚不知道訓練資料集 ARN，在此指定萬用字元。

13. 選擇下一步。
14. 選取您建立的政策名稱旁的核取方塊，然後選擇下一步。
15. 針對名稱、檢閱和建立，輸入角色名稱和描述。

 Note

角色名稱必須符合授予成員passRole許可中的模式，該成員可以查詢和接收結果和成員角色。

- a. 檢閱選取信任的實體，並視需要編輯。
- b. 檢閱新增許可中的許可，並視需要編輯。
- c. 檢閱標籤，並視需要新增標籤。
- d. 選擇建立角色。

您已為 建立 服務角色 AWS Clean Rooms。

建立服務角色以建立設定的資料表關聯

AWS Clean Rooms 使用服務角色來控制誰可以建立設定的資料表關聯。如果您有必要的 IAM 許可，您可以使用 主控台 建立此角色。如果您沒有CreateRole許可，請要求管理員建立服務角色。

此角色可讓您使用 CreateConfiguredTableAssociation 動作。

建立服務角色以允許建立設定的資料表關聯

1. 使用您的管理員帳戶登入 IAM 主控台 (<https://console.aws.amazon.com/iam/> : //)。
2. 在 Access management (存取管理) 下，請選擇 Policies (政策)。
3. 選擇 Create policy (建立政策)。
4. 在政策編輯器中，選取 JSON 索引標籤，然後複製並貼上下列政策。

 Note

下列範例政策支援建立設定的資料表關聯。不過，您可能需要根據設定 Amazon S3 資料的方式修改此政策。此政策不包含用於解密資料的 KMS 金鑰。

您的 Amazon S3 資源必須與 AWS Clean Rooms 協同合作 AWS 區域 位於相同的 中。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "kms:Decrypt",
        "kms:DescribeKey"
      ],
      "Resource": "arn:aws:kms:us-east-1:111122223333:key/KMS-key-ID",
      "Effect": "Allow"
    },
    {
      "Action": [
        "s3:ListBucket",
        "s3:GetBucketLocation"
      ],
      "Resource": "arn:aws:s3:::bucket-name",
      "Effect": "Allow"
    },
    {
      "Action": "s3:GetObject",
      "Resource": "arn:aws:s3:::bucket-name/*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetTable",
        "glue:GetTables",
        "glue:GetPartitions",
```

```

        "glue:GetPartition",
        "glue:BatchGetPartition"
    ],
    "Resource": [
        "arn:aws:glue:us-east-1:111122223333:catalog",
        "arn:aws:glue:us-east-1:111122223333:database/Glue database name",
        "arn:aws:glue:us-east-1:111122223333:table/Glue database name/Glue table name"
    ],
    "Effect": "Allow"
},
{
    "Action": [
        "glue:GetSchema",
        "glue:GetSchemaVersion"
    ],
    "Resource": "*",
    "Effect": "Allow"
}
]
}

```

取代預留位置資源 ARNs

使用此政策時，您必須將預留位置資源識別符取代為您資源的實際 ARNs：

- AWS KMS 金鑰資源：將 *KMS-key-ID* 取代為加密 Amazon S3 資料的實際 AWS KMS 金鑰 ID。金鑰必須位於擁有 AWS Glue 目錄資源的相同帳戶 ((111122223333) 中。
- Amazon S3 儲存貯體資源：將 *bucket-name* 取代為包含 AWS Glue 資料表資料的 Amazon S3 儲存貯體實際名稱。請注意，Amazon S3 儲存貯體 ARNs 不包含帳戶 IDs 因為儲存貯體名稱是全域唯一的。
- AWS Glue 資源：將下列預留位置取代為您的實際資源名稱：
 - *Glue #####* - AWS Glue 資料庫的名稱
 - *Glue #####* - AWS Glue 資料表的名稱

所有 AWS Glue 資源（目錄、資料庫和資料表）必須位於相同的 AWS 帳戶（111122223333），以確保一致的存取許可。此帳戶應該與擁有用於資料加密之 AWS KMS 金鑰的帳戶相同，為您的 AWS Clean Rooms 資料資源建立統一的安全界限。

5. 使用您自己的資訊取代每個####：

- #### *Amazon S3* #### *KMS* ## – 用來加密 Amazon S3 資料的 KMS 金鑰。若要解密資料，您需要提供用來加密資料的相同 KMS 金鑰。
- *AWS Glue* #### *Amazon S3* #### – 包含您資料之 AWS Glue 資料表的 Amazon S3 儲存貯體名稱。
- *region* – 的名稱 AWS 區域。例如 **us-east-1**。
- *accountId* – 擁有資料的帳戶 AWS 帳戶 ID。
- *AWS Glue* ##### – 包含您資料的 AWS Glue 資料庫名稱。
- *AWS Glue* ##### – 包含您資料的 AWS Glue 資料表名稱。

6. 選擇下一步。

7. 針對檢閱和建立，輸入政策名稱和描述，然後檢閱摘要。

8. 選擇建立政策。

您已為 建立政策 AWS Clean Rooms。

9. 在 Access management (存取管理) 下，請選擇 Roles (角色)。

透過 角色，您可以建立短期登入資料，為提高安全性而建議這麼做。您也可以選擇使用者來建立長期登入資料。

10. 選擇建立角色。

11. 在建立角色精靈中，針對信任的實體類型，選擇自訂信任政策。

12. 將下列自訂信任政策複製並貼到 JSON 編輯器。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
```

```
        "Principal": {
            "Service": "cleanrooms-ml.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }
}
```

一律SourceAccount是您的 AWS 帳戶 SourceArn可以限制在特定訓練資料集內，但僅限於建立該資料集之後。由於您尚不知道訓練資料集 ARN，在此指定萬用字元。

13. 選擇下一步。
14. 選取您建立的政策名稱旁的核取方塊，然後選擇下一步。
15. 針對名稱、檢閱和建立，輸入角色名稱和描述。

 Note

角色名稱必須符合授予成員passRole許可中的模式，該成員可以查詢和接收結果和成員角色。

- a. 檢閱選取信任的實體，並視需要編輯。
- b. 檢閱新增許可中的許可，並視需要編輯。
- c. 檢閱標籤，並視需要新增標籤。
- d. 選擇建立角色。

您已為 建立 服務角色 AWS Clean Rooms。

中的協作和成員資格 AWS Clean Rooms

協同合作是安全的邏輯界限，成員可以在其中 AWS Clean Rooms 對設定的資料表進行分析。

中的任何成員 AWS Clean Rooms 都可以建立協同合作。

協同合作建立者可以指定單一成員來分析設定的資料表並接收結果。不過，協同合作建立者可能想要防止可執行分析的成員存取查詢結果。在這種情況下，協同合作建立者可以指定一個[成員給誰可以查詢](#)，或一個[成員可以執行查詢和任務](#)，另一個[成員可以接收結果](#)。

在大多數情況下，可以查詢的成員或可以查詢和執行任務的成員也是[支付運算成本的成員](#)。不過，協同合作建立者可以設定不同的成員來負責支付查詢運算成本。

如需有關如何使用 AWS SDKs 建立協同合作的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

主題

- [建立協同合作](#)
- [建立成員資格並加入協同合作](#)
- [編輯協同合作](#)
- [在 中變更請求AWS Clean Rooms](#)
- [刪除協同合作](#)
- [檢視協同合作](#)
- [邀請成員參與協作](#)
- [監控成員](#)
- [將成員新增至協同合作](#)
- [從協同合作中移除成員](#)
- [離開協同合作](#)

建立協同合作

有三種方式可在其中建立協同合作 AWS Clean Rooms。

最基本的形式是[查詢的協同合作](#)。此協同合作著重於 SQL 查詢分析，並維護具有兩個主要角色的簡單結構：一個成員可以執行查詢，另一個成員可以接收結果。此基本協同作業設定非常適合簡單的資料分析任務。

第二種形式是[查詢和任務的協同合作](#)，透過整合 SQL 查詢和 PySpark 任務來擴展功能。此協同合作設定會維護相同的基本角色結構，但會展開包含任務執行的許可。值得注意的要求是，建立 PySpark 分析範本的成員也必須是接收結果的成員，以確保分析過程中的明確責任。

第三個形式是[ML 建模的協作](#)，專為機器學習工作流程而建置。此協同合作設定會新增兩個角色：一個用於需要訓練模型結果的使用者，另一個用於需要使用這些模型結果進行預測的使用者。此協同合作設定可協助協同合作成員共同處理複雜的資料專案，同時保持每個人的角色和許可清晰。

下列主題說明如何建立查詢、任務和 ML 建模的協同合作。

主題

- [建立查詢的協同合作](#)
- [為查詢和任務建立協同合作](#)
- [為 ML 建模建立協作](#)

建立查詢的協同合作

在此程序中，身為[協作建立者](#)的您執行下列任務：

- [建立協同合作](#)。
- 邀請一或多個[成員](#)參與[協作](#)。
- 指派能力給成員，例如[可查詢的成員](#)和[可接收結果的成員](#)。

如果協作建立者也是可以接收結果的成員，他們會指定結果目的地和格式。他們也提供服務角色 Amazon Resource Name (ARN)，將結果寫入結果目的地。

- 設定哪些[成員負責支付協同合作中的運算成本](#)。

在開始前，請確定您具有以下先決條件：

- 您擁有要邀請到協同合作的每個成員的名稱和 AWS 帳戶 ID。
- 您有權與協同合作的所有成員共用每個成員的名稱和 AWS 帳戶 ID。

Note

您無法在建立協同合作之後新增更多成員。

如需有關如何使用 AWS SDKs 建立協同合作的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

建立查詢的協同合作

1. 登入 AWS 管理主控台，並使用開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶，以做為協同合作建立者。
2. 在左側導覽窗格中，選擇協同合作。
3. 在右上角，選擇建立協同合作。
4. 針對步驟 1：定義協同合作，執行下列動作：

- a. 如需詳細資訊，請輸入協同合作的名稱和描述。

受邀參與協作的協作成員可以看到此資訊。名稱和描述有助於他們了解協同合作的內容。

- b. 對於成員：

- i. 針對成員 1：輸入成員顯示名稱，如同您希望在協同合作時顯示的名稱。

Note

您的 AWS 帳戶 ID 會自動包含在成員 AWS 帳戶 ID 中。

- ii. 針對成員 2，輸入您要邀請加入協同合作之成員的成員顯示名稱和成員 AWS 帳戶 ID。

受邀參與協同合作的每個人都可以看到成員顯示名稱和成員 AWS 帳戶 ID。輸入並儲存這些欄位的值之後，您就無法對其進行編輯。

Note

您必須通知協同合作成員，協同合作中所有受邀和作用中的協同合作者都可看見其成員 AWS 帳戶 ID 和成員顯示名稱。

- iii. 如果您想要新增其他成員，請選擇新增其他成員。然後輸入每個成員的成員顯示名稱和成員 AWS 帳戶 ID，這些成員可以提供您要邀請加入協同合作的資料。

- c. 如果您想要啟用分析記錄，請選取啟用分析記錄核取方塊。

- 選擇支援日誌類型下的查詢日誌核取方塊。

您將會在 Amazon CloudWatch Logs 帳戶中收到從 SQL 查詢產生的日誌。

- d. 如果您想要啟用詳細監控，請選取啟用詳細監控核取方塊。

分析執行器和設定的付款人可以選擇在建立其成員資格時啟用詳細指標。啟用時，詳細的監控指標將發佈至 CloudWatch 以進行協同合作的操作監控，包括查詢效能和資源使用率。這些指標可供分析執行器和各自 AWS 帳戶中設定的付款人使用。

如需了解有關 CloudWatch 定價的詳細資訊，請參閱 [CloudWatch 定價](#)。

- e. 在允許的查詢結果區域下，選取您要傳送查詢結果的一或多個 AWS 區域 位置。

根據預設，只會選取目前的區域（例如維吉尼亞北部 us-east-1）。

 Important

當您啟用跨區域查詢結果交付時，您的結果可能會處理並存放在來源區域之外。

如需區域的詳細資訊，請參閱《》中的 [區域和端點](#) AWS 一般參考。

- f. （選用）設定哪些設定可自動變更，無需手動核准變更請求，即可透過自動變更請求核准來管理對資料的存取。根據預設，某些設定只能透過提交變更請求來變更，該請求必須經過所有成員的核准才能生效。
- 授予成員能力 – 選擇可授予協作成員的功能，無需手動核准。成員一律可以貢獻資料。
 - 選擇功能：
 - 貢獻資料（一律啟用）
 - 接收結果
 - 自動核准具有這些功能的新成員 – 如果允許，任何具有上述選取功能新增的成員都會立即加入協作。新增其他功能的成員仍需要手動核准才能加入。
 - 可自動撤銷的功能 – 選擇無需手動核准即可撤銷的功能。成員一律可以貢獻資料。
 - 選擇功能：
 - 貢獻資料（一律啟用）
 - 接收結果

如果您選擇此選項，您可以透過位於協同作業詳細資訊頁面詳細資訊索引標籤的變更請求歷史記錄來追蹤所有協同作業組態修改。

- g. （選用）如果您想要啟用密碼編譯運算功能，請選取啟用密碼編譯運算核取方塊。

- i. 選擇下列密碼編譯涵蓋範圍參數：

- 允許plaintext資料欄

如果您需要完全加密的資料表，請選擇否。

如果您想要加密資料表中允許cleartext的資料欄，請選擇是。

若要在特定資料欄AVG上執行 SUM或 ，資料欄必須位於 中cleartext。

- 保留NULL值

如果您不想保留NULL值，請選擇否。 NULL值不會在加密資料表NULL中顯示為 。

如果您想要保留NULL值，請選擇是。 NULL值會在加密的資料表NULL中顯示為 。

ii. 選擇下列指紋參數：

- 允許重複項目

如果您不希望資料fingerprint欄中允許重複的項目，請選擇否。

如果您想要在資料fingerprint欄中允許重複的項目，請選擇是。

- 允許具有不同名稱JOIN的資料欄

如果您不想使用不同的名稱聯結fingerprint資料欄，請選擇否。

如果您想要使用不同的名稱聯結fingerprint資料欄，請選擇是。

如需密碼編譯運算參數的詳細資訊，請參閱 [密碼編譯運算參數](#)。

如需如何加密資料以用於 的詳細資訊 AWS Clean Rooms，請參閱 [使用的密碼編譯運算準備加密資料表 Clean Rooms](#)。

 Note

在完成下一個步驟之前，請仔細驗證這些組態。建立協同合作之後，您只能編輯協同合作名稱、描述，以及日誌是否存放在 Amazon CloudWatch Logs 中。

h. 如果您想要為協同合作資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。

i. 選擇下一步。

5. 對於步驟 2：指定成員能力，對於使用查詢和任務的分析，在支援的分析類型下，會保持選取查詢核取方塊，並根據目標採取建議的動作。

您的目標	建議的動作
查詢協同合作中的資料並接收結果	1. 選擇您自己做為可執行查詢的成員。 2. 選擇自己做為成員，從下拉式清單中接收分析結果。
查詢協同合作中的資料，並指派不同的成員來接收結果	1. 選擇您自己做為可執行查詢的成員。 2. 從下拉式清單中選取可從分析中接收結果的成員。
接收協同合作中查詢的結果，並指派不同的成員來查詢資料	1. 從下拉式清單中選取可執行查詢的成員。 2. 選擇自己做為成員，從下拉式清單中接收來自分析的結果。
建立和管理協同合作、指派不同的成員來查詢資料，以及指派不同的成員來接收結果	1. 從下拉式清單中選取可執行查詢的成員。 2. 從下拉式清單中選取可從分析中接收結果的成員。

- a. 如果您使用 Clean Rooms ML，對於使用專用工作流程的 ML 建模，
 - i. （選用）從下拉式清單中選取可以從訓練模型接收輸出的成員。
 - ii. （選用）從下拉式清單中選取可以從模型推論接收輸出的成員。
 - b. 使用 檢視 ID 解析 AWS Entity Resolution 下的成員功能。
 - c. 選擇下一步。
6. 針對步驟 3：設定付款，針對使用查詢的分析，請根據您的目標採取下列其中一個動作。

您的目標	建議的動作
將可執行查詢的成員指派為支付查詢運算成本的成員	1. 對於使用查詢的分析，選擇將支付查詢費用的成員與可執行查詢的成員相同。 2. 選擇下一步。
指派不同的成員來支付查詢運算成本	1. 對於使用查詢的分析，選擇您自己作為將支付查詢費用的成員。

您的目標	建議的動作
	2. 選擇下一步。

對於使用專用工作流程的 ML 建模，已設定外觀模型的建立者是將支付外觀建模費用的成員。

對於使用 ID 解析 AWS Entity Resolution，ID 映射表的建立者是將為 ID 映射表付費的成員。

7. 針對步驟 4：設定成員資格，選擇下列其中一個選項：

Yes, join by creating membership now

1. 對於結果設定預設值，對於查詢結果設定，如果您是可以接收結果的成員，
 - a. 針對 Amazon S3 中的結果目的地，輸入 Amazon S3 目的地，或選擇瀏覽 S3 以選取 S3 儲存貯體。
 - b. 針對查詢結果格式，選擇 CSV 或 PARQUET。
 - c. (僅限 Spark) 對於結果檔案，選擇多個或單一。
 - d. (選用) 對於服務存取，如果您想要將最多需要 24 小時的查詢交付到 S3 目的地，請選取新增服務角色以支援最多需要 24 小時才能完成的查詢核取方塊。

最多需要 24 小時才能完成的大型查詢將會傳送到您的 S3 目的地。

如果您未選取核取方塊，則只會將 12 小時內完成的查詢傳送到您的 S3 位置。

- e. 選取建立和使用新的服務角色或使用現有的服務角色，以指定服務存取許可。

如果您選擇 ...	然後 ...
建立和使用新的服務角色	• AWS Clean Rooms 會建立具有此資料表所需政策的服務角色。 • 預設的服務角色名稱為 <code>cleanrooms-result-receiver- <timestamp></code> • 您必須具有建立角色和連接政策的許可。
使用現有的服務角色	i. 從下拉式清單中選擇現有的服務角色名稱。

如果您選擇 ...	然後 ...
	如果您有列出角色的許可，則會顯示角色清單。 如果您沒有列出角色的許可，您可以輸入要使用的角色的 Amazon Resource Name (ARN)。 ii. 選擇 IAM 外部連結中的檢視，以檢視服務角色。 如果沒有現有的服務角色，則無法使用使用現有服務角色的選項。 根據預設，AWS Clean Rooms 不會嘗試更新現有的角色政策來新增必要的許可。

 Note

- AWS Clean Rooms 需要根據分析規則查詢的許可。如需許可的詳細資訊 AWS Clean Rooms，請參閱 [AWS 的 受管政策 AWS Clean Rooms](#)。
- 如果角色沒有足夠的許可 AWS Clean Rooms，您會收到錯誤訊息，指出角色沒有足夠的許可 AWS Clean Rooms。必須先新增角色政策，才能繼續。
- 如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

2. 針對日誌設定，為 Amazon CloudWatch Logs 中的日誌儲存選擇下列其中一個選項：

 Note

如果您選擇啟用查詢記錄，則會顯示日誌設定區段。

a. 選擇開啟，與您相關的查詢日誌將存放在您的 Amazon CloudWatch Logs 帳戶中。

每個成員只能收到其啟動或包含其資料的查詢的日誌。

可以接收結果的成員也會收到在協同合作中執行之所有查詢的日誌，即使未在查詢中存取其資料。

在支援的日誌類型下，預設會開啟查詢日誌核取方塊。

 Note

開啟查詢記錄後，可能需要幾分鐘的時間才能設定日誌儲存，並開始在 Amazon CloudWatch Logs 中接收日誌。在此短暫期間內，可以查詢的成員可能會執行未實際傳送日誌的查詢。

- b. 選擇關閉，與您相關的查詢日誌將不會存放在您的 Amazon CloudWatch Logs 帳戶中。
3. 如果您想要為成員資格資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
4. 如果您是支付查詢運算費用的成員，請選取我同意在此協同合作核取方塊中支付運算成本，以表示您接受。

 Note

您必須選取此核取方塊才能繼續。

如需如何計算定價的詳細資訊，請參閱 [的定價 AWS Clean Rooms](#)。

如果您是[支付查詢運算成本的成員](#)，但不是[可以查詢的成員](#)，建議您使用 AWS Budgets 來設定預算，AWS Clean Rooms 並在達到預算上限時接收通知。如需設定預算的詳細資訊，請參閱AWS Cost Management 《使用者指南》中的[使用 管理您的成本 AWS Budgets](#)。如需設定通知的詳細資訊，請參閱[《使用者指南》中的為預算通知建立 Amazon SNS 主題](#)。AWS Cost Management 如果已達到預算上限，您可以聯絡可執行查詢或[離開協同合作](#)的成員。如果您離開協同合作，將不再允許執行任何查詢，因此您不再需要支付查詢運算成本。

5. 選擇下一步。

協同合作和您的成員資格都會建立。

協同合作中的狀態為作用中。

No, I will create a membership later

1. 選擇下一步。

只會建立協同合作。

您在協同合作中的狀態為非作用中。

8. 針對步驟 5：檢閱和建立，執行下列動作：

- a. 檢閱您針對先前步驟所做的選擇，並視需要編輯。
- b. 選擇其中一個選項。

如果您已選擇 ...	然後選擇 ...
使用協同合作建立成員資格 (是，立即建立成員資格來加入)	建立協作和成員資格
建立協同合作，而不是目前建立成員資格 (否，稍後我會建立成員資格)	建立協同合作

成功建立協同合作後，您可以在協同合作下看到協同合作詳細資訊頁面。

您現在已準備好：

- [準備要在 中分析的資料表 AWS Clean Rooms](#)。(如果您想要分析自己的事件資料，或想要查詢身分資料，則為選用。)
- [將設定的資料表與協同合作建立關聯](#)。(如果您想要分析自己的事件資料，則為選用。)
- [新增已設定資料表的分析規則](#)。(如果您想要分析自己的事件資料，則為選用。)
- [建立成員資格並加入協作](#)。(如果您已建立成員資格，則為選用。)
- [邀請成員加入協作](#)。

為查詢和任務建立協同合作

在此程序中，身為[協作建立者](#)的您執行下列任務：

- [建立協同合作](#)。
- 邀請一或多個[成員](#)加入[協同合作](#)。
- 指派能力給成員，例如[可執行查詢和任務的成員](#)，以及[可接收結果的成員](#)。

如果協作建立者也是可以接收結果的成員，他們會指定結果目的地和格式。他們也提供服務角色 Amazon Resource Name (ARN)，將結果寫入結果目的地。

- 設定哪些[成員負責支付協同合作中的查詢和任務運算成本](#)。

在開始前，請確定您具有以下先決條件：

- 您擁有要邀請到協同合作的每個成員的名稱和 AWS 帳戶 ID。
- 您有權與協同合作的所有成員共用每個成員的名稱和 AWS 帳戶 ID。

 Note

您無法在建立協同合作之後新增更多成員。

如需有關如何使用 AWS SDKs 建立協同合作的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

為查詢和任務建立協同合作

1. 登入 AWS 管理主控台，並使用 開啟[AWS Clean Rooms 主控台](#) AWS 帳戶，以做為協同合作建立者。
2. 在左側導覽窗格中，選擇協同合作。
3. 在右上角，選擇建立協同合作。
4. 針對步驟 1：定義協同合作，執行下列動作：
 - a. 如需詳細資訊，請輸入協同合作的名稱和描述。

受邀參與協作的協作成員可以看到此資訊。名稱和描述有助於他們了解協同合作的內容。

- b. 對於成員：

- i. 針對成員 1：輸入成員顯示名稱，如同您希望在協同合作時顯示的名稱。

 Note

您的 AWS 帳戶 ID 會自動包含在成員 AWS 帳戶 ID 中。

- ii. 針對成員 2，輸入您要邀請加入協同合作之成員的成員顯示名稱和成員 AWS 帳戶 ID。

受邀參與協同合作的每個人都可以看到成員顯示名稱和成員 AWS 帳戶 ID。輸入並儲存這些欄位的值之後，您就無法對其進行編輯。

 Note

您必須通知協同合作成員，協同合作中所有受邀和作用中的協同合作者都可看見其成員 AWS 帳戶 ID 和成員顯示名稱。

- iii. 如果您想要新增其他成員，請選擇新增其他成員。然後輸入每個成員的成員顯示名稱和成員 AWS 帳戶 ID，這些成員可以提供您要邀請加入協同合作的資料。
- c. 如果您想要啟用分析記錄，請選取啟用分析記錄核取方塊，然後選擇支援的日誌類型。
 - 如果您想要接收從 SQL 查詢產生的日誌，請選擇查詢的日誌核取方塊。
 - 如果您想要使用 PySpark 接收任務產生的日誌，請選擇任務的日誌核取方塊。
- d. 如果您想要啟用詳細監控，請選取啟用詳細監控核取方塊。

分析執行器和設定的付款人可以選擇在建立其成員資格時啟用詳細指標。啟用時，詳細的監控指標將發佈至 CloudWatch 以進行協同合作的操作監控，包括查詢效能和資源使用率。這些指標可供分析執行器和各自 AWS 帳戶中設定的付款人使用。

如需了解有關 CloudWatch 定價的詳細資訊，請參閱 [CloudWatch 定價](#)。

- e. 在允許的查詢結果區域下，選取您要傳送查詢結果的一或多個 AWS 區域 位置。

根據預設，只會選取目前的區域（例如維吉尼亞北部 us-east-1）。

 Important

當您啟用跨區域查詢結果交付時，您的結果可能會處理並存放在來源區域之外。

如需區域的詳細資訊，請參閱《》中的 [區域和端點](#) AWS 一般參考。

- f. （選用）設定哪些設定可自動變更，無需手動核准變更請求，即可透過自動變更請求核准來管理對資料的存取。根據預設，某些設定只能透過提交變更請求來變更，該請求必須經過所有成員的核准才能生效。
 - 授予成員能力 – 選擇可授予協作成員的功能，無需手動核准。成員一律可以貢獻資料。
 - 選擇功能：

- 貢獻資料（一律啟用）
- 接收結果
- 自動核准具有這些功能的新成員 – 如果允許，任何具有上述選取功能新增的成員都會立即加入協作。新增其他功能的成員仍需要手動核准才能加入。
- 可自動撤銷的功能 - 選擇無需手動核准即可撤銷的功能。成員一律可以貢獻資料。
- 選擇功能：
 - 貢獻資料（一律啟用）
 - 接收結果

如果您選擇此選項，您可以透過位於協同作業詳細資訊頁面詳細資訊索引標籤的變更請求歷史記錄來追蹤所有協同作業組態修改。

g. （選用）如果您想要啟用密碼編譯運算功能，請選取啟用密碼編譯運算核取方塊。

i. 選擇下列密碼編譯涵蓋範圍參數：

- 允許plaintext資料欄

如果您需要完全加密的資料表，請選擇否。

如果您想要加密資料表中允許cleartext的資料欄，請選擇是。

若要在特定資料欄AVG上執行 SUM或，資料欄必須位於 中cleartext。

- 保留NULL值

如果您不想保留NULL值，請選擇否。 NULL值不會在加密的資料表NULL中顯示為。

如果您想要保留NULL值，請選擇是。 NULL值會在加密的資料表NULL中顯示為。

ii. 選擇下列指紋參數：

- 允許重複項目

如果您不希望資料fingerprint欄中允許重複的項目，請選擇否。

如果您想要資料fingerprint欄中允許重複的項目，請選擇是。

- 允許具有不同名稱JOIN的資料欄

如果您不想加入具有不同名稱的資料fingerprint欄，請選擇否。

如果您想要使用不同的名稱聯結fingerprint資料欄，請選擇是。

如需密碼編譯運算參數的詳細資訊，請參閱 [密碼編譯運算參數](#)。

如需如何加密資料以用於 的詳細資訊 AWS Clean Rooms，請參閱 [使用的密碼編譯運算準備加密資料表 Clean Rooms](#)。

 Note

在完成下一個步驟之前，請仔細驗證這些組態。建立協同合作之後，您只能編輯協同合作名稱、描述，以及日誌是否存放在 Amazon CloudWatch Logs 中。

- h. 如果您想要為協同合作資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
 - i. 選擇下一步。
5. 針對步驟 2：指定成員能力，執行下列動作：
- a. 對於使用查詢和任務的分析，在支援的分析類型下，選擇任務核取方塊。

預設會選取查詢核取方塊。

- i. 從下拉式清單中選取可執行查詢和任務的成員。
- ii. 從下拉式清單中選取可從分析中接收結果的成員。

 Note

建立 PySpark 分析範本的成員也必須是收到結果的成員。

- b. 如果您使用 Clean Rooms ML，對於使用專用工作流程的 ML 建模，
 - i. （選用）從下拉式清單中選取可以從訓練模型接收輸出的成員。
 - ii. （選用）從下拉式清單中選取可以從模型推論接收輸出的成員。
 - c. 使用 檢視 ID 解析 AWS Entity Resolution 下的成員功能。
 - d. 選擇下一步。
6. 對於步驟 3：設定付款，
- a. 針對使用查詢和任務的分析，選擇將支付查詢和任務費用的成員。

您可以指派可執行查詢和任務的成員，成為支付查詢和任務運算成本的成員。

您可以指派不同的成員來支付查詢和任務運算成本。

- b. 對於使用專用工作流程的 ML 建模，已設定外觀模型的建立者是將支付外觀建模費用的成員。
 - c. 對於使用 ID 解析 AWS Entity Resolution，ID 映射表的建立者是將為 ID 映射表付費的成員。
 - d. 選擇下一步。
7. 針對步驟 4：設定成員資格，選擇下列其中一個選項：

Yes, join by creating membership now

1. 對於結果設定預設值，對於查詢結果設定，如果您是可以接收結果的成員，
 - a. 選擇設定查詢的預設設定核取方塊。針對 Amazon S3 中的結果目的地，輸入 Amazon S3 目的地，或選擇瀏覽 S3 以選取 S3 儲存貯體。
 - b. 針對查詢結果格式，選擇 CSV 或 PARQUET。
 - c. （僅限 Spark）對於結果檔案，選擇多個或單一。
 - d. （選用）對於服務存取，如果您想要將最多需要 24 小時的查詢交付至 S3 目的地，請選取新增服務角色以支援最多需要 24 小時才能完成的查詢核取方塊。

最多需要 24 小時才能完成的大型查詢將會傳送到您的 S3 目的地。

如果您未選取核取方塊，則只會將 12 小時內完成的查詢傳送到您的 S3 位置。

- e. 選取建立和使用新的服務角色或使用現有的服務角色，以指定服務存取許可。

如果您選擇 ...	然後 ...
建立和使用新的服務角色	• AWS Clean Rooms 會建立具有此資料表所需政策的服務角色。 • 預設的服務角色名稱為 <code>cleanrooms-result-receiver- <timestamp></code> • 您必須具有建立角色和連接政策的許可。
使用現有的服務角色	i. 從下拉式清單中選擇現有的服務角色名稱。

如果您選擇 ...	然後 ...
	如果您有列出角色的許可，則會顯示角色清單。 如果您沒有列出角色的許可，您可以輸入要使用的角色的 Amazon Resource Name (ARN)。 ii. 選擇在 IAM 外部連結中檢視，以檢視服務角色。 如果沒有現有的服務角色，則無法使用使用現有服務角色的選項。 根據預設，AWS Clean Rooms 不會嘗試更新現有的角色政策來新增必要的許可。

 Note

- AWS Clean Rooms 需要根據分析規則查詢的許可。如需 許可的詳細資訊 AWS Clean Rooms，請參閱 [AWS 的 受管政策 AWS Clean Rooms](#)。
- 如果角色沒有足夠的許可 AWS Clean Rooms，您會收到錯誤訊息，指出角色沒有足夠的許可 AWS Clean Rooms。必須先新增角色政策，才能繼續。
- 如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

2. 對於任務結果，

Example

例如：`s3://bucket/prefix`

- 選擇設定任務的預設設定核取方塊，然後輸入 S3 目的地，在 Amazon S3 中指定結果目的地，或選擇瀏覽 S3 從可用的 S3 儲存貯體清單中選擇。
- 從下拉式清單中選擇現有的服務角色名稱，以指定服務存取許可。

3. 針對日誌設定，為 Amazon CloudWatch Logs 中的日誌儲存選擇下列其中一個選項：

 Note

如果您選擇啟用查詢記錄，則會顯示日誌設定區段。

- a. 選擇開啟，與您相關的查詢日誌將存放在您的 Amazon CloudWatch Logs 帳戶中。

每個成員只能收到其啟動或包含其資料的查詢的日誌。

可以接收結果的成員也會收到在協同合作中執行之所有查詢的日誌，即使未在查詢中存取其資料。

在支援的日誌類型下，從協同合作建立者選擇支援的日誌類型中選擇：

在支援的日誌類型下，預設會開啟查詢日誌和任務日誌核取方塊。

 Note

開啟分析記錄後，可能需要幾分鐘的時間才能設定日誌儲存，並開始在 Amazon CloudWatch Logs 中接收日誌。在此短暫期間內，可以查詢的成員可能會執行未實際傳送日誌的查詢。

- b. 選擇關閉，與您相關的查詢日誌將不會存放在您的 Amazon CloudWatch Logs 帳戶中。
4. 如果您想要啟用成員資格資源的成員資格標籤，請選擇新增標籤，然後輸入金鑰和值對。
5. 如果您是支付查詢運算、任務運算或兩者的成員，請選取我同意支付此協同合作核取方塊中的運算成本，以表示您接受。

 Note

您必須選取此核取方塊才能繼續。

如需如何計算定價的詳細資訊，請參閱 [的定價 AWS Clean Rooms](#)。

如果您是[支付查詢運算成本的成員](#)，但不是[可以查詢的成員](#)，建議您使用 AWS Budgets 來設定預算，AWS Clean Rooms 並在達到預算上限時接收通知。如需設定預算的詳細資訊，請參閱 AWS Cost Management 《使用者指南》中的[使用 管理您的成本 AWS Budgets](#)。如需設定通知的詳細資訊，請參閱 [《使用者指南》中的為預算通知建立 Amazon SNS 主題](#)。

AWS Cost Management 如果已達到預算上限，您可以聯絡可執行查詢或[離開協同合作](#)的成員。如果您離開協同合作，將不再允許執行任何查詢，因此您不再需要支付查詢運算成本。

6. 選擇下一步。

協同合作和您的成員資格都會建立。

協同合作中的狀態為作用中。

No, I will create a membership later

1. 選擇下一步。

只會建立協同合作。

您在協同合作中的狀態為非作用中。

8. 針對步驟 5：檢閱和建立，執行下列動作：

- a. 檢閱您針對先前步驟所做的選擇，並視需要編輯。
- b. 選擇其中一個選項。

如果您已選擇 ...	然後選擇 ...
使用協同合作建立成員資格 (是，立即建立成員資格來加入)	建立協作和成員資格
建立協同合作，而不是目前建立成員資格 (否，我稍後會建立成員資格)	建立協同合作

成功建立協同合作之後，您可以在協同合作下看到協同合作詳細資訊頁面。

您現在已準備好：

- [準備要在 中分析的資料表 AWS Clean Rooms](#)。(如果您想要分析自己的事件資料，或想要查詢身分資料，則為選用。)
- [將設定的資料表與協同合作建立關聯](#)。(如果您想要分析自己的事件資料，則為選用。)
- [新增已設定資料表的分析規則](#)。(如果您想要分析自己的事件資料，則為選用。)
- [建立成員資格並加入協作](#)。(如果您已建立成員資格，則為選用。)
- [邀請成員加入協作](#)。

為 ML 建模建立協作

在此程序中，身為[協作建立者](#)的您執行下列任務：

- [建立協同合作](#)。
- 邀請一或多個[成員](#)參與[協作](#)。
- 指派能力給成員，例如
 - [可查詢的成員](#)
 - [可以接收結果的成員](#)
 - 可以從訓練模型接收輸出的成員
 - 可以從模型推論輸出的成員

如果協作建立者也是可以接收結果的成員，他們會指定結果目的地和格式。他們也提供服務角色 Amazon Resource Name (ARN)，將結果寫入結果目的地。

- 設定哪些[成員負責支付協作中的運算成本、模型訓練和模型推論成本](#)。

在開始前，請確定您具有以下先決條件：

- 您擁有要邀請到協同合作的每個成員的名稱和 AWS 帳戶 ID。
- 您有權與協同合作的所有成員共用每個成員的名稱和 AWS 帳戶 ID。

Note

您無法在建立協同合作之後新增更多成員。

如需有關如何使用 AWS SDKs 建立協同合作的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

建立 ML 建模的協同合作

1. 登入 AWS 管理主控台，並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶，以做為協同合作建立者。
2. 在左側導覽窗格中，選擇協同合作。
3. 在右上角，選擇建立協同合作。
4. 針對步驟 1：定義協同合作，執行下列動作：

- a. 如需詳細資訊，請輸入協同合作的名稱和描述。

受邀參與協作的協作成員可以看到此資訊。名稱和描述有助於他們了解協同合作的內容。

- b. 對於成員：

- i. 針對成員 1：輸入成員顯示名稱，如同您希望在協同合作時顯示的名稱。

 Note

您的 AWS 帳戶 ID 會自動包含在成員 AWS 帳戶 ID 中。

- ii. 對於成員 2，輸入您要邀請加入協同合作的成員的成員顯示名稱和成員 AWS 帳戶 ID。

受邀參與協作的每個人都可以看到成員顯示名稱和成員 AWS 帳戶 ID。輸入並儲存這些欄位的值之後，您就無法對其進行編輯。

 Note

您必須通知協同合作成員，協同合作中所有受邀和作用中的協同合作者都可看見其成員 AWS 帳戶 ID 和成員顯示名稱。

- iii. 如果您想要新增其他成員，請選擇新增其他成員。然後輸入每個成員的成員顯示名稱和成員 AWS 帳戶 ID，這些成員可以提供您要邀請加入協同合作的資料。
- c. 如果您想要啟用分析記錄，請選取啟用分析記錄核取方塊，然後在支援的日誌類型下，從查詢中選擇日誌。
- d. 如果您想要啟用詳細監控，請選取啟用詳細監控核取方塊。

分析執行器和設定的付款人可以選擇在建立其成員資格時啟用詳細指標。啟用時，詳細的監控指標將發佈至 CloudWatch 以進行協同合作的操作監控，包括查詢效能和資源使用率。這些指標可供分析執行器和各自 AWS 帳戶中設定的付款人使用。

如需了解有關 CloudWatch 定價的詳細資訊，請參閱 [CloudWatch 定價](#)。

- e. 在允許的查詢結果區域下，選取您要傳送查詢結果的一或多個 AWS 區域 位置。

根據預設，只會選取目前的區域（例如維吉尼亞北部 us-east-1）。

⚠ Important

當您啟用跨區域查詢結果交付時，您的結果可能會處理並存放在來源區域之外。

如需區域的詳細資訊，請參閱《》中的[區域和端點](#) AWS 一般參考。

- f. (選用) 透過設定哪些設定可自動變更，而無需手動核准變更請求，以自動變更請求核准來管理對資料的存取。根據預設，某些設定只能透過提交變更請求來變更，該請求必須經過所有成員的核准才能生效。
 - 授予成員能力 – 選擇可授予協作成員的功能，無需手動核准。成員一律可以貢獻資料。
 - 選擇功能：
 - 貢獻資料（一律啟用）
 - 接收結果
 - 自動核准具有這些功能的新成員 – 如果允許，任何具有上述選取功能新增的成員都會立即加入協作。新增其他功能的成員仍需要手動核准才能加入。
 - 可自動撤銷的功能 – 選擇無需手動核准即可撤銷的功能。成員一律可以貢獻資料。
 - 選擇功能：
 - 貢獻資料（一律啟用）
 - 接收結果

如果您選擇此選項，您可以透過位於協作詳細資訊頁面詳細資訊索引標籤的變更請求歷史記錄來追蹤所有協作組態修改。

- g. (選用) 如果您想要啟用密碼編譯運算功能，請選取啟用密碼編譯運算核取方塊。
 - i. 選擇下列密碼編譯涵蓋範圍參數：
 - 允許plaintext資料欄

如果您需要完全加密的資料表，請選擇否。

如果您想要加密資料表中允許cleartext的資料欄，請選擇是。

若要在特定資料欄AVG上執行 SUM或，資料欄必須位於 中cleartext。
 - 保留NULL值

如果您不想保留NULL值，請選擇否。 NULL值不會在加密資料表NULL中顯示為。

如果您想要保留NULL值，請選擇是。 NULL值會在加密的資料表NULL中顯示為。

ii. 選擇下列指紋參數：

- 允許重複項目

如果您不希望資料fingerprint欄中允許重複的項目，請選擇否。

如果您想要在資料fingerprint欄中允許重複的項目，請選擇是。

- 允許具有不同名稱JOIN的資料欄

如果您不想加入具有不同名稱的資料fingerprint欄，請選擇否。

如果您想要使用不同的名稱聯結fingerprint資料欄，請選擇是。

如需密碼編譯運算參數的詳細資訊，請參閱 [密碼編譯運算參數](#)。

如需如何加密資料以用於 的詳細資訊 AWS Clean Rooms，請參閱 [使用的密碼編譯運算準備加密資料表 Clean Rooms](#)。

 Note

在完成下一個步驟之前，請仔細驗證這些組態。建立協同合作之後，您只能編輯協同合作名稱、描述，以及日誌是否存放在 Amazon CloudWatch Logs 中。

h. 如果您想要為協同合作資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。

i. 選擇下一步。

5. 對於步驟 2：指定成員能力，

a. 對於使用查詢和任務的分析，在支援的分析類型下，保持選取查詢核取方塊。

b. 針對執行查詢，選擇將啟動模型訓練的成員

c. 針對從分析接收結果，選擇一或多個將接收查詢結果的成員。

d. 對於使用專用工作流程的 ML 建模，

i. 針對從訓練模型接收輸出，選擇將接收訓練模型結果的成員，包括模型成品和指標。

ii. 針對從模型推論接收輸出，選擇將接收模型推論結果的成員。

- e. 使用 檢視 ID 解析 AWS Entity Resolution 下的成員功能。
6. 對於步驟 3：設定付款，
 - a. 在使用查詢的分析下，針對查詢的付款，執行下列其中一個動作：
 - 若要讓相同的成員支付並執行查詢，請選取您為執行查詢選擇的相同成員。
 - 若要讓不同的成員支付查詢費用，請選取您的成員帳戶。
 - b. 對於使用專用工作流程的 ML 建模，
 - 選擇將為模型訓練付費的成員。
 - c. 選擇將支付推論任務費用的成員。
 - d. 對於支付外觀建模，不需要任何動作。已設定外觀模型的建立者是將支付外觀模型費用的成員。
 - e. （選用）選擇將為產生合成資料付費的成員。
 - f. 對於使用的 ID 解析 AWS Entity Resolution，不需要任何動作。ID 映射表的建立者是將為 ID 映射表付費的成員。
7. 選擇下一步。
8. 針對步驟 4：設定成員資格，在協同合作成員資格下，選擇下列其中一個選項：

Yes, join by creating membership now

1. 對於結果設定預設值，對於查詢結果設定，如果您是可以接收結果的成員，
 - a. 選取設定查詢的預設設定核取方塊。
 - b. 針對 Amazon S3 中的結果目的地，輸入 Amazon S3 目的地，或選擇瀏覽 S3 以選取 S3 儲存貯體。
 - c. 針對查詢結果格式，選擇 CSV 或 PARQUET。
 - d. （僅限 Spark）對於結果檔案，選擇多個或單一。
 - e. （選用）如果您想要將最多需要 24 小時的查詢交付至 S3 目的地，請選取新增服務角色以支援最多需要 24 小時才能完成的查詢核取方塊。

最多需要 24 小時才能完成的大型查詢將會傳送到您的 S3 目的地。

如果您未選取核取方塊，則只會將 12 小時內完成的查詢傳送到您的 S3 位置。

- f. 選取建立和使用新的服務角色或使用現有的服務角色，以指定服務存取許可。

如果您選擇 ...	然後 ...
建立和使用新的服務角色	• AWS Clean Rooms 會建立具有此資料表所需政策的服務角色。 • 預設的服務角色名稱為 <code>cleanrooms-result-receiver-<timestamp></code> • 您必須具有建立角色和連接政策的許可。
使用現有的服務角色	- 從下拉式清單中選擇現有的服務角色名稱。 如果您有列出角色的許可，則會顯示角色清單。 如果您沒有列出角色的許可，您可以輸入要使用的角色的 Amazon Resource Name (ARN)。 - 選擇 IAM 外部連結中的檢視，以檢視服務角色。 如果沒有現有的服務角色，則無法使用使用現有服務角色的選項。 根據預設，AWS Clean Rooms 不會嘗試更新現有的角色政策來新增必要的許可。

 Note

- AWS Clean Rooms 需要根據分析規則查詢的許可。如需 許可的詳細資訊 AWS Clean Rooms，請參閱 [AWS 的 受管政策 AWS Clean Rooms](#)。
- 如果角色沒有足夠的許可 AWS Clean Rooms，您會收到錯誤訊息，指出角色沒有足夠的許可 AWS Clean Rooms。必須先新增角色政策，才能繼續。

- 如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

2. 對於 ML 組態，
 - a. 選擇建立 ML 組態核取方塊，然後輸入 S3 目的地，在 Amazon S3 上指定模型輸出目的地，或選擇瀏覽 S3 從可用的 S3 儲存貯體清單中選擇。
 - b. 選擇建立和使用新的服務角色或使用現有的服務角色，以指定服務存取許可。
 - c. 如果 S3 儲存貯體已加密，請選取目的地儲存貯體已使用 KMS 金鑰加密核取方塊，然後輸入 AWS KMS key 或選取建立 AWS KMS key 以建立新的 KMS 金鑰。
3. 如果您想要啟用成員資格資源的成員資格標籤，請選擇新增標籤，然後輸入金鑰和值對。
4. 如果您是支付查詢運算費用的成員，請選取我同意支付此協同合作核取方塊中的運算成本，以表示您接受。

 Note

您必須選取此核取方塊才能繼續。

如需如何計算定價的詳細資訊，請參閱 [的定價 AWS Clean Rooms](#)。

如果您是[支付查詢運算成本的成員](#)，但不是[可以查詢的成員](#)，建議您使用 AWS Budgets 來設定預算，AWS Clean Rooms 並在達到預算上限時接收通知。如需設定預算的詳細資訊，請參閱 AWS Cost Management 《使用者指南》中的[使用 管理您的成本 AWS Budgets](#)。如需設定通知的詳細資訊，請參閱《使用者指南》中的[為預算通知建立 Amazon SNS 主題](#)。AWS Cost Management 如果已達到預算上限，您可以聯絡可執行查詢或[離開協同合作](#)的成員。如果您離開協同合作，將不再允許執行任何查詢，因此您不再需要支付查詢運算成本。

5. 選擇下一步。

協同合作和您的成員資格都會建立。

協同合作中的狀態為作用中。

No, I will create a membership later

1. 選擇下一步。

只會建立協同合作。

您在協同合作中的狀態為非作用中。

9. 針對步驟 5：檢閱和建立，執行下列動作：

- a. 檢閱您針對先前步驟所做的選擇，並視需要編輯。
- b. 選擇其中一個選項。

如果您已選擇 ...	然後選擇 ...
使用協同合作建立成員資格 (是，立即建立成員資格來加入)	建立協作和成員資格
建立協同合作，而不是目前建立成員資格 (否，我稍後會建立成員資格)	建立協同合作

建立成員資格並加入協同合作

成員資格是在成員加入協同合作時建立的資源 AWS Clean Rooms。

您可以將協同合作加入為

- [可查詢的成員](#)
- [可執行查詢和任務的成員](#)
- [可接收查詢或任務結果的成員](#)
- [成員支付查詢運算成本](#)
- [支付查詢和任務費用的成員](#)

所有成員都可以貢獻資料。

如需有關如何使用 SDKs 建立成員資格和加入協同合作 AWS 的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

在此程序中，受邀成員會透過[建立成員資格資源來加入協同合作](#)。

如果受邀成員是可以接收結果的成員，他們會指定結果目的地和格式。他們也提供服務角色 ARN 來寫入結果目的地。

如果受邀成員是負責支付運算成本的成員，他們在加入協同合作之前會接受其付款責任。

建立成員資格並加入協同合作

1. 登入 AWS 管理主控台，並使用您的成員 開啟[AWS Clean Rooms 主控台](#) AWS 帳戶。
2. 在左側導覽窗格中，選擇協同合作。
3. 在可加入索引標籤上，針對可加入的協同合作，選擇協同合作的名稱。
4. 在協同合作詳細資訊頁面的概觀區段中，檢視協同合作詳細資訊，包括您的成員詳細資訊和其他成員的清單。

確認每個協同合作成員的 AWS 帳戶 IDs 是您打算加入協同合作的 ID。

5. 選擇建立成員資格。
6. 在建立成員資格頁面上的概觀中，檢視協作名稱、協作描述、協作建立者的 AWS 帳戶 ID、您的成員詳細資訊，以及將支付查詢費用的成員 AWS 帳戶 ID。
7. 如果協同合作建立者已選擇啟用分析記錄，請在 Amazon CloudWatch Logs 中為日誌儲存選擇下列其中一個選項：

如果選擇...	然後 ...
開啟	與您相關的日誌會存放在 Amazon CloudWatch Logs 中。 每個成員只能收到其啟動或包含其資料的查詢的日誌。 可以接收結果的成員也會收到在協同合作中執行之所有分析的日誌，即使未在分析中存取其資料。 在支援的日誌類型下，從協同合作建立者選擇支援的日誌類型中選擇： 1. 如果您想要接收從 SQL 查詢產生的日誌，請選擇查詢的日誌核取方塊。 2. 如果您想要使用 PySpark 接收任務產生的日誌，請選擇任務的日誌核取方塊。
關閉	與您相關的查詢日誌不會存放在 Amazon CloudWatch Logs 帳戶中。

Note

開啟分析記錄後，可能需要幾分鐘的時間才能設定日誌儲存，並開始在 Amazon CloudWatch Logs 中接收日誌。在此短暫期間內，可以查詢的成員可能會執行未實際傳送日誌的查詢。

8. 如果協同合作建立者已啟用此協同合作的詳細監控，請選擇是否要在 CloudWatch 帳戶中接收詳細的可觀測性指標。

如需詳細監控：

詳細監控選項

選項	Description
開啟	AWS Clean Rooms 會在您的帳戶中發佈此協同合作的詳細監控指標至 CloudWatch。您可以使用這些指標進行操作監控，包括查詢效能和資源使用率。 需支付額外的 CloudWatch 費用。如需詳細資訊，請參閱 CloudWatch 定價。
關閉	不會將詳細指標匯出到您的 CloudWatch 帳戶。如果其他成員已啟用此選項，仍然可以在自己的帳戶中檢視詳細的監控指標。

Note

詳細的監控指標僅適用於可執行查詢（分析執行器）的成員，以及設定為協同合作付款人的成員。

9. 如果您的成員能力包含接收結果，則結果設定預設為：
 - a. 針對查詢結果，選擇設定查詢的預設設定核取方塊，然後在 Amazon S3 中輸入 S3 目的地，或在可用的 S3 儲存貯體清單中選擇瀏覽 S3 以選取結果目的地。S3

Example

例如：**s3://bucket/prefix**

- i. 針對結果格式，選擇 CSV 或 PARQUET。
- ii. (僅限 Spark) 對於結果檔案，選擇多個或單一。
- iii. (選用) 對於服務存取，如果您想要將最多需要 24 小時的查詢交付到 S3 目的地，請選取新增服務角色以支援最多需要 24 小時才能完成的查詢核取方塊。

最多需要 24 小時才能完成的大型查詢將會傳送到您的 S3 目的地。

如果您未選取核取方塊，則只會將 12 小時內完成的查詢傳送到您的 S3 位置。

 Note

您必須選取現有的服務角色，或具有建立新角色的許可。如需詳細資訊，請參閱[建立服務角色以接收結果](#)。

- iv. 選取建立和使用新的服務角色或使用現有的服務角色，以指定服務存取許可。

Create and use a new service role

- AWS Clean Rooms 會建立具有此資料表所需政策的服務角色。
- 預設的服務角色名稱為 `cleanrooms-result-receiver-<timestamp>`
- 您必須具有建立角色和連接政策的許可。

Use an existing service role

1. 從下拉式清單中選擇現有的服務角色名稱。

如果您具有列出角色的許可，則會顯示角色清單。

如果您沒有列出角色的許可，您可以輸入要使用的角色的 Amazon Resource Name (ARN)。

2. 選擇在 IAM 外部連結中檢視，以檢視服務角色。

如果沒有現有的服務角色，則無法使用使用現有服務角色的選項。

根據預設，AWS Clean Rooms 不會嘗試更新現有的角色政策來新增必要的許可。

Note

- AWS Clean Rooms 需要根據分析規則查詢的許可。如需 許可的詳細資訊 AWS Clean Rooms，請參閱 [AWS 的 受管政策 AWS Clean Rooms](#)。
- 如果角色沒有足夠的許可 AWS Clean Rooms，您會收到錯誤訊息，指出角色沒有足夠的許可 AWS Clean Rooms。必須先新增角色政策，才能繼續。
- 如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

- b. 針對任務結果，選擇設定任務的預設設定核取方塊，然後在 Amazon S3 中輸入 S3 目的地，或在可用的 S3 儲存貯體清單中選擇瀏覽 S3 以選取結果目的地。S3

Example

例如：**s3://bucket/prefix**

- 從下拉式清單中選擇現有的服務角色名稱，以指定服務存取許可。

10. 如果您想要為成員資格資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
11. 如果協同合作建立者已將您指定為將支付查詢費用或支付查詢和任務費用的成員，請選取我同意為此協同合作核取方塊中的運算成本來表示接受。

Note

您必須選取此核取方塊才能繼續。

如需如何計算定價的詳細資訊，請參閱 [的定價 AWS Clean Rooms](#)。

如果您是[支付查詢運算成本的成員](#)，或[支付查詢和任務運算成本的成員](#)，但不是[可以查詢的成員](#)，建議您使用 AWS Budgets 來設定預算，AWS Clean Rooms 並在達到預算上限時接收通知。如需設定預算的詳細資訊，請參閱 AWS Cost Management 《使用者指南》中的[使用 管理您的成本 AWS Budgets](#)。如需設定通知的詳細資訊，請參閱 [《使用者指南》中的為預算通知建立 Amazon SNS 主題](#)。AWS Cost Management 如果已達到預算上限，您可以聯絡可執行查詢和任務或[離開協同合作](#)的成員。如果您離開協同合作，將不再允許執行任何查詢，因此您將不再支付查詢運算成本。

12. 如果您確定要建立成員資格並加入協同合作，請選擇建立成員資格。

您獲得協同合作中繼資料的讀取存取權。除了其他成員的所有名稱和 AWS 帳戶 IDs 之外，還包括協同合作的顯示名稱和描述等資訊。

您現在已準備好：

- [準備要在 中查詢的資料表 AWS Clean Rooms](#)。（如果您想要查詢自己的事件資料，或想要查詢身分資料，則為選用。）
- 如果您想要查詢事件資料，[請將設定的資料表與協同合作建立關聯](#)。
- 若要查詢事件資料，[請新增已設定資料表的分析規則](#)。
- [建立新的 ID 命名空間並建立關聯](#) – 如果您想要建立 ID 映射表來查詢身分資料。

如需如何離開協同合作的詳細資訊，請參閱 [離開協同合作](#)。

編輯協同合作

身為協作建立者，您可以編輯協作的不同部分。

如需有關如何使用 AWS SDKs 編輯協同合作的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

主題

- [編輯協同合作名稱和描述](#)
- [更新協同合作分析引擎](#)
- [關閉日誌儲存](#)
- [編輯協同合作日誌設定](#)
- [編輯協同合作標籤](#)
- [編輯成員資格標籤](#)
- [新增成員](#)
- [編輯現有的成員功能](#)
- [編輯協同合作自動核准設定](#)
- [編輯相關聯的資料表標籤](#)
- [編輯分析範本標籤](#)
- [編輯差異隱私權政策標籤](#)

編輯協同合作名稱和描述

建立協同合作之後，您只能編輯協同合作名稱和描述。

編輯協同合作名稱和描述

1. 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#)。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您建立的協同合作。
4. 在協同合作詳細資訊頁面上，選擇動作，然後選擇編輯協同合作。
5. 在編輯協同合作頁面上，如需詳細資訊，請編輯協同合作的名稱和描述。
6. 選擇儲存變更。

更新協同合作分析引擎

建立協同合作之後，您可以將分析引擎從 AWS Clean Rooms SQL 變更為 Spark。

Note

將分析引擎從 AWS Clean Rooms SQL 變更為 Spark 可能會中斷現有的工作流程。

更新協同合作分析引擎

1. 登入 AWS 管理主控台 並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您建立的協同合作。
4. 在協同合作詳細資訊頁面上，選擇動作，然後選擇編輯協同合作。
5. 選擇儲存變更。

關閉日誌儲存

如果您已啟用分析記錄，則可以編輯分析日誌是否存放在 Amazon CloudWatch Logs 帳戶中。

關閉日誌儲存

1. 登入 AWS 管理主控台 並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇已開啟分析記錄的協同合作。
4. 在協同合作詳細資訊頁面上，選擇動作，然後選擇關閉日誌儲存。

Note

出現警告，指出下列事項：

- 新的查詢將不再記錄到您的 CloudWatch 帳戶。
- 現有的日誌將根據您目前的保留設定保留。
- 如果您未來重新啟用記錄，它將僅適用於重新啟用後提出的查詢。
- 此變更只會影響您的日誌 - 其他團隊成員的日誌設定保持不變。

5. 選擇關閉。

編輯協同合作日誌設定

如果您已啟用查詢記錄，則可以編輯查詢日誌是否存放在您的 Amazon CloudWatch Logs 帳戶中。

編輯協同合作日誌設定

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您建立的協同合作。
4. 在協同合作詳細資訊頁面上，執行下列其中一項操作：
 - 選擇動作，然後選擇編輯日誌設定。
 - 在日誌索引標籤上，選擇編輯日誌設定。
5. 在編輯日誌設定模式上，Amazon Amazon CloudWatch Logs 中的日誌儲存：
 - 如果您不希望與您相關的日誌存放在 Amazon CloudWatch Logs 帳戶中，請選擇關閉。

- 如果您想要將與您相關的日誌存放在 Amazon CloudWatch Logs 帳戶中，請選擇開啟。

您只能接收您啟動或包含資料的查詢日誌。

可以接收結果的成員也會收到在協同合作中執行之所有查詢的日誌，即使查詢中未存取其資料。

1. 在支援的日誌類型下，從協同合作建立者選擇支援的日誌類型中選擇：

- 如果您想要接收從 SQL 查詢產生的日誌，請選擇查詢的日誌核取方塊。
- 如果您想要使用 PySpark 接收任務產生的日誌，請選擇任務的日誌核取方塊。

6. 選擇儲存變更。

Note

開啟記錄後，可能需要幾分鐘的時間才能設定日誌儲存，並開始在 Amazon CloudWatch Logs 中接收日誌。在此短暫期間內，可以查詢的成員可能會執行未實際傳送日誌的查詢。

編輯協同合作標籤

身為協同合作建立者，在建立協同合作之後，您可以管理協同合作資源上的標籤。

編輯協同合作標籤

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您建立的協同合作。
4. 選擇下列其中一項：

如果您是...	然後 ...
協同合作建立者和協同合作的成員	選擇詳細資訊索引標籤。
協作建立者，但不是協作的成員	向下捲動頁面至標籤區段。

5. 如需協同合作詳細資訊，請選擇管理標籤。
6. 在 Manage tags (管理標籤) 頁面上，可以執行下列操作：

- 若要移除標籤，請選擇 Remove (移除)。
- 若要新增標籤，請選擇 Add new tag (新增新標籤)。
- 若要儲存變更，請選擇儲存變更

編輯成員資格標籤

身為協同合作建立者，在建立協同合作之後，您可以管理成員資格資源上的標籤。

編輯成員資格標籤

1. 登入 AWS 管理主控台 並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您建立的協同合作。
4. 選擇詳細資訊索引標籤。
5. 如需成員資格詳細資訊，請選擇管理標籤。
6. 在管理成員資格標籤頁面上，您可以執行下列動作：
 - 若要移除標籤，請選擇 Remove (移除)。
 - 若要新增標籤，請選擇 Add new tag (新增新標籤)。
 - 若要儲存您所做的變更，請選擇 Save changes (儲存變更)。

新增成員

如需詳細資訊，請參閱[將新成員新增至協同合作](#)。

編輯現有的成員功能

如需詳細資訊，請參閱[更新現有的成員能力](#)。

編輯協同合作自動核准設定

如需詳細資訊，請參閱[編輯協同合作自動核准設定](#)。

編輯相關聯的資料表標籤

身為協同合作建立者，在將資料表與協同合作建立關聯之後，您可以管理相關聯資料表資源上的標籤。

編輯相關聯的資料表標籤

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您建立的協同合作。
4. 選擇 Tables (資料表) 索引標籤。
5. 對於與您相關聯的資料表，選擇資料表。
6. 在設定的資料表詳細資訊頁面上，針對標籤選擇管理標籤。

在 Manage tags (管理標籤) 頁面上，可以執行下列操作：

- 若要移除標籤，請選擇 Remove (移除)。
- 若要新增標籤，請選擇 Add new tag (新增新標籤)。
- 若要儲存您所做的變更，請選擇 Save changes (儲存變更)。

編輯分析範本標籤

身為協同合作建立者，在建立協同合作之後，您可以管理分析範本資源上的標籤。

編輯成員資格標籤

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您建立的協同合作。
4. 選擇範本索引標籤。
5. 在由您建立的分析範本區段中，選擇分析範本。
6. 在分析範本資料表詳細資訊頁面上，向下捲動至標籤區段。
7. 選擇管理標籤。
8. 在 Manage tags (管理標籤) 頁面上，可以執行下列操作：

- 若要移除標籤，請選擇 Remove (移除)。
- 若要新增標籤，請選擇 Add new tag (新增新標籤)。
- 若要儲存您所做的變更，請選擇 Save changes (儲存變更)。

編輯差異隱私權政策標籤

身為協同合作建立者，在建立協同合作之後，您可以管理分析範本資源上的標籤。

編輯成員資格標籤

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇包含您要編輯之差異隱私權政策的協同合作。
4. 選擇 Tables (資料表) 索引標籤。
5. 在資料表索引標籤上，選擇管理標籤。
6. 在 Manage tags (管理標籤) 頁面上，可以執行下列操作：
 - 若要移除標籤，請選擇 Remove (移除)。
 - 若要新增標籤，請選擇 Add new tag (新增新標籤)。
 - 若要儲存您所做的變更，請選擇 Save changes (儲存變更)。

在 中變更請求AWS Clean Rooms

變更請求可讓您提議變更現有的協同合作設定，以供其他協同合作成員核准。透過變更請求，您可以提交請求來新增新成員、更新現有成員功能，以及修改協同合作自動核准設定。所有協同合作成員都必須核准變更請求，提議的變更才會生效。

變更請求特定於協同合作，可由協同合作建立者提交。

您可以透過下列方式提交變更請求：

- 將新成員新增至協同合作
- 更新現有的成員能力
- 編輯協同合作自動核准設定

Note

您必須是協作建立者，才能提交變更請求。

將新成員新增至協同合作

若要將新成員新增至協同合作，您必須是協同合作建立者。將新成員新增至協同合作需要現有協同合作成員的手動核准，並提交變更請求。

您可以依照下列步驟，將新成員新增至協同合作：

將新成員新增至協同合作

1. 登入 AWS 管理主控台並與您的成員開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶。
2. 在左側導覽窗格中，選擇協同合作。
3. 選取協同合作以導覽至協同合作詳細資訊。
4. 在協同合作頁面上，導覽至成員索引標籤。
5. 在成員表格中，選取編輯成員。
6. 選擇新增其他成員。
7. 輸入新的成員資訊：
 - 成員顯示名稱
 - 成員 AWS 帳戶ID
 - 指定成員是否可以接收結果。勾選方塊以授予此成員能力。
8. 選擇儲存變更。
9. 確認您的變更請求提交。在確認模式中，確認變更，然後選取提交變更請求。

Note

如果您的協同合作支援自動核准的變更類型，則可能不需要手動核准變更請求。您可以在協同合作的「概觀」區段中檢閱哪些變更類型不需要變更請求。如需詳細資訊，請參閱 [the section called “編輯協同合作自動核准設定”](#)。

更新現有的成員能力

若要更新現有的協同合作成員能力，您必須是協同合作建立者。現有協同合作成員的更新功能需要現有成員的手動核准，並提交變更請求。

可更新的成員功能包括：

- 可以接收結果

您可以依照下列步驟更新現有協同合作成員的成員能力：

更新現有的成員能力

1. 登入 AWS 管理主控台，並與您的成員開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶。
2. 在左側導覽窗格中，選擇協同合作。
3. 選取協同合作以導覽至協同合作詳細資訊。
4. 在協同合作頁面上，導覽至成員索引標籤。
5. 在成員表格中，選取編輯成員。
6. 指定要變更的成員功能。
7. 選擇儲存變更。
8. 確認您的變更請求提交。在確認模式中，確認變更，然後選取提交變更請求。

Note

如果您的協同合作支援自動核准的變更類型，則可能不需要手動核准變更請求。您可以在協同合作的「概觀」區段中檢閱哪些變更類型不需要變更請求。如需詳細資訊，請參閱[the section called “編輯協同合作自動核准設定”](#)。

編輯協同合作自動核准設定

若要編輯自動核准的協同合作設定，您必須是協同合作建立者，並提交變更請求以供其他協同合作成員核准。

您可以依照下列步驟編輯協同合作的自動核准設定：

編輯協同合作自動核准設定

1. 登入 AWS 管理主控台，並與您的成員開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶。
2. 在左側導覽窗格中，選擇協同合作。
3. 選取協同合作以導覽至協同合作詳細資訊。
4. 在協同合作頁面上，選擇動作按鈕，然後選取編輯自動核准。
5. 若要將成員能力授予現有成員，而不手動核准變更請求：
 1. 導覽至授予成員能力區段。
 2. 指定可自動授予哪些成員能力。

Note

根據預設，所有協同合作成員都可以「貢獻資料」。

3. (選用) 若要允許新成員立即加入具有指定能力的協同合作，而無需手動核准變更請求，請選取自動核准具有這些能力的新成員。
6. 若要允許在沒有變更請求的情況下從現有成員移除功能：
 1. 導覽至可自動撤銷的功能區段。
 2. 指定可以自動移除哪些成員功能。
7. 選擇儲存變更。
8. 確認您的變更請求提交。在確認模式中，確認變更，然後選取提交變更請求。

刪除協同合作

身為協同合作建立者，您可以刪除您建立的協同合作。

Note

當您刪除協同合作時，您和所有成員都無法執行查詢、接收結果或貢獻資料。作為成員資格的一部分，每個協作成員都可以繼續存取自己的資料。

刪除協同合作

1. 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#)。

2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您要刪除的協同合作。
4. 在動作下，選擇刪除協同合作。
5. 確認刪除，然後選擇刪除。

檢視協同合作

身為協作建立者，您可以檢視您建立的所有協作。

檢視協同合作

1. 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#)。
2. 在左側導覽窗格中，選擇協同合作。
3. 在協同合作頁面的上次使用下，檢視最後使用的 5 個協同合作。
4. 在具有作用中成員資格索引標籤上，檢視具有作用中成員資格的協同合作清單。

您可以依名稱、成員資格建立日期和成員詳細資訊排序。

您可以使用搜尋列來搜尋協同合作。

5. 在可加入索引標籤上，檢視可加入的協作清單。
6. 在不再可用索引標籤上，檢視已刪除的協同合作和不再可用的協同合作成員資格清單（已移除的成員資格）。

邀請成員參與協作

身為協同合作建立者，在建立協同合作之後，您可以傳送邀請連結給成員索引標籤上列出的成員。

邀請成員加入協同合作

1. 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#)。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您建立的協同合作。
4. 選擇成員索引標籤。
5. 在成員表格中，選擇複製邀請連結按鈕。

邀請連結已複製。

- 將邀請連結貼到您選擇的安全通訊方法中，並將其傳送給協同合作的每個成員。

監控成員

身為協同合作建立者，在建立協同合作之後，您可以在成員索引標籤上監控所有成員的狀態。這有助於確保在整個協同合作中採取適當的存取控制和成本管理。

監控成員的狀態

- 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#)。
- 在左側導覽窗格中，選擇協同合作。
- 選擇您建立的協同合作。
- 選擇成員索引標籤。
- 在成員表格中，檢閱每個成員的狀態。
- 在成員能力表格中，檢閱哪些成員可以查詢、接收結果、貢獻資料，以及執行其他任務。
- 在付款組態表格中，檢閱哪些成員要為查詢、ID 映射表和 ML 建模付費。

將成員新增至協同合作

先決條件

- AWS 帳戶具有管理協同合作許可的
- 您要新增的成員 AWS 帳戶IDs

將成員新增至協同合作

- 登入 AWS 管理主控台並開啟 [AWS Clean Rooms主控台](#)。
- 在左側導覽窗格中，選擇協同合作。
- 選取您要新增成員的協同合作。
- 選擇成員索引標籤。
- 選擇 Edit members (編輯成員)。
- 選擇新增其他成員，然後輸入下列資訊：
 - 成員顯示名稱

- 成員 AWS 帳戶 ID
 - 指定他們是否可以接收結果
7. 選擇儲存變更。
 8. 在確認模式中，驗證資訊是否正確，然後選擇提交變更請求。
 - 如果變更請求需要其他成員的核准，則所有現有成員都必須在新增新成員之前核准變更請求。如需變更請求的詳細資訊，請參閱 [the section called “變更請求”](#)。
 - 如果具有指定功能的新成員支援自動核准變更類型，則此變更將立即生效。如需詳細資訊，請參閱 [the section called “編輯協同合作自動核准設定”](#)。
 9. 在協同合作詳細資訊頁面的成員索引標籤下，確認新增成員的成員狀態顯示為已邀請。

完成這些步驟後，受邀的成員可以加入協作。如需加入協同合作的詳細資訊，請參閱 [建立成員資格並加入協同合作](#)

從協同合作中移除成員

Note

開始之前，請注意移除成員：

- 從協同合作中移除所有相關聯的資料集
- 如果 [成員支付查詢運算成本](#)，此動作會停止協同合作中的所有查詢執行。

移除成員也會從協同合作中移除所有相關聯的資料集。

先決條件

- 您必須是協作建立者
- 您無法移除自己的帳戶

從協同合作中移除成員

1. 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#)。
2. 在左側導覽窗格中，選擇協同合作。

3. 選取您要修改的協同合作。
4. 選擇成員索引標籤。
5. 選取您要移除的成員旁的選項按鈕。
6. 選擇移除。
7. 在確認對話方塊中，輸入 **confirm** 來驗證移除。

Note

移除成員後，與其帳戶相關聯的所有資料集也會從協同合作中移除。

Important

如果您移除支付查詢運算成本的成員，在您指定新的付費成員之前，無法在協同合作中執行其他查詢。

離開協同合作

身為協同合作成員，您可以透過刪除您的成員資格來離開協同合作。如果您是協同合作建立者，您只能透過[刪除協同合作來離開協同合作](#)。

Note

刪除成員資格時，您會離開協同合作，無法重新加入。如果您是[支付查詢運算成本的成員](#)，而且您刪除了成員資格，則不允許再執行任何查詢。

離開協同合作

1. 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#)。
2. 在左側導覽窗格中，選擇協同合作。
3. 對於具有作用中成員資格的，請選擇您成為成員的協同合作。
4. 選擇動作。
5. 選擇刪除成員資格。

6. 在對話方塊中，**confirm**在文字輸入欄位中輸入，確認離開協同合作的決定，然後選擇空白並刪除成員資格。

您在主控台上看到一則訊息，指出成員資格已刪除。

協同合作建立者會將成員狀態視為左。

在 中準備資料表 AWS Clean Rooms

Note

準備資料表可以在您加入協同合作之前或之後進行。準備好資料表之後，只要該資料表的隱私權需求相同，您就可以在多個協同合作之間重複使用該資料表。

身為協同合作的成員，您必須先準備資料表，才能 AWS Clean Rooms 讓可查詢的協同合作成員查詢資料表。

您在 中用於查詢的資料表 AWS Clean Rooms 通常與您用於其他應用程式的資料表類型相同。例如，相同的資料集類型會與 Amazon Athena、Amazon EMR、Amazon Redshift Spectrum 和 Amazon Quick 搭配使用。

您可以直接從下列任何資料來源以原始格式查詢資料：

- Amazon Simple Storage Service (Amazon S3)
- Amazon Athena
- Snowflake

AWS Clean Rooms 在查詢執行時間存取資料集，確保可查詢的成員一律存取up-to-date資料。任何暫時讀取到 AWS Clean Rooms 協同合作的資料都會在查詢完成後刪除。查詢結果會寫入您的 Amazon S3 儲存貯體。

如果您的使用案例涉及查詢身分資料，請參閱 [AWS Entity Resolution 在 中 AWS Clean Rooms](#)。

主題

- [的資料格式 AWS Clean Rooms](#)
- [Apache Iceberg 中的資料表 AWS Clean Rooms](#)
- [為 中的查詢準備資料表 AWS Clean Rooms](#)
- [使用 的密碼編譯運算準備加密資料表 Clean Rooms](#)
- [使用 C3R 加密用戶端解密資料表](#)

的資料格式 AWS Clean Rooms

若要分析資料，資料集必須採用 AWS Clean Rooms 支援的格式。

主題

- [PySpark 任務支援的資料格式](#)
- [SQL 查詢支援的資料格式](#)
- [支援的資料類型](#)
- [的檔案壓縮類型 AWS Clean Rooms](#)
- [的伺服器端加密 AWS Clean Rooms](#)

PySpark 任務支援的資料格式

AWS Clean Rooms 支援執行 PySpark 任務的下列結構化格式。

- Parquet
- OpenCSV
- JSON

SQL 查詢支援的資料格式

AWS Clean Rooms 支援執行 SQL 查詢的下列結構化格式。

- [Apache Iceberg 資料表](#)
- Parquet
- OpenCSV
- JSON

Note

文字檔案中timestamp的值必須採用 格式yyyy-MM-dd HH:mm:ss.SSSSSS。例如：2017-05-01 11:30:59.000000。

建議使用單欄式儲存檔案格式，例如 Apache Parquet。使用單欄式儲存檔案格式，您只需選取所需的資料欄，即可將資料移動降至最低。為了獲得最佳效能，大型物件應分割為 100mb–1gb 物件。

支援的資料類型

AWS Clean Rooms 支援下列資料類型。

- ARRAY
- BIGINT
- BOOLEAN
- BYTE
- CHAR
- DATE
- DECIMAL
- FLOAT
- INTEGER
- INTERVAL
- LONG
- MAP
- REAL
- SHORT
- SMALLINT
- STRUCT
- TIME
- TIMESTAMP_LTZ
- TIMESTAMP_NTZ
- TINYINT
- VARCHAR

如需詳細資訊，請參閱 AWS Clean Rooms SQL 參考中的[資料類型](#)。

的檔案壓縮類型 AWS Clean Rooms

為了減少儲存空間、改善效能並將成本降至最低，強烈建議您壓縮資料集。

AWS Clean Rooms 根據副檔名識別檔案壓縮類型，並支援下表所示的壓縮類型和副檔名。

壓縮演算法	副檔名
GZIP	.gz
Bzip2	.bz2
Snappy	.snappy

您可以套用不同層級的壓縮。最常見的是，您可以壓縮整個檔案或壓縮檔案中的個別區塊。在檔案層級壓縮單欄格式不會產生效能優勢。

的伺服器端加密 AWS Clean Rooms

Note

伺服器端加密不會取代那些需要密碼編譯運算的使用案例。

AWS Clean Rooms 以透明方式解密使用下列加密選項加密的資料集：

- SSE-S3 – 使用由 Amazon S3 管理的 AES-256 加密金鑰進行伺服器端加密
- SSE-KMS – 使用管理的金鑰進行伺服器端加密 AWS Key Management Service

若要使用 SSE-S3，用於將設定的資料表與協同合作建立關聯的 AWS Clean Rooms 服務角色必須具有 KMS-decrypt 許可。若要使用 SSE-KMS，KMS 金鑰政策也必須允許 AWS Clean Rooms 服務角色解密。

AWS Clean Rooms 不支援 Amazon S3 用戶端加密。如需伺服器端加密的詳細資訊，請參閱《Amazon Simple Storage Service 使用者指南》中的[使用伺服器端加密保護資料](#)。

Apache Iceberg 中的資料表 AWS Clean Rooms

Apache Iceberg 是資料湖的開放原始碼資料表格式。AWS Clean Rooms 可以使用 Apache Iceberg 中繼資料中存放的統計資料來最佳化查詢計劃，並減少無塵室查詢處理期間的檔案掃描。如需詳細資訊，請參閱 [Apache Iceberg](#) 文件。

AWS Clean Rooms 搭配 Iceberg 資料表使用時，請考慮下列事項：

- 適用於 S3 的 Apache Iceberg 資料表 – 資料表必須在 AWS Glue Data Catalog 根據 [開放原始碼膠水目錄實作](#) 來定義。
- 適用於 Athena 的 Apache Iceberg 資料表 – 如需詳細資訊，請參閱 <https://docs.aws.amazon.com/athena/latest/ug/querying-iceberg.html>
- Snowflake 的 Apache Iceberg 資料表 – 如需詳細資訊，請參閱 <https://docs.snowflake.com/en/user-guide/tables-iceberg>
- Parquet 檔案格式 – AWS Clean Rooms 僅支援 Parquet 資料檔案格式的 Iceberg 資料表。
- GZIP 和 Snappy 壓縮 – AWS Clean Rooms 支援具有 GZIP 和 Snappy 壓縮的 Parquet。
- Iceberg 版本 – AWS Clean Rooms 支援針對第 1 版和第 2 版 Iceberg 資料表執行查詢。
- 分割區 – 您不需要手動新增 Apache Iceberg 資料表的分割區 AWS Glue。會自動 AWS Clean Rooms 偵測 Apache Iceberg 資料表中的新分割區，而且不需要手動操作即可更新資料表定義中的分割區。Iceberg 分割區在 AWS Clean Rooms 資料表結構描述中顯示為一般資料欄，而不是在設定的資料表結構描述中單獨顯示為分割區索引鍵。
- 限制

Apache Iceberg 不支援從資料表轉換的 Apache Parquet 資料表。

- 時間歷程查詢

AWS Clean Rooms 不支援具有 Apache Iceberg 資料表的時間行程查詢。

- Athena 引擎版本 2

Iceberg 不支援使用 Athena 引擎版本 2 建立的資料表。

- 檔案格式

Avro 不支援 和 Optimized Row Columnar (ORC) 檔案格式。

- 壓縮

Zstandard Parquet 不支援的 (Zstd) 壓縮。

Iceberg 資料表支援的資料類型

AWS Clean Rooms 可以查詢包含下列資料類型的 Iceberg 資料表：

- BOOLEAN
- DATE
- DECIMAL
- DOUBLE
- FLOAT
- INT
- LIST
- LONG
- MAP
- STRING
- STRUCT
- TIMESTAMP WITHOUT TIME ZONE

如需 Iceberg 資料類型的相關資訊，請參閱 Apache Iceberg 文件中的 [Iceberg 結構描述](#)。

為 中的查詢準備資料表 AWS Clean Rooms

如果您的使用案例不需要您自備資料，您可以略過此程序。

如果您的使用案例涉及查詢身分資料，請參閱 [AWS Entity Resolution 在 中 AWS Clean Rooms](#)。

如需您可以使用之資料格式的詳細資訊，請參閱 [的資料格式 AWS Clean Rooms](#)。

主題

- [在 Amazon S3 中準備資料表](#)
- [在 Amazon Athena 中準備資料表](#)
- [在 Snowflake 中準備資料表](#)

在 Amazon S3 中準備資料表

您可以分析已在 中編製目錄 AWS Glue 並存放在 Amazon S3 中的資料表。如果您的資料表已在 中編製目錄 AWS Glue，請跳至 [在 中建立設定的資料表 AWS Clean Rooms](#)。

Important

在 Amazon S3 中準備資料以搭配使用時 AWS Clean Rooms，您必須確保 AWS Glue 資料表位置與儲存資料檔案的目錄路徑完全相符。

例如：如果您的資料存放在 `s3://mybucket/folder/subfolder/data.parquet` 中，您的 AWS Glue 資料表位置必須指向 `s3://mybucket/folder/subfolder/`。將資料表位置設定為父目錄 (`s3://mybucket/folder/`) 會導致查詢時資料表顯示空白。

在 Amazon S3 中準備資料表需要下列步驟：

主題

- [步驟 1：完成先決條件](#)
- [步驟 2：\(選用\) 準備資料以進行密碼編譯運算](#)
- [步驟 3：將資料表上傳至 Amazon S3](#)
- [步驟 4：建立 AWS Glue 資料表](#)
- [步驟 5：後續步驟](#)

步驟 1：完成先決條件

若要準備資料表以搭配使用 AWS Clean Rooms，您必須完成下列先決條件：

- 您的資料資料表會儲存為其中一個[支援的資料格式 AWS Clean Rooms](#)。
- 您的資料表會在 中編製目錄，AWS Glue 並使用[支援的資料類型 AWS Clean Rooms](#)。
- 您的所有資料表都會儲存在建立協同合作的相同 AWS 區域 Amazon Simple Storage Service (Amazon S3) 中。
- AWS Glue Data Catalog 必須與協同合作位於相同的區域。
- AWS Glue Data Catalog 與成員資格 AWS 帳戶 位於相同的 。
- Amazon S3 儲存貯體未註冊 AWS Lake Formation。

步驟 2：(選用) 準備資料以進行密碼編譯運算

(選用) 如果您使用密碼編譯運算，且資料表包含您要加密的敏感資訊，則必須使用 C3R 加密用戶端來加密資料表。

若要準備資料以進行密碼編譯運算，請遵循中的程序[使用的密碼編譯運算準備加密資料表 Clean Rooms](#)。

步驟 3：將資料表上傳至 Amazon S3

Note

如果您想要在協同合作中使用加密資料表，您必須先加密資料以進行密碼編譯運算，才能將資料表上傳至 Amazon S3。如需詳細資訊，請參閱[使用的密碼編譯運算準備加密資料表 Clean Rooms](#)。

將資料表上傳至 Amazon S3

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/s3/> 開啟 Amazon S3 主控台。
2. 選擇儲存貯體，然後選擇您要存放資料表的儲存貯體。
3. 選擇上傳，然後依照提示操作。
4. 選擇物件索引標籤，以檢視儲存資料的字首。記下資料夾的名稱。

您可以選取要檢視資料的資料夾。

步驟 4：建立 AWS Glue 資料表

如果您已有 AWS Glue 資料表，可以略過此步驟。

在此步驟中，您會在 [中](#) 設定爬蟲程式 AWS Glue，以編目 S3 儲存貯體中的所有檔案並建立 AWS Glue 資料表。如需詳細資訊，請參閱 AWS Glue 《使用者指南》中的[在 中定義爬蟲程式 AWS Glue](#)。

如需支援的 AWS Glue Data Catalog 資料類型的詳細資訊，請參閱 [支援的資料類型](#)。

Note

AWS Clean Rooms 目前不支援向 註冊的 S3 儲存貯體 AWS Lake Formation。

下列程序說明如何建立 AWS Glue 資料表。如果您想要使用具有 AWS Key Management Service (AWS KMS) 金鑰的加密 AWS Glue Data Catalog 物件，您需要設定 KMS 金鑰許可政策，以允許存取

該加密的資料表。如需詳細資訊，請參閱《AWS Glue 開發人員指南》中的在 [AWS Glue 中設定加密](#)。

建立 AWS Glue 資料表

1. 遵循AWS Glue 《使用者指南》中的在 [AWS Glue 主控台程序上使用爬蟲程式](#)。
2. 請記下 AWS Glue 資料庫名稱和 AWS Glue 資料表名稱。

步驟 5：後續步驟

現在您已在 Amazon S3 中準備資料表，您已準備好：

- [建立設定的資料表](#)
- [建立 ML 模型](#)

您可以在以下時間之後查詢資料表：

- 協同合作建立者已在其中設定協同合作 AWS Clean Rooms。如需詳細資訊，請參閱[建立協同合作](#)。
- 協同合作建立者已將協同合作 ID 傳送給身為協同合作參與者的您。

在 Amazon Athena 中準備資料表

您可以在 Amazon Athena 中查詢已建立為 AWS Glue Data Catalog (GDC) 檢視的資料表。

GDC 檢視是從一或多個基礎資料表建立的虛擬 AWS Glue 資料表。它必須使用 Athena AwsGlueCatalog目錄中的 Athena SQL 建立。

在 Amazon Athena 中準備資料表需要下列步驟：

主題

- [步驟 1：完成先決條件](#)
- [步驟 2：（選用）準備資料以進行密碼編譯運算](#)
- [步驟 3：後續步驟](#)

步驟 1：完成先決條件

若要準備資料表以搭配使用 AWS Clean Rooms，您必須完成下列先決條件：

- 您的資料資料表會儲存為其中一個[支援的資料格式 AWS Clean Rooms](#)。
- 您的資料表使用[支援的資料類型 AWS Clean Rooms](#)。
- 您已在 Athena AwsDataCatalog目錄中使用 Athena SQL 在 AWS Glue 資料表上建立 GDC 檢視。

檢視將顯示在：

- Athena 主控台（在下AwsDataCatalog) 做為檢視：<https://console.aws.amazon.com/athena/>
- AWS Glue 主控台做為 AWS Glue 資料表：<https://console.aws.amazon.com/glue/>

如需詳細資訊，請參閱《Amazon Athena [Athena 使用者指南](#)》中的在 Athena 中使用 Data Catalog 檢視。

Note

您需要適當的許可，才能在 Athena 和 中建立檢視 AWS Glue。此外，請確定您能夠存取檢視定義中參考的基礎資料表。

AWS Clean Rooms 僅支援 Athena 的 AWS Glue 目錄類型，不支援 Lambda 或 Hive 目錄類型。

- 您的資料資料表或 GDC 檢視會在 中編製目錄，AWS Glue 並向其註冊 AWS Lake Formation。
- 您已在 Amazon S3 中建立單獨的輸出儲存貯體，以接收 Athena 結果。
- 您已設定服務角色從 Amazon Athena 讀取資料。如需詳細資訊，請參閱[建立服務角色以從 Amazon Athena 讀取資料](#)。
- 服務角色具有 GDC 檢視或資料表上的 Lake Formation Select 和 Describe 存取許可。

步驟 2：（選用）準備資料以進行密碼編譯運算

（選用）如果您使用密碼編譯運算，且資料表包含您要加密的敏感資訊，則必須使用 C3R 加密用戶端來加密資料表。

若要準備資料以進行密碼編譯運算，請遵循中的程序[使用的密碼編譯運算準備加密資料表 Clean Rooms](#)。

步驟 3：後續步驟

現在您已在 Amazon Athena 中準備資料表，您已準備好：

- [建立設定的資料表](#)
- [建立 ML 模型](#)

您可以在以下時間之後查詢資料表：

- 協同合作建立者已在其中設定協同合作 AWS Clean Rooms。如需詳細資訊，請參閱[建立協同合作](#)。
- 協同合作建立者已將協同合作 ID 傳送給身為協同合作參與者的您。

在 Snowflake 中準備資料表

您可以查詢存放在 Snowflake 資料倉儲中的資料表。

在 Snowflake 中準備資料表需要下列步驟：

主題

- [步驟 1：完成先決條件](#)
- [步驟 2：（選用）準備資料以進行密碼編譯運算](#)
- [步驟 3：建立 AWS Secrets Manager 秘密](#)
- [步驟 4：後續步驟](#)

步驟 1：完成先決條件

若要準備資料表以搭配使用 AWS Clean Rooms，您必須完成下列先決條件：

- 您擁有授予適當許可來讀取資料表 AWS 帳戶的。如需詳細資訊，請參閱[建立服務角色以從 Snowflake 讀取資料](#)。
- 您的資料資料表會儲存為其中一個[支援的資料格式 AWS Clean Rooms](#)。
- 您的資料表使用[支援的資料類型 AWS Clean Rooms](#)。
- 您的資料資料表存放在 Snowflake 倉儲中。如需詳細資訊，請參閱[Snowflake 文件](#)。
- 您已設定新的 Snowflake 使用者，具有要與協同合作建立關聯的 Snowflake 資料表的唯讀權限。

步驟 2：（選用）準備資料以進行密碼編譯運算

（選用）如果您使用密碼編譯運算，且資料表包含您要加密的敏感資訊，則必須使用 C3R 加密用戶端來加密資料表。

若要準備資料以進行密碼編譯運算，請遵循中的程序[使用的密碼編譯運算準備加密資料表 Clean Rooms](#)。

步驟 3：建立 AWS Secrets Manager 秘密

若要從連線至 Snowflake AWS Clean Rooms，您需要在 AWS Secrets Manager 秘密中建立和存放 Snowflake 登入資料，然後將該秘密與 Snowflake 資料表建立關聯 AWS Clean Rooms。

Note

我們建議您建立專門用於的新使用者 AWS Clean Rooms。該使用者應只具有 AWS Clean Rooms 您要存取之資料的讀取許可角色。

建立 AWS Secrets Manager 秘密

1. 在 Snowflake 中，產生使用者 `snowflakeUser` 並設定金鑰對身分驗證。

Note

在 2025 年 11 月，Snowflake 將轉換為僅支援金鑰對身分驗證。此變更會影響目前 AWS Clean Rooms 與 Snowflake 的整合，而 Snowflake 會使用使用者名稱和密碼身分驗證。在此日期之後，中的 Snowflake 連線 AWS Clean Rooms 將需要使用 Snowflake 隱私權增強型郵件 (PEM) 私有金鑰進行金鑰對身分驗證。

2. 決定此使用者將與哪個 Snowflake 倉儲互動 `snowflakeWarehouse`。在 Snowflake `snowflakeUser` 中將其設定為 `DEFAULT_WAREHOUSE` 的，或在下一個步驟中記住它。
3. 在 [AWS Secrets Manager](#) 中，使用您的 Snowflake 憑證建立機密。若要在 Secrets Manager 中建立秘密，請遵循 AWS Secrets Manager 《使用者指南》中[建立 AWS Secrets Manager 秘密](#)中提供的教學課程。建立秘密之後，請保留秘密名稱 `secretName` 以進行下一個步驟。
 - 選取鍵/值對時，`snowflakeUser` 請使用鍵 為 建立對 `sfUser`。
 - 選取金鑰/值對時，請使用金鑰 為您的 Snowflake PEM 私有金鑰建立對 `pem_private_key`。
 - 選取鍵/值對時，`snowflakeWarehouse` 請使用鍵 為 建立對 `sfWarehouse`。

如果在 Snowflake 中設定預設值，則不需要這麼做。

- 選取鍵/值對時，`snowflakeRole` 請使用鍵 為 建立對 `sfRole`。

步驟 4：後續步驟

現在您已在 Snowflake 中準備資料表，您已準備好：

- [建立設定的資料表](#)
- [建立 ML 模型](#)

您可以在以下時間之後查詢資料表：

- 協同合作建立者已在其中設定協同合作 AWS Clean Rooms。如需詳細資訊，請參閱[建立協同合作](#)。
- 協同合作建立者已將協同合作 ID 傳送給身為協同合作參與者的您。

使用 的密碼編譯運算準備加密資料表 Clean Rooms

Clean Rooms (C3R) 的加密運算功能。AWS Clean Rooms 您可以使用 C3R 以密碼編譯方式限制任一方和 AWS AWS Clean Rooms 協同合作可以學習的內容。

您可以使用用戶端加密工具 C3R 加密用戶端來加密資料表，再將資料表上傳至資料來源：Amazon Simple Storage Service (Amazon S3)、Amazon Athena 或 Snowflake。

如需詳細資訊，請參閱[的加密運算 Clean Rooms](#)。

使用 C3R 準備加密資料表需要下列步驟：

步驟

- [步驟 1：完成先決條件](#)
- [步驟 2：下載 C3R 加密用戶端](#)
- [步驟 3：（選用）在 C3R 加密用戶端中檢視可用的命令](#)
- [步驟 4：產生表格式檔案的加密結構描述](#)
- [步驟 5：建立共用私密金鑰](#)
- [步驟 6：將共用私密金鑰存放在 環境變數中](#)
- [步驟 7：加密資料](#)
- [步驟 8：驗證資料加密](#)
- [（選用）建立結構描述（進階使用者）](#)

步驟 1：完成先決條件

若要準備資料表以搭配 C3R 使用，您必須完成下列先決條件：

- 您可以在 GitHub 上存取儲存Clean Rooms庫的加密運算：

<https://github.com/aws/c3r>

- 您已設定 AWS 登入資料來使用 C3R 加密用戶端。C3R 加密用戶端會將這些登入資料用於對 進行唯讀 API 呼叫 AWS Clean Rooms，以擷取協同合作中繼資料。如需詳細資訊，請參閱《第 AWS Command Line Interface 2 版使用者指南》中的[設定 AWS CLI](#)。
- 您的機器已安裝 Java Runtime Environment(JRE) 11 或更新版本。
 - 建議的 Java Runtime EnvironmentAmazon Corretto 11 或更新版本可從 <https://aws.amazon.com/corretto> 下載。
 - Java Development Kit (JDK) 包含JRE相同版本的對應。不過，執行 Cryptographic Computing for Clean Rooms(C3R) 加密用戶端JDK不需要的額外功能。
- 您的表格式資料檔案 (.csv) 或Parquet檔案 (.parquet) 會儲存在本機。
- 您或協同合作中的其他成員能夠建立共用私密金鑰。如需詳細資訊，請參閱[步驟 5：建立共用私密金鑰](#)。
- 協同合作建立者已在 中建立協同合作 AWS Clean Rooms，並為協同合作啟用密碼編譯運算。如需詳細資訊，請參閱[建立協同合作](#)。
- 協同合作建立者已將協同合作 ID 傳送給身為協同合作參與者的您。協同合作 Amazon Resource Name (ARN) 包含在傳送的邀請中，其中包含協同合作 ID。

步驟 2：下載 C3R 加密用戶端

從 GitHub 下載 C3R 加密用戶端

- 前往 儲存Clean Rooms AWS GitHub庫的加密運算：<https://github.com/aws/c3r>
- 選取並下載檔案。

原始碼、授權和相關資料可以複製或從GitHub儲存庫的登陸頁面下載為 .zip 檔案。（請參閱儲存庫內容清單右上角的程式碼按鈕）。

最新簽署的 C3R 加密用戶端 Java Executable File（即命令列界面應用程式）位於GitHub儲存庫的發行頁面上。

Apache Spark (c3r-cli-spark) 的 C3R 加密用戶端套件是 c3r-cli 的版本，必須以工作形式提交至執行中的 Apache Spark 伺服器。如需詳細資訊，請參閱[在 Apache Spark 上執行 C3R](#)。

步驟 3：(選用) 在 C3R 加密用戶端中檢視可用的命令

使用此程序來熟悉 C3R 加密用戶端中可用的命令。

在 C3R 加密用戶端中檢視所有可用的命令

1. 從命令列界面 (CLI)，導覽至包含下載 c3r-cli.jar 檔案的資料夾。
2. 執行下列命令：`java -jar c3r-cli.jar`
3. 檢視可用命令和選項的清單。

步驟 4：產生表格式檔案的加密結構描述

若要加密資料，需要描述如何使用資料的加密結構描述。本節說明 C3R 加密用戶端如何協助產生具有標頭列或檔案之 CSV Parquet 檔案的加密結構描述。

每個檔案只需要執行此動作一次。結構描述存在之後，可以重複使用它來加密相同的檔案（或任何具有相同資料欄名稱的檔案）。如果資料欄名稱或所需的加密結構描述變更，您必須更新結構描述檔案。如需詳細資訊，請參閱[\(選用\) 建立結構描述 \(進階使用者\)](#)。

Important

重要的是，所有協作方都使用相同的共用私密金鑰。協作方也應該協調資料欄名稱，以便在查詢中 JOIN 進行編輯或以其他方式比較相等性時相符。否則，SQL 查詢可能會產生非預期或不正確的結果。不過，如果協同合作建立者在協同合作建立期間啟用 `allowJoinsOnColumnsWithDifferentNames` 加密設定，則不需要這麼做。如需加密相關設定的詳細資訊，請參閱[密碼編譯運算參數](#)。

在結構描述模式下執行時，C3R 加密用戶端會依資料欄逐一查看輸入檔案資料欄，提示您是否應處理該資料欄以及如何處理該資料欄。如果檔案包含加密輸出不需要的許多資料欄，則互動式結構描述產生可能會變得繁瑣，因為您必須略過每個不需要的資料欄。若要避免這種情況，您可以手動撰寫結構描述，或建立僅具有所需資料欄的簡化輸入檔案版本。然後，互動式結構描述產生器可以在減少的檔案上執行。C3R 加密用戶端會輸出結構描述檔案的相關資訊，並詢問如何在目標輸出中包含或加密來源資料欄（如果有的話）。

對於輸入檔案中的每個來源資料欄，系統會提示您：

1. 應該產生多少目標資料欄
2. 如何加密每個目標資料欄（如果有的話）
3. 每個目標資料欄的名稱
4. 如果資料欄正在加密為sealed資料欄，如何在加密之前填入資料

Note

當您加密已加密為sealed資料欄的資料時，您必須判斷哪些資料需要填補。C3R 加密用戶端會在產生結構描述期間建議預設填補，以相同長度填補資料欄中的所有項目。

判斷的長度時fixed，請注意填補是以位元組為單位，而不是位元。

以下是建立結構描述的決策表。

結構描述決策表

決策	來源資料欄的目標資料欄數 <'name-of-column'> ?	目標欄類型：【c】cleartext、【f】fingerprint或【s】sealed ?	目標欄標頭名稱 <default 'name-of-column'>	將尾碼 <suffix> 新增至標頭，以指出加密方式、【y】是或【n】否 <default 'yes'>	<'name-of-column_sealed'> 填補類型：【n】1、【f】固定或【m】max <default 'max'>
保持資料欄未加密。	1	c	不適用	不適用	不適用
將資料欄加密為資料 fingerprint 欄。	1	f	選擇預設或輸入新的標頭名稱。	輸入 y 以選擇預設 (_fingerprint) 或輸入 n。	不適用

決策	來源資料欄的目標資料欄數 <'name-of-column'> ?	目標欄類型：【c】cleartext、【f】fingerprint或【s】sealed ?	目標欄標頭名稱 <default 'name-of-column'>	將尾碼 <suffix> 新增至標頭，以指出加密方式、【y】是或【n】否 <default 'yes'>	<'name-of-column_sealed'> 填補類型：【n】1、【f】固定或【m】max <default 'max'>
將資料欄加密為資料sealed欄。	1	s	選擇預設或輸入新的標頭名稱。	輸入 y 以選擇預設 (_sealed) 或輸入 n。	選擇填補類型。 如需詳細資訊，請參閱 (選用) 建立結構描述 (進階使用者) 。
將資料欄同時加密為 fingerprint 和 sealed。	2	輸入第一個目標欄：f。 輸入第二個目標欄：s。	為每個目標資料欄選擇目標標頭。	輸入 y 以選擇預設或輸入 n。	選擇填補類型 (僅適用於 sealed 欄)。 如需詳細資訊，請參閱 (選用) 建立結構描述 (進階使用者) 。

以下是如何建立加密結構描述的兩個範例。互動的確切內容取決於輸入檔案和您提供的回應。

範例

- [範例：產生資料fingerprint欄和cleartext資料欄的加密結構描述](#)
- [範例：使用 sealed、fingerprint和 cleartext資料欄產生加密結構描述](#)

範例：產生資料fingerprint欄和cleartext資料欄的加密結構描述

在此範例中，對於 `ads.csv`，只有兩個資料欄：`username`和`ad_variant`。對於這些資料欄，我們需要下列項目：

- 將資料`username`欄加密為資料`fingerprint`欄
- 讓資料`ad_variant`欄成為資料`cleartext`欄

為資料`fingerprint`欄和`cleartext`資料欄產生加密結構描述

1. (選用) 為確保要加密`c3r-cli.jar`的檔案存在：
 - a. 導覽至所需的目錄並執行 `ls` (如果使用 Mac或 Unix/Linux)，或者`dir`如果使用 Windows)。
 - b. 檢視表格式資料檔案清單 (例如 `.csv`)，然後選擇要加密的檔案。

在此範例中，`ads.csv` 是要加密的檔案。

2. 從 CLI 執行下列命令，以互動方式建立結構描述。

```
java -jar c3r-cli.jar schema ads.csv --interactive --output=ads.json
```

Note

- 您可以執行 `java --jar PATH/T0/c3r-cli.jar`。或者，如果您已將 `PATH/T0/c3r-cli.jar` 新增至 `CLASSPATH` 環境變數，您也可以執行類別名稱。C3R 加密用戶端會在 `CLASSPATH` 中尋找它 (例如，`java com.amazon.psion.cli.Main`)。
- `--interactive` 旗標會選取開發結構描述的互動式模式。這會引導使用者完成建立結構描述的精靈。具備進階技能的使用者無需使用精靈即可建立自己的結構描述 JSON。如需詳細資訊，請參閱 [\(選用\) 建立結構描述 \(進階使用者\)](#)。
- `--output` 旗標會設定輸出名稱。如果您未包含 `--output` 旗標，C3R 加密用戶端會嘗試挑選預設輸出名稱 (例如 `<input>.out.csv` 或結構描述的 `<input>.json`)。

3. 對於 Number of target columns from source column 'username'?, 輸入 **1**，然後按 Enter。
4. 對於 Target column type: [c]leartext, [f]ingerprint, or [s]ealed?, 輸入 **f**，然後按 Enter。
5. 對於 Target column headername <default 'username'>, 請按 Enter。

使用預設名稱「username」。

6. 對於 Add suffix '_fingerprint' to header to indicate how it was encrypted, [y]es or [n]o <default 'yes'>, 輸入 **y**, 然後按 Enter。

 Note

互動式模式建議將尾碼新增至加密的資料欄標頭 (_fingerprint 代表fingerprint資料欄, _sealed 代表資料sealed欄)。當您執行上傳資料 AWS 服務 或建立 AWS Clean Rooms 協同合作等任務時, 字尾可能很有幫助。這些字尾有助於指出如何處理每個資料欄中的加密資料。例如, 如果您將資料欄加密為資料欄 (_sealed), 並嘗試在資料sealed欄 JOIN上進行加密或反向嘗試, 則實物將無法運作。

7. 對於 Number of target columns from source column 'ad_variant'?, 輸入 **1**, 然後按 Enter。
8. 對於 Target column type: [c]leartext, [f]ingerprint, or [s]ealed?, 輸入 **c**, 然後按 Enter。
9. 對於 Target column headername <default 'username'>, 請按 Enter。

使用預設名稱「ad_variant」。

結構描述會寫入名為 的新檔案ads.json。

 Note

您可以在任何文字編輯器中開啟結構描述, 例如 Notepad 上的 Windows或 TextEdit上的 macOS。

10. 您現在已準備好[加密資料](#)。

範例：使用 sealed、fingerprint和 cleartext資料欄產生加密結構描述

在此範例中, 對於 sales.csv, 有三個資料欄: username、purchased和 product。對於這些資料欄, 我們需要下列項目:

- 讓資料product欄成為資料sealed欄
- 將資料username欄加密為資料fingerprint欄
- 讓資料purchased欄成為資料cleartext欄

使用 sealed、fingerprint 和 cleartext 資料欄產生加密結構描述

1. (選用) 為確保要加密 c3r-cli.jar 的檔案存在：

- 導覽至所需的目錄並執行 `ls` (如果使用 Mac 或 Unix/Linux) , 或者 `dir` 如果使用 Windows)。
- 檢視表格式資料檔案 (.csv) 的清單 , 然後選擇要加密的檔案。

在此範例中 , `sales.csv` 是我們想要加密的檔案。

2. 從 CLI 執行下列命令 , 以互動方式建立結構描述。

```
java -jar c3r-cli.jar schema sales.csv --interactive --  
output=sales.json
```

Note

- `--interactive` 旗標會選取開發結構描述的互動式模式。這會引導使用者完成建立結構描述的引導式工作流程。
- 如果您是進階使用者 , 則可以建立自己的結構描述 JSON , 而無需使用引導式工作流程。如需詳細資訊 , 請參閱 [\(選用\) 建立結構描述 \(進階使用者\)](#)。
- 對於沒有資料欄標頭的 .csv 檔案 , 請參閱 CLI 中可用的結構描述命令 `--noHeaders` 旗標。
- `--output` 旗標會設定輸出名稱。如果您未包含 `--output` 旗標 , C3R 加密用戶端會嘗試挑選預設輸出名稱 (例如 `<input>.out` 或結構描述的 `<input>.json`)。

- 對於 Number of target columns from source column 'username'? , 輸入 **1** , 然後按 Enter。
- 對於 Target column type: [c]leartext, [f]ingerprint, or [s]ealed? , 輸入 **f** , 然後按 Enter。
- 對於 Target column headername <default 'username'> , 請按 Enter。

使用預設名稱「username」。

- 對於 Add suffix '_fingerprint' to header to indicate how it was encrypted, [y]es or [n]o <default 'yes'> , 輸入 **y** , 然後按 Enter。
- 對於 Number of target columns from source column 'purchased'? , 輸入 **1** , 然後按 Enter。

8. 對於 Target column type: [c]leartext, [f]ingerprint, or [s]ealed?, 輸入 **c**, 然後按 Enter。
9. 對於 Target column headername <default 'purchased'>, 請按 Enter。
使用預設名稱「purchased」。
10. 對於 Number of target columns from source column 'product'?, 輸入 **1**, 然後按 Enter。
11. 對於 Target column type: [c]leartext, [f]ingerprint, or [s]ealed?, 輸入 **s**, 然後按 Enter。
12. 對於 Target column headername <default 'product'>, 請按 Enter。
使用預設名稱「product」。
13. 針對 'product_sealed' padding type: [n]one, [f]ixed, or [m]ax <default 'max'>?, 按 Enter 選擇預設值。
14. 針對 , Byte-length beyond max length to pad cleartext to in 'product_sealed' <default '0'>? 按 Enter 選擇預設值。
結構描述會寫入名為 `sales.json` 的新檔案。
15. 您現在已準備好[加密資料](#)。

步驟 5：建立共用私密金鑰

若要加密資料表，協同合作參與者必須同意並安全地共用共用私密金鑰。

共用私密金鑰必須至少為 256 位元 (32 位元組)。您可以指定較大的金鑰，但它不會為您提供任何額外的安全性。

Important

請記住，所有協同合作參與者用於加密和解密的金鑰和協同合作 ID 必須相同。

下列各節提供主控台命令的範例，用於產生儲存在 `secret.key` 各自終端機目前工作目錄中的共用私密金鑰。

主題

- [範例：使用產生金鑰 OpenSSL](#)

- [範例：Windows使用 在 上產生金鑰 PowerShell](#)

範例：使用 產生金鑰 OpenSSL

對於常見的一般用途密碼編譯程式庫，請執行下列命令來建立共用私密金鑰。

```
openssl rand 32 > secret.key
```

如果您使用 Windows且尚未OpenSSL安裝，您可以使用範例：[Windows使用 在 上產生金鑰中描述的範例來產生金鑰PowerShell](#)。

範例：Windows使用 在 上產生金鑰 PowerShell

對於 上可用的PowerShell終端機應用程式Windows，請執行下列命令來建立共用私密金鑰。

```
$bs = New-Object Byte[](32);  
[Security.Cryptography.RandomNumberGenerator]::Create().GetBytes($bs); Set-  
Content 'secret.key' -Encoding Byte -Value $bs
```

步驟 6：將共用私密金鑰存放在 環境變數中

環境變數是方便且可擴展的方式，可讓使用者從各種金鑰存放區提供私密金鑰，例如，AWS Secrets Manager 並將其傳遞給 C3R 加密用戶端。

AWS 服務 如果您使用 AWS CLI 將這些金鑰存放在相關環境變數中，C3R 加密用戶端可以使用 中存放的金鑰。例如，C3R 加密用戶端可以使用來自 的金鑰 AWS Secrets Manager。如需詳細資訊，請參閱AWS Secrets Manager 《使用者指南》中的[使用 建立和管理秘密 AWS Secrets Manager](#)。

Note

不過，在您使用 AWS 服務 之類的 AWS Secrets Manager 來保存 C3R 金鑰之前，請確認您的使用案例允許。某些使用案例可能需要從中保留金鑰 AWS。這是為了確保加密的資料和金鑰永遠不會由相同的第三方持有。

共用私密金鑰的唯一要求是共用私密金鑰是 base64編碼並存放在環境變數中C3R_SHARED_SECRET。

下列各節說明將secret.key檔案轉換為 base64 並將其儲存為環境變數的主控台命令。secret.key 檔案可能已從 中列出的任何命令產生，[步驟 5：建立共用私密金鑰](#)而且只是範例來源。

Windows 使用 在 的環境變數中存放金鑰 PowerShell

若要Windows使用 在 上轉換為 base64並設定環境變數PowerShell，請執行下列命令。

```
$Bytes=[IO.File]::ReadAllBytes((Get-Location).ToString()+'\secret.key');  
$env:C3R_SHARED_SECRET=[Convert]::ToBase64String($Bytes)
```

在 Linux或 的環境變數中存放金鑰 macOS

若要在 Linux或 上轉換為 base64並設定環境變數macOS，請執行下列命令。

```
export C3R_SHARED_SECRET="$(cat secret.key | base64)"
```

步驟 7：加密資料

若要執行此步驟，您必須取得 AWS Clean Rooms 協作 ID 和共用私密金鑰。如需詳細資訊，請參閱[必要條件](#)。

在下列範例中，我們使用我們建立名為 的結構描述ads.csv，在 上執行加密ads.json。

加密資料

1. 在 中存放協同合作的共用私密金鑰[步驟 6：將共用私密金鑰存放在 環境變數中](#)。
2. 從命令列輸入下列命令。

```
java -jar c3r-cli.jar encrypt <name of input .csv file> --schema=<name  
of schema .json file> --id=<collaboration id> --output=<name of  
output.csv file> <optional flags>
```

3. 針對 <## .csv #####>，輸入輸入 .csv 檔案的名稱。
4. 針對 schema=，輸入 .json 加密結構描述檔案的名稱。
5. 針對 id=，輸入協同合作 ID。
6. 針對 output=，輸入輸出檔案的名稱（例如 ads-output.csv）。
7. 包含 [密碼編譯運算參數](#)和 中所述的任何命令列旗標[適用於 的密碼編譯運算中的選用旗標 Clean Rooms](#)。
8. 執行 命令。

在 範例中ads.csv，我們會執行下列命令。

```
java -jar c3r-cli.jar encrypt ads.csv --schema=ads.json --id=123e4567-e89b-42d3-a456-556642440000 --output=ads-output.csv
```

在範例中sales.csv，我們會執行下列命令。

```
java -jar c3r-cli.jar encrypt sales.csv --schema=sales.json --id=123e4567-e89b-42d3-a456-556642440000
```

Note

在此範例中，我們不會指定輸出檔案名稱 (--output=**sales-output.csv**)。因此，name-of-file.out.csv會產生預設輸出檔案名稱。

您現在已準備好驗證加密的資料。

步驟 8：驗證資料加密

驗證資料已加密

1. 檢視加密的資料檔案（例如 sales-output.csv）。
2. 驗證下列資料欄：
 - a. 欄 1 – 加密（例如，username_fingerprint）。

對於fingerprint資料欄 (HMAC)，在版本和類型字首（例如 01:hmac:）之後，有 44 個字元的 base64 編碼資料。

- b. 資料欄 2 – 未加密（例如 purchased）。
- c. 第 3 欄 – 加密（例如，product_sealed）。

對於加密的 (SELECT) 資料欄，版本和類型字首（例如 01:enc:）後面加上cleartext任何填補的長度與cleartext加密的長度直接成正比。也就是說，由於編碼，長度是輸入的大小加上大約 33% 的額外負荷。

您現在已準備好：

1. [將加密的資料上傳到 S3](#)。
2. [建立 AWS Glue 資料表](#)。

3. 在 中建立設定的資料表 AWS Clean Rooms。

C3R 加密用戶端會建立不包含未加密資料的暫存檔案（除非最終輸出中也會未加密該資料）。不過，某些加密的值可能無法正確填入。即使協同合作設定為 `allowRepeatedFingerprintValue`，指紋資料欄也可能包含重複的值 `false`。此問題是因為在檢查適當的填補長度和重複移除屬性之前寫入暫存檔案。

如果 C3R 加密用戶端在加密期間失敗或中斷，它可能會在寫入暫存檔案之後，但在檢查這些屬性並刪除暫存檔案之前停止。因此，這些暫存檔案可能仍在磁碟上。如果是這種情況，這些檔案中的內容不會保護純文字資料到與輸出相同的層級。特別是，這些暫存檔案可能會向無法對最終輸出運作的統計分析顯示純文字資料。使用者應刪除這些檔案（特別是 SQLite 資料庫），以防止這些檔案落入未經授權的手中。

（選用）建立結構描述（進階使用者）

手動建立結構描述適用於進階使用者。

以下是具有或沒有資料欄標頭之輸入檔案的 JSON 結構描述檔案格式說明。進階使用者可以視需要直接寫入或修改結構描述。

Note

C3R 加密用戶端可以透過 中所述的互動式程序 [範例：使用 sealed、fingerprint 和 cleartext 資料欄產生加密結構描述](#) 或透過建立 stub 範本來協助您建立結構描述。

映射和位置資料表結構描述

下節說明兩種資料表結構描述：

- 映射的資料表結構描述 – 此結構描述用於加密具有標頭列和檔案的 .csv Apache Parquet 檔案。
- 位置資料表結構描述 – 此結構描述用於加密沒有標頭列的 .csv 檔案。

C3R 加密用戶端可以加密協作的表格式檔案。若要這樣做，它必須具有對應的結構描述檔案，指定應如何從輸入衍生加密的輸出。

C3R 加密用戶端可以透過在命令列執行 C3R 加密用戶端結構描述命令，協助產生 INPUT 檔案的結構描述。命令的範例為 `java -jar c3r-cli.jar schema --interactive INPUT`。

結構描述會指定下列資訊：

1. 哪些來源資料欄透過其標頭名稱（映射結構描述）或位置（位置結構描述）映射至輸出檔案中轉換的資料欄
2. 要保留的目標欄 cleartext
3. 要針對SELECT查詢加密的目標資料欄
4. 要針對JOIN查詢加密的目標資料欄

此資訊在資料表特定的 JSON 結構描述檔案中編碼，其中包含單一物件，其headerRow欄位為布林值。值必須是true具有標頭列Parquet的檔案和.csv 檔案，false否則為。

映射的資料表結構描述

映射的結構描述具有下列形狀。

```
{
  "headerRow": true,
  "columns": [
    {
      "sourceHeader": STRING,
      "targetHeader": STRING,
      "type": TYPE,
      "pad": PAD
    },
    ...
  ]
}
```

如果 headerRow 為 true，則物件中的下一個欄位為 columns，其中包含將來源標頭映射至目標標頭的欄結構描述陣列（即描述輸出欄應包含內容的 JSON 物件）。

- sourceHeader – 資料衍生來源欄的STRING標頭名稱。

Note

相同的來源資料欄可用於多個目標資料欄。

來自輸入檔案的資料欄未列在結構描述的任何sourceHeader位置，不會出現在輸出檔案中。

- targetHeader – 輸出檔案中對應欄的STRING標頭名稱。

Note

對於映射的結構描述，此欄位是選用的。如果省略此欄位，sourceHeader則會針對輸出中的標頭名稱重複使用。如果輸出資料欄分別是fingerprint資料欄或sealed資料欄，_sealed則會附加_fingerprint 或。

- type – 輸出檔案中目標欄TYPE的。也就是說，cleartext、sealed或之一，fingerprint取決於資料欄在協同合作中的使用方式。
- pad – 資料欄結構描述物件的欄位，只有在TYPE為時才存在sealed。其對應值PAD為物件，描述資料在加密之前應如何填入。

```
{
  "type": PAD_TYPE,
  "length": INT
}
```

若要指定預先加密填補，type length 使用方式如下：

- PAD_TYPE as none – 不會將填補套用至資料欄的資料，且 length 欄位不適用（即省略）。
- PAD_TYPE as fixed – 資料欄的資料會填入指定的length位元組。
- PAD_TYPE as max – 資料欄的資料會填入最長值位元組長度加上額外length位元組的大小。

以下是映射結構描述的範例，其中包含每種類型的資料欄。

```
{
  "headerRow": true,
  "columns": [
    {
      "sourceHeader": "FullName",
      "targetHeader": "name",
      "type": "cleartext"
    },
    {
      "sourceHeader": "City",
      "targetHeader": "city_sealed",
      "type": "sealed",
      "pad": {
        "type": "max",
        "length": 16
      }
    }
  ]
}
```

```

    }
  },
  {
    "sourceHeader": "PhoneNumber",
    "targetHeader": "phone_number_fingerprint",
    "type": "fingerprint"
  },
  {
    "sourceHeader": "PhoneNumber",
    "targetHeader": "phone_number_sealed",
    "type": "sealed",
    "pad": {
      "type": "fixed",
      "length": 20
    }
  }
]
}

```

作為更複雜的範例，以下是具有標頭的 .csv 檔案範例。

```

FirstName,LastName,Address,City,State,PhoneNumber,Title,Level,Notes
Jorge,Souza,12345 Mills Rd,Anytown,SC,703-555-1234,CEO,10,
Paulo,Santos,0 Street,Anytown,MD,404-555-111,CIO,9,This is a really long note that
could really be a paragraph
Mateo,Jackson,1 Two St,Anytown,NY,304-555-1324,COO,9,""
Terry,Whitlock,4 N St,Anytown,VA,407-555-8888,EA,7,Secret notes
Diego,Ramirez,9 Hollows Rd,Anytown,VA,407-555-1222,SDE I,4,null
John,Doe,8 Hollows Rd,Anytown,VA,407-555-4321,SDE I,4,Jane's younger brother
Jane,Doe,8 Hollows Rd,Anytown,VA,407-555-4322,SDE II,5,John's older sister

```

在下列映射結構描述範例中，資料欄 `FirstName` 和 `LastName` 是資料 `cleartext` 欄。State 資料欄會加密為資料 `fingerprint` 欄和填補為的資料 `sealed` 欄 `none`。剩餘的資料欄會被省略。

```

{
  "headerRow": true,
  "columns": [
    {
      "sourceHeader": "FirstName",
      "targetHeader": "GivenName",
      "type": "cleartext"
    },
    {

```

```

    "sourceHeader": "LastName",
    "targetHeader": "Surname",
    "type": "cleartext"
  },
  {
    "sourceHeader": "State",
    "targetHeader": "State_Join",
    "type": "fingerprint"
  },
  {
    "sourceHeader": "State",
    "targetHeader": "State",
    "type": "sealed",
    "pad": {
      "type": "none"
    }
  }
]
}

```

以下是映射結構描述產生的 .csv 檔案。

```

givenname,surname,state_fingerprint,state
John,Doe,01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv
+1Mk=,01:enc:FQ3n3Ahv9BQQNWQGcugeHzHYzEZE1vapHa2Uu4SRgSatZ3q0bjPA4TcsHt
+B0kMKBcnHWI13BeGG/SBqmj7vKpI=
Paulo,Santos,01:hmac:CHF4eIrtTNgAooU9v4h9Qjc
+txBnMidQTjdjWuaDTTA=,01:enc:KZ5n5GtaXACco65AXk48BQ02durDNR2ULc4YxmMC8NaZZKKJiksU1IwFadAvV4iBQ1
Mateo,Jackson,01:hmac:iIRnjfNBzryusIJ1w35lgNzeY1RQ1bSfq6PDHW8Xrbk=,01:enc:mLKpS5HIOSgphdEsrzhdEd
eN9nB02gAbIygt40Fn4La1Yn9Xyj/XUWXlmn8zFe2T4kyDTD8kG0vpQEUGxAUFk=
Diego,Ramirez,01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:rmZhT98Zm
+IIGw1UTjMIJP4IrW/AA1tBLMXcHvnYfRgmWP623VFQ6aUnhsb2MDqEw4G5Uwg5rKKZepUxx5uKbfbk=
Jorge,Souza,01:hmac:3BxJdXiFFyZ8HBbYNqqEhBVqhN0d7s2ZiKUe7QiTy08=,01:enc:vVaqWC1VRbhvkf8gnuR7q0z
Terry,Whitlock01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:3c9VEWb0D0/
xbQjdGuccLvI7oZTBdPU+SyrJIyr2kudfAxbuMQ2uRdU/q7rbgyJjxZS8M2U35ILJf/1DgTyg7cM=
Jane,Doe,01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:9RWv46YLveykeNZ/
G0Nd1YFg+AVD0nu05hHyAYTQkPLHnyX+0/jbzD/g9ZT8GCgVE9aB5bV4ooJIXHGBVMXcjrQ=

```

位置資料表結構描述

位置結構描述的形狀如下。

```

{

```

```
"headerRow": false,
"columns": [
  [
    {
      "targetHeader": STRING,
      "type": TYPE,
      "pad": PAD
    },
    {
      "targetHeader": STRING,
      "type": TYPE,
      "pad": PAD
    }
  ],
  [],
  ...
]
```

如果 `headerRow` 為 `false`，則物件中的下一個欄位為 `columns`，其中包含項目陣列。每個項目本身都是零個或多個位置資料欄結構描述（無 `sourceHeader` 欄位）的陣列，這些結構描述輸出應包含的內容。

- `sourceHeader` – 資料衍生來源欄的 `STRING` 標頭名稱。

 Note

必須在位置結構描述中省略此欄位。在位置結構描述中，來源資料欄由結構描述檔案中資料欄的對應索引推斷。

- `targetHeader` – 輸出檔案中對應欄的 `STRING` 標頭名稱。

 Note

位置結構描述需要此欄位。

- `type` – 輸出檔案中目標欄 `TYPE` 的。也就是說，`cleartext`、`sealed` 或 之一，`fingerprint` 取決於資料欄在協同合作中的使用方式。
- `pad` – 資料欄結構描述物件的欄位，只有在 `TYPE` 為 時才存在 `sealed`。其對應值 `PAD` 為 物件，描述資料在加密之前應如何填入。

```
{
  "type": PAD_TYPE,
  "length": INT
}
```

若要指定預先加密填補，type length 使用方式如下：

- PAD_TYPE as none – 不會將填補套用至資料欄的資料，且 length 欄位不適用（即省略）。
- PAD_TYPE as fixed – 資料欄的資料會填入指定的length位元組。
- PAD_TYPE as max – 資料欄的資料會填入最長值位元組長度加上額外length位元組的大小。

Note

fixed 如果您事先知道資料欄資料的位元組大小有上限，則 很有用。如果該欄中的任何資料超過指定的，就會引發錯誤length。

max 當輸入資料的確切大小不明時很方便，因為它無論資料的大小為何都能運作。不過，max需要額外的處理時間，因為它會加密資料兩次。會在讀取到暫存檔案時max加密資料一次，並在已知資料欄中最長的項目之後加密一次。

此外，在用戶端的叫用之間不會儲存最長值的長度。如果您打算批次加密資料，或定期加密新資料，請注意產生的加密文字長度可能因批次而異。

以下是位置結構描述的範例。

```
{
  "headerRow": false,
  "columns": [
    [
      {
        "targetHeader": "name",
        "type": "cleartext"
      }
    ],
    [
      {
        "targetHeader": "city_sealed",
        "type": "sealed",
        "pad": {
          "type": "max",

```

```

        "length": 16
      }
    }
  ],
  [
    {
      "targetHeader": "phone_number_fingerprint",
      "type": "fingerprint"
    },
    {
      "targetHeader": "phone_number_sealed",
      "type": "sealed",
      "pad": {
        "type": "fixed",
        "length": 20
      }
    }
  ]
]
}

```

作為複雜的範例，如果沒有具有 標頭的第一列，則以下是範例 .csv 檔案。

```

Jorge,Souza,12345 Mills Rd,Anytown,SC, 703 -555 -1234,CEO, 10,
Paulo,Santos, 0 Street,Anytown,MD, 404-555-111,CIO, 9,This is a really long note that
  could really be a paragraph
Mateo,Jackson, 1 Two St,Anytown,NY, 304-555-1324,C00, 9, ""
Terry,Whitlock, 4 N St,Anytown,VA, 407-555-8888,EA, 7,Secret notes
Diego,Ramirez, 9 Hollows Rd,Anytown,VA, 407-555-1222,SDE I, 4,null
John,Doe, 8 Hollows Rd,Anytown,VA, 407-555-4321,SDE I, 4,Jane's younger brother
Jane,Doe, 8 Hollows Rd,Anytown,VA, 407-555-4322,SDE II, 5,John's older sister

```

位置結構描述的格式如下。

```

{
  "headerRow": false,
  "columns": [
    [
      {
        "targetHeader": "GivenName",
        "type": "cleartext"
      }
    ]
  ],

```

```
[
  {
    "targetHeader": "Surname",
    "type": "cleartext"
  }
],
[],
[],
[
  {
    "targetHeader": "State_Join",
    "type": "fingerprint"
  },
  {
    "targetHeader": "State",
    "type": "sealed",
    "pad": {
      "type": "none"
    }
  }
],
[],
[],
[],
[]
]
```

上述結構描述會產生下列輸出檔案，其中包含指定目標標頭的標頭列。

```
givenname,surname,state_fingerprint,state
Mateo,Jackson,01:hmac:iIRnjfNBzryusIJ1w35lgNzeY1RQ1bSfq6PDHW8Xrbk=,01:enc:ENS6QD3cMV19vQEGfe9MM
Q8m/Y5SA89dJwKpT5rGPP8e36h6klwDoslpFzGvU0=
Jorge,Souza,01:hmac:3BxJdXiFFyZ8HBbYNqqEhBVqhN0d7s2ZiKUe7QiTy08=,01:enc:LKo0zirq2+
+XEIIIMNRjAsGMdyWUDwYaum0B+IFP+rUf1BNeZDJjtFe1Z+zbZfXQWwJy52Rt7HqvAb2WIK1oMmk=
Paulo,Santos,01:hmac:CHF4eIrtTNgAooU9v4h9Qjc
+txBnMidQTjdjWuaDTTA=,01:enc:MyQKyWxJ9kvK1xDQQtX1UNwv3F+yRBRr0xrUY/1BGg5KFg0n9pK+MZ7g
+ZNqZEPcPz4lht1u0t/wbTaqz0CLXFQ=
Jane,Doe,01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:Pd8sbITBfb0/
ttUB4svVsgoYkDfnDvgkvxzeci0Yxq54rLSwccy1o3/B50C3cpkkn56dovCwzgmPNwrmCmYtb4=
Terry,Whitlock01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv
+1Mk=,01:enc:Qmtzu3B3GAXKh2KkRYTiEAaMopYedsSdF2e/
ADUiBQ9kv2CxCpZWyYTD3ztmKPMka19dHre5VhUHNp030+j1AQ8=
```



```
Diego,Ramirez,01:hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:ysdg
+GHKdeZrS/geBIoo0EPLHG68MsWpx1dh3xjb+fG5rmFmqUcJLNuuYBHhHA1xchM2WVeV1fmHkBX3mvZNvkc=
John,Doe,01:hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:9uX0wZu07kAPAx
+Hf6uvQownkQFSKtWS7gQIJSe5aXFquKWCK6yZN0X5Ea2N3bn03Uj1kh0agDwoiP9FRZGJA4=
```

使用 C3R 加密用戶端解密資料表

對於使用 密碼編譯運算的協同合作Clean Rooms，以及使用 C3R 加密用戶端來加密資料表，請遵循此程序。在協同[合作中查詢資料之後，請使用](#)此程序。

此程序需要共用私密金鑰和協同合作 ID。

可以接收結果的成員會使用用來加密協同合作資料的相同共用私密金鑰和協同合作 ID 來解密資料。

Note

AWS Clean Rooms 協同合作已限制可執行和檢視查詢結果的人員。若要執行解密，有權存取這些結果的人員，都需要用來加密資料的相同共用私密金鑰和協同合作 ID。

解密加密的資料表

1. (選用) 在 [C3R 加密用戶端中檢視可用的命令](#)。
2. (選用) 導覽至所需的目錄並執行 `ls(macOS)` 或 `dir()`Windows。
 - 確認 `c3r-cli.jar` 檔案和加密的查詢結果資料檔案位於所需的目錄中。

Note

如果查詢結果是從 AWS Clean Rooms 主控台界面下載，則它們可能位於您使用者帳戶的下載資料夾中。(例如，Windows 和 macOS 上的使用者目錄中的下載資料夾 `macOS`。)建議您將查詢結果檔案移至與相同的資料夾 `c3r-cli.jar`。

3. 將共用私密金鑰存放在 `C3R_SHARED_SECRET` 環境變數中。如需詳細資訊，請參閱 [步驟 6：將共用私密金鑰存放在環境變數中](#)。
4. 從 AWS Command Line Interface (AWS CLI) 執行下列命令。

```
java -jar c3r-cli.jar decrypt <name of input .csv file> --id=<collaboration id> --
output=<output file name>
```

5. 將每個#####取代為您自己的資訊：

- a. 針對 id=，輸入協同合作 ID。
- b. 針對 output=，輸入輸出檔案的名稱（例如 results-decrypted.csv）。

如果您未指定輸出名稱，則預設名稱會顯示在終端機中。

- c. 使用您偏好的 CSV 或檢視應用程式（例如 Microsoft Excel、文字編輯器或其他應用程式），在指定的輸出檔案中Parquet檢視解密的資料。

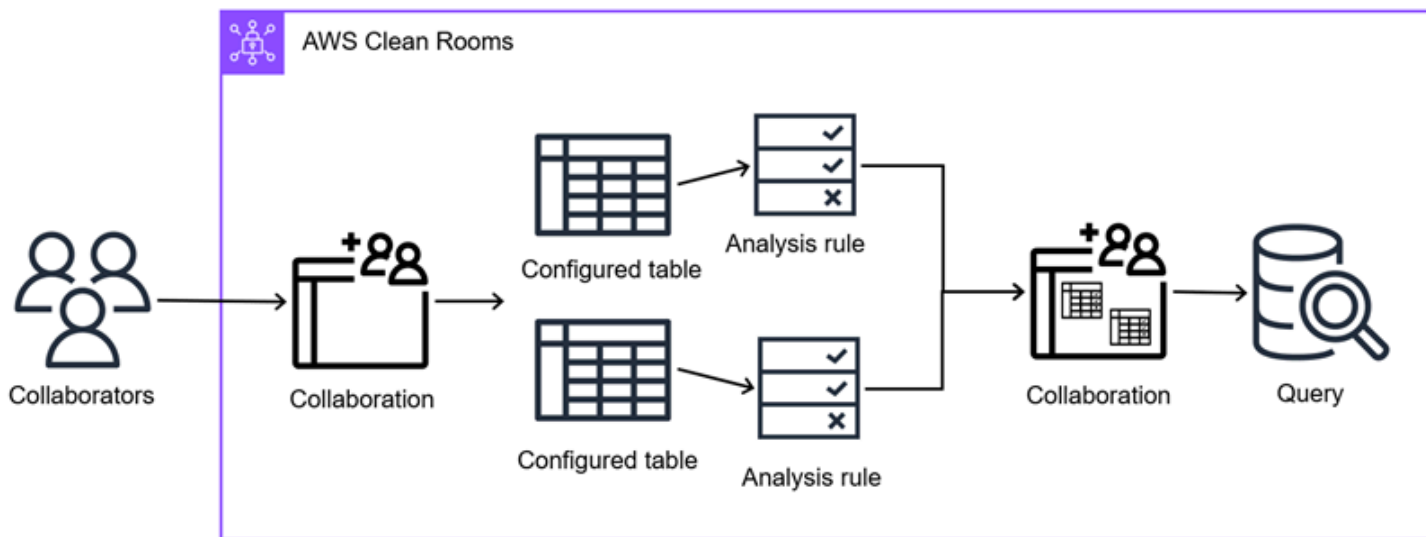
在 中設定的資料表 AWS Clean Rooms

設定的資料表是資料來源中現有資料表的參考。它包含分析規則，可決定如何在 中 AWS Clean Rooms 查詢資料，並可包含資料存取預算來控制資料表用量。設定的資料表可以與一或多個協同合作相關聯。

透過 AWS Clean Rooms，您可以對事件資料執行彙總分析，例如與購買次數相比的購買次數。您也可以對事件資料執行清單分析，例如擴充從區段資料到 CRM 資料的重疊客戶資料。您也可以對事件資料執行自訂查詢並設定差異隱私權，例如檢視資料和客群屬性。對於任何這些分析類型，您可以設定資料存取預算，以監控和控制透過查詢存取多少資料。

首先，您在 中建立協同合作，AWS Clean Rooms 並新增 AWS 帳戶 您要邀請的，或透過建立成員資格來加入您受邀的協同合作。接著，您和協同合作中的其他成員會建立設定的資料表。您都會將分析規則新增至設定的資料表（彙總、清單或自訂），並選擇性地設定資料存取預算。然後，您將設定的資料表與協同合作建立關聯。最後，可以查詢的成員會跨兩個資料表執行查詢，並在執行查詢時耗用資料存取預算。

下圖摘要說明如何在 中使用事件資料 AWS Clean Rooms。



主題

- [在 中建立設定的資料表 AWS Clean Rooms](#)
- [將分析規則新增至設定的資料表](#)
- [將已設定的資料表與協作建立關聯](#)
- [設定資料存取預算](#)
- [將協同合作分析規則新增至設定的資料表](#)

- [設定差異隱私權政策（選用）](#)
- [檢視資料表和分析規則](#)
- [編輯設定的資料表](#)
- [編輯設定的資料表標籤](#)
- [編輯設定的資料表分析規則](#)
- [刪除設定的資料表分析規則](#)
- [設定的資料表不允許的資料欄](#)
- [編輯設定的資料表關聯](#)
- [取消已設定資料表的關聯](#)

在 中建立設定的資料表 AWS Clean Rooms

設定的資料表是資料來源中現有資料表的參考。它包含分析規則，可決定如何查詢資料 AWS Clean Rooms。設定的資料表可以與一或多個協同合作相關聯。

如需如何使用 AWS SDKs 建立已設定資料表的詳細資訊，請參閱 [AWS Clean Rooms API 參考](#)。

主題

- [建立設定的資料表 – Amazon S3 資料來源](#)
- [建立設定的資料表 – Amazon Athena 資料來源](#)
- [建立設定的資料表 – Snowflake 資料來源](#)

建立設定的資料表 – Amazon S3 資料來源

在此程序中，[成員](#)會執行下列任務：

- 設定現有 AWS Glue 資料表以供使用 AWS Clean Rooms。（除非使用適用於 的密碼編譯運算，否則此步驟可以在加入協同合作之前或之後完成Clean Rooms。）

Note

AWS Clean Rooms 支援 AWS Glue 資料表。如需在 中取得資料的詳細資訊 AWS Glue，請參閱 [步驟 3：將資料表上傳至 Amazon S3](#)。

- 為[設定的資料表](#)命名，並選擇要在協同合作中使用的資料欄。

下列程序假設：

- 協同合作成員已將其資料表上傳至 [Amazon S3](#)，並 [建立 AWS Glue 資料表](#)。

 Note

Amazon S3 中的結果目的地不能與任何資料來源位於相同的 S3 儲存貯體中。

- (選用) 僅針對 [加密](#) 資料表，協同合作成員已使用 C3R [加密用戶端準備加密資料表](#)。

您可以使用提供的統計資料產生 AWS Glue 來運算 AWS Glue Data Catalog 資料表的資料欄層級統計資料。在 AWS Glue 產生 Data Catalog 中資料表的統計資料後，Amazon Redshift Spectrum 會自動使用這些統計資料來最佳化查詢計劃。如需使用運算資料欄層級統計資料的詳細資訊 AWS Glue，請參閱 AWS Glue 《使用者指南》中的 [使用資料欄統計資料最佳化查詢效能](#)。如需的詳細資訊 AWS Glue，請參閱 [AWS Glue 開發人員指南](#)。

建立設定的資料表 – Amazon S3 資料來源

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 在右上角，選擇設定新資料表。
4. 對於資料來源，在 AWS 資料來源下，選擇 Amazon S3。
5. 在 Amazon S3 資料表下：
 - a. 選取託管 S3 資料表的區域。

根據預設，會選取目前的區域（例如維吉尼亞北部 us-east-1）。

 Warning

當您的 Amazon S3 資料來源與處理位置位於不同的區域時，資料處理可能會在來源區域之外暫時進行。在繼續之前，請確認跨區域資料移動符合您的資料主權要求、法規合規政策和資料控管標準。

如需區域的詳細資訊，請參閱《》中的 [區域和端點](#) AWS 一般參考。

- b. 從下拉式清單中選擇資料庫。

- c. 從下拉式清單中選擇您要設定的資料表。

 Note

若要驗證這是正確的資料表，請執行下列其中一項操作：

- 選擇檢視 AWS Glue。
- 開啟從 檢視結構描述 AWS Glue 以檢視結構描述。

 Important

對於資料為 CSV 格式的 AWS Glue 資料表，Glue 結構描述中的資料欄名稱和順序必須完全符合 CSV 資料。如果不相符，可能無法正確強制執行已設定資料表的允許資料欄清單。

6. 對於協同合作中允許的資料欄和分析方法，
 - a. 對於您希望在協同合作中允許哪些資料欄？
 - 選擇所有資料欄，以允許在協同合作中查詢所有資料欄。
 - 選擇自訂清單，以允許在協同合作中查詢指定允許的資料欄下拉式清單中的一或多個資料欄。
 - b. 對於允許的分析方法，
 - i. 選擇直接查詢以允許 SQL 查詢直接在此資料表上執行
 - ii. 選擇直接任務以允許 PySpark 任務直接在此資料表上執行。

Example 範例

例如，如果您想要允許協同合作成員在所有資料欄上執行直接 SQL 查詢和 PySpark 任務，請選擇所有資料欄、直接查詢和直接任務。

7. 對於已設定的資料表詳細資訊，
 - a. 輸入已設定資料表的名稱。

您可以使用預設名稱或重新命名此資料表。

b. 輸入資料表的描述。

描述有助於區分具有類似名稱的其他已設定資料表。

8. 如果您想要為設定的資料表資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
9. 選擇設定新資料表。

現在您已建立已設定的資料表，即可：

- [將分析規則新增至設定的資料表](#)
- [將設定的資料表與協同合作建立關聯](#)

建立設定的資料表 – Amazon Athena 資料來源

Amazon Athena 資料來源選項可讓您查詢存放在 Amazon S3 中的資料、目錄在 AWS Glue 資料目錄或聯合目錄，以及透過控制的存取 AWS Lake Formation。同時支援資料表和 AWS Glue Data Catalog 檢視。Lake Formation 資源連結可用來跨成員 AWS 區域 AWS Clean Rooms 帳戶共用資料表 AWS 帳戶和檢視，並將其加入 AWS Clean Rooms 協同合作。

Note

只有 Amazon S3 型資料集可以透過 Athena 資料來源整合進行查詢。

在此程序中，[成員](#)會執行下列任務：

- 在 中設定現有的資料表或檢視 AWS Glue Data Catalog 以使用 AWS Clean Rooms
- 為[設定的資料表](#)命名，並選擇要在協同合作中使用的資料欄。

下列程序假設：

- 協同合作成員已建立 AWS Glue Data Catalog 資料庫和資料表或 GDC 檢視。

建立設定的資料表 – Athena 資料來源

1. 登入 AWS 管理主控台 並開啟位於 <https://console.aws.amazon.com/cleanrooms> 的 AWS Clean Rooms 主控台。

2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 在右上角，選擇設定新資料表。
4. 對於資料來源，在AWS 資料來源下，選擇 Amazon Athena。
5. 在 Amazon Athena 資料表下：

- a. 選取託管 Amazon Athena 資料表的區域。

根據預設，會選取目前的區域（例如維吉尼亞北部 us-east-1）。

 Warning

當您的 Amazon Athena 資料來源與處理位置位於不同的區域時，資料處理可能會在來源區域之外暫時進行。在繼續之前，請確認跨區域資料移動符合您的資料主權要求、法規合規政策和資料控管標準。

如需區域的詳細資訊，請參閱《》中的[區域和端點](#) AWS 一般參考。

- b. 從下拉式清單中選擇目錄。

根據預設，會選取 AWS Glue Data Catalog。

- AWS Glue Data Catalog – 中資料表的預設目錄 AWS Glue。
- 聯合目錄 – 如果您已設定 AWS Glue 目錄聯合來連線至遠端 Apache Iceberg REST 目錄，則可供使用。如需詳細資訊，請參閱《AWS Lake Formation 開發人員指南》中的[目錄聯合](#)。

- c. 從下拉式清單中選擇資料庫。
- d. 從下拉式清單中選擇您要設定的資料表。

 Note

若要驗證這是正確的資料表，請執行下列其中一項操作：

- 選擇在 中檢視 AWS Glue或在 中檢視 AWS Lake Formation（取決於您的目錄類型）。
- 開啟從 檢視結構描述 AWS Glue以檢視結構描述。

6. 對於 Amazon Athena 組態，

- a. 從下拉式清單中選擇工作群組。
- b. 針對 S3 輸出位置，根據下列其中一個案例選擇建議的動作。

案例	建議的動作
您的工作群組沒有預設的輸出位置。	輸入 S3 輸出位置，或選擇瀏覽 S3。
您的工作群組會強制執行您的預設輸出位置。	S3 輸出位置會自動選擇，您無法變更。
您的工作群組不會強制執行您的預設輸出位置。	輸入 S3 輸出位置，或選擇瀏覽 S3。

7. 對於協同合作中允許的欄，根據您的目標選擇一個選項。

您的目標	建議選項
允許所有資料欄用於 AWS Clean Rooms（視分析規則而定）	所有資料欄
從指定允許的資料欄下拉式清單中允許一或多個資料欄	自訂清單

8. 對於已設定的資料表詳細資訊，

- a. 輸入已設定資料表的名稱。

您可以使用預設名稱或重新命名此資料表。

- b. 輸入資料表的描述。

描述有助於區分具有類似名稱的其他已設定資料表。

- c. 如果您想要為設定的資料表資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。

9. 選擇設定新資料表。

現在您已建立已設定的資料表，即可：

- [將分析規則新增至設定的資料表](#)

- [將設定的資料表與協同合作建立關聯](#)

建立設定的資料表 – Snowflake 資料來源

在此程序中，[成員](#)會執行下列任務：

- 設定現有的 Snowflake 資料表以用於 AWS Clean Rooms。（除非使用適用於 的加密運算，否則此步驟可以在加入協同合作之前或之後完成Clean Rooms。）
- 為[設定的資料表](#)命名，並選擇要在協同合作中使用的資料欄。

下列程序假設：

- 協同合作成員已將其資料表上傳至 Snowflake。
- （選用）僅針對[加密](#)資料表，協同合作成員已使用 C3R [加密用戶端準備加密資料表](#)。

建立設定的資料表 – Snowflake 資料來源

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 在右上角，選擇設定新資料表。
4. 對於資料來源，在第三方雲端和資料來源下，選擇 Snowflake。
5. 使用現有的秘密 ARN 指定 Snowflake 登入資料，或儲存此資料表的新秘密。

Use existing secret ARN

1. 如果您有秘密 ARN，請在秘密 ARN 欄位中輸入它。

您可以選擇 Go to AWS Secrets Manager 來查詢秘密 ARN。

2. 如果您有來自另一個資料表的現有秘密，請選擇從現有資料表匯入秘密 ARN。

Note

秘密 ARN 可以是跨帳戶。

Store a new secret for this table

1. 輸入下列 Snowflake 登入資料：
 - Snowflake 使用者名稱
 - Snowflake 倉儲
 - Snowflake 角色
 - Snowflake 隱私權增強型郵件 (PEM) 私有金鑰
 2. 對於加密，請執行下列其中一項操作：
 - 若要使用 AWS 受管金鑰（預設），請清除自訂加密設定核取方塊。
 - 若要使用自訂 AWS KMS key：
 - 選取自訂加密設定核取方塊。
 - 對於 KMS 金鑰，輸入金鑰 ARN，或從清單中選擇一個金鑰 ARN。
 3. 輸入秘密名稱以協助您稍後尋找登入資料。
6. 對於 Snowflake 資料表和結構描述詳細資訊，請手動輸入詳細資訊或自動匯入詳細資訊。

Enter the details manually

1. 輸入 Snowflake 帳戶識別符。

如需詳細資訊，請參閱 Snowflake 文件中的[帳戶識別符](#)。

您的帳戶識別符必須採用 Snowflake 驅動程式所使用的格式。您需要以連字號 (-) 取代句點 (.)，讓識別符格式化為 **<orgname>-<account_name>**。

2. 輸入 Snowflake 資料庫。

如需詳細資訊，請參閱 [Snowflake 文件](#) 中的 Snowflake 資料庫。

3. 輸入 Snowflake 結構描述名稱。
4. 輸入 Snowflake 資料表名稱。

如需詳細資訊，請參閱 [Snowflake 文件中的了解 Snowflake 資料表結構](#)。

5. 針對結構描述，輸入資料欄名稱，然後從下拉式清單中選擇資料類型。
6. 選擇新增資料欄以新增更多資料欄。
 - 如果您選擇物件資料類型，請指定物件結構描述。

Example物件結構描述範例

```
name STRING,  
location OBJECT(  
  x INT,  
  y INT,  
  metadata OBJECT(uuid STRING)  
),  
history ARRAY(TEXT)
```

- 如果您選擇陣列資料類型，請指定陣列結構描述。

Example陣列結構描述範例

```
OBJECT(x INT, y INT)
```

- 如果您選擇映射資料類型，請指定映射結構描述。

Example範例映射結構描述

```
STRING, OBJECT(x INT, y INT)
```

Automatically import the details

1. 從 Snowflake 匯出您的 COLUMNS 檢視做為 CSV 檔案。

如需 Snowflake COLUMNS 檢視的詳細資訊，請參閱 Snowflake 文件中的 [COLUMNS 檢視](#)。

2. 選擇從檔案匯入以匯入 CSV 檔案，並指定任何其他資訊。

系統會自動匯入資料庫名稱、結構描述名稱、資料表名稱、資料欄名稱和資料類型。

- 如果您選擇物件資料類型，請指定物件結構描述。
- 如果您選擇陣列資料類型，請指定陣列結構描述。
- 如果您選擇映射資料類型，請指定映射結構描述。

3. 輸入 Snowflake 帳戶識別符。

如需詳細資訊，請參閱 Snowflake 文件中的 [帳戶識別碼](#)。

Note

只有中編製目錄的 S3 資料表 AWS Glue 才能自動擷取資料表結構描述。

7. 對於協同合作中允許的欄，根據您的目標選擇一個選項。

您的目標	建議選項
允許所有資料欄用於 AWS Clean Rooms（視分析規則而定）	所有資料欄
從指定允許的資料欄下拉式清單中允許一或多個資料欄	自訂清單

8. 對於已設定的資料表詳細資訊，

- a. 輸入已設定資料表的名稱。

您可以使用預設名稱或重新命名此資料表。

- b. 輸入資料表的描述。

描述有助於區分具有類似名稱的其他已設定資料表。

- c. 如果您想要為設定的資料表資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。

9. 選擇設定新資料表。

現在您已建立已設定的資料表，即可：

- [將分析規則新增至設定的資料表](#)
- [將設定的資料表與協同合作建立關聯](#)

將分析規則新增至設定的資料表

下列各節說明如何將分析規則新增至您設定的資料表。透過定義分析規則，您可以授權可查詢的成員執行符合支援之特定分析規則的查詢 AWS Clean Rooms。

AWS Clean Rooms 支援下列類型的分析規則：

- [彙總分析規則](#)
- [列出分析規則](#)
- [中的自訂分析規則 AWS Clean Rooms](#)

每個設定的資料表只能有一個分析規則。您可以在將設定的資料表與協同合作建立關聯之前，隨時設定分析規則。

Important

如果您針對 使用密碼編譯運算，Clean Rooms並在協同合作中加密資料表，則新增至加密設定資料表的分析規則應與資料加密方式一致。例如，如果您已加密 SELECT（彙總分析規則）的資料，則不應新增 JOIN（列出分析規則）的分析規則。

主題

- [將彙總分析規則新增至資料表（引導流程）](#)
- [將清單分析規則新增至資料表（引導流程）](#)
- [將自訂分析規則新增至資料表（引導流程）](#)
- [將分析規則新增至資料表 \(JSON 編輯器\)](#)
- [後續步驟](#)

將彙總分析規則新增至資料表（引導流程）

彙總分析規則允許彙總統計資料的查詢SUM，而不會沿著選用維度使用 COUNT、和 AVG函數顯示資料列層級資訊。

此程序說明使用 主控台中的 AWS Clean Rooms 引導流程選項，將彙總分析規則新增至設定資料表的程序。

Note

使用non-S3 資料來源設定的資料表僅支援[自訂分析規則](#)。

將彙總分析規則新增至資料表（引導流程）

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 選擇設定的資料表。
4. 在設定的資料表詳細資訊頁面上，選擇設定分析規則。
5. 在步驟 1：選擇分析規則類型下，在分析規則類型下，選擇彙總選項。
6. 在建立方法下，選取引導流程，然後選擇下一步。
7. 在步驟 2：為彙總函數指定查詢控制項：

- a. 從下拉式清單中選擇彙總函數：

- COUNT
- COUNT DISTINCT
- SUM
- SUM DISTINCT
- AVG

- b. 從資料欄下拉式清單中選擇可在彙總函數中使用的資料欄。

- c. （選用）選擇新增另一個函數以新增另一個彙總函數，並將一或多個資料欄與該函數建立關聯。

Note

至少需要一個彙總函數。

- d. （選用）選擇移除以移除彙總函數。

8. 對於聯結控制項，

- a. 選擇允許資料表自行查詢的一個選項：

如果選擇...	然後 ...
否，只能查詢重疊	只有在加入可查詢之成員擁有的資料表時，才能查詢資料表。

如果選擇...	然後 ...
是	資料表可以自行查詢，也可以在加入其他資料表時查詢。

- b. 在指定聯結資料欄下，選擇您要允許在INNERJOIN陳述式中使用的資料欄。

如果您在上一個步驟中選取是，這是選用的。

- c. 在指定允許相符的運算子下，選擇哪些運算子可用於多個聯結資料欄上的相符項目。如果您選取兩個或多個JOIN資料欄，則需要其中一個運算子。

如果選擇...	然後 ...
和	您可以在INNER JOIN比對條件AND中包含，將一個資料欄加入資料表之間的另一個資料欄。
或	您可以在INNER JOIN比對條件OR中包含，以在資料表之間合併多個資料欄比對。此邏輯運算子有助於取得更高的相符率。

9. （選用）對於維度控制項，在指定維度資料欄下拉式清單中，選擇要允許在 SELECT 陳述式中使用的資料欄，以及查詢的 WHERE、GROUP BY和 ORDERBY部分。

 Note

彙總函數或聯結資料欄無法用作維度資料欄。

10. 對於純量函數，請選擇您要允許哪些純量函數的一個選項？

如果選擇...	然後 ...
目前支援的所有 AWS Clean Rooms	您可以允許目前支援的所有純量函數 AWS Clean Rooms。 您可以選擇檢視清單，以查看 中支援的純量函數 AWS Clean Rooms完整清單。

如果選擇...	然後 ...
自訂清單	您可以自訂要允許哪些純量函數。 • 從指定允許的純量函數下拉式清單中選擇一或多個選項。
無	您不想允許任何純量函數。

如需詳細資訊，請參閱[純量函數](#)。

11. 選擇下一步。

12. 在步驟 3：指定查詢結果控制項下，針對彙總限制：

- a. 選取每個資料欄名稱的下拉式清單。
- b. 將COUNT DISTINCT函數套用到每個輸出資料列後，為每個必須符合的相異值數目下限，選取下拉式清單。
- c. 選擇新增限制條件以新增更多彙總限制條件。
- d. （選用）選擇移除以移除彙總限制。

13. 針對套用至輸出的其他分析，請根據您的目標選取選項。

您的目標	建議選項
僅允許此資料表上的直接查詢。拒絕在查詢結果上執行其他分析。資料表只能用於直接查詢。	不允許
允許，但不需要對此資料表進行直接查詢和其他分析。	允許
要求資料表只能用於使用其中一項必要額外分析處理的直接查詢。必須先進一步處理此資料表上的直接查詢，才能傳回。	必要

14. 選擇下一步。

15. 在步驟 4：檢閱和設定下，檢閱您為先前步驟所做的選擇，視需要編輯，然後選擇設定分析規則。

您會看到確認訊息，指出您已成功設定資料表的彙總分析規則。

將清單分析規則新增至資料表（引導流程）

清單分析規則允許輸出相關聯資料表與可查詢之成員資料表之間重疊資料列層級清單的查詢。

此程序說明使用 AWS Clean Rooms 主控台中的引導流程選項，將清單分析規則新增至已設定資料表的程序。

Note

使用non-S3 資料來源設定的資料表僅支援[自訂分析規則](#)。

將清單分析規則新增至資料表（引導流程）

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 選擇設定的資料表。
4. 在設定的資料表詳細資訊頁面上，選擇設定分析規則。
5. 在步驟 1：選擇分析規則類型下，在分析規則類型下，選擇清單選項。
6. 在建立方法下，選取引導流程，然後選擇下一步。
7. 在步驟 2：指定加入控制項的查詢控制項下：
 - a. 在指定聯結資料欄下，選擇您要允許在INNERJOIN陳述式中使用的資料欄。
 - b. 在指定允許相符的運算子下，選擇哪個運算子可用於多個聯結資料欄上的相符。如果您選取兩個或多個JOIN資料欄，則需要其中一個運算子。

如果選擇...	然後 ...
和	您可以在INNER JOIN比對條件AND中包含，將一個資料欄加入資料表之間的另一個資料欄。
或	您可以在INNER JOIN比對條件OR中包含，以在資料表之間合併多個資料欄比

如果選擇...	然後 ...
	對。此邏輯運算子有助於取得更高的相符率。

8. (選用) 對於清單控制項，在指定清單資料欄下拉式清單中，選擇您要允許在查詢輸出中使用的資料欄 (即在SELECT陳述式中使用)，或用於篩選結果 (即WHERE陳述式)。
9. 選擇下一步。
10. 在步驟 3：指定查詢結果控制項下，針對套用至輸出的其他分析，請根據您的目標選取選項。

您的目標	建議選項
僅允許此資料表上的直接查詢。拒絕在查詢結果上執行其他分析。資料表只能用於直接查詢。	不允許
允許，但不需要對此資料表進行直接查詢和其他分析。	允許
要求資料表只能用於使用其中一項必要額外分析處理的直接查詢。必須先進一步處理此資料表上的直接查詢，才能傳回。	必要

11. 在步驟 4：檢閱和設定下，檢閱您為先前步驟所做的選擇，視需要編輯，然後選擇設定分析規則。

您會看到確認訊息，指出您已成功設定資料表的清單分析規則。

將自訂分析規則新增至資料表 (引導流程)

自訂分析規則會在設定的資料表上啟用自訂 SQL 查詢或 PySpark 任務。如果您使用的是：

- [分析範本](#)，以允許一組特定的預先核准 SQL 查詢或 PySpark 任務，或一組可提供使用您資料之查詢的特定帳戶。
- 防止使用者識別嘗試[AWS Clean Rooms 的不同隱私權](#)。
- Non-S3 資料來源，例如 Amazon Athena 或 Snowflake。

此程序說明使用 AWS Clean Rooms 主控台內的引導流程選項，將自訂分析規則新增至設定資料表的程序。

將自訂分析規則新增至資料表（引導流程）

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 選擇設定的資料表。
4. 在設定的資料表詳細資訊頁面上，選擇設定分析規則。
5. 在步驟 1：選擇分析規則類型下，在分析規則類型下，選擇自訂選項。
6. 在建立方法下，選取引導流程，然後選擇下一步。
7. 在步驟 2：指定分析控制項下，針對直接分析控制項，根據您的目標選擇一個選項。

您的目標	建議的動作
檢閱每個新分析，然後才允許在此設定的資料表上執行	1. 在允許執行的分析範本下，選擇新增分析範本。 2. 從下拉式清單中選擇適當的協同合作和分析範本。 3. 選擇下一步。
允許特定協作者執行所選類型的任何分析，而無需檢閱此表格	1. 在分析類型下， a. 選擇任何查詢，以允許 AWS 帳戶 您指定的 建立的任何查詢。 b. 選擇任何查詢，以允許 AWS 帳戶 您指定的 建立的任何任務。 2. 在 AWS 帳戶 允許建立任何分析下，選擇新增 AWS 帳戶。 3. 輸入 AWS 帳戶 或從下拉式清單中選擇 AWS 帳戶 ID。 4. （選用）選擇新增另一個 AWS 帳戶 以新增另一個 AWS 帳戶。 5. 選擇下一步。

8. 在步驟 3：指定分析結果控制項下，
 - a. 對於任務結果控制項，請注意不支援其他結果控制項。

- b. 在查詢結果控制項下，對於輸出中不允許的資料欄，根據您的目標選擇要在查詢輸出中允許的資料欄。

您的目標	建議的動作
允許在查詢輸出中傳回所有資料欄	1. 選擇無 2. 繼續進行套用至輸出的其他分析。
不允許在查詢輸出中傳回特定資料欄	1. 選擇自訂清單 2. 在指定不允許的資料欄下，選擇您要從查詢輸出中移除的資料欄。

- c. 針對套用至輸出的其他分析，根據您的目標，選擇是否可以將其他分析套用至查詢輸出。

您的目標	建議選項
• 僅允許此資料表上的直接查詢。 • 拒絕在查詢結果上執行其他分析。 • 資料表只能用於直接查詢。	不允許
允許，但不需要對此資料表進行直接查詢和其他分析。	Allow
• 要求資料表只能用於使用其中一項必要額外分析處理的直接查詢。 • 必須先進一步處理此資料表上的直接查詢，才能傳回。	必要

- d. 選擇下一步。

9. (選用) 在步驟 4：設定差異隱私權下，判斷您是否要開啟或關閉差異隱私權。

差異隱私權是一種經過數學驗證的技術，可保護您的資料免於重新識別攻擊。

 Note

AWS Clean Rooms 差異隱私權僅適用於資料存放在 Amazon S3 中的協同合作。

對於差異隱私權，根據您的目標選擇是否開啟或關閉差異隱私權。

您的目標	建議的動作
- 您不需要防止重新識別嘗試 - 您的資料表沒有使用者層級資料	- 選擇關閉。 - 選擇下一步。
- 您需要防止重新識別嘗試 - 您的資料表具有使用者層級資料	- 選擇 Turn on (開啟)。 - 選取使用者識別符欄，其中包含您要保護其隱私權的使用者唯一識別符，例如 <code>user_id</code> 欄。 若要開啟協同合作中兩個或多個資料表的差異隱私權，您必須在兩個分析規則中設定與使用者識別符資料欄相同的資料欄，以維持資料表之間使用者的一致定義。如果組態錯誤，可以查詢的成員會收到錯誤訊息，指出有兩個資料欄可供選擇，以便在執行查詢時計算使用者貢獻的數量（例如，使用者所做的廣告曝光次數）。 - 選擇下一步。

10. 在步驟 5：檢閱和設定下，檢閱您為先前步驟所做的選擇，視需要編輯，然後選擇設定分析規則。

您會看到確認訊息，指出您已成功設定資料表的自訂分析規則。

將分析規則新增至資料表 (JSON 編輯器)

下列程序說明如何使用 AWS Clean Rooms 主控台中的 JSON 編輯器選項，將分析規則新增至資料表。

Note

使用 non-S3 資料來源設定的資料表僅支援 [自訂分析規則](#)。

將彙總、清單或自訂分析規則新增至資料表 (JSON 編輯器)

1. 登入 AWS 管理主控台 並開啟位於 <https://console.aws.amazon.com/cleanrooms> 的 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 選擇設定的資料表。
4. 在設定的資料表詳細資訊頁面上，選擇設定分析規則。
5. 在步驟 1：選擇分析規則類型下，在分析規則類型下，選擇彙總、清單或自訂選項。
6. 在建立方法下，選取 JSON 編輯器，然後選擇下一步。
7. 在步驟 2：指定控制項下，您可以選擇插入查詢結構 (插入範本) 或插入檔案 (從檔案匯入)。

如果選擇...	然後 ...
插入範本	1. 在分析規則定義中指定所選分析規則的參數。 2. 您可以按 Ctrl + 空格鍵來啟用自動完成。 如需彙總分析規則參數的詳細資訊，請參閱 彙總分析規則 - 查詢控制項。 如需列出分析規則參數的詳細資訊，請參閱 列出分析規則 - 查詢控制項。
從 檔案匯入	1. 從本機磁碟機選取您的 JSON 檔案。 2. 選擇 Open (開啟)。 分析規則定義會顯示上傳檔案的分析規則。

8. 選擇下一步。
9. 在步驟 3：檢閱和設定下，檢閱您為先前步驟所做的選擇，視需要編輯，然後選擇設定分析規則。

您會收到確認訊息，指出您已成功設定資料表的分析規則。

後續步驟

現在您已將分析規則設定為已設定的資料表，您就可以：

- [將設定的資料表與協同合作建立關聯](#)
- [查詢資料表](#) (做為可查詢的成員)

將已設定的資料表與協作建立關聯

在您建立已設定的資料表並將其新增分析規則之後，您可以將其與協同合作建立關聯 AWS Clean Rooms，並讓服務角色存取您的 AWS Glue 資料表。

Note

此服務角色具有資料表的許可。服務角色只能由 擔任 AWS Clean Rooms，以代表可查詢的成員執行允許的查詢。協作成員（資料擁有者除外）無法存取協作中的基礎資料表。資料擁有者可以開啟差異隱私權，使其資料表可供其他成員查詢。

資料存取預算

將設定的資料表建立關聯時，您可以套用資料存取預算。資料存取預算控制資料表在協同合作中可用於查詢、任務和 ML 輸入通道的次數。這些預算透過限制資料表使用，協助組織管理資源使用率和控制成本。

每次在查詢、任務或 ML 輸入通道中使用資料表時，該資料表的預算都會減少一個。當預算達到零時，資料表無法用於 SQL 查詢、Pyspark 任務，也無法做為衍生自資料表的 ML 輸入通道的一部分。

您可以建立每個期間定期重新整理的預算、整體用量的生命週期預算，或兩者。根據預設，資料表用量不受限制。

- 每個期間預算 – 可限制在指定期間內使用此資料表之時間量的可再生能源配置。您可以將期間設定為每日、每週或每月。此預算可設定為每天、每週或每月自動重新整理。
- 生命週期預算 – 執行中的配置，限制此資料表可使用的總次數。

關聯已設定的資料表

下列主題說明如何將已設定的資料表建立關聯，並使用 主控台將 AWS Clean Rooms 資料存取預算套用到協同合作。

如需有關如何使用 AWS SDKs 將已設定的資料表與協同合作建立關聯的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

步驟 1：完成先決條件

若要關聯設定的資料表，您必須完成下列先決條件：

- 指向 Amazon S3 資料夾位置的 AWS Glue 資料表（非單一檔案）
- 對於加密 AWS Glue 的資料表：
 - 具有使用 AWS KMS 金鑰解密 AWS Glue 資料表之許可的服務角色
 - 針對 AWS KMS 加密的 Amazon S3 資料集：服務角色也必須具有使用 AWS KMS 金鑰解密 Amazon S3 資料的許可

如需有關設定加密的資訊，請參閱《AWS Glue 開發人員指南》中的[在中設定加密 AWS Glue](#)。

若要驗證 AWS Glue 資料表位置：

1. 在 <https://console.aws.amazon.com/glue/> 開啟 AWS Glue 主控台
2. 檢視資料表詳細資訊並確認位置指向 S3 資料夾

步驟 2：關聯已設定的資料表

建立已設定資料表的關聯

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 選擇要與資料表建立關聯的方法：
 - a. 從設定的資料表詳細資訊頁面：
 - i. 在左側導覽窗格中，選擇 Tables (資料表)。
 - ii. 選擇設定的資料表。
 - iii. 在設定的資料表詳細資訊頁面上，選擇關聯以協同合作。
 - iv. 針對關聯資料表至協同合作對話方塊，從下拉式清單中選擇協同合作。
 - b. 從協同合作詳細資訊頁面：
 - i. 在左側導覽窗格中，選擇協同合作。
 - ii. 選擇協同合作。
 - iii. 在資料表索引標籤上，選擇關聯資料表。

3. 在關聯資料表頁面上，執行下列其中一項操作：

- 選擇現有的已設定資料表 – 從下拉式清單中選擇您要與協同合作建立關聯的已設定資料表名稱。
- 設定新資料表 – 選擇設定新資料表，然後遵循設定新資料表頁面上的提示操作。
- 檢視已設定資料表的結構描述和分析規則 – 開啟檢視結構描述和分析規則。

4. 如需資料表關聯詳細資訊，

a. 輸入關聯資料表的名稱。

您可以使用預設名稱或重新命名此資料表。

b. (選用) 輸入資料表的描述。

描述有助於撰寫查詢。

5. 選取建立和使用新的服務角色或使用現有的服務角色，以指定服務存取許可。

Note

如果您要關聯由 Amazon Athena 支援的已設定資料表，請從下拉式清單中選擇現有服務角色名稱。確保服務角色具有 IAM，並視需要具有資料集的 Lake Formation 許可。

如果選擇...	然後 ...
建立和使用新的服務角色	• AWS Clean Rooms 會建立具有此資料表所需政策的服務角色。 • 預設的服務角色名稱為 <code>cleanrooms- <timestamp></code> • 您必須具有建立角色和連接政策的許可。 • 如果您的輸入資料已加密，您可以選取 此資料是使用 KMS 金鑰加密，然後輸入將用於解密您資料輸入的 AWS KMS key。
使用現有的服務角色	1. 從下拉式清單中選擇現有的服務角色名稱。 如果您具有列出角色的許可，則會顯示角色清單。

如果選擇...	然後 ...
	如果您沒有列出角色的許可，您可以輸入要使用的角色的 Amazon Resource Name (ARN)。 2. 選擇 IAM 外部連結中的檢視，以檢視服務角色。 如果沒有現有的服務角色，則無法使用使用現有服務角色的選項。 根據預設，AWS Clean Rooms 不會嘗試更新現有的角色政策來新增必要的許可。 3. (選用) 選取新增具有此角色必要許可的預先設定政策核取方塊，以將必要許可新增至角色。您必須擁有修改角色和建立政策的許可。

Note

- AWS Clean Rooms 需要根據分析規則查詢的許可。如需許可的詳細資訊 AWS Clean Rooms，請參閱 [AWS 的 受管政策 AWS Clean Rooms](#)。
- 如果角色沒有足夠的許可 AWS Clean Rooms，您會收到錯誤訊息，指出角色沒有足夠的許可 AWS Clean Rooms。必須先新增角色政策，才能繼續。
- 如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

6. 如果您想要為已設定的資料表關聯資源啟用已設定的資料表關聯標籤，請選擇新增標籤，然後輸入金鑰和值對。
7. 選擇下一步。
8. 在設定協同合作分析規則頁面上，選擇下列其中一項：
 - 是，立即建立協同合作分析規則 – 將您的資料表與此協同合作建立關聯，並建立協同合作分析規則

- 否，我稍後會建立協作分析規則 – 僅將您的資料表與此協作建立關聯。您可以稍後建立協同合作分析規則。
9. 如果您選擇是，請現在建立協作分析規則，對於結果交付，請選擇允許從下拉式清單接收查詢輸出結果的成員。
 10. 選擇下一步。
 11. 在新增資料存取預算頁面上，針對資料存取預算組態，選擇下列其中一項：
 - 是，立即新增資料存取預算 – 將您的資料表與此協同合作建立關聯，並新增資料存取預算。您可以選取期間預算、生命週期預算或兩者。
 - 否，我稍後會新增資料存取預算 – 僅將您的資料表與此協同合作建立關聯。您可以稍後新增資料存取預算。

如果您選取否，我稍後會新增資料存取預算，請跳到步驟 15。

12. 如果您選擇是，請立即新增資料存取預算，請選擇下列其中一個預算組態：

僅限每個期間預算	僅限生命週期預算	每個期間和生命週期預算
1. 保持選取每個期間新增預算。 2. 輸入介於 1 到 1,000,000 之間的每期期間預算金額。 3. 針對期間，選擇每日、每週或每月。 4. (選用) 保持每週選取自動重新整理預算以續約配置。 5. 清除新增生命週期預算。	1. 清除每期間新增預算。 2. 選取新增生命週期預算。 3. 輸入介於 1 到 1,000,000 之間的終身預算金額。	1. 保留選取每個期間新增預算。 2. 輸入介於 1 到 1,000,000 之間的每期期間預算金額。 3. 針對期間，選擇每日、每週或每月。 4. 保持選取每週自動重新整理預算。 5. 選取新增生命週期預算。 6. 輸入介於 1 到 1,000,000 之間的終身預算金額。

13. 在資料存取預算摘要下檢閱您的選擇。

Example範例

例如，如果您已選擇每期間預算金額 1,000、將期間設定為每週、讓自動重新整理預算每週核取方塊保持選取狀態，並將生命週期預算設定為 1,000,000，則存取預算摘要會顯示下列訊息：每週最多可以使用此表格執行查詢或任務 1,000 次。此預算設定為每週日 00:00 UTC 自動重新整理，並繼續重新整理，直到此資料表已達到其 1,000,000 次使用的生命週期預算為止。

14. (選用) 如果您想要為存取預算資源啟用資料存取預算標籤，請選擇新增標籤，然後輸入金鑰和值對。
15. 選擇下一步。
16. 檢閱檢閱和建立頁面上的資訊。
 - a. 如果您需要編輯任何區段，請選擇編輯。
 - b. 編輯您的組態，然後選擇下一步。
17. 選擇關聯資料表。

步驟 3：後續步驟

現在您已將設定的資料表與協同合作建立關聯，您已準備好：

- 將[協同合作分析規則](#)新增至設定的資料表
- 如果您是[協同合作建立者](#)，請編輯協同合作
- [查詢資料表](#) (做為可查詢的成員)

設定資料存取預算

協作者可以檢視、新增、編輯和刪除資料存取預算，以設定資料表可用於工作流程的次數限制。使用這些預算來管理資料和成本。

每次查詢資料表或使用衍生自資料表的 ML 輸入通道執行任務時，該資料表的預算都會減少一個。當預算達到零時，就無法查詢資料表，也無法使用衍生自資料表的 ML 輸入通道來執行 ML 任務。

您可以建立每個期間定期重新整理的預算、整體用量的生命週期預算，或兩者。根據預設，資料表用量不受限制。

- 每個期間預算 – 可限制在指定期間內使用此資料表之時間量的可再生能源配置。您可以將期間設定為每日、每週或每月。此預算可設定為每天、每週或每月自動重新整理。

- 生命週期預算 – 執行中的配置，限制此資料表可使用的總次數。

主題

- [檢視資料存取預算](#)
- [將資料存取預算新增至現有的關聯資料表](#)
- [編輯資料存取預算](#)
- [刪除資料存取預算](#)

檢視資料存取預算

您可以從資料表索引標籤或資料表詳細資訊頁面檢視資料存取預算。

檢視資料存取預算

1. 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇 Tables (資料表) 索引標籤。
5. 執行以下任意一項：
 - 在剩餘資料存取預算欄下，選取預算以檢視詳細資訊。
 - 選擇資料表，然後在資料表詳細資訊頁面上，向下捲動以檢視資料存取預算詳細資訊區段。

將資料存取預算新增至現有的關聯資料表

身為協同合作成員，您可以將資料存取預算新增至現有的關聯資料表。

將資料存取預算新增至現有的關聯資料表

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。

4. 選擇 Tables (資料表) 索引標籤。
5. 選取您要新增資料存取預算之資料表旁的選項按鈕。
6. 從動作下拉式清單的資料存取預算下，選取新增（如果還沒有預算）。
7. 選擇下列其中一個預算組態：

僅限每個期間預算	僅限生命週期預算	每個期間和生命週期預算
1. 保留選取每個期間新增預算。 2. 輸入介於 1 到 1,000,000 之間的每期間預算金額。 3. 針對期間，選擇每日、每週或每月。 4. （選用）保持每週選取自動重新整理預算以續約配置。 5. 清除新增生命週期預算。	1. 清除每期間新增預算。 2. 選取新增生命週期預算。 3. 輸入介於 1 到 1,000,000 之間的終身預算金額。	1. 保留選取每個期間新增預算。 2. 輸入介於 1 到 1,000,000 之間的每期間預算金額。 3. 針對期間，選擇每日、每週或每月。 4. 保持選取每週自動重新整理預算。 5. 選取新增生命週期預算。 6. 輸入介於 1 到 1,000,000 之間的終身預算金額。

8. 在資料存取預算摘要下檢閱您的選擇。
9. Example 範例

例如，如果您已選擇每期間預算金額 1,000、將期間設定為每週、讓自動重新整理預算每週核取方塊保持選取狀態，並將生命週期預算設定為 1,000,000，則存取預算摘要會顯示下列訊息：每週最多可以使用此表格執行查詢或任務 1,000 次。此預算設定為每週日 00:00 UTC 自動重新整理，並繼續重新整理，直到此資料表已達到其 1,000,000 次使用的生命週期預算為止。

10. （選用）如果您想要為存取預算資源啟用資料存取預算標籤，請選擇新增標籤，然後輸入金鑰和值對。
11. 選擇新增資料存取預算。

編輯資料存取預算

身為協同合作成員，您可以編輯資料存取預算。當您編輯資料存取預算時，它會重設目前的預算餘額。

您可以從資料表索引標籤或資料表詳細資訊頁面編輯資料存取預算。

Tables tab

從資料表索引標籤編輯資料存取預算

1. 登入 AWS 管理主控台 並在 [https : //https://console.aws.amazon.com/cleanrooms](https://console.aws.amazon.com/cleanrooms) 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇 Tables (資料表) 索引標籤。
5. 選取您要編輯之資料表旁的選項按鈕。
6. 從動作下拉式清單的資料存取預算下，選取編輯。
7. 在編輯資料存取預算頁面上，更新每個期間預算或生命週期預算資訊。
8. 檢視資料存取預算摘要，以確認您所做的編輯正確無誤。
9. 選擇儲存變更。

Table details page

從資料表詳細資訊頁面編輯資料存取預算

1. 登入 AWS 管理主控台 ，並在 [https : //https://console.aws.amazon.com/cleanrooms](https://console.aws.amazon.com/cleanrooms) 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇 Tables (資料表) 索引標籤。
5. 選擇 表格。
6. 在資料表詳細資訊頁面上，向下捲動至資料存取預算詳細資訊區段。
7. 從動作下拉式清單中，選擇編輯。
8. 在編輯資料存取預算頁面上，更新每個期間預算或生命週期預算資訊。
9. 選擇儲存變更。

刪除資料存取預算

您可以從資料表索引標籤或資料表詳細資訊頁面刪除資料存取預算。

Tables tab

從資料表索引標籤刪除資料存取預算

1. 登入 AWS 管理主控台 並在 [https : //https://console.aws.amazon.com/cleanrooms](https://console.aws.amazon.com/cleanrooms) 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇 Tables (資料表) 索引標籤。
5. 選取您要刪除之資料表旁的選項按鈕。
6. 從動作下拉式清單的資料存取預算下，選取刪除。

Important

您無法復原此動作，您的資料存取預算將重設為無限制。

7. 如果您確定要刪除資料存取預算，請選擇刪除。

Table details page

從資料表詳細資訊頁面刪除資料存取預算

1. 登入 AWS 管理主控台，並在 [https : //https://console.aws.amazon.com/cleanrooms](https://console.aws.amazon.com/cleanrooms) 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇 Tables (資料表) 索引標籤。
5. 選擇 表格。
6. 在資料表詳細資訊頁面上，向下捲動至資料存取預算詳細資訊區段。
7. 從動作下拉式選單中，選擇刪除。

 Important

您無法復原此動作，您的資料存取預算將重設為無限制。

8. 如果您確定要刪除資料存取預算，請選擇刪除。

將協同合作分析規則新增至設定的資料表

協同合作分析規則可讓您指定此協同合作特有的控制項。這些控制項與設定的資料表分析規則搭配使用，以決定如何在此協同合作中分析此資料表。

建立已設定的資料表、[新增分析規則並將其與協同合作建立關聯之後](#)，您可以將協同合作分析規則新增至[已設定的資料表](#)。[將已設定的資料表與協作建立關聯](#)如果資料表設定為支援直接分析或允許其他分析，則需要新增協同合作分析規則。

- 直接分析 – 資料表可用於直接分析它的查詢。例如，在輸出彙總測量分析或用於啟用之識別符清單的查詢中。
- 其他分析 – 除了直接分析的查詢之外，資料表也可以用作其他分析的輸入。例如，資料表可用於類似 ML 模型的種子查詢，或自訂 ML 模型的 ML 輸入通道。

將協同合作分析規則新增至資料表

1. 登入 AWS 管理主控台 並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在資料表索引標籤的與您相關聯的資料表下，檢視您已與協同合作相關聯的已設定資料表。
 - 如果直接分析狀態或其他分析狀態為就緒，則資料表已準備好進行查詢。
 - 如果直接分析狀態或其他分析狀態的狀態為未就緒，請選取狀態，然後在對話方塊中選擇設定。
5. 在設定協同分析規則頁面上，展開檢視設定的資料表分析規則以檢視詳細資訊。
6. 針對允許的額外分析，根據您的目標選擇 選項。

您的目標	建議選項
允許對資料表進行任何額外的分析。	任何
僅允許特定成員對資料表進行其他分析。	特定成員的任何
僅允許對資料表進行特定分析。	自訂清單

- 對於結果交付，指定誰可以接收允許接收查詢輸出結果的成員的結果下拉式清單。
- 選擇設定分析規則。

設定差異隱私權政策（選用）

Note

AWS Clean Rooms 差異隱私權僅適用於資料存放在 Amazon S3 中的協同合作。

此程序說明使用 AWS Clean Rooms 主控台中的引導流程選項，在協同合作中設定差異隱私權政策的程序。這是具有差異隱私權保護的所有資料表的一次性步驟。

設定差異隱私權設定（引導流程）

- 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
- 在左側導覽窗格中，選擇協同合作。
- 選擇協同合作。
- 在協同合作頁面的資料表索引標籤上，選擇設定差異隱私權政策。
- 在設定差異隱私權政策頁面上，選擇下列屬性的值：
 - 隱私權預算
 - 每月重新整理隱私權預算
 - 每個查詢新增的雜訊

您可以使用預設值或輸入支援特定使用案例的自訂值。選擇每個查詢新增的隱私權預算和雜訊值後，您可以根據資料上所有查詢中可能的彙總數目來預覽產生的公用程式。

6. 選擇設定。

您將看到確認訊息，指出您已成功為協同合作設定差異隱私權政策。

現在您已設定差異隱私權，您已準備好：

- [查詢資料表](#)（做為可查詢的成員）
- [協同合作](#)（如果您是協同合作建立者）

檢視差異隱私權用量日誌

身為使用差異隱私權保護資料的協同合作成員，在建立與差異隱私權的協同合作之後，您可以監控隱私權預算的使用情況。

檢視執行了多少彙總，以及使用了多少隱私權預算

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇 Tables (資料表) 索引標籤。
5. 選擇檢視用量日誌（藍色文字）。
6. 檢視用量詳細資訊，包括隱私權預算和提供的公用程式數量。

編輯差異隱私權政策

設定差異隱私權政策之後，您可以隨時更新該政策，以更好地反映您的隱私權需求。

編輯差異隱私權政策

1. 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。

2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在協同合作頁面的資料表索引標籤上，在您相關聯的資料表下，選擇編輯。
5. 在編輯差異隱私權頁面上，選擇下列屬性的新值：
 - 隱私權預算 – 移動滑桿以在協同合作期間的任何時候增加或減少預算。在可以查詢的成員開始查詢您的資料之後，您無法減少預算。如果隱私權預算增加，AWS Clean Rooms 會繼續使用現有的預算，直到完全用盡，再使用新增的隱私權預算。
 - 每個查詢新增的雜訊 – 在協同合作期間的任何時間點，移動滑桿來增加或減少每個查詢新增的雜訊。

Note

您可以選擇互動式範例，探索每個查詢新增的不同隱私權預算和雜訊值如何影響您可以執行的彙總函數數量。

您無法變更隱私權預算重新整理的值。若要變更您的選擇，您必須刪除差異隱私權政策，並建立新的政策。

6. 選擇儲存變更。

您會看到確認訊息，指出您已成功編輯差異隱私權政策。

刪除差異隱私權政策

您可以從協同合作的資料表索引標籤中刪除差異隱私權政策。

刪除差異隱私權政策

1. 登入 AWS 管理主控台 並開啟位於 <https://console.aws.amazon.com/cleanrooms> 的 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在協同合作頁面的資料表索引標籤上，選取差異隱私權政策旁的刪除。
5. 如果您確定要刪除差異隱私權政策，請選擇刪除。

刪除差異隱私權政策後，您無法從該政策存取隱私權預算用量日誌。如果刪除差異隱私權政策，則無法查詢開啟差異隱私權的資料表。

檢視計算的差異隱私權參數

對於具備差異隱私權專業知識的使用者，您可以從協同合作的分析索引標籤檢視計算的差異隱私權參數。

檢視計算的差異隱私權參數

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在分析索引標籤的結果區段中，選取檢視計算的差異隱私權參數。

在計算差異隱私權參數表中，您可以看到彙總函數的敏感度值，其定義為在新增、移除或修改單一使用者的記錄時，函數結果可以變更的最大數量。此清單包含下列差異隱私權參數：

- 使用者貢獻限制 (UCL) 是 SQL 查詢中使用使用者貢獻的最大資料列數。例如，如果您想要計算每個使用者可以有多個曝光的指定行銷活動中的相符曝光總數，AWS Clean Rooms 則差異隱私權需要限制單一使用者的曝光數，以確保差異隱私權計算的準確性。換言之，如果任何使用者具有比邊界更多的印模，則會根據計算的 UCL 值 AWS Clean Rooms 自動取得該使用者印模的統一隨機樣本，並在執行查詢時排除該使用者的剩餘印模。如果您計算唯一使用者的數量，UCL 值等於 1。這是因為新增、移除或修改單一使用者最多可以將不同使用者的計數變更為 1。
- 最小值是彙總函數中使用的表達式下限，例如 `sum()`。例如，如果表達式是稱為 `purchase_value` 的資料欄，則最小值是資料欄的下限。
- 最大值是彙總函數中使用的表達式上限，例如 `sum()`。例如，如果表達式是稱為 `purchase_value` 的資料欄，則最大值是資料欄的上限。

在計算差異隱私權參數表中，您可以使用這些參數來更好地了解查詢結果中的雜訊總量。例如，當每個查詢新增的已設定雜訊為 30 個使用者並執行 `COUNT DISTINCT (user_id)` 查詢時，AWS Clean Rooms 差異隱私會新增介於 -30 和 30 之間的隨機雜訊，可能性很高，因為的敏感度 `COUNT DISTINCT` 為 1。在具有相同組態的 `COUNT` 查詢中，AWS Clean Rooms 差異隱私權會新增由使用者貢獻限制擴展的統計雜訊，因為單一使用者可以將多個資料列貢獻給查詢結果。在像是 `SUM (purchase_value)` 所有資料欄值都是正值的 `SUM` 查詢中，總雜訊會依使用者貢獻限制乘以最大值來

擴展。AWS Clean Rooms 差異隱私權會自動計算敏感參數，以在查詢執行時間執行雜訊新增，並耗盡隱私權預算。隱私權預算的耗盡是必要的，因為敏感參數取決於資料。

檢視資料表和分析規則

檢視與協同合作和分析規則相關聯的資料表

1. 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇 Tables (資料表) 索引標籤。
5. 選擇下列其中一項：
 - a. 若要檢視與協同合作相關聯的資料表，請針對與您相關聯的資料表選擇資料表（藍色文字）。
 - b. 若要檢視與協同合作相關聯的其他資料表，針對與協同合作者相關聯的資料表，請選擇資料表（藍色文字）。
6. 在資料表詳細資訊頁面上檢視資料表詳細資訊和分析規則。

編輯設定的資料表

事前準備：

- AWS 帳戶 可存取 的 AWS Clean Rooms

下列各節說明如何編輯 Amazon S3、Amazon Athena 和 Snowflake 資料來源資料表的名稱、描述和組態詳細資訊。

如需如何使用 AWS SDKs 編輯已設定資料表的詳細資訊，請參閱 [AWS Clean Rooms API 參考](#)。

編輯設定的資料表

1. 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。

3. 選擇您建立的已設定資料表。
4. 在設定的資料表詳細資訊頁面上，選擇編輯。
5. 編輯您的組態。
6. 選擇儲存變更。

編輯設定的資料表標籤

身為協同合作成員，在建立已設定的資料表之後，您可以在已設定的資料表索引標籤上管理已設定的資料表資源上的標籤。

編輯設定的資料表標籤

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 選擇您建立的已設定資料表。
4. 在設定的資料表詳細資訊頁面上，向下捲動至標籤區段。
5. 選擇管理標籤。
6. 在 Manage tags (管理標籤) 頁面上，可以執行下列操作：
 - 若要移除標籤，請選擇 Remove (移除)。
 - 若要新增標籤，請選擇 Add new tag (新增新標籤)。
 - 若要儲存您所做的變更，請選擇 Save changes (儲存變更)。

編輯設定的資料表分析規則

編輯設定的資料表分析規則

1. 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 選擇您建立的已設定資料表。
4. 在設定的資料表詳細資訊頁面上，向下捲動至彙總分析規則、清單分析規則或自訂分析規則區段。
(您的選擇取決於您為設定的資料表選擇的分析規則類型。)

5. 選擇編輯。
6. 在編輯分析規則頁面上，您可以：
 - 透過下列方式修改分析規則定義：
 - 修改 JSON 編輯器。
 - 選擇從檔案匯入以上傳新的分析規則定義。
 - 從下列選項中選取，預覽成員將在協同合作中看到的內容：
 - 資料表檢視
 - JSON
 - 查詢範例
7. 選擇儲存變更，以儲存您所做的變更。

刪除設定的資料表分析規則

Warning

此動作無法復原並影響所有相關資源。

刪除設定的資料表分析規則

1. 登入 AWS 管理主控台 並開啟位於 <https://console.aws.amazon.com/cleanrooms> 的 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 選擇您建立的已設定資料表。
4. 在設定的資料表詳細資訊頁面上，向下捲動至彙總分析規則、清單分析規則或自訂分析規則區段。
(您的選擇取決於您為設定的資料表選擇的分析規則類型。)
5. 選擇 刪除。
6. 如果您確定要刪除分析規則，請選擇刪除。

設定的資料表不允許的資料欄

不允許的輸出資料欄組態是 AWS Clean Rooms 自訂分析規則中的控制項，可讓您定義不允許在查詢結果中投影的資料欄清單 (如果有的話)。此清單中參考的資料欄會被視為「不允許的輸出資

料欄」。這表示透過轉換、別名或其他方式對此類資料欄的任何參考，可能不會出現在查詢的最終 SELECT（投影）中。

雖然功能禁止資料欄直接投影在輸出中，但無法完全防止透過其他機制間接推斷基礎值。這些欄仍然可用於投影子句（例如子查詢或通用資料表表達式 (CTE)），只要它們在最終投影中沒有參考即可。

不允許的輸出資料欄組態可讓您靈活地在資料表上套用和編纂控制，並結合根據使用案例和對應的隱私權需求的分析範本層級檢閱。

如需如何設定此組態的詳細資訊，請參閱 [將自訂分析規則新增至資料表（引導流程）](#)。

範例

下列範例顯示如何套用不允許的輸出資料欄控制。

- 成員 A 正在與成員 B 合作。
- 成員 B 是可執行查詢的成員。
- 成員 A 使用欄年齡、性別、電子郵件和名稱來定義資料表使用者。資料欄存留期和名稱是不允許的輸出資料欄。
- 成員 B 使用類似的資料欄存留期、性別和 owner_name 來定義資料表寵物。不過，它們不會對輸出資料欄設定任何限制，這表示資料表中的所有資料欄都可以在查詢中自由投影。

如果成員 B 執行下列查詢，則會遭到封鎖，因為無法直接投影不允許的輸出資料欄：

```
SELECT
  age
FROM
  users
```

如果成員 B 執行下列查詢，則會遭到封鎖，因為無法透過專案星隱含投影不允許的輸出資料欄：

```
SELECT
  *
FROM
  users
```

如果成員 B 執行下列查詢，則會遭到封鎖，因為無法投影不允許的輸出資料欄轉換：

```
SELECT
  COUNT(age)
FROM
  users
```

如果成員 B 執行下列查詢，則會遭到封鎖，因為無法使用別名在最終投影中參考不允許的輸出資料欄：

```
SELECT
  count_age
FROM
  (SELECT COUNT(age) AS count_age FROM users)
```

如果成員 B 執行下列查詢，則會遭到封鎖，因為轉換的限制資料欄會投影在輸出中：

```
SELECT
  CONCAT(name, email)
FROM
  users
```

如果成員 B 執行下列查詢，則會遭到封鎖，因為無法在最終投影中參考 CTE 中定義的不允許輸出資料欄：

```
WITH cte AS (
  SELECT
    age AS age_alias
  FROM
    users
)
SELECT age_alias FROM cte
```

如果成員 B 執行下列查詢，則會遭到封鎖，因為不允許的輸出資料欄無法用作最終投影中的排序或分割區索引鍵：

```
SELECT
  LISTAGG(gender) WITHIN GROUP (ORDER BY age) OVER (PARTITION BY age)
```

```
FROM
  users
```

如果成員 B 執行下列查詢，它會成功，因為屬於不允許輸出資料欄一部分的資料欄仍可用於查詢中的其他建構，例如聯結或篩選子句。

```
SELECT
  u.name,
  p.gender,
  p.age
FROM
  users AS u
JOIN
  pets AS p
ON
  u.name = p.owner_name
```

在相同的案例中，成員 B 也可以使用使用者中的名稱欄做為篩選條件或排序索引鍵：

```
SELECT
  u.email,
  u.gender
FROM
  users AS u
WHERE
  u.name = 'Mike'
ORDER BY
  u.name
```

此外，來自使用者的不允許輸出資料欄可用於中繼投影，例如子查詢和 CTEs，例如：

```
WITH cte AS (
  SELECT
    u.gender,
    u.id,
    u.first_name
FROM
  users AS u
```

```
)  
SELECT  
    first_name  
FROM  
    (SELECT cte.gender, cte.id, cte.first_name FROM cte)
```

編輯設定的資料表關聯

身為協同合作成員，您可以編輯已建立的已設定資料表關聯。

編輯設定的資料表關聯

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇資料表索引標籤。
5. 對於與您相關聯的資料表，選擇資料表。
6. 在資料表詳細資訊頁面上，向下捲動以檢視資料表關聯詳細資訊。
7. 選擇編輯。
8. 在編輯設定的資料表關聯頁面上，更新描述或服務存取資訊。
9. 選擇儲存變更。

取消已設定資料表的關聯

身為協同合作成員，您可以取消已設定的資料表與協同合作的關聯。此動作可防止可查詢的成員查詢資料表。

取消已設定資料表的關聯

1. 登入 AWS 管理主控台 並開啟位於 <https://console.aws.amazon.com/cleanrooms> 的 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇資料表索引標籤。

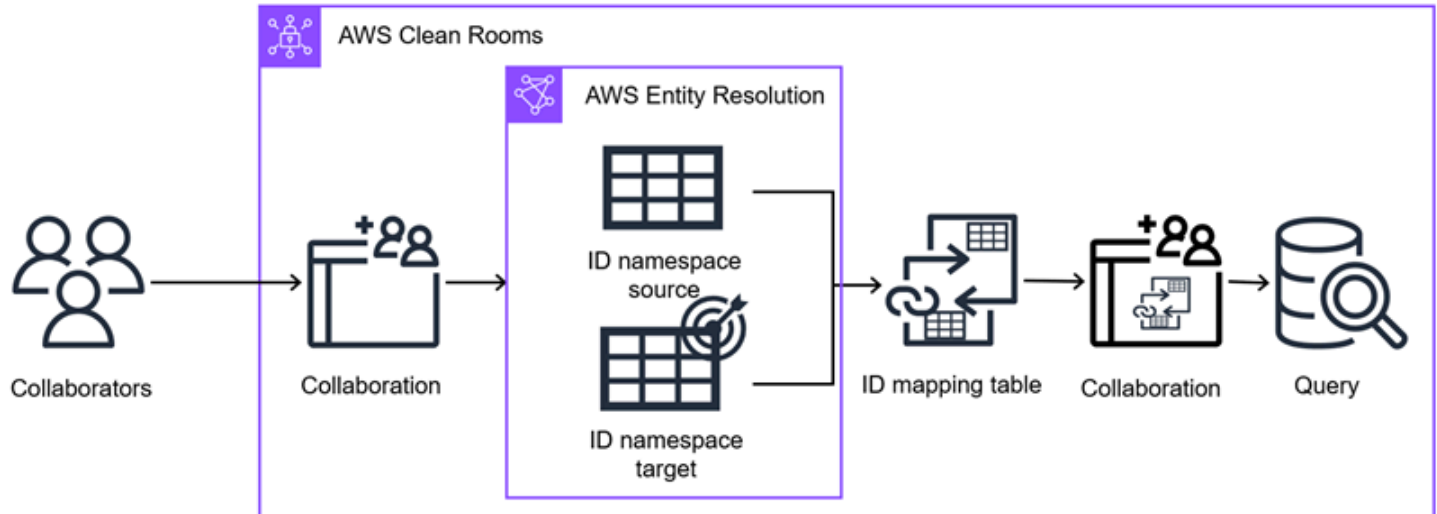
5. 對於與您相關聯的資料表，選取您要取消關聯的資料表旁的選項按鈕。
6. 選擇取消關聯。
7. 在對話方塊中，確認取消已設定資料表關聯的決定，並藉由選擇取消關聯來防止可查詢資料表的成員查詢資料表。

AWS Entity Resolution 在 AWS Clean Rooms

使用 AWS Entity Resolution 中的 AWS Clean Rooms，您可以將資料從來源轉譯到目標、將翻譯的資料填入 ID 映射表，以及查詢資料。

首先，您在 中建立協作，AWS Clean Rooms 並新增 AWS 帳戶 您要邀請的，或藉由建立成員資格來加入您受邀的協作。接下來，您會在兩個資料表上執行 ID 映射。您可以透過關聯現有的 ID 命名空間來源或在其中建立新的 ID 命名空間來源來執行此操作 AWS Entity Resolution。協同合作的另一個成員會建立現有 ID 命名空間目標的關聯，或建立新的 ID 命名空間目標。然後，您可以從兩個相關聯的 ID 命名空間建立並填入 ID 映射表。最後，可以查詢的成員透過加入 ID 映射表，跨兩個資料表執行查詢。

下圖摘要說明如何 AWS Entity Resolution 使用 AWS Clean Rooms。



Note

目前支援的轉碼服務提供者為 LiveRamp，可在下列位置取得 AWS 區域：美國東部（維吉尼亞北部）、美國東部（俄亥俄）和美國西部（奧勒岡）。

主題

- [中的 ID 命名空間 AWS Clean Rooms](#)
- [中的 ID 映射表 AWS Clean Rooms](#)

中的 ID 命名空間 AWS Clean Rooms

ID 命名空間是身分資料表的包裝函式，可讓您提供中繼資料，說明資料集以及如何在 ID 映射工作流程中使用資料集。ID 映射工作流程是一種資料處理任務，根據指定的 ID 映射方法，將資料從輸入資料來源映射到輸入資料目標。它會產生 ID 映射表。

ID 命名空間有兩種類型：來源和目標。來源包含將在 ID 映射工作流程中處理的來源資料的組態。目標包含所有來源將解析的目標資料的組態。若要定義您要跨兩個項目解析的輸入資料 AWS 帳戶，請建立 ID 命名空間來源和 ID 命名空間目標，將您的資料從一組 (來源) 轉譯為另一組 (目標)。

您可以建立新的 ID 命名空間，也可以關聯現有的 ID 命名空間。如需如何在 中建立 ID 命名空間的詳細資訊 AWS Entity Resolution，請參閱AWS Entity Resolution 《使用者指南》中的[使用 ID 命名空間定義輸入資料](#)。

主題

- [建立和關聯新的 ID 命名空間](#)
- [關聯現有的 ID 命名空間](#)
- [編輯 ID 命名空間關聯](#)
- [取消 ID 命名空間關聯的關聯](#)

建立和關聯新的 ID 命名空間

協作的每個成員都必須建立 ID 命名空間來源或 ID 命名空間目標並建立關聯，才能建立 ID 映射資料表來查詢身分資料。

如果您已在 中建立 ID 命名空間 AWS Entity Resolution，請跳至 [關聯現有的 ID 命名空間](#)。

建立新的 ID 命名空間並建立關聯

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms/> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在實體解析索引標籤上，選擇關聯 ID 命名空間。
5. 在關聯 ID 命名空間頁面上，針對實體解析資料，選擇建立 ID 命名空間。

AWS Entity Resolution 主控台會出現在新的索引標籤中。

6. 遵循主控台中建立 ID 命名空間頁面上 AWS Entity Resolution 的提示。
 - a. 如需詳細資訊，請輸入 ID 命名空間名稱、描述，然後選取 ID 命名空間類型 (來源或目標)。
 - b. 針對 ID 命名空間方法，選擇規則型比對的規則型方法，或選擇第三方轉碼的提供者服務。
 - c. 根據您選擇的 ID 命名空間方法，指定資料輸入類型。
 - d. 選擇建立 ID 命名空間。
7. 返回 AWS Clean Rooms 主控台。
8. 在關聯 ID 命名空間頁面上，針對實體解析資料，從下拉式清單中選擇您要與協同合作建立關聯的 AWS Entity Resolution ID 命名空間來源或目標。
9. 如需關聯詳細資訊，請執行下列步驟。
 - a. 輸入相關聯 ID 命名空間的名稱。

您可以使用預設名稱或重新命名此 ID 命名空間。
 - b. (選用) 輸入 ID 命名空間的描述。

描述有助於撰寫查詢。
10. 選取 選項，然後採取建議的動作，以指定AWS Clean Rooms 存取許可。

選項	建議的動作
允許 AWS Clean Rooms 新增和管理許可政策	AWS Clean Rooms 會建立具有此關聯所需政策的服務角色。
手動新增和管理許可	執行以下任意一項： - 檢閱資源政策，並將必要的許可新增至政策。 - 選擇新增政策陳述式來使用現有政策。 您必須具有修改角色和建立政策的許可。

選項	建議的動作
	 Note 如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

11. (選用) 對於進階 ID 映射資料表組態，請修改來自 ID 名稱之資料欄的預設保護。

根據預設，ID 映射表設定為僅允許 sourceID 欄和 targetID 欄 INNER JOIN 上的。您可以修改此組態，以便在查詢中的任何位置允許來自此 ID 命名空間 (sourceID 或 targetID) 的資料欄。

您的目標	建議選項
將資料欄分類為「加入資料欄」，並只允許在 INNER JOIN 子句中	是
將資料欄分類為「維度資料欄」，並允許它在查詢中的任何位置，包括查詢的 JOIN 子句 SELECT、WHERE 和 GROUP BY 陳述式。	否，允許查詢中的任何位置

12. (選用) 如果您想要為 ID 命名程式資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
13. 選擇關聯。
14. 在實體解析索引標籤的關聯 ID 命名空間資料表下，檢視關聯的 ID 命名空間，並確認 ID 命名空間類型正確 (來源或目標)。

在協同合作中的所有成員都與其 ID 命名空間建立關聯後，您可以[建立 ID 映射表](#)並查詢資料。

關聯現有的 ID 命名空間

在此程序中，每個成員都會在協同合作中建立其現有 ID 命名空間來源或其 ID 命名空間目標的關聯。

建立現有 ID 命名空間的關聯

- 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms/> 開啟 AWS Clean Rooms 主控台。

2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在實體解析索引標籤上，選擇關聯 ID 命名空間。
5. 在關聯 ID 命名空間頁面上，針對實體解析資料，從下拉式清單中選擇您要與協同合作關聯的 AWS Entity Resolution ID 命名空間來源或目標。
6. 如需關聯詳細資訊，請執行下列步驟。
 - a. 輸入相關聯 ID 命名空間的名稱。

您可以使用預設名稱或重新命名此 ID 命名空間。

- b. (選用) 輸入 ID 命名空間的描述。

描述有助於撰寫查詢。

7. 選取 選項，然後採取建議的動作，以指定AWS Clean Rooms 存取許可。

選項	建議的動作
允許 AWS Clean Rooms 新增和管理許可政策	AWS Clean Rooms 會建立具有此關聯所需政策的服務角色。
手動新增和管理許可	執行以下任意一項： - 檢閱資源政策，並將必要的許可新增至政策。 - 選擇新增政策陳述式，以使用現有的政策。 您必須具有修改角色和建立政策的許可。  Note 如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

8. (選用) 對於進階 ID 映射資料表組態，請修改來自 ID 名稱之資料欄的預設保護。

根據預設，ID 映射表設定為僅允許 sourceID 欄和 targetID 欄 INNER JOIN 上的。您可以修改此組態，以便在查詢中的任何位置允許來自此 ID 命名空間 (sourceID 或 targetID) 的資料欄。

您的目標	建議選項
將資料欄分類為「加入資料欄」，並只允許在 INNER JOIN 子句中。	是
將資料欄分類為「維度資料欄」，並允許它在查詢中的任何位置，包括查詢的 JOIN 子句、WHERE、SELECT 和 GROUP BY 陳述式。	否，允許查詢中的任何位置

9. (選用) 如果您想要為 ID 命名程式資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
10. 選擇關聯。
11. 在實體解析索引標籤的關聯 ID 命名空間資料表下，檢視關聯的 ID 命名空間，並確認 ID 命名空間類型正確 (來源或目標)。

在協同合作中的所有成員都與其 ID 命名空間建立關聯後，您可以[建立 ID 映射表](#)並查詢資料。

編輯 ID 命名空間關聯

身為協同合作成員，您可以編輯已建立的 ID 命名空間關聯。

編輯 ID 命名空間關聯

1. 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms/> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇實體解析索引標籤。
5. 針對關聯的 ID 命名空間，選擇 ID 命名空間。
6. 在 ID 命名空間詳細資訊頁面上，向下捲動以檢視 ID 命名空間關聯詳細資訊。
7. 選擇編輯。
8. 在編輯 ID 命名空間關聯頁面上，編輯下列任何項目：

- a. 如需關聯詳細資訊，請更新名稱或描述。
- b. (選用) 對於進階 ID 映射資料表組態，請修改來自 ID 命名集的資料欄的預設保護。

根據預設，ID 映射表設定為僅允許 sourceID 欄和 targetID 欄 INNER JOIN 上的。您可以修改此組態，以便在查詢中的任何位置允許來自此 ID 命名空間 (sourceID 或 targetID) 的資料欄。

您的目標	建議選項
將資料欄分類為「加入資料欄」，並只允許在 INNER JOIN 子句中	是
將資料欄分類為「維度資料欄」，並允許它在查詢中的任何位置，包括查詢的 JOIN 子句 SELECT、WHERE 和 GROUP BY 陳述式。	否，允許查詢中的任何位置

9. 選擇儲存變更。

取消 ID 命名空間關聯的關聯

身為協同合作成員，您可以取消 ID 命名空間與協同合作的關聯。此動作可防止可查詢的成員查詢資料表。

Warning

取消 ID 命名空間與協同合作的關聯會刪除衍生 ID 映射資料表中的任何資料，使其無法查詢。例如，如果您的 ID 命名空間關聯在三個不同的 ID 映射表中用作 SOURCE，則取消 ID 命名空間關聯的關聯時，將會刪除來自這些 ID 映射表的所有資料。

取消 ID 命名空間關聯的關聯

- 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms/> 開啟 AWS Clean Rooms 主控台。
- 在左側導覽窗格中，選擇協同合作。
- 選擇協同合作。
- 選擇實體解析索引標籤。

5. 針對關聯的 ID 命名空間，選取您要取消關聯的 ID 命名空間旁的選項按鈕。
6. 選擇取消關聯。
7. 在對話方塊中，選擇取消關聯，以確認您中斷連接 ID 命名空間的決定。此動作可防止任何可查詢的成員存取 ID 映射表。

如果協同合作的成員移除其中一個 ID 命名空間，如果來源已離開協同合作，您就無法重新填入 ID 映射表。

即使先前已填入 ID 映射表，取消 ID 命名空間的關聯表示您無法再在該資料表上執行查詢。

中的 ID 映射表 AWS Clean Rooms

ID 映射表是 中的資源 AWS Clean Rooms，可在協同合作中啟用多方身分映射。

建立 ID 映射表之前，您必須先將來源和目標資料都設定為 ID 命名空間。

建立 ID 映射表之後，您可以使用 ID 映射工作流程，將來源 ID 命名空間轉譯為目標 ID 命名空間。您可以使用規則型方法或提供者服務轉碼方法來執行此操作。

ID 映射工作流程是一種資料處理任務，會根據指定的 ID 映射工作流程方法，將資料從輸入資料來源映射至輸入資料目標。此工作流程會填入 ID 映射表。

Note

ID 映射資料表只能從存放在 Amazon S3 中的資料集建立，並編排到 AWS Glue 資料表中。

有兩種 ID 映射工作流程方法：規則型 ID 映射或提供者服務 ID 映射：

- 規則型 ID 映射 – 您可以使用比對規則將第一方資料從來源轉換至目標。
- 提供者服務 ID 映射 – 您可以使用 LiveRamp 提供者服務將第三方資料從來源轉譯到目標。

Note

目前支援的轉碼服務提供者是 LiveRamp。協同合作中透過 訂閱 LiveRamp 的任何成員 AWS Data Exchange 都可以建立 ID 映射表。如果您已訂閱 LiveRamp，但未透過 訂閱 AWS Data Exchange，請聯絡 LiveRamp 以取得私有優惠。如需詳細資訊，請參閱 AWS Entity Resolution 《使用者指南》中的 [在上訂閱供應商服務 AWS Data Exchange](#)。

主題

- [建立和填入新的 ID 映射表](#)
- [填入現有的 ID 映射表](#)
- [編輯 ID 映射表](#)
- [刪除 ID 映射表](#)

建立和填入新的 ID 映射表

先決條件

建立 ID 映射表之前，請確定您有：

- 相關聯的 ID 命名空間來源和目標
- 針對規則型 ID 映射或提供者服務 ID 映射設定的 ID 命名空間

您可以建立兩種類型的 ID 映射表：

- 規則型 – 使用相符的規則來翻譯第一方資料
- 提供者服務 – 使用 LiveRamp 轉譯 RampIDs

建立 ID 映射表之後，您可以立即執行 ID 映射工作流程來填入它，也可以等到稍後才填入資料表。

成功填入 ID 映射表後，您可以在 ID 映射表上執行多資料表聯結查詢，以將 `sourceId` 與 聯結 `targetId` 並分析資料。

主題

- [建立 ID 映射表（規則型）](#)
- [建立 ID 映射表（提供者服務）](#)

建立 ID 映射表（規則型）

本主題說明建立 ID 映射表的程序，該資料表使用相符的規則將第一方資料從來源轉譯到目標。

當您建立規則型 ID 映射表時，您可以選擇只處理工作流程中新的、已更新或刪除的記錄，方法是開啟增量處理。

使用規則型方法建立和填入新的 ID 映射表

1. 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms/> 開啟 AWS Clean Rooms 主控台。
2. 在導覽窗格中，選擇協同合作。
3. 選擇協同合作，然後選擇實體解析索引標籤。
4. 選擇建立 ID 映射表。
5. 在 ID 映射設定下，執行下列其中一項操作：
 - 若要建立新的工作流程，請保持選取建立新的 ID 映射工作流程。
 - 若要使用現有的工作流程，請清除核取方塊，然後從下拉式清單中選取工作流程，然後跳到步驟 9。
6. 在身分資料下，檢視或設定來源和目標。
 - 對於單一 ID 命名空間對：檢閱預先選取的來源和目標。
 - 對於多個 ID 命名空間：從下拉式清單中選取來源和目標。
7. (選用) 選取開啟增量處理，以僅處理工作流程中新的、已更新或刪除的記錄。

AWS Entity Resolution 只會處理來源或目標 ID 命名空間中的新記錄、更新記錄或刪除記錄，而不是重新建立整個 ID 映射表。

如果您未選取此選項，AWS Entity Resolution 會在 ID 映射表上執行預設批次處理 ID 映射工作流程。

8. 在規則參數下，設定下列項目：
 - 規則控制 – 選擇目標或來源是否提供相符的規則。

您可以開啟顯示規則來檢視規則。

規則控制必須在來源與目標 ID 命名空間之間相容，才能用於 ID 映射工作流程。例如，如果來源 ID 命名空間將規則限制為目標，但目標 ID 命名空間將規則限制為來源，則會導致錯誤。

- 比較類型會自動設定為多個輸入欄位。

這是因為兩位參與者先前都已選取此選項。

- 記錄比對
 - 一個來源到一個目標 – 為每個目標儲存一個相符記錄
 - 一個目標的許多來源 – 儲存每個目標的所有相符記錄

 Note

為來源和目標 ID 命名空間指定的限制必須相容。

9. 如需 ID 映射詳細資訊，請設定下列項目：

- a. 輸入 ID 映射資料表名稱或保留預設名稱。
- b. (選用) 輸入 ID 映射表的描述。

描述有助於撰寫查詢。

10. 針對AWS Clean Rooms 存取，請選擇一項：

- 允許 AWS Clean Rooms 新增和管理許可政策 – 自動建立服務角色。
- 手動新增和管理許可 – 檢閱和修改資源政策，或選擇新增政策陳述式。

 Note

如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

11. 針對AWS Entity Resolution 存取，請選擇一項：

只有在您建立新的 ID 映射資料表時，才會顯示此區段。

- 建立和使用新的服務角色
 - 預設的服務角色名稱為 `entityresolution-id-mapping-workflow-<timestamp>`
 - (選用) 對於加密的資料，選取此資料由 KMS 金鑰加密，然後輸入AWS KMS 金鑰。
- 使用現有的服務角色
 - 從下拉式清單中選擇現有的服務角色名稱，或輸入角色 ARN。

如果您有列出角色的許可，則會顯示角色清單。

選擇 IAM 外部連結中的檢視，以檢視服務角色。

如果沒有現有的服務角色，則無法使用使用現有服務角色的選項。

根據預設，AWS Clean Rooms 不會嘗試更新現有的角色政策來新增必要的許可。

- (選用) 選取將具有必要許可的預先設定政策新增至此角色，以將必要許可連接至角色。

您必須具有修改角色和建立政策的許可。

12. (選用) 在其他設定下，設定：

a. ID 映射資料表設定

- 若要啟用自訂加密，請選擇自訂加密設定並輸入 AWS KMS 金鑰。

 Note

此 KMS 金鑰需要授予必要的許可，才能使用 KMS 金鑰政策 AWS Entity Resolution 在 `cleanrooms.amazonaws.com` 內使用。如需使用 ID 映射工作流程進行加密所需許可的詳細資訊，請參閱 AWS Entity Resolution 《使用者指南》中的 [為 建立工作流程任務角色 AWS Entity Resolution](#)。

- 若要新增標籤，請選擇新增標籤，然後輸入鍵/值對

b. ID 映射工作流程設定（僅限新工作流程）：

- 若要使用不同的名稱，請清除保留相同的 ID 映射資料表名稱和描述，然後輸入新值。
- 若要新增標籤，請選擇新增標籤，然後輸入鍵/值對

13. 選擇下列其中一項：

- 建立 ID 映射表 – 建立您可以稍後填入的空白資料表 ([填入現有的 ID 映射表](#))
- 建立並填入 ID 映射表 – 建立並立即填入資料表（可能需要幾個小時）

ID 映射工作流程程序開始。在此過程中，ID 映射表會填入翻譯的 IDs。ID 映射工作流程可能需要幾個小時的時間來處理。

成功填入 ID 映射表後，您可以[查詢 ID 映射表](#)，將 `sourceId` 與 聯結 `targetId` 並分析資料。

建立 ID 映射表（提供者服務）

本主題說明建立使用提供者服務 (LiveRamp) 的 ID 映射表的程序。LiveRamp 提供者服務會使用維護或衍生 RampIDs，將一組來源 RampID 轉譯為另一個來源 RampIDs。

使用提供者服務方法建立新的 ID 映射表

1. 登入 AWS 管理主控台 並開啟位於 <https://console.aws.amazon.com/cleanrooms/> 的 AWS Clean Rooms 主控台。
2. 在導覽窗格中，選擇協同合作。
3. 選擇協同合作，然後選擇實體解析索引標籤。
4. 選擇建立 ID 映射表。
5. 在 ID 映射設定下，執行下列其中一項操作：
 - 若要建立新的工作流程，請保持選取建立新的 ID 映射工作流程。
 - 若要使用現有的工作流程，請清除核取方塊，然後從下拉式清單中選取工作流程，然後跳到步驟 9。
6. 在身分資料下，檢視或設定來源和目標。
 - 對於單一 ID 命名空間對：檢閱預先選取的來源和目標。
 - 對於多個 ID 命名空間：從下拉式清單中選取來源和目標。
7. 在方法下，確認選取的 ID 映射工作流程方法為 LiveRamp 轉碼。
8. 對於 LiveRamp 組態，請執行下列其中一項操作：
 - 輸入 LiveRamp ID Manager ARN 和 LiveRamp 秘密管理員 ARN。
 - 選擇從現有工作流程匯入。
9. 如需 ID 映射詳細資訊，請設定下列項目：
 - a. 輸入 ID 映射資料表名稱或保留預設名稱。
 - b. （選用）輸入 ID 映射表的描述。

描述有助於撰寫查詢。
10. 針對AWS Clean Rooms 存取，請選擇一項：
 - 允許 AWS Clean Rooms 新增和管理許可政策 – 自動建立服務角色。
 - 手動新增和管理許可 – 檢閱和修改資源政策，或選擇新增政策陳述式。

Note

如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

11. 針對AWS Entity Resolution 存取，請選擇一項：

只有在您建立新的 ID 映射資料表時，才會顯示此區段。

- 建立和使用新的服務角色
 - 預設的服務角色名稱為 `entityresolution-id-mapping-workflow-<timestamp>`
 - （選用）對於加密的資料，選取此資料由 KMS 金鑰加密，然後輸入AWS KMS 金鑰。
- 使用現有的服務角色
 - 從下拉式清單中選擇現有的服務角色名稱，或輸入角色 ARN。

如果您有列出角色的許可，則會顯示角色清單。

選擇 IAM 外部連結中的檢視，以檢視服務角色。

如果沒有現有的服務角色，則無法使用使用現有服務角色的選項。

根據預設，AWS Clean Rooms 不會嘗試更新現有的角色政策來新增必要的許可。

- （選用）選取將具有必要許可的預先設定政策新增至此角色，以將必要許可連接至角色。

您必須具有修改角色和建立政策的許可。

12. （選用）在其他設定下設定：

a. ID 映射資料表設定

- 若要啟用自訂加密，請選擇自訂加密設定並輸入 AWS KMS 金鑰。

Note

此 KMS 金鑰需要授予必要的許可，才能使用 KMS 金鑰政策 AWS Entity Resolution 在 `cleanrooms.amazonaws.com`內使用。如需使用 ID 映射工作流程進行加密所需許可的詳細資訊，請參閱AWS Entity Resolution 《使用者指南》中的[為 建立工作流程任務角色 AWS Entity Resolution](#)。

- 若要新增標籤，請選擇新增標籤，然後輸入鍵/值對。

b. ID 映射工作流程設定（僅限新工作流程）：

- 若要使用不同的名稱，請清除保留相同的 ID 映射資料表名稱和描述，然後輸入新值。
- 若要新增標籤，請選擇新增標籤，然後輸入鍵/值對。

13. 選擇下列其中一項：

- 建立 ID 映射表 – 建立您可以稍後填入的空白資料表 ([填入現有的 ID 映射表](#))
- 建立並填入 ID 映射表 – 建立並立即填入資料表 (可能需要幾個小時)

ID 映射工作流程程序開始。在此過程中，ID 映射表會填入翻譯的 IDs。ID 映射工作流程可能需要幾個小時的時間來處理。

成功填入 ID 映射表後，您可以[查詢 ID 映射表](#)，將 sourceId 與 聯結targetId並分析資料。

填入現有的 ID 映射表

將新資料新增至 ID 命名空間時，請使用此工作流程。如果您已選擇在[建立 ID 映射表](#)時開啟增量處理，則能夠在來源或目標 ID 命名空間中僅處理新的、已更新或刪除的記錄，而不是重新建立整個 ID 映射表。

填入現有的 ID 映射表

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms/> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 前往實體解析索引標籤。
5. 在 ID 映射表區段下，選取 ID 映射表。
6. 如果您尚未開啟此 ID 映射表的增量處理，請選擇 ID 映射表，然後選擇填入。
7. 如果您已開啟此 ID 映射表的增量處理，請選擇填入，然後選擇下列其中一項：

- 增量處理 – 僅處理來源或目標 ID 命名空間中的新記錄、更新記錄或刪除的記錄。

建議用於頻繁更新、每日執行或即時資料同步。

- 批次處理 – 處理整個 ID 映射表。

建議用於初始設定、定期完整重新整理，或來源和目標 ID 命名空間發生重大變更時。

- 僅刪除處理 — 僅處理來源 ID 命名空間中已刪除的記錄，並相應地更新目標 ID 命名空間。

建議用於快速同步移除。

8. ID 映射工作流程程序開始。

在此過程中，ID 映射表會填入轉碼 IDs。ID 映射工作流程可能需要幾個小時的時間來處理。

成功填入 ID 映射表後，您可以[查詢 ID 映射表](#)，將 sourceId 與 聯結targetId。

編輯 ID 映射表

身為協同合作成員，您可以編輯已建立的 ID 映射表。

編輯 ID 映射表

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 選擇實體解析索引標籤。
5. 針對 ID 映射資料表，選擇資料表。
6. 在 ID 映射表詳細資訊頁面上，向下捲動以檢視 ID 映射表詳細資訊。
7. 選擇編輯。
8. 在編輯 ID 映射表頁面上，更新描述或服務存取資訊。
9. 選擇儲存變更。

刪除 ID 映射表

身為協同合作成員，您可以刪除已建立的 ID 映射表。此動作可防止可查詢 的成員查詢資料表。

Warning

刪除映射表會永久移除任何填入的資料。

刪除 ID 映射表

1. 登入 AWS 管理主控台 並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。

3. 選擇協同合作。
4. 選擇實體解析索引標籤。
5. 針對 ID 映射資料表，選擇資料表。
6. 在 ID 映射表詳細資訊頁面上，向下捲動以檢視 ID 映射表。
7. 選擇 和 ID 映射表，然後選擇刪除。
8. 如果您確定要刪除 ID 映射表，請選擇刪除。

中的分析範本 AWS Clean Rooms

分析範本可與 搭配使用 [中的自訂分析規則 AWS Clean Rooms](#)。透過分析範本，您可以定義參數，以協助您重複使用相同的查詢。AWS Clean Rooms 支援具有常值的參數化子集。

分析範本是特定於協同合作。對於每個協同合作，成員只能查看該協同合作中的查詢。如果您計劃在協同合作中使用差異隱私權，您應該確保您的分析範本與差異隱私權 [的一般用途查詢結構](#) AWS Clean Rooms 相容。

您可以透過兩種方式建立分析範本：使用 SQL 程式碼或使用適用於 Spark 的 Python 程式碼。

主題

- [SQL 分析範本](#)
- [PySpark 分析範本](#)
- [故障診斷 PySpark 分析範本](#)

SQL 分析範本

SQL 分析範本可讓您查詢和分析協同合作中不同資料集的資料。您可以使用這些範本來執行各種類型的分析，例如識別對象重疊和計算彙總指標。

使用 SQL 分析範本，您可以：

- 寫入標準 SQL 查詢
- 新增參數，讓您的查詢成為動態
- 控制對特定資料欄和資料表的存取
- 設定敏感資料的彙總需求
- 為自訂機器學習 (ML) 模型的隱私權增強合成資料集的產生定義輸入資料

主題

- [建立 SQL 分析範本](#)
- [檢閱 SQL 分析範本](#)

建立 SQL 分析範本

先決條件

建立 SQL 分析範本之前，您必須擁有：

- 主動 AWS Clean Rooms 協同合作
- 在協同合作中存取至少一個已設定的資料表

如需在 中設定資料表的詳細資訊 AWS Clean Rooms，請參閱 [在 中建立設定的資料表 AWS Clean Rooms](#)。

- 建立分析範本的許可
- SQL 查詢語法的基本知識

下列程序說明使用 [AWS Clean Rooms 主控台](#) 建立 SQL 分析範本的程序。

如需有關如何使用 AWS SDKs 建立 SQL 分析範本的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

建立 SQL 分析範本

1. 登入 AWS 管理主控台，並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶，以做為協作建立者。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在範本索引標籤上，前往您建立的分析範本區段。
5. 選擇建立分析範本。
6. 在建立分析範本頁面上，如需詳細資訊，
 - a. 輸入分析範本的名稱。
 - b. （選用）輸入描述。
 - c. 對於格式，保留選取 SQL 選項。
7. 對於資料表，檢視與協同合作相關聯的已設定資料表。
8. 對於定義，
 - a. 輸入分析範本的定義。
 - b. 選擇從 匯入以匯入定義。

- c. (選用) 在 SQL 編輯器中指定參數，方法是在參數名稱前面輸入冒號 (:)。

例如：

```
WHERE table1.date + :date_period > table1.date
```

9. 如果您先前已新增參數，請在參數 - 選用下，為每個參數名稱選擇類型和預設值（選用）。
10. 對於合成資料，如果您想要為模型訓練產生合成資料，請選取需要分析範本輸出為合成核取方塊。

如需詳細資訊，請參閱[隱私權增強合成資料集產生](#)。

- a. 對於資料欄分類，請從下拉式清單中選擇資料欄。至少需要五個資料欄。

- i. 從下拉式清單中選擇分類。這會識別每個資料欄的資料類型。

分類類型包括：

- 數值 – 連續數值，例如測量或計數
- 分類 – 離散值或類別，例如標籤或類型

- ii. 若要移除資料欄，請選取移除。

- iii. 若要新增另一個資料欄，請選取新增另一個資料欄。從下拉式清單中選擇資料欄和分類。

- iv. 對於預測值，請從下拉式清單中選擇資料欄。這是自訂模型在訓練合成資料集之後用於預測的資料欄。

- b. 進階設定可讓您設定隱私權層級和隱私權閾值。調整設定以符合您的需求。

- 針對隱私權層級，輸入 epsilon 值，以判斷合成模型為保護所產生資料集的隱私權所新增的雜訊。值必須介於 0.0001 和 10 之間。
 - 較低的值會增加更多雜訊，提供更強大的隱私權保護，但可能會降低對此資料訓練的下游自訂模型的公用程式。
 - 較高的值會加入較少雜訊，提供較高的準確性，但可能會降低隱私權保護的強度。

針對隱私權閾值，輸入成員資格推論攻擊可以識別原始資料集成員的最高允許機率。值必須介於 50.0 和 100 之間。

- 分數為 50% 表示成員推論攻擊無法成功區分成員與非成員，比隨機猜測更好。
- 對於沒有隱私權限制，請輸入 100%。

最佳值取決於您的特定使用案例和隱私權需求。如果超過隱私權閾值，ML 輸入通道建立會失敗，而且您無法使用合成資料集來訓練模型。

Warning

合成資料產生可防止推斷個別屬性，無論原始資料集中是否存在特定個人，或是否存在這些個人的學習屬性。不過，它不會阻止原始資料集的常值出現，包括個人身分識別資訊 (PII) 出現在合成資料集中。

建議您避免輸入資料集中僅與一個資料主體相關聯的值，因為這些值可能會重新識別資料主體。例如，如果只有一個使用者住在郵遞區號中，則合成資料集中是否存在該郵遞區號，會確認使用者位於原始資料集中。可以使用截斷高精確度值或將不常見目錄取代為其他等技術來降低此風險。這些轉換可以用來建立 ML 輸入通道之查詢的一部分。

11. 如果您想要為資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
12. 選擇建立。
13. 您現在可以通知協同合作成員，他們可以[檢閱分析範本](#)。（如果您想要查詢自己的資料，則為選用。）

檢閱 SQL 分析範本

協同合作成員建立 SQLanalysis範本後，您可以檢閱並核准該範本。在分析範本和核准後，它可用於中的查詢 AWS Clean Rooms。

Note

當您將分析程式碼帶入協同合作時，請注意下列事項：

- AWS Clean Rooms 不會驗證或保證分析程式碼的行為。
 - 如果您需要確保特定行為，請直接檢閱協同合作夥伴的程式碼，或與信任的第三方稽核人員合作進行檢閱。
- 在共用安全模型中：
 - 您（客戶）負責保護環境中執行之程式碼的安全性。
 - AWS Clean Rooms 負責環境的安全，確保
 - 只有核准的程式碼會執行

- 只能存取指定的已設定資料表
- 唯一的輸出目的地是結果接收者的 S3 儲存貯體。

使用 AWS Clean Rooms 主控台檢閱 SQL 分析範本

1. 登入 AWS 管理主控台，並使用將做為協作建立者的 AWS 帳戶 開啟 [AWS Clean Rooms 主控台](#)。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在範本索引標籤上，前往其他成員建立的分析範本區段。
5. 選擇可以執行狀態為否的分析範本需要您的檢閱。
6. 選擇檢閱。
7. 檢閱分析規則概觀、定義和參數（如果有的話）。
8. 檢閱定義中參考的資料表下列出的已設定資料表。

每個資料表旁的狀態將不允許讀取範本。

9. 選擇 表格。

如果您	然後選擇
核准分析範本	允許資料表上的範本。選擇允許，確認您的核准。
不要核准分析範本	不允許

您現在可以使用 SQL 分析範本查詢設定的資料表。如需詳細資訊，請參閱[執行 SQL 查詢](#)。

PySpark 分析範本

PySpark 分析範本需要 Python 使用者指令碼和選用的虛擬環境，才能使用自訂和開放原始碼程式庫。這些檔案稱為成品。

建立分析範本之前，請先建立成品，然後將成品存放在 Amazon S3 儲存貯體中。AWS Clean Rooms 會在執行分析任務時使用這些成品。AWS Clean Rooms 只會在執行任務時存取成品。

在 PySpark 分析範本上執行任何程式碼之前，會透過以下方式 AWS Clean Rooms 驗證成品：

- 檢查建立範本時使用的特定 S3 物件版本
- 驗證成品的 SHA-256 雜湊
- 失敗修改或移除成品的任何任務

 Note

中指定 PySpark 分析範本的所有合併成品大小上限為 AWS Clean Rooms 1 GB。

PySpark 分析範本的安全性

為了保留安全的運算環境，AWS Clean Rooms 會使用雙層運算架構來隔離使用者程式碼與系統操作。此架構是以 Amazon EMR Serverless Fine Grained Access Control 技術為基礎，也稱為 Membrane。如需詳細資訊，請參閱[在具有必要程式碼的情況下，在 Apache Spark 中安全且高效能的資料存取控制](#)。

運算環境元件分為單獨的使用者空間和系統空間。使用者空間會在 PySpark 分析範本中執行 PySpark 程式碼。AWS Clean Rooms 會使用系統空間來執行任務，包括使用客戶提供的服務角色來讀取資料以執行任務並實作資料欄允許清單。由於此架構，會封鎖影響系統空間的客戶 PySpark 程式碼，其中可能包含少量 Spark SQL 和 PySpark DataFrames APIs。

中的 PySpark 限制 AWS Clean Rooms

當客戶提交核准的 PySpark 分析範本時，會在自己的安全運算環境中 AWS Clean Rooms 執行，而客戶無法存取該環境。運算環境實作具有使用者空間和系統空間的運算架構，以保留安全的運算環境。如需詳細資訊，請參閱[PySpark 分析範本的安全性](#)。

在 中使用 PySpark 之前，請考慮下列限制 AWS Clean Rooms。

限制

- 僅支援 DataFrame 輸出
- 每個任務執行的單一 Spark 工作階段

不支援的功能

- 資料管理
 - Iceberg 資料表格式
 - LakeFormation 受管資料表
 - 彈性分散式資料集 (RDD)
 - Spark 串流
 - 巢狀資料欄的存取控制
- 自訂函數和擴充功能
 - 使用者定義的資料表函數 UDTFs)
 - HiveUDFs
 - 使用者定義函數中的自訂類別
 - 自訂資料來源
 - 適用於以下項目的其他 JAR 檔案：
 - Spark 延伸模組
 - 連接器
 - 中繼存放區組態
- 監控和分析
 - Spark 記錄
 - Spark 使用者介面
 - ANALYZE TABLE 命令

Important

這些限制適用於維持使用者和系統空間之間的安全隔離。
無論協同合作組態為何，都適用所有限制。
未來的更新可能會根據安全性評估新增對其他功能的支援。

最佳實務

我們建議在建立 PySpark 分析範本時採用下列最佳實務。

- [中的 PySpark 限制 AWS Clean Rooms](#) 以 為考量來設計您的分析範本。
- 首先在開發環境中測試您的程式碼。

- 僅使用支援的 DataFrame 操作。
- 規劃您的輸出結構以使用 DataFrame 限制。

我們建議您使用下列最佳實務來管理成品

- 將所有 PySpark 分析範本成品保留在專用 S3 儲存貯體或字首中。
- 針對不同的成品版本使用透明版本命名。
- 需要成品更新時建立新的分析範本。
- 維護哪些範本使用哪些成品版本的庫存。

如需如何撰寫 Spark 程式碼的詳細資訊，請參閱下列內容：

- [Apache Spark 範例](#)
- 在 Amazon EMR 版本指南中 [撰寫 Spark 應用程式](#)
- [使用者指南中的教學課程：撰寫 AWS Glue 適用於 Spark 的指令碼](#) AWS Glue

下列主題說明如何在建立和檢閱分析範本之前建立 Python 使用者指令碼和程式庫。

主題

- [建立使用者指令碼](#)
- [在 PySpark 分析範本中使用參數](#)
- [建立虛擬環境（選用）](#)
- [在 S3 中存放使用者指令碼和虛擬環境](#)
- [建立 PySpark 分析範本](#)
- [檢閱 PySpark 分析範本](#)

建立使用者指令碼

使用者指令碼必須包含進入點函數（也就是處理常式）。您可以使用任何有效的 Python 檔案名稱來命名使用者指令碼檔案。

下列程序說明如何建立使用者指令碼來定義 PySpark 分析的核心功能。

先決條件

- PySpark 1.0 (對應至 Python 3.11 和 Spark 3.5.3)
- Amazon S3 中的資料集只能在您定義的 Spark 工作階段中讀取為已設定的資料表關聯。
- 您的程式碼無法直接呼叫 Amazon S3 和 AWS Glue
- 您的程式碼無法進行網路呼叫

建立使用者指令碼

1. 開啟您選擇的文字編輯器或整合式開發環境 (IDE)。

您可以使用支援 Python 檔案的任何文字編輯器或 IDE (例如 Visual Studio Code、PyCharm 或 Notepad++)。

2. 使用您選擇的名稱建立新的 Python 檔案 (例如 **my_analysis.py**)。
3. 定義接受內容物件參數的進入點函數。

```
def entrypoint(context)
```

`context` 物件參數是一種字典，可讓您存取重要的 Spark 元件、參考的資料表和分析參數。它包含以下內容：

透過 存取 Spark 工作階段 `context['sparkSession']`

透過 參考的資料表 `context['referencedTables']`

透過 分析參數 `context['analysisParameters']` (如果在範本中定義參數)

4. 定義進入點函數的結果：

```
return results
```

`results` 必須將包含檔案名稱結果字典的物件傳回至輸出 DataFrame。

Note

AWS Clean Rooms 會自動將 DataFrame 物件寫入結果接收器的 S3 儲存貯體。

5. 您現在已準備好：

- a. 將此使用者指令碼儲存在 S3 中。如需詳細資訊，請參閱[在 S3 中存放使用者指令碼和虛擬環境](#)。
- b. 建立選用的虛擬環境，以支援使用者指令碼所需的任何其他程式庫。如需詳細資訊，請參閱[建立虛擬環境（選用）](#)。

Example範例 1

<caption>The following example demonstrates a generic user script for a PySpark analysis template.</caption>

```
# File name: my_analysis.py

def entrypoint(context):
    try:
        # Access Spark session
        spark = context['sparkSession']

        # Access input tables
        input_table1 = context['referencedTables']['table1_name']
        input_table2 = context['referencedTables']['table2_name']

        # Example data processing operations
        output_df1 = input_table1.select("column1", "column2")
        output_df2 = input_table2.join(input_table1, "join_key")
        output_df3 = input_table1.groupBy("category").count()

        # Return results - each key creates a separate output folder
        return {
            "results": {
                "output1": output_df1,          # Creates output1/ folder
                "output2": output_df2,          # Creates output2/ folder
                "analysis_summary": output_df3 # Creates analysis_summary/ folder
            }
        }

    except Exception as e:
        print(f"Error in main function: {str(e)}")
        raise e
```

此範例的資料夾結構如下：

```
analysis_results/
#
### output1/ # Basic selected columns
# ### part-00000.parquet
# ### _SUCCESS
#
### output2/ # Joined data
# ### part-00000.parquet
# ### _SUCCESS
#
### analysis_summary/ # Aggregated results
### part-00000.parquet
### _SUCCESS
```

Example範例 2

<caption>The following example demonstrates a more complex user script for a PySpark analysis template.</caption>

```
def entrypoint(context):
    try:
        # Get DataFrames from context
        emp_df = context['referencedTables']['employees']
        dept_df = context['referencedTables']['departments']

        # Apply Transformations
        emp_dept_df = emp_df.join(
            dept_df,
            on="dept_id",
            how="left"
        ).select(
            "emp_id",
            "name",
            "salary",
            "dept_name"
        )

        # Return Dataframes
        return {
            "results": {
                "outputTable": emp_dept_df
            }
        }
    }
```

```
except Exception as e:
    print(f"Error in entrypoint function: {str(e)}")
    raise e
```

在 PySpark 分析範本中使用參數

參數允許在任務提交時提供不同的值，以增加 PySpark 分析範本的彈性。您可以透過傳遞至進入點函數的內容物件來存取參數。

Note

參數是使用者提供的字串，可包含任意內容。

- 檢閱程式碼以確保安全地處理參數，以防止分析中的意外行為。
- 無論提交時提供哪些參數值，設計參數處理都能安全地運作。

存取參數

參數可在 `context['analysisParameters']` 字典中使用。所有參數值都是字串。

Example安全地存取參數

```
def entrypoint(context):
    # Access parameters from context
    parameters = context['analysisParameters']
    threshold = parameters['threshold']
    table_name = parameters['table_name']

    # Continue with analysis using parameters
    spark = context['sparkSession']
    input_df = context['referencedTables'][table_name]

    # Convert threshold value
    threshold_val = int(threshold)

    # Use parameter in DataFrame operation
    filtered_df = input_df.filter(input_df.amount > threshold_val)

    return {
```

```
"results": {
  "output": filtered_df
}
```

參數安全最佳實務

Warning

參數是使用者提供的字串，可包含任意內容。您必須安全地處理參數，以防止分析程式碼中的安全漏洞。

要避免的不安全參數處理模式：

- 以程式碼形式執行參數 – 切勿在參數值 `exec()` 上使用 `eval()` 或

```
# UNSAFE - Don't do this
eval(parameters['expression']) # Can execute arbitrary code
```

- SQL 字串插補 – 切勿將參數直接串連至 SQL 字串

```
# UNSAFE - Don't do this
sql = f"SELECT * FROM table WHERE column = '{parameters['value']}'" # SQL injection risk
```

- 不安全的檔案路徑操作 – 請勿在未驗證的情況下直接在檔案系統操作中使用參數

```
# UNSAFE - Don't do this
file_path = f"/data/{parameters['filename']}" # Path traversal risk
```

安全參數處理模式：

- 在 DataFrame 操作中使用參數 – Spark DataFrames 安全地處理參數值

```
# SAFE - Use parameters in DataFrame operations
threshold = int(parameters['threshold'])
filtered_df = input_df.filter(input_df.value > threshold)
```

- 驗證參數值 – 在使用前檢查參數是否符合預期的格式

```
# SAFE - Validate parameters before use
def validate_date(date_str):
    try:
        from datetime import datetime
        datetime.strptime(date_str, '%Y-%m-%d')
        return True
    except ValueError:
        return False

date_param = parameters['date_filter'] or '2024-01-01'
if not validate_date(date_param):
    raise ValueError(f"Invalid date format: {date_param}")
```

- 使用參數值的允許清單 – 如果可能，請根據已知的良好值驗證參數

```
# SAFE - Use allowlists
allowed_columns = ['column1', 'column2', 'column3']
column_param = parameters['column_name']
if column_param not in allowed_columns:
    raise ValueError(f"Invalid column: {column_param}")
```

- 使用錯誤處理進行類型轉換 – 將字串參數安全地轉換為預期類型

```
# SAFE - Convert with error handling
try:
    batch_size = int(parameters['batch_size'] or '1000')
    if batch_size <= 0 or batch_size > 10000:
        raise ValueError(f"Batch size must be between 1 and 10000")
except ValueError as e:
    print(f"Invalid parameter: {e}")
    raise
```

Important

請記住，當任務執行器提供不同的值時，參數會略過程式碼檢閱。設計您的參數處理方式，無論提供哪些參數值，都能安全地運作。

完成參數範例

Example在 PySpark 指令碼中安全地使用參數

```
def entrypoint(context):
    try:
        # Access Spark session and tables
        spark = context['sparkSession']
        input_table = context['referencedTables']['sales_data']

        # Access parameters - fail fast if analysisParameters missing
        parameters = context['analysisParameters']

        # Validate and convert numeric parameter (handles empty strings with default)
        try:
            threshold = int(parameters['threshold'] or '100')
            if threshold <= 0:
                raise ValueError("Threshold must be positive")
        except (ValueError, TypeError) as e:
            print(f"Invalid threshold parameter: {e}")
            raise

        # Validate date parameter (handles empty strings with default)
        date_filter = parameters['start_date'] or '2024-01-01'
        from datetime import datetime
        try:
            datetime.strptime(date_filter, '%Y-%m-%d')
        except ValueError:
            raise ValueError(f"Invalid date format: {date_filter}")

        # Use parameters safely in DataFrame operations
        filtered_df = input_table.filter(
            (input_table.amount > threshold) &
            (input_table.date >= date_filter)
        )

        result_df = filtered_df.groupBy("category").agg(
            {"amount": "sum"}
        )

        return {
            "results": {
                "filtered_results": result_df
            }
        }
```

```
}

except Exception as e:
    print(f"Error in analysis: {str(e)}")
    raise
```

建立虛擬環境（選用）

如果您有使用者指令碼所需的任何其他程式庫，您可以選擇建立虛擬環境來存放這些程式庫。如果您不需要其他程式庫，可以略過此步驟。

使用具有原生擴充功能的程式庫時，請記住 中的 PySpark 可在具有 ARM64 架構的 Linux 上 AWS Clean Rooms 運作。

下列程序示範如何使用基本 CLI 命令建立虛擬環境。

建立虛擬環境

1. 開啟終端機或命令提示字元。
2. 新增下列內容：

```
# create and activate a python virtual environment
python3 -m venv pyspark_venvsource
source pyspark_venvsource/bin/activate

# install the python packages
pip3 install pandas # add packages here

# package the virtual environment into an archive
pip3 install venv-pack
venv-pack -f -o pyspark_venv.tar.gz

# optionally, remove the virtual environment directory
deactivate
rm -fr pyspark_venvsource
```

3. 您現在已準備好將此虛擬環境存放在 S3 中。如需詳細資訊，請參閱[在 S3 中存放使用者指令碼和虛擬環境](#)。

如需使用 Docker 和 Amazon ECR 的詳細資訊，請參閱 [Amazon ECR User Guide](#)。

在 S3 中存放使用者指令碼和虛擬環境

下列程序說明如何在 Amazon S3 中存放使用者指令碼和選用的虛擬環境。在建立 PySpark 分析範本之前，請先完成此步驟。

Important

建立分析範本後，請勿修改或移除成品（使用者指令碼或虛擬環境）。這樣做將：

- 導致使用此範本的所有未來分析任務失敗。
- 需要建立具有新成品的新分析範本。
- 不會影響先前完成的分析任務

先決條件

- AWS 帳戶 具有適當許可的
- 使用者指令碼檔案（例如 `my_analysis.py`）
- （選用，如果有的話）虛擬環境套件（`.tar.gz` 檔案）
- 建立或修改 IAM 角色的存取權

Console

若要使用主控台將使用者指令碼和虛擬環境存放在 S3 中：

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/s3/>：// 開啟 Amazon S3 主控台。
2. 建立新的 S3 儲存貯體或使用現有的儲存貯體。
3. 啟用儲存貯體的版本控制。
 - a. 選取您的儲存貯體。
 - b. 選擇 Properties (屬性)。
 - c. 在儲存貯體版本控制區段中，選擇編輯。
 - d. 選取啟用並儲存變更。
4. 上傳成品並啟用 SHA-256 雜湊。

- a. 導覽至您的儲存貯體。
 - b. 選擇上傳。
 - c. 選擇新增檔案並新增您的使用者指令碼檔案。
 - d. （選用，如果有的話）新增您的 .tar.gz 檔案。
 - e. 展開屬性。
 - f. 在檢查總和下，針對檢查總和函數，選取 SHA256。
 - g. 選擇上傳。
5. 您現在已準備好建立 PySpark 分析範本。

CLI

若要使用 將使用者指令碼和虛擬環境存放在 S3 中 AWS CLI：

1. 執行以下命令：

```
aws s3 cp --checksum-algorithm sha256 pyspark_venv.tar.gz s3://ARTIFACT-BUCKET/EXAMPLE-PREFIX/
```

2. 您現在已準備好建立 PySpark 分析範本。

Note

如果您需要更新指令碼或虛擬環境：

1. 將新版本上傳為單獨的物件。
2. 使用新的成品建立新的分析範本。
3. 棄用舊範本。
4. 如果仍然需要舊範本，請將原始成品保留在 S3 中。

建立 PySpark 分析範本

Note

參數是使用者提供的字串，可包含任意內容。

- 檢閱程式碼以確保安全地處理參數，以防止分析中的意外行為。
- 無論提交時提供哪些參數值，設計參數處理都能安全地運作。

先決條件

建立 PySpark 分析範本之前，您必須具有：

- 作用中 AWS Clean Rooms 協同合作的成員資格
- 存取作用中協同合作中至少一個已設定的資料表
- 建立分析範本的許可
- 在 S3 中建立和存放的 Python 使用者指令碼和虛擬環境
 - S3 儲存貯體已啟用版本控制。如需詳細資訊，請參閱[在 S3 儲存貯體中使用版本控制](#)
 - S3 儲存貯體可以計算上傳成品的 SHA-256 檢查總和。如需詳細資訊，請參閱[使用檢查總和](#)
- 從 S3 儲存貯體讀取程式碼的許可

如需建立所需服務角色的詳細資訊，請參閱 [建立服務角色以從 S3 儲存貯體讀取程式碼 \(PySpark 分析範本角色\)](#)。

下列程序說明使用[AWS Clean Rooms 主控台](#)建立 PySpark 分析範本的程序。其假設您已建立使用者指令碼和虛擬環境檔案，並將使用者指令碼和虛擬環境檔案存放在 Amazon S3 儲存貯體中。

Note

建立 PySpark 分析範本的成員也必須是收到結果的成員。

如需有關如何使用 AWS SDKs 建立 PySpark 分析範本的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

建立 PySpark 分析範本

1. 登入 AWS 管理主控台，並使用開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶，以做為協作建立者。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。

4. 在範本索引標籤上，前往您建立的分析範本區段。
5. 選擇建立分析範本。
6. 在建立分析範本頁面上，如需詳細資訊，
 - a. 輸入分析範本的名稱。
 - b. (選用) 輸入描述。
 - c. 針對格式，選擇 PySpark 選項。
7. 對於定義，
 - a. 檢閱先決條件，並確保符合每個先決條件，然後再繼續。
 - b. 對於進入點檔案，輸入 S3 儲存貯體或選擇瀏覽 S3。
 - c. (選用) 對於程式庫檔案，輸入 S3 儲存貯體或選擇瀏覽 S3。
8. 對於參數 - 選用，如果您想要新增參數以使分析範本可重複使用：
 - a. 選擇 Add parameter (新增參數)。
 - b. 輸入參數名稱。

參數名稱必須以字母或底線開頭，後面接著英數字元或底線。

- c. 對於類型，會自動選取 STRING 作為 PySpark 分析範本唯一支援的類型。
- d. (選用) 輸入參數的預設值。

如果您提供預設值，任務執行器可以在執行任務時使用此值，而無需明確提供參數值。

- e. 若要新增更多參數，請選擇新增另一個參數並重複上述步驟。

Note

每個 PySpark 分析範本最多可以定義 50 個參數。每個參數值最多可達 1,000 個字元。

9. 對於定義中參考的資料表，
 - 如果定義中參考的所有資料表都已與協同合作相關聯：
 - 保留定義中參考的所有資料表已與選取的協同合作核取方塊相關聯。
 - 在與協同合作相關聯的資料表下，選擇定義中參考的所有相關資料表。
 - 如果定義中參考的所有資料表尚未與協同合作相關聯：
 - 清除定義中參考的所有資料表都已與協同合作核取方塊相關聯。

- 在與協同合作相關聯的資料表下，選擇定義中參考的所有相關資料表。
- 在稍後相關聯的資料表下，輸入資料表名稱。
- 選擇列出另一個資料表以列出另一個資料表。

10. 針對錯誤訊息組態，選擇下列其中一項：

- 基本錯誤訊息 – 傳回基本錯誤訊息，而不會公開基礎資料。建議用於生產工作負載。
- 詳細的錯誤訊息 – 傳回詳細的錯誤訊息，以加快故障診斷速度。建議在開發和測試環境中使用。可能會公開敏感資料，包括個人身分識別資訊 (PII)。

 Note

使用詳細錯誤訊息時，所有資料提供者成員都必須核准範本的此設定。

11. 從下拉式清單中選取現有的服務角色名稱，以指定服務存取許可。

1. 如果您具有列出角色的許可，則會顯示角色清單。

如果您沒有列出角色的許可，您可以輸入要使用的角色的 Amazon Resource Name (ARN)。

2. 選擇 IAM 外部連結中的檢視，以檢視服務角色。

如果沒有現有的服務角色，則無法使用使用現有服務角色的選項。

根據預設，AWS Clean Rooms 不會嘗試更新現有的角色政策來新增必要的許可。

 Note

- AWS Clean Rooms 需要根據分析規則查詢的許可。如需許可的詳細資訊 AWS Clean Rooms，請參閱 [AWS 的 受管政策 AWS Clean Rooms](#)。
- 如果角色沒有足夠的許可 AWS Clean Rooms，您會收到錯誤訊息，指出角色沒有足夠的許可給 AWS Clean Rooms。必須先新增角色政策，才能繼續。
- 如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

12. 如果您想要為設定的資料表資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。

13. 選擇建立。

14. 您現在可以通知協同合作成員，他們可以[檢閱分析範本](#)。（如果您想要查詢自己的資料，則為選用。）

Important

建立分析範本後，請勿修改或移除成品（使用者指令碼或虛擬環境）。這樣做將：

- 導致使用此範本的所有未來分析任務失敗。
- 需要建立具有新成品的新分析範本。
- 不會影響先前完成的分析任務。

檢閱 PySpark 分析範本

當其他成員在協同合作中建立分析範本時，您必須先檢閱並核准，才能使用它。

下列程序說明如何檢閱 PySpark 分析範本，包括其規則、參數和參考資料表。身為協同合作成員，您將評估範本是否符合您的資料共用協議和安全要求。

分析範本和核准後，即可用於中的任務 AWS Clean Rooms。

Note

當您將分析程式碼帶入協同合作時，請注意下列事項：

- AWS Clean Rooms 不會驗證或保證分析程式碼的行為。
 - 如果您需要確保特定行為，請直接檢閱協同合作夥伴的程式碼，或與信任的第三方稽核人員合作進行檢閱。
- AWS Clean Rooms 保證 PySpark 分析範本中所列程式碼的 SHA-256 雜湊符合 PySpark 分析環境中執行的程式碼。
- AWS Clean Rooms 對於您帶入環境的其他程式庫，不會執行任何稽核或安全分析。
- 在共用安全模型中：
 - 您（客戶）負責保護環境中執行之程式碼的安全性。
 - 您（客戶）負責為環境設定適當的錯誤訊息組態。
 - AWS Clean Rooms 負責環境的安全，確保

- 只有核准的程式碼會執行
- 只能存取指定的已設定資料表
- 唯一的輸出目的地是結果接收者的 S3 儲存貯體。

AWS Clean Rooms 會產生使用者指令碼和虛擬環境的 SHA-256 雜湊，供您檢閱。不過，實際的使用者指令碼和程式庫無法直接在內存取 AWS Clean Rooms。

若要驗證共用的使用者指令碼和程式庫與分析範本中參考的指令碼和程式庫相同，您可以建立共用檔案的 SHA-256 雜湊，並將其與建立的分析範本雜湊進行比較 AWS Clean Rooms。程式碼執行的雜湊也會在任務日誌中。

先決條件

- Linux/Unix 作業系統或適用於 Linux 的 Windows 子系統 (WSL)
- 您要雜湊的使用者指令碼檔案
 - 請求分析範本建立者透過安全管道共用檔案。
- 建立的分析範本雜湊 AWS Clean Rooms

使用 AWS Clean Rooms 主控台檢閱 PySpark 分析範本

1. 登入 AWS 管理主控台，並使用開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶，以做為協同合作建立者。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在範本索引標籤上，前往其他成員建立的分析範本區段。
5. 選擇可以執行狀態為否的分析範本需要您的檢閱。
6. 選擇檢閱。
7. 檢閱分析規則概觀、定義和參數（如果有的話）。

Note

參數允許分析執行器在提交時提交不同的值。如果分析範本支援參數，請檢閱如何在協同合作夥伴的程式碼中使用參數值，以確保其符合您的需求。

8. 驗證共用的使用者指令碼和程式庫是否與分析範本中參考的使用者指令碼和程式庫相同。

- a. 建立共用檔案的 SHA-256 雜湊，並將其與 建立的分析範本雜湊進行比較 AWS Clean Rooms。

您可以導覽至包含使用者指令碼檔案的目錄，然後執行下列命令來產生雜湊：

```
sha256sum your_script_filename.py
```

輸出範例：

```
e3b0c44298fc1c149afb4c8996fb92427ae41e4649b934ca495991b7852b855 my_analysis.py
```

- b. 或者，您可以使用 Amazon S3 檢查總和功能。如需詳細資訊，請參閱 [《Amazon S3 使用者指南》中的在 Amazon S3 中檢查物件完整性](#)。Amazon S3
 - c. 另一個替代方法是檢視任務日誌中已執行程式碼的雜湊。
9. 檢閱定義中參考的資料表下列出的已設定資料表。

每個資料表旁的狀態將讀取不允許的範本。
 10. 選擇 表格。
 - a. 若要核准分析範本，請選擇允許資料表上的範本。選擇允許，確認您的核准。
 - b. 若要拒絕核准，請選擇不允許。

如果您選擇核准分析範本，可執行任務的成員現在可以使用 PySpark 分析範本，在設定的資料表上執行 PySpark 任務。如需詳細資訊，請參閱[執行 PySpark 任務](#)。

故障診斷 PySpark 分析範本

使用 PySpark 分析範本執行任務時，您可能會在任務初始化或執行期間遇到失敗。這些失敗通常與指令碼組態、資料存取許可或環境設定相關。

如需 PySpark 限制的詳細資訊，請參閱 [中的 PySpark 限制 AWS Clean Rooms](#)。

主題

- [對程式碼進行故障診斷](#)
- [分析範本任務未啟動](#)
- [分析範本任務開始，但在處理期間失敗](#)

- [虛擬環境設定失敗](#)

對程式碼進行故障診斷

為了協助您開發和疑難排解程式碼，我們建議您啟用詳細的錯誤訊息並使用自己的測試資料執行任務 AWS Clean Rooms AWS 帳戶，以自行模擬。

您也可以透過下列步驟，在 Amazon EMR Serverless 中模擬 AWS Clean Rooms 中的 PySpark。它與 AWS Clean Rooms 中的 PySpark 會有些微差異，但主要涵蓋程式碼的執行方式。

在 EMR Serverless 中模擬 AWS Clean Rooms 中的 PySpark

1. 在 Amazon S3 中建立資料集，將其編目在 AWS Glue Data Catalog，並設定 Lake Formation 許可。
2. 使用自訂角色向 Lake Formation 註冊 S3 位置。
3. 如果您還沒有 Amazon EMR Studio 執行個體，請建立執行個體（需要 Amazon EMR Studio 才能使用 Amazon EMR Serverless）。
4. 建立 EMR Serverless 應用程式
 - 選取發行版本 emr-7.7.0。
 - 選取 ARM64 架構。
 - 選擇使用自訂設定。
 - 停用預先初始化的容量。
 - 如果您打算執行互動式工作，請選取互動式端點 > 為 EMR Studio 啟用端點。
 - 選取其他組態 > 使用 Lake Formation 進行精細存取控制。
 - 建立應用程式。
5. 透過 EMR-Studio 筆記本或 StartJobRun API 使用 EMR-S。

分析範本任務未啟動

常見原因

由於三個主要組態問題，分析範本任務可能會在啟動時立即失敗：

- 不正確的指令碼命名不符合所需的格式

- 使用者指令碼中遺失或格式不正確的進入點函數
- 虛擬環境中不相容的 Python 版本

Resolution

若要解決問題：

1. 驗證您的使用者指令碼：

- 檢查您的使用者指令碼是否具有有效的 Python 檔案名稱。

有效的 Python 檔案名稱使用小寫字母、底線分隔單字和 .py 副檔名。

2. 驗證進入點函數。如果您的使用者指令碼沒有進入點函數，請新增一個。

- a. 開啟您的使用者指令碼。
- b. 新增此進入點函數：

```
def entrypoint(context):  
    # Your analysis code here
```

- c. 確保函數名稱的拼寫與 完全相同entrypoint。
- d. 確認函數接受 context 參數。

3. 檢查 Python 版本相容性：

- a. 驗證您的虛擬環境是否使用 Python 3.9 或 3.11。
- b. 若要檢查您的版本，請執行：python --version
- c. 如有需要，請更新您的虛擬環境：

```
conda create -n analysis-env python=3.9  
conda activate analysis-env
```

預防

- 使用提供的分析範本入門程式碼，其中包含正確的檔案結構。
- 使用 Python 3.9 或 3.11 為所有分析範本設定專用虛擬環境。
- 在提交任務之前，使用範本驗證工具在本機測試您的分析範本。

- 實作 CI/CD 檢查來驗證指令碼命名和進入點函數需求。

分析範本任務開始，但在處理期間失敗

常見原因

基於這些安全性和格式原因，分析任務可能會在執行期間失敗：

- 對 Amazon S3 或等 AWS 服務進行未經授權的直接存取嘗試 AWS Glue
- 以不符合所需 DataFrame 規格的不正確格式傳回輸出
- 由於執行環境中的安全限制而封鎖網路呼叫

Resolution

若要解決

1. 移除直接 AWS 服務存取：
 - a. 搜尋您的程式碼以取得直接 AWS 服務匯入和呼叫。
 - b. 使用提供的 Spark 工作階段方法取代直接 S3 存取。
 - c. 只能透過協作界面使用預先設定的資料表。
2. 正確格式化輸出：
 - a. 確認所有輸出都是 Spark DataFrames
 - b. 更新您的傳回陳述式以符合此格式：

```
return {  
  "results": {  
    "output1": dataframe1  
  }  
}
```

- c. 移除任何非 DataFrame 傳回物件。
3. 移除網路呼叫：
 - a. 識別並移除任何外部 API 呼叫。
 - b. 移除任何 urllib、請求或類似的網路程式庫。

- c. 移除任何通訊端連線或 HTTP 用戶端程式碼。

預防

- 使用提供的程式碼包來檢查未經授權的 AWS 匯入和網路呼叫。
- 在安全限制符合生產的開發環境中測試任務。
- 部署任務之前，請遵循輸出結構描述驗證程序。
- 檢閱已核准服務存取模式的安全準則。

虛擬環境設定失敗

常見原因

虛擬環境組態失敗通常原因如下：

- 開發與執行環境之間的 CPU 架構不相符
- Python 程式碼格式化問題會阻止適當的環境初始化
- 容器設定中的基礎映像組態不正確

Resolution

若要解決

1. 設定正確的架構：

- a. 使用 檢查您目前的架構 `uname -m`。
- b. 更新您的 Dockerfile 以指定 ARM64：

```
FROM --platform=linux/arm64 public.ecr.aws/amazonlinux/amazonlinux:2023-minimal
```

- c. 使用 重建您的容器 `docker build --platform=linux/arm64`。

2. 修正 Python 縮排：

- a. 在您的程式碼檔案 `black` 上執行 Python 程式碼格式化程式，如。
- b. 驗證是否一致地使用空格或標籤（非兩者）。
- c. 檢查所有程式碼區塊的縮排：

```
def my_function():  
    if condition:  
        do_something()  
    return result
```

- d. 使用 IDE 搭配 Python 縮排反白顯示。
3. 驗證環境組態：
 - a. 執行 `python -m py_compile your_script.py` 以檢查語法錯誤。
 - b. 在部署之前在本機測試環境。
 - c. 確認所有相依性都列在 `requirements.txt` 中。

預防

- 使用 Visual Studio Code 或 PyCharm 搭配 Python 格式外掛程式
- 設定預先遞交掛鉤以自動執行程式碼格式化工具
- 使用提供的 ARM64 基礎映像在本機建置和測試環境
- 在 CI/CD 管道中實作自動化程式碼樣式檢查

分析協同合作中的資料

在 [AWS Clean Rooms](#)，您可以透過執行查詢或任務來分析資料。

查詢是一種方法，可使用支援的一組函數、類別和變數來存取和分析協同合作中設定的資料表。目前支援的查詢語言 [AWS Clean Rooms](#) 為 SQL。有 3 種方式可在 [AWS Clean Rooms](#) 中執行查詢 [AWS Clean Rooms](#)：撰寫 SQL 程式碼、使用核准的 SQL 分析範本，或使用分析建置器 UI。

任務是一種使用支援的一組函數、類別和變數，在協同合作中存取和分析已設定資料表的方法。目前支援的任務類型 [AWS Clean Rooms](#) 為 PySpark。有一種方式可在 [AWS Clean Rooms](#) 中執行任務 [AWS Clean Rooms](#)：使用核准的 PySpark 分析範本。

下列主題說明如何 [AWS Clean Rooms](#) 透過執行 SQL 查詢或 PySpark 任務來分析 [AWS Clean Rooms](#) 中的資料。

主題

- [執行 SQL 查詢](#)
- [執行 PySpark 任務](#)

執行 SQL 查詢

Note

只有在負責支付查詢運算成本的成員以作用中成員的身分加入協同合作時，您才能執行查詢。

身為[可以查詢的成員](#)，您可以透過下列方式執行 SQL 查詢：

- 使用 SQL 程式碼編輯器手動建置 SQL 查詢。
- 使用核准的 SQL [分析範本](#)。
- 使用分析建置器 UI 來建置查詢，而不必撰寫 SQL 程式碼。

當可查詢的成員在協同合作中的資料表上執行 SQL 查詢時，[AWS Clean Rooms](#) 擔任相關角色來代表他們存取資料表。[AWS Clean Rooms](#) 會視需要將分析規則 [AWS Clean Rooms](#) 套用至輸入查詢及其輸出。

分析規則和輸出限制條件會自動強制執行。[AWS Clean Rooms](#) 只會傳回符合定義分析規則的結果。

AWS Clean Rooms 支援與其他查詢引擎不同的 SQL 查詢。如需規格，請參閱 [AWS Clean Rooms SQL 參考](#)。如果您想要對受差異隱私權保護的資料表執行查詢，則應確保您的查詢與差異隱私權的[一般用途查詢結構](#) AWS Clean Rooms 相容。

Note

使用適用於 [加密運算時Clean Rooms](#)，並非所有 SQL 操作都會產生有效的結果。例如，您可以在加密資料欄COUNT上執行，但在加密數字SUM上執行會導致錯誤。此外，查詢也可能產生不正確的結果。例如，SUM密封資料欄的查詢會產生錯誤。不過，對密封資料欄的GROUPBY查詢似乎成功，但產生的群組與透過純文字GROUPBY查詢產生的群組不同。

[支付查詢運算成本的成員](#)會針對協同合作中執行的查詢付費。

可以查詢的成員可以選取多個[可以接收結果的成員](#)，以接收來自單一查詢的結果。如需詳細資訊，請參閱[使用 SQL 程式碼編輯器查詢設定的資料表](#)。如需接收查詢結果的一般資訊，請參閱[接收和使用分析結果](#)。

先決條件

執行 SQL 查詢之前，請確定您有下列項目：

- AWS Clean Rooms 協作中的作用中成員資格
- 在協同合作中存取至少一個已設定的資料表
- 確認負責查詢運算成本的成員是作用中的協同合作成員

如需有關如何直接呼叫 AWS Clean Rooms [StartProtectedQuery API](#) 操作或使用 AWS SDKs 查詢資料或檢視查詢的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

如需查詢記錄的資訊，請參閱 [分析登入 AWS Clean Rooms](#)。

Note

如果您在[加密](#)的資料表上執行查詢，加密資料欄的結果會加密。

SQL 查詢的 Spark 屬性組態

AWS Clean Rooms 可讓您設定 SQL 查詢支援的 Spark 屬性，選擇性地自訂 Spark 執行時間行為。這些屬性可讓您微調效能、記憶體用量和查詢執行參數。透過此功能，您可以更好地控制 Spark 型查詢的處理方式，從而根據您的特定工作負載需求進行最佳化。

您現在可以直接從 AWS Clean Rooms 主控台調整隨機播放分割區、廣播聯結閾值和自適應查詢執行參數等設定。此功能在預設組態可能不是最佳的複雜查詢或大型資料集中特別有用。透過微調這些 Spark 屬性，您可以改善查詢效能、減少資源消耗，以及更好地管理 Spark 型協同合作分析的記憶體用量。

若要利用此功能，您可以在查詢界面中找到新的 Spark 屬性區段。您可以從支援的屬性清單中選取，並指定自訂值。您也可以使用 [StartProtectedQuery API，以程式設計方式設定 Spark 屬性](#)。此進階組態選項可讓資料分析師和工程師最佳化其查詢，以提升效率和可擴展性。

如需 Spark 屬性的詳細資訊，包括預設值，請參閱 Apache Spark 文件中的 Spark [屬性](#)。

下列主題說明如何使用 AWS Clean Rooms 主控台查詢協同合作中的資料。

主題

- [使用 SQL 程式碼編輯器查詢設定的資料表](#)
- [使用 SQL 程式碼編輯器查詢 ID 映射表](#)
- [使用 SQL 分析範本查詢設定的資料表](#)
- [使用分析建置器查詢](#)
- [檢視差異隱私權的影響](#)
- [檢視近期查詢](#)
- [檢視查詢詳細資訊](#)

使用 SQL 程式碼編輯器查詢設定的資料表

身為可以查詢的成員，您可以在 SQL 程式碼編輯器中撰寫 SQL 程式碼，手動建置查詢。SQL 程式碼編輯器位於 AWS Clean Rooms 主控台分析索引標籤的分析區段。

預設會顯示 SQL 程式碼編輯器。如果您想要使用分析建置器來建置查詢，請參閱 [使用分析建置器查詢](#)。

⚠ Important

如果您開始在程式碼編輯器中寫入 SQL 查詢，然後開啟分析建置器 UI，則不會儲存您的查詢。

AWS Clean Rooms 支援許多 SQL 命令、函數和條件。如需詳細資訊，請參閱 [AWS Clean Rooms SQL 參考](#)。

💡 Tip

如果在查詢執行時發生排程維護，查詢會終止並復原。您必須重新啟動查詢。

使用 SQL 程式碼編輯器查詢設定的資料表

1. 登入 AWS 管理主控台 並開啟位於 <https://console.aws.amazon.com/cleanrooms> 的 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇具有執行查詢成員能力狀態的協同合作。
4. 在分析索引標籤的資料表下，檢視資料表清單及其相關聯的分析規則類型（彙總分析規則、清單分析規則或自訂分析規則）。

💡 Note

如果您沒有在清單中看到預期的資料表，可能是因為下列原因：

- 資料表尚未 [建立關聯](#)。
- 資料表未 [設定分析規則](#)。

5. （選用）若要檢視資料表的結構描述和分析規則控制項，請選取加號圖示 (+) 展開資料表。
6. 在分析區段下，針對分析模式，選取寫入 SQL 程式碼。

💡 Note

分析區段只會在可接收結果的成員，以及負責支付查詢運算成本的成員以作用中成員的身分加入協同合作時顯示。

7. 在 SQL 程式碼編輯器中輸入查詢來建置查詢。

如需支援的 SQL 命令和函數的詳細資訊，請參閱 [AWS Clean Rooms SQL 參考](#)。

您也可以使用下列選項來建置查詢。

Use an example query

使用範例查詢

1. 選取資料表旁的三個垂直點。
2. 在插入編輯器下，選擇範例查詢。

Note

插入範例查詢會將其附加至編輯器中已有的查詢。

查詢範例隨即出現。資料表下列出的所有資料表都會包含在查詢中。

3. 編輯查詢中的預留位置值。

Insert column names or functions

插入資料欄名稱或函數

1. 選取資料欄旁的三個垂直點。
2. 在插入編輯器下，選擇資料欄名稱。
3. 若要手動插入資料欄上允許的函數，
 - a. 選取資料欄旁的三個垂直點。
 - b. 選取在編輯器中插入。
 - c. 選取允許的函數名稱（例如 INNER JOIN、SUM DISTINCT、SUM 或 COUNT）。
4. 按 Ctrl + Space 在程式碼編輯器中檢視資料表結構描述。

Note

可以查詢的成員可以檢視和使用每個已設定資料表關聯中的分割區資料欄。確保分割區資料欄在已設定 AWS Glue 資料表的基礎資料表中標記為分割區資料欄。

5. 編輯查詢中的預留位置值。
8. 指定支援的工作者類型和工作者數量。

您可以選擇執行個體類型和執行個體數量（工作者）來執行 SQL 查詢。

對於 CR.1X，您最多可以選取 128 個工作者或至少 4 個工作者。

對於 CR.4X，您最多可以選取 32 個工作者或至少 4 個工作者。

使用下表來判斷使用案例所需的類型和數量或工作者。

工作者類型	vCPU	記憶體 (GB)	儲存體 (GB)	工作者數目	無塵室處理單元總數 (CRPU)
CR.1X（預設）	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

Note

不同的工作者類型和工作者數量都有相關聯的成本。若要進一步了解定價，請參閱 [AWS Clean Rooms 定價](#)。

9. 針對傳送結果至，指定誰可以接收結果。

Note

若要接收結果，協作成員必須設定為結果接收者，而且必須是協作中的作用中參與者（狀態：作用中）

10. （只能查詢的成員）預設會選取使用您的預設結果設定核取方塊。如果您想要保留預設結果設定，請保持選取此選項。

如果您想要為此查詢指定不同的結果設定，請清除使用您的預設結果設定核取方塊，然後選擇下列選項。

- a. 結果格式 (CSV 或 PARQUET)
- b. 結果檔案 (單一或多個)
- c. Amazon S3 中的結果目的地

每個可以接收結果的成員可以在 Amazon S3 中指定不同的結果格式、結果檔案和結果目的地。
Amazon S3

11. 若要指定 Spark 屬性：

- a. 展開 Spark 屬性。
- b. 選擇新增 Spark 屬性。
- c. 在 Spark 屬性對話方塊中，從下拉式清單中選擇屬性名稱，然後輸入值。

下表提供每個屬性的定義。

如需 Spark 屬性的詳細資訊，請參閱 Apache Spark 文件中的 Spark [屬性](#)。

屬性名稱	Description	預設值
spark.task.maxFailures	控制任務在任務失敗之前可以失敗的連續次數。需要大於或等於 1 的值。允許重試的次數等於此值減去 1。如果任何嘗試成功，就會重設失敗計數。不同任務的故障不會累積到此限制。	4
spark.sql.files.maxPartitionBytes	設定從 Parquet、JSON 和 ORC 等檔案型來源讀取時封裝到單一分割區的最大位元組數。	128MB
spark.hadoop.fs.s3.maxRetries	設定 Amazon S3 檔案操作的重試嘗試次數上限。	
spark.network.timeout	設定所有網路互動的預設逾時。如果未設定，則覆寫下列逾時設定：	120 秒

屬性名稱	Description	預設值
	- spark.storage.blockManagerHeartbeatTimeoutMs - spark.shuffle.io.connectionTimeout - spark.rpc.askTimeout - spark.rpc.lookupTimeout	
spark.rdd.compress	指定是否使用 spark.io.compression.codec 壓縮序列化 RDD 分割區。適用於 Java 和 Scala 中的 StorageLevel.MEMORY_ONLY_SER，或 Python 中的 StorageLevel.MEMORY_ONLY。減少儲存空間，但需要額外的 CPU 處理時間。	FALSE
spark.shuffle.spill.compress	指定是否使用 spark.io.compression.codec 壓縮隨機溢出資料。	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	在 spark.sql.adaptive.enabled 為 true 的適應性最佳化期間，設定隨機播放分割區的目標大小，以位元組為單位。合併小型分割區或分割偏斜分割區時，控制分割區大小。	(spark.sql.adaptive.shuffle.targetPostShuffleInputSize 的值)
spark.sql.adaptive.autoBroadcastJoinThreshold	設定在聯結期間廣播至工作者節點的資料表大小上限，以位元組為單位。僅適用於自適應架構。使用與 spark.sql.autoBroadcastJoinThreshold 相同的預設值。設定為 -1 以停用廣播。	(無)
spark.sql.adaptive.coalescePartitions.enabled	指定是否要根據 spark.sql.adaptive.advisoryPartitionSizeInBytes 來合併連續隨機播放分割區，以最佳化任務大小。需要 spark.sql.adaptive.enabled 才能成立。	TRUE

屬性名稱	Description	預設值
spark.sql.adaptive.coalescePartitions.initialPartitionNum	在合併之前定義隨機播放分割區的初始數量。同時需要 spark.sql.adaptive.enabled 和 spark.sql.adaptive.coalescePartitions.enabled 為 true。預設為 spark.sql.shuffle.partitions 的值。	(無)
spark.sql.adaptive.coalescePartitions.minPartitionSize	設定合併隨機播放分割區的大小下限，以防止分割區在適應性最佳化期間變得太小。	1 MB
spark.sql.adaptive.coalescePartitions.parallelismFirst	指定是否要在分割區合併期間根據叢集平行處理而非 spark.sql.adaptive.advisoryPartitionSizeInBytes 計算分割區大小。產生的分割區大小小於設定的目標大小，以最大化平行處理。我們建議在忙碌叢集上將此設定為 false，以透過防止過多的小型任務來改善資源使用率。	TRUE
spark.sql.adaptive.enabled	指定是否啟用自適應查詢執行，以在查詢執行期間根據準確的執行時間統計資料重新最佳化查詢計劃。	TRUE
spark.sql.adaptive.forceOptimizeSkewedJoin	指定是否強制啟用 OptimizeSkewedJoin，即使它引入額外的隨機播放。	FALSE
spark.sql.adaptive.localShuffleReader.enabled	指定是否在不需要隨機分割時使用本機隨機讀取器，例如從排序合併聯結轉換為廣播雜湊聯結之後。需要 spark.sql.adaptive.enabled 才能成立。	TRUE

屬性名稱	Description	預設值
spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold	設定用於建置本機雜湊映射的分割區大小上限，以位元組為單位。在下列情況下，優先考慮排序合併聯結的隨機雜湊聯結： - 此值等於或超過 spark.sql.adaptive.advisoryPartitionSizeInBytes - 所有分割區大小都在此限制內 覆寫 spark.sql.join.preferSortMergeJoin 設定。	0 個位元組
spark.sql.adaptive.optimizeSkewsInRebalancePartitions.enabled	指定是否根據 spark.sql.adaptive.advisoryPartitionSizeInBytes 將扭曲隨機播放分割區分割為較小的分割區，以最佳化扭曲隨機播放分割區。需要 spark.sql.adaptive.enabled 才能成立。	TRUE
spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor	定義分割期間合併分割區的大小閾值因素。小於此因素的分割區乘以 spark.sql.adaptive.advisoryPartitionSizeInBytes 會合併。	0.2
spark.sql.adaptive.skewJoin.enabled	指定是否透過分割和選擇性複寫扭曲的分割區來處理隨機聯結中的資料扭曲。適用於排序合併和隨機雜湊聯結。需要 spark.sql.adaptive.enabled 才能成立。	TRUE
spark.sql.adaptive.skewJoin.skewedPartitionFactor	決定決定分割區扭曲的大小因素。當分割區的大小超過以下兩者時，就會扭曲： - 此係數乘以中位數分割區大小 - spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes 的值	5

屬性名稱	Description	預設值
spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes	設定識別扭曲分割區的大小閾值，以位元組為單位。當分割區的大小超過以下兩者時，就會扭曲： - 此閾值 - 中位數分割區大小乘以 spark.sql.adaptive.skewJoin.skewedPartitionFactor 我們建議將此值設定為大於 spark.sql.adaptive.advisoryPartitionSizeInBytes。	256MB
spark.sql.autoBroadcastJoinThreshold	設定在聯結期間廣播至工作者節點的資料表大小上限，以位元組為單位。設定為 -1 以停用廣播。	10MB
spark.sql.broadcastTimeout	控制廣播聯結期間廣播操作的逾時期間，以秒為單位。	300 秒
spark.sql.cbo.enabled	指定是否要為計劃統計資料估算啟用成本型最佳化 (CBO)。	FALSE
spark.sql.cbo.joinReorder.dp.star.filter	指定是否要在成本型聯結列舉期間套用星狀聯結篩選條件啟發式。	FALSE
spark.sql.cbo.joinReorder.dp.threshold	設定動態程式設計演算法中允許的聯結節點數目上限。	12
spark.sql.cbo.joinReorder.enabled	指定是否要在成本型最佳化 (CBO) 中啟用聯結重新排序。	FALSE
spark.sql.cbo.planStats.enabled	指定是否要在邏輯計畫產生期間從目錄中擷取資料列計數和資料欄統計資料。	FALSE
spark.sql.cbo.starSchemaDetection	指定是否根據星狀結構描述偵測啟用聯結重新排序。	FALSE

屬性名稱	Description	預設值
spark.sql.files.maxPartitionNum	設定檔案型來源 (Parquet、JSON 和 ORC) 的目標分割檔案分割區數量上限。當初始計數超過此值時，重新擴展分割區。這是建議的目標，而不是保證的限制。	(無)
spark.sql.files.maxRecordsPerFile	設定寫入單一檔案的記錄數目上限。設定為零或負值時不會套用任何限制。	0
spark.sql.files.minPartitionNum	設定檔案型來源 (Parquet、JSON 和 ORC) 的目標分割檔案分割區數量下限。預設為 spark.sql.leafNodeDefaultParallelism。這是建議的目標，而不是保證的限制。	(無)
spark.sql.inMemoryColumnarStorage.batchSize	控制單欄式快取的批次大小。增加大小可改善記憶體使用率和壓縮，但會增加out-of-memory錯誤的風險。	10000
spark.sql.inMemoryColumnarStorage.compressed	指定是否根據資料統計資料自動選取資料欄的壓縮轉碼器。	TRUE
spark.sql.inMemoryColumnarStorage.enableVectorizedReader	指定是否要為單欄式快取啟用向量化讀取。	TRUE
spark.sql.legacy.allowHashOnMapType	指定是否允許對映射類型資料結構執行雜湊操作。此舊版設定可維持與舊版 Spark 版本的映射類型處理的相容性。	
spark.sql.legacy.allowNegativeScaleOfDecimal	指定是否在十進位類型定義中允許負縮放值。此舊版設定可維持與支援負小數位數擴展之較舊 Spark 版本的相容性。	

屬性名稱	Description	預設值
spark.sql.legacy.castComplexTypesToString.enabled	指定是否啟用舊版行為，將複雜類型轉換為字串。維持與舊版 Spark 版本類型轉換規則的相容性。	
spark.sql.legacy.charVarcharAsString	指定是否將 CHAR 和 VARCHAR 類型視為 STRING 類型。此舊版設定提供與較舊 Spark 版本的字串類型處理相容性。	
spark.sql.legacy.createEmptyCollectionUsingStringType	指定是否使用字串類型元素建立空集合。此舊版設定可維持與較舊 Spark 版本的集合初始化行為的相容性。	
spark.sql.legacy.exponentLiteralAsDecimal.enabled	指定是否要將指數常值解譯為十進位類型。此舊版設定可維持與舊版 Spark 版本數值常值處理的相容性。	
spark.sql.legacy.json.allowEmptyString.enabled	指定是否要在 JSON 處理中允許空字串。此舊版設定可維持與舊版 Spark 版本的 JSON 剖析行為的相容性。	
spark.sql.legacy.parquet.int96RebaseModelRead	指定是否在讀取 Parquet 檔案時使用舊版 INT96 時間戳記重新基礎模式。此舊版設定可維持與較舊 Spark 版本時間戳記處理的相容性。	
spark.sql.legacy.timeParserPolicy	控制時間剖析行為以實現回溯相容性。此舊版設定會決定如何從字串剖析時間戳記和日期。	
spark.sql.legacy.typeCoercion.datetimeToString.enabled	指定是否在將日期時間值轉換為字串時啟用舊版類型強制行為。維持與較舊 Spark 版本日期時間轉換規則的相容性。	
spark.sql.maxSinglePartitionBytes	設定以位元組為單位的分割區大小上限。規劃器為較大的分割區引進隨機播放操作，以改善平行處理。	128 公尺

屬性名稱	Description	預設值
spark.sql.metadataCacheTTLSeconds	控制中繼資料快取的time-to-live (TTL)。適用於分割區檔案中繼資料和工作階段目錄快取。需要： - 大於零的正值 - spark.sql.catalogImplementation 設定為 hive - spark.sql.hive.filesourcePartitionFileCacheSize 大於零 - spark.sql.hive.manageFilesourcePartitions 設定為 true	-1000 毫秒
spark.sql.optimizer.collapseProjectAlwaysInline	指定是否要摺疊相鄰的投影和內嵌表達式，即使它會導致重複。	FALSE
spark.sql.optimizer.dynamicPartitionPruning.enabled	指定是否要為用作聯結索引鍵的分割區資料欄產生述詞。	TRUE
spark.sql.optimizer.enableCsvExpressionOptimization	指定是否要透過從_csv 操作剪除不必要的資料欄來最佳化 SQL 最佳化工具中的 CSV 表達式。	TRUE
spark.sql.optimizer.enableJsonExpressionOptimization	指定是否要透過下列方式最佳化 SQL 最佳化工具中的 JSON 表達式： - 從_json 操作刪除不必要的資料欄 - 簡化從_json 和 到_json 的組合 - 最佳化 named_struct 操作	TRUE
spark.sql.optimizer.excludedRules	定義要停用的最佳化工具規則，以逗號分隔的規則名稱識別。某些規則無法停用，因為正確性需要它們。最佳化工具會記錄哪些規則已成功停用。	(無)

屬性名稱	Description	預設值
spark.sql.optimize r.runtime.bloomFilter.applicationSideScanSizeThreshold	設定在應用程式端插入 Bloom 篩選條件所需的最小彙總掃描大小，以位元組為單位。	10GB
spark.sql.optimize r.runtime.bloomFilter.creationSideThreshold	定義在建立端注入 Bloom 篩選條件的大小閾值上限。	10MB
spark.sql.optimize r.runtime.bloomFilter.enabled	指定是否要插入 Bloom 篩選條件，以便在隨機聯結的一側具有選擇性述詞時減少隨機播放資料。	TRUE
spark.sql.optimize r.runtime.bloomFilter.expectedNumItems	定義執行時間 Bloom 篩選條件中預期項目的預設數量。	1000000
spark.sql.optimize r.runtime.bloomFilter.maxNumBits	設定執行時間 Bloom 篩選條件中允許的位元數上限。	67108864
spark.sql.optimize r.runtime.bloomFilter.maxNumItems	設定執行時間 Bloom 篩選條件中允許的預期項目數量上限。	4000000
spark.sql.optimize r.runtime.bloomFilter.number.threshold	限制每個查詢允許的非 DPP 執行時間篩選條件數量上限，以防止out-of-memory錯誤。	10
spark.sql.optimize r.runtime.bloomFilter.numBits	定義執行時間 Bloom 篩選條件中使用的預設位元數。	8388608

屬性名稱	Description	預設值
spark.sql.optimize r.runtime.rowlevel OperationGroupFilter.enabled	指定是否啟用資料列層級操作的執行時間群組篩選。允許資料來源： - 使用資料來源篩選條件刪除整個資料群組（例如檔案或分割區） - 執行執行時間查詢以識別相符的記錄 - 捨棄不必要的群組，以避免昂貴的重寫 限制： - 並非所有表達式都可以轉換為資料來源篩選條件 - 有些表達式需要 Spark 評估（例如子查詢）	TRUE
spark.sql.optimize r.runtimeFilter.se miJoinReduction.en abled	指定是否要插入半聯結，以便在隨機聯結的一側具有選擇性述詞時減少隨機資料。	FALSE
spark.sql.parquet. aggregatePushdown	指定是否將彙總下推至 Parquet 以進行最佳化。支援： - 布林值、整數、浮點數和日期類型的 MIN 和 MAX - 所有資料類型的 COUNT 如果任何 Parquet 檔案頁尾缺少統計資料，則擲回例外狀況。	FALSE
spark.sql.parquet. columnarReaderBatc hSize	控制每個 Parquet 向量化讀取器批次中的資料列數。選擇平衡效能額外負荷和記憶體用量的值，以防止out-of-memory錯誤。	4096

屬性名稱	Description	預設值
spark.sql.session.timeZone	定義工作階段時區，以處理字串常值和 Java 物件轉換中的時間戳記。接受： - 區域/城市格式的區域型 IDs (例如 America/Los_Angeles) - 區域位移，以 (+/-)HH、(+/-)HH : mm 或 (+/-)HH : mm : ss 格式顯示 (例如 -08 或 +01 : 00) +00 : 00 的 UTC 或 Z 別名	(本機時區的值)
spark.sql.shuffle.partitions	設定聯結或彙總期間資料隨機播放的預設分割區數量。結構式串流查詢從相同檢查點位置重新啟動之間無法修改。	200
spark.sql.shuffledHashJoinFactor	定義用於判斷隨機雜湊聯結資格的乘數係數。當小端資料大小乘以此因素小於大型端資料大小時，會選取隨機雜湊聯結。	3
spark.sql.sources.parallelPartitionDiscovery.threshold	使用檔案型來源 (Parquet、JSON 和 ORC) 設定驅動程式端檔案清單的路徑數目上限。在分割區探索期間超過 時，會使用個別的 Spark 分散式任務列出檔案。	32
spark.sql.statistics.histogram.enabled	指定是否要在資料欄統計資料運算期間產生等高長條圖，以改善估算準確度。需要額外的資料表掃描，超過基本資料欄統計資料所需的掃描。	FALSE

12. 選擇執行。

Note

如果可以接收結果的成員尚未設定查詢結果設定，則無法執行查詢。

13. 檢視結果。

如需詳細資訊，請參閱[接收和使用分析結果](#)。

14. 繼續調整參數並再次執行查詢，或選擇 + 按鈕在新索引標籤中啟動新查詢。

 Note

AWS Clean Rooms 旨在提供明確的錯誤訊息。如果錯誤訊息沒有足夠的詳細資訊來協助您進行疑難排解，請聯絡客戶團隊。向他們提供錯誤發生方式和錯誤訊息（包括任何識別符）的說明。如需詳細資訊，請參閱[故障診斷 AWS Clean Rooms](#)。

使用 SQL 程式碼編輯器查詢 ID 映射表

下列程序說明如何在 ID 映射資料表上執行多資料表聯結查詢，以將 `sourceId` 與 聯結 `targetId`。

查詢 ID 映射表之前，必須成功填入 ID 映射表。

使用 SQL 程式碼編輯器查詢 ID 映射表

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇具有執行查詢成員能力狀態的協同合作。
4. 在分析索引標籤上，前往分析區段。

 Note

分析區段只會在可接收結果的成員和負責支付查詢運算成本的成員以作用中成員的身分加入協同合作時顯示。

5. 在分析索引標籤的資料表下，檢視 ID 映射資料表清單 (受管者 AWS Clean Rooms 下) 及其相關聯的分析規則類型 (ID 映射資料表分析規則)。

 Note

如果您沒有在清單中看到預期的 ID 映射表，可能是因為 ID 映射表尚未成功填入。如需詳細資訊，請參閱[填入現有的 ID 映射表](#)。

6. 在 SQL 程式碼編輯器中輸入查詢來建置查詢。

(選用) 如果您想要使用範例查詢

1. 選取資料表旁的三個垂直點。
2. 在插入編輯器下，選擇範例 JOIN 陳述式。

 Note

插入範例 JOIN 陳述式會附加已在編輯器中的查詢。

範例 JOIN 陳述式隨即出現。

3. 編輯查詢中的預留位置值。

(選用) 如果您想要插入資料表名稱

1. 選取資料欄旁的三個垂直點。
2. 在插入編輯器下，選擇資料表名稱。
3. 編輯查詢中的預留位置值。

7. 指定支援的工作者類型和工作者數量。

使用下表來判斷使用案例所需的類型和數量或工作者。

工作者類型	vCPU	記憶體 (GB)	儲存體 (GB)	工作者數目	無塵室處理單元總數 (CRPU)
CR.1X (預設)	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

 Note

不同的工作者類型和工作者數量都有相關聯的成本。若要進一步了解定價，請參閱 [AWS Clean Rooms 定價](#)。

8. 選擇執行。

Note

如果可以接收結果的成員尚未設定查詢結果設定，則無法執行查詢。

9. 檢視結果。

如需詳細資訊，請參閱[接收和使用分析結果](#)。

10. 繼續調整參數並再次執行查詢，或選擇 + 按鈕在新索引標籤中啟動新查詢。

Note

AWS Clean Rooms 旨在提供明確的錯誤訊息。如果錯誤訊息沒有足夠的詳細資訊來協助您進行疑難排解，請聯絡客戶團隊。向他們提供錯誤發生方式和錯誤訊息（包括任何識別符）的說明。如需詳細資訊，請參閱[故障診斷 AWS Clean Rooms](#)。

使用 SQL 分析範本查詢設定的資料表

此程序示範如何使用 AWS Clean Rooms 主控台的分析範本，查詢具有自訂分析規則的已設定資料表。

使用 SQL 分析範本查詢具有自訂分析規則的已設定資料表

1. 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇具有執行查詢成員能力狀態的協同合作。
4. 在分析索引標籤的資料表區段下，檢視資料表及其相關聯的分析規則類型（自訂分析規則）。

Note

如果您沒有在清單中看到預期的資料表，可能是因為下列原因：

- 資料表尚未[建立關聯](#)。
- 資料表未[設定分析規則](#)。

5. 在分析區段下，針對分析模式，選取執行分析範本，然後從下拉式清單中選擇分析範本。

6. SQL 分析範本中的參數會自動填入 定義中。
7. 指定支援的工作者類型和工作者數量。

使用下表來判斷使用案例所需的類型和數量或工作者。

工作者類型	vCPU	記憶體 (GB)	儲存體 (GB)	工作者數目	無塵室處理單元總數 (CRPU)
CR.1X (預設)	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

 Note

不同的工作者類型和工作者數量都有相關聯的成本。若要進一步了解定價，請參閱 [AWS Clean Rooms 定價](#)。

8. 指定支援的 Spark 屬性。
 - a. 選取新增 Spark 屬性。
 - b. 在 Spark 屬性對話方塊中，從下拉式清單中選擇屬性名稱，然後輸入值。

下表提供每個屬性的定義。

如需 Spark 屬性的詳細資訊，請參閱 Apache Spark 文件中的 Spark [屬性](#)。

屬性名稱	Description	預設值
spark.task.maxFailures	控制任務在任務失敗之前可以失敗的連續次數。需要大於或等於 1 的值。允許重試的次數等於此值減去 1。如果任何嘗試成功，就會重	4

屬性名稱	Description	預設值
	設失敗計數。不同任務的故障不會累積到此限制。	
spark.sql.files.maxPartitionBytes	設定從 Parquet、JSON 和 ORC 等檔案型來源讀取時封裝到單一分割區的最大位元組數。	128MB
spark.hadoop.fs.s3.maxRetries	設定 Amazon S3 檔案操作的重試嘗試次數上限。	
spark.network.timeout	設定所有網路互動的預設逾時。如果未設定，則覆寫下列逾時設定： - spark.storage.blockManagerHeartbeatTimeoutMs - spark.shuffle.io.connectionTimeout - spark.rpc.askTimeout - spark.rpc.lookupTimeout	120 秒
spark.rdd.compress	指定是否使用 spark.io.compression.codec 壓縮序列化 RDD 分割區。適用於 Java 和 Scala 中的 StorageLevel.MEMORY_ONLY_SER，或 Python 中的 StorageLevel.MEMORY_ONLY。減少儲存空間，但需要額外的 CPU 處理時間。	FALSE
spark.shuffle.spill.compress	指定是否使用 spark.io.compression.codec 壓縮隨機溢出資料。	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	在 spark.sql.adaptive.enabled 為 true 的適應性最佳化期間，設定隨機播放分割區的目標大小，以位元組為單位。合併小型分割區或分割偏斜分割區時，控制分割區大小。	(spark.sql.adaptive.shuffle.targetPostShuffleInputSize 的值)

屬性名稱	Description	預設值
spark.sql.adaptive.autoBroadcastJoinThreshold	設定在聯結期間廣播至工作者節點的資料表大小上限，以位元組為單位。僅適用於自適應架構。使用與 spark.sql.adaptive.autoBroadcastJoinThreshold 相同的預設值。設定為 -1 以停用廣播。	(無)
spark.sql.adaptive.coalescePartitions.enabled	指定是否要根據 spark.sql.adaptive.advisoryPartitionSizeInBytes 來合併連續隨機播放分割區，以最佳化任務大小。需要 spark.sql.adaptive.enabled 才能成立。	TRUE
spark.sql.adaptive.coalescePartitions.initialPartitionNum	在合併之前定義隨機播放分割區的初始數量。同時需要 spark.sql.adaptive.enabled 和 spark.sql.adaptive.coalescePartitions.enabled 為 true。預設為 spark.sql.shuffle.partitions 的值。	(無)
spark.sql.adaptive.coalescePartitions.minPartitionSize	設定合併隨機播放分割區的大小下限，以防止分割區在適應性最佳化期間變得太小。	1 MB
spark.sql.adaptive.coalescePartitions.parallelismFirst	指定是否要在分割區合併期間根據叢集平行處理而非 spark.sql.adaptive.advisoryPartitionSizeInBytes 計算分割區大小。產生的分割區大小小於設定的目標大小，以最大化平行處理。我們建議在忙碌叢集上將此設定為 false，以透過防止過多的小型任務來改善資源使用率。	TRUE
spark.sql.adaptive.enabled	指定是否啟用自適應查詢執行，以在查詢執行期間根據準確的執行時間統計資料重新最佳化查詢計劃。	TRUE
spark.sql.adaptive.forceOptimizeSkewedJoin	指定是否強制啟用 OptimizeSkewedJoin，即使它引入額外的隨機播放。	FALSE

屬性名稱	Description	預設值
spark.sql.adaptive.localShuffleReader.enabled	指定是否在不需要隨機分割時使用本機隨機讀取器，例如從排序合併聯結轉換為廣播雜湊聯結之後。需要 spark.sql.adaptive.enabled 才能成立。	TRUE
spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold	設定用於建置本機雜湊映射的分割區大小上限，以位元組為單位。在下列情況下，優先考慮排序合併聯結的隨機雜湊聯結： - 此值等於或超過 spark.sql.adaptive.advisoryPartitionSizeInBytes - 所有分割區大小都在此限制內 覆寫 spark.sql.join.preferSortMergeJoin 設定。	0 個位元組
spark.sql.adaptive.optimizeSkewsInRebalancePartitions.enabled	指定是否根據 spark.sql.adaptive.advisoryPartitionSizeInBytes 將扭曲隨機播放分割區分割為較小的分割區，以最佳化扭曲隨機播放分割區。需要 spark.sql.adaptive.enabled 才能成立。	TRUE
spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor	定義分割期間合併分割區的大小閾值因素。小於此因素的分割區乘以 spark.sql.adaptive.advisoryPartitionSizeInBytes 會合併。	0.2
spark.sql.adaptive.skewJoin.enabled	指定是否透過分割和選擇性複寫扭曲的分割區來處理隨機聯結中的資料扭曲。適用於排序合併和隨機雜湊聯結。需要 spark.sql.adaptive.enabled 才能成立。	TRUE

屬性名稱	Description	預設值
spark.sql.adaptive.skewJoin.skewedPartitionFactor	決定決定分割區扭曲的大小因素。當分割區的大小超過以下兩者時，就會扭曲： - 此係數乘以中位數分割區大小 - spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes 的值	5
spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes	設定識別扭曲分割區的大小閾值，以位元組為單位。當分割區的大小超過以下兩者時，就會扭曲： - 此閾值 - 中位數分割區大小乘以 spark.sql.adaptive.skewJoin.skewedPartitionFactor 我們建議將此值設定為大於 spark.sql.adaptive.advisoryPartitionSizeInBytes。	256MB
spark.sql.autoBroadcastJoinThreshold	設定在聯結期間廣播至工作者節點的資料表大小上限，以位元組為單位。設定為 -1 以停用廣播。	10MB
spark.sql.broadcastTimeout	控制廣播聯結期間廣播操作的逾時期間，以秒為單位。	300 秒
spark.sql.cbo.enabled	指定是否要為計劃統計資料估算啟用成本型最佳化 (CBO)。	FALSE
spark.sql.cbo.joinReorder.dp.star.filter	指定是否要在成本型聯結列舉期間套用星狀聯結篩選條件啟發式。	FALSE
spark.sql.cbo.joinReorder.dp.threshold	設定動態程式設計演算法中允許的聯結節點數目上限。	12
spark.sql.cbo.joinReorder.enabled	指定是否要在成本型最佳化 (CBO) 中啟用聯結重新排序。	FALSE

屬性名稱	Description	預設值
spark.sql.cbo.planStats.enabled	指定是否要在邏輯計畫產生期間從目錄中擷取資料列計數和資料欄統計資料。	FALSE
spark.sql.cbo.starSchemaDetection	指定是否根據星狀結構描述偵測啟用聯結重新排序。	FALSE
spark.sql.files.maxPartitionNum	設定檔案型來源 (Parquet、JSON 和 ORC) 的目標分割檔案分割區數量上限。當初始計數超過此值時，重新擴展分割區。這是建議的目標，而不是保證的限制。	(無)
spark.sql.files.maxRecordsPerFile	設定寫入單一檔案的記錄數目上限。設定為零或負值時不會套用任何限制。	0
spark.sql.files.minPartitionNum	設定檔案型來源 (Parquet、JSON 和 ORC) 的目標分割檔案分割區數量下限。預設為 spark.sql.leafNodeDefaultParallelism。這是建議的目標，而不是保證的限制。	(無)
spark.sql.inMemoryColumnarStorage.batchSize	控制單欄式快取的批次大小。增加大小可改善記憶體使用率和壓縮，但會增加out-of-memory錯誤的風險。	10000
spark.sql.inMemoryColumnarStorage.compressed	指定是否根據資料統計資料自動選取資料欄的壓縮轉碼器。	TRUE
spark.sql.inMemoryColumnarStorage.enableVectorizedReader	指定是否要為單欄式快取啟用向量化讀取。	TRUE
spark.sql.legacy.allowHashOnMapType	指定是否允許對映射類型資料結構執行雜湊操作。此舊版設定可維持與舊版 Spark 版本的映射類型處理的相容性。	

屬性名稱	Description	預設值
<code>spark.sql.legacy.allowNegativeScaleOfDecimal</code>	指定是否在十進位類型定義中允許負縮放值。此舊版設定可維持與支援負小數位數擴展之較舊 Spark 版本的相容性。	
<code>spark.sql.legacy.castComplexTypesToString.enabled</code>	指定是否啟用舊版行為，將複雜類型轉換為字串。維持與舊版 Spark 版本類型轉換規則的相容性。	
<code>spark.sql.legacy.charVarcharAsString</code>	指定是否將 CHAR 和 VARCHAR 類型視為 STRING 類型。此舊版設定提供與較舊 Spark 版本的字串類型處理相容性。	
<code>spark.sql.legacy.createEmptyCollectionUsingStringType</code>	指定是否使用字串類型元素建立空集合。此舊版設定可維持與較舊 Spark 版本的集合初始化行為的相容性。	
<code>spark.sql.legacy.exponentLiteralAsDecimal.enabled</code>	指定是否要將指數常值解譯為十進位類型。此舊版設定可維持與舊版 Spark 版本數值常值處理的相容性。	
<code>spark.sql.legacy.json.allowEmptyString.enabled</code>	指定是否要在 JSON 處理中允許空字串。此舊版設定可維持與舊版 Spark 版本的 JSON 剖析行為的相容性。	
<code>spark.sql.legacy.parquet.int96RebaseModelRead</code>	指定是否在讀取 Parquet 檔案時使用舊版 INT96 時間戳記重新基礎模式。此舊版設定可維持與較舊 Spark 版本時間戳記處理的相容性。	
<code>spark.sql.legacy.timeParserPolicy</code>	控制時間剖析行為以實現回溯相容性。此舊版設定會決定如何從字串剖析時間戳記和日期。	
<code>spark.sql.legacy.typeCoercion.datetimeToString.enabled</code>	指定是否在將日期時間值轉換為字串時啟用舊版類型強制行為。維持與較舊 Spark 版本日期時間轉換規則的相容性。	

屬性名稱	Description	預設值
spark.sql.maxSinglePartitionBytes	設定以位元組為單位的分割區大小上限。規劃器為較大的分割區引進隨機播放操作，以改善平行處理。	128 公尺
spark.sql.metadataCacheTTLSeconds	控制中繼資料快取的time-to-live (TTL)。適用於分割區檔案中繼資料和工作階段目錄快取。需要： • 大於零的正值 • spark.sql.catalogImplementation 設定為 hive • spark.sql.hive.filesourcePartitionFileCacheSize 大於零 • spark.sql.hive.manageFilesourcePartitions 設定為 true	-1000 毫秒
spark.sql.optimizer.collapseProjectAlwaysInline	指定是否要摺疊相鄰的投影和內嵌表達式，即使它會導致重複。	FALSE
spark.sql.optimizer.dynamicPartitionPruning.enabled	指定是否要為用作聯結索引鍵的分割區資料欄產生述詞。	TRUE
spark.sql.optimizer.enableCsvExpressionOptimization	指定是否要透過從 csv 操作剪除不必要的資料欄來最佳化 SQL 最佳化工具中的 CSV 表達式。	TRUE
spark.sql.optimizer.enableJsonExpressionOptimization	指定是否要透過下列方式最佳化 SQL 最佳化工具中的 JSON 表達式： • 從_json 操作刪除不必要的資料欄 • 簡化從_json 和 到_json 的組合 • 最佳化 named_struct 操作	TRUE

屬性名稱	Description	預設值
spark.sql.optimizer.excludedRules	定義要停用的最佳化工具規則，以逗號分隔的規則名稱識別。某些規則無法停用，因為正確性需要它們。最佳化工具會記錄哪些規則已成功停用。	(無)
spark.sql.optimizer.runtime.bloomFilter.applicationSideScanSizeThreshold	設定在應用程式端注入 Bloom 篩選條件所需的最小彙總掃描大小，以位元組為單位。	10GB
spark.sql.optimizer.runtime.bloomFilter.creationSideThreshold	定義在建立端插入 Bloom 篩選條件的大小閾值上限。	10MB
spark.sql.optimizer.runtime.bloomFilter.enabled	指定是否要插入 Bloom 篩選條件，以便在隨機聯結的一側具有選擇性述詞時減少隨機播放資料。	TRUE
spark.sql.optimizer.runtime.bloomFilter.expectedNumItems	定義執行時間 Bloom 篩選條件中預期項目的預設數量。	1000000
spark.sql.optimizer.runtime.bloomFilter.maxNumBits	設定執行時間 Bloom 篩選條件中允許的位元數上限。	67108864
spark.sql.optimizer.runtime.bloomFilter.maxNumItems	設定執行時間 Bloom 篩選條件中允許的預期項目數量上限。	4000000
spark.sql.optimizer.runtime.bloomFilter.number.threshold	限制每個查詢允許的非 DPP 執行時間篩選條件數量上限，以防止out-of-memory錯誤。	10

屬性名稱	Description	預設值
spark.sql.optimize r.runtime.bloomFilter.numBits	定義執行時間 Bloom 篩選條件中使用的預設位元數。	8388608
spark.sql.optimize r.runtime.rowlevel OperationGroupFilter.enabled	指定是否要為資料列層級操作啟用執行時間群組篩選。允許資料來源： - 使用資料來源篩選條件刪除整個資料群組（例如檔案或分割區） - 執行執行時間查詢以識別相符的記錄 - 捨棄不必要的群組，以避免昂貴的重寫 限制： - 並非所有表達式都可以轉換為資料來源篩選條件 - 有些表達式需要 Spark 評估（例如子查詢）	TRUE
spark.sql.optimize r.runtimeFilter.semiJoinReduction.enabled	指定是否要插入半聯結，以便在隨機聯結的一側具有選擇性述詞時減少隨機資料。	FALSE
spark.sql.parquet. aggregatePushdown	指定是否將彙總下推至 Parquet 以進行最佳化。支援： - 布林值、整數、浮點數和日期類型的 MIN 和 MAX - 所有資料類型的 COUNT 如果任何 Parquet 檔案頁尾缺少統計資料，則擲回例外狀況。	FALSE

屬性名稱	Description	預設值
spark.sql.parquet.columnarReaderBatchSize	控制每個 Parquet 向量化讀取器批次中的資料列數。選擇平衡效能額外負荷和記憶體用量的值，以防止out-of-memory錯誤。	4096
spark.sql.session.timeZone	定義工作階段時區，以處理字串常值和 Java 物件轉換中的時間戳記。接受： - 區域/城市格式的區域型 IDs (例如 America/Los_Angeles) - 區域位移，以 (+/-)HH、(+/-)HH : mm 或 (+/-)HH : mm : ss 格式顯示 (例如 -08 或 +01 : 00) +00 : 00 的 UTC 或 Z 別名	(本機時區的值)
spark.sql.shuffle.partitions	設定聯結或彙總期間資料隨機播放的預設分割區數量。結構式串流查詢從相同檢查點位置重新啟動之間無法修改。	200
spark.sql.shuffledHashJoinFactor	定義用於判斷隨機雜湊聯結資格的乘數係數。當小端資料大小乘以此因素時，會選取隨機雜湊聯結，小於大端資料大小。	3
spark.sql.sources.parallelPartitionDiscovery.threshold	使用檔案型來源 (Parquet、JSON 和 ORC) 設定驅動程式端檔案清單的路徑數目上限。在分割區探索期間超過 時，會使用個別的 Spark 分散式任務列出檔案。	32
spark.sql.statistics.histogram.enabled	指定是否要在資料欄統計資料計算期間產生等高長條圖，以改善估算準確度。需要額外的資料表掃描，超過基本資料欄統計資料所需的掃描。	FALSE

9. 選擇執行。

 Note

如果可以接收結果的成員尚未設定查詢結果設定，則無法執行查詢。

10. 繼續調整參數並再次執行查詢，或選擇 + 按鈕在新索引標籤中啟動新查詢。

使用分析建置器查詢

您可以使用分析建置器來建置查詢，而不必撰寫 SQL 程式碼。使用分析建置器，您可以為具有下列項目的協同合作建立查詢：

- 使用[彙總分析規則](#)且不需要 JOIN 的單一資料表
- 兩個使用[彙總分析規則](#)的資料表（每個成員各一個）
- 兩個使用[清單分析規則](#)的資料表（每個成員各一個）
- 兩個資料表（每個成員各一個）都使用彙總分析規則，兩個資料表（每個成員各一個）都使用清單分析規則

如果您想要手動寫入 SQL 查詢，請參閱 [使用 SQL 程式碼編輯器查詢設定的資料表](#)。

分析建置器會在 AWS Clean Rooms 主控台分析索引標籤的分析區段中顯示為分析建置器 UI 選項。

 Important

如果您開啟分析建置器 UI，開始在分析建置器中建置查詢，然後關閉分析建置器 UI，則不會儲存您的查詢。

 Tip

如果在查詢執行時發生排程維護，查詢會終止並復原。您必須重新啟動查詢。

下列主題說明如何使用分析建置器。

主題

- [使用分析建置器查詢單一資料表（彙總）](#)

- [使用分析建置器查詢兩個資料表（彙總或清單）](#)

使用分析建置器查詢單一資料表（彙總）

此程序示範如何在 AWS Clean Rooms 主控台中使用分析建置器 UI 來建置查詢。查詢適用於具有單一資料表的協同合作，該資料表使用[彙總分析規則](#)，而不需要JOIN。

使用分析建置器查詢單一資料表

登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。

1. 在左側導覽窗格中，選擇協同合作。
2. 選擇具有執行查詢成員能力狀態的協同合作。
3. 在分析索引標籤的資料表下，檢視資料表及其相關聯的分析規則類型。（分析規則類型應該是彙總分析規則。）

Note

如果您沒有看到預期的資料表，可能原因如下：

- 資料表尚未[建立關聯](#)。
- 資料表未[設定分析規則](#)。

4. 在分析區段下，開啟分析建置器 UI。
5. 建置查詢。

如果您想要查看所有彙總指標，請跳至步驟 9。

- a. 對於選擇指標，請檢閱預設預先選取的彙總指標，並視需要移除任何指標。
- b. （選用）對於新增客群 – 選用，請選擇一或多個參數。

Note

新增客群 – 只有在為資料表指定維度時，才會顯示選用項目。

- c. （選用）對於新增篩選條件 – 選用，選擇新增篩選條件，然後選擇參數、運算子和值。

若要新增更多篩選條件，請選擇新增另一個篩選條件。

若要移除篩選條件，請選擇移除。

 Note

ORDER BY 不支援彙總查詢。
篩選條件中僅支援 AND 運算子。

- d. （選用）對於新增描述 – 選用，輸入描述以協助識別查詢清單中的查詢。
6. 展開預覽 SQL 程式碼。
 - a. 檢視從分析建置器產生的 SQL 程式碼。
 - b. 若要複製 SQL 程式碼，請選擇複製。
 - c. 若要編輯 SQL 程式碼，請在 SQL 程式碼編輯器中選擇編輯。
 7. 選擇執行。

 Note

如果可以接收結果的成員尚未設定查詢結果設定，則無法執行查詢。

8. 繼續調整參數並再次執行查詢，或選擇 + 按鈕在新索引標籤中啟動新查詢。

 Note

AWS Clean Rooms 旨在提供明確的錯誤訊息。如果錯誤訊息沒有足夠的詳細資訊來協助您進行疑難排解，請聯絡客戶團隊。向他們提供錯誤發生方式和錯誤訊息（包括任何識別符）的說明。如需詳細資訊，請參閱[故障診斷 AWS Clean Rooms](#)。

使用分析建置器查詢兩個資料表（彙總或清單）

此程序說明如何在 AWS Clean Rooms 主控台中使用分析建置器，為具有下列特性的協同合作建立查詢：

- 兩個使用[彙總分析規則](#)的資料表（每個成員各一個）
- 兩個使用[清單分析規則](#)的資料表（每個成員各一個）

- 兩個資料表（每個成員各一個）都使用彙總分析規則，兩個資料表（每個成員各一個）都使用清單分析規則

使用分析建置器查詢兩個資料表

登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。

1. 在左側導覽窗格中，選擇協同合作。
2. 選擇具有執行查詢成員能力狀態的協同合作。
3. 在分析索引標籤的資料表下，檢視兩個資料表及其相關聯的分析規則類型（彙總分析規則或清單分析規則）。

Note

如果您沒有在清單中看到預期的資料表，可能是因為下列原因：

- 資料表尚未[建立關聯](#)。
- 資料表未[設定分析規則](#)。

4. 在分析區段下，開啟分析建置器 UI。
5. 建置查詢。

如果協同合作包含兩個使用彙總分析規則的資料表，以及兩個使用清單分析規則的資料表，請先選擇彙總或清單，然後根據選取的分析規則遵循提示。

如果兩個資料表使用彙總分析規則

1. 對於選擇指標，請檢閱預設預先選取的彙總指標，並視需要移除任何指標。
2. 針對相符記錄，選擇一或多個記錄。

Note

使用分析建置器時，您只能比對單對資料欄。

如果兩個資料表使用清單分析規則

1. 對於選擇屬性，請檢閱預設已預先選取的清單屬性，並視需要移除任何指標。
2. 針對相符記錄，選擇一或多個記錄。

Note

使用分析建置器時，您只能比對單對資料欄。

如果兩個資料表使用彙總分析規則

3. (選用) 對於新增客群 – 選用，請選擇一或多個參數。

 Note

新增區段 – 只有在資料表指定維度時，才會顯示選用項目。

4. (選用) 對於新增篩選條件 – 選用，選擇新增篩選條件，然後選擇參數、運算子和值。

若要新增更多篩選條件，請選擇新增另一個篩選條件。

若要移除篩選條件，請選擇移除。

 Note

ORDER BY 不支援彙總查詢。
篩選條件中僅支援 AND 運算子。

5. (選用) 對於新增描述 – 選用，輸入描述以協助識別最近查詢清單中的查詢。

如果兩個資料表使用清單分析規則

3. (選用) 對於新增篩選條件 – 選用，選擇新增篩選條件，然後選擇參數、運算子和值。

若要新增更多篩選條件，請選擇新增另一個篩選條件。

若要移除篩選條件，請選擇移除。

 Note

LIMIT 清單查詢不支援。
篩選條件中僅支援 AND 運算子。

4. (選用) 對於新增描述 – 選用，輸入描述以協助識別最近查詢清單中的查詢。

6. 展開預覽 SQL 程式碼。

- a. 檢視從分析建置器產生的 SQL 程式碼。
- b. 若要複製 SQL 程式碼，請選擇複製。
- c. 若要編輯 SQL 程式碼，請在 SQL 程式碼編輯器中選擇編輯。

7. 選擇執行。

 Note

如果可以接收結果的成員尚未設定查詢結果設定，則無法執行查詢

8. 繼續調整參數並再次執行查詢，或選擇 + 按鈕在新索引標籤中啟動新查詢。

Note

AWS Clean Rooms 旨在提供明確的錯誤訊息。如果錯誤訊息沒有足夠的詳細資訊來協助您進行疑難排解，請聯絡客戶團隊。向他們提供錯誤發生方式和錯誤訊息（包括任何識別符）的說明。如需詳細資訊，請參閱[故障診斷 AWS Clean Rooms](#)。

檢視差異隱私權的影響

一般而言，開啟差異隱私權時，寫入和執行查詢不會變更。不過，如果剩餘的隱私權預算不足，您就無法執行查詢。當您執行查詢並使用隱私權預算時，您可以看到可以執行多少彙總，以及這如何影響未來的查詢。

檢視協同合作中差異隱私權的影響

1. 登入 AWS 管理主控台 並開啟位於 <https://console.aws.amazon.com/cleanrooms> 的 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇具有您的成員詳細資訊執行查詢狀態的協同合作。
4. 在分析索引標籤的資料表下，檢視剩餘的隱私權預算。這會顯示為剩餘的彙總函數估計數量和使用的公用程式（以百分比呈現）。

Note

剩餘的彙總函數估計數量和使用的公用程式百分比只會顯示給可查詢的成員。

5. 選擇檢視影響，以檢視將多少雜訊注入結果，以及您可以執行大約多少彙總函數。

檢視近期查詢

您可以在分析索引標籤上檢視過去 90 天內執行的查詢。

Note

如果您唯一的成員能力是貢獻資料，而且您不是[支付查詢運算成本的成員](#)，則分析索引標籤不會出現在主控台上。

檢視最近的查詢

登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。

1. 在左側導覽窗格中，選擇協同合作。
2. 選擇協同合作。
3. 在分析索引標籤的分析下，從下拉式清單中選取所有查詢，並檢視過去 90 天內已執行的查詢。
4. 若要依狀態排序最近的查詢，請從所有狀態下拉式清單中選取狀態。

狀態為：已提交、已開始、已取消、成功、失敗和逾時。

檢視查詢詳細資訊

您可以將查詢詳細資訊檢視為可執行查詢的成員，或檢視為可接收結果的成員。

檢視查詢的詳細資訊

1. 登入 AWS 管理主控台 並開啟位於 <https://console.aws.amazon.com/cleanrooms> 的 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在分析索引標籤上，執行下列其中一項操作：
 - 選擇您要檢視之特定查詢的選項按鈕，然後選擇檢視詳細資訊。
 - 選擇受保護的查詢 ID。
5. 在查詢詳細資訊頁面上，
 - 如果您是可執行查詢的成員，請檢視查詢詳細資訊、SQL 文字和結果。

您看到一則訊息，確認查詢結果已交付給可接收結果的成員。

- 如果您是可以接收結果的成員，請檢視查詢詳細資訊和結果。

執行 PySpark 任務

身為[可以查詢的成員](#)，您可以使用核准的 PySpark [分析範本](#)，在設定的資料表上執行 PySpark 任務。

先決條件

執行 PySpark 任務之前，您必須擁有：

- AWS Clean Rooms 協作中的作用中成員資格
- 在協同合作中存取至少一個分析範本
- 在協同合作中存取至少一個已設定的資料表
- 將 PySpark 任務的結果寫入指定 S3 儲存貯體的許可

如需建立所需服務角色的詳細資訊，請參閱 [建立服務角色以寫入 PySpark 任務的結果](#)。

- 負責支付運算成本的成員已以作用中成員的身分加入協同合作

如需有關如何直接呼叫 AWS Clean Rooms StartProtectedJob API 操作或使用 AWS SDKs 查詢資料或檢視查詢的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

如需任務記錄的相關資訊，請參閱 [分析登入 AWS Clean Rooms](#)。

如需接收任務結果的資訊，請參閱 [接收和使用分析結果](#)。

下列主題說明如何在使用 AWS Clean Rooms 主控台的協同合作中，在設定的資料表上執行 PySpark 任務。

主題

- [使用 PySpark 分析範本在設定的資料表上執行 PySpark 任務](#)
- [檢視最近的任務](#)
- [檢視任務詳細資訊](#)

使用 PySpark 分析範本在設定的資料表上執行 PySpark 任務

此程序示範如何在主控台中使用 AWS Clean Rooms PySpark 分析範本，以自訂分析規則分析設定的資料表。

使用 PySpark 分析範本在設定的資料表上執行 PySpark 任務

登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。

1. 在左側導覽窗格中，選擇協同合作。

2. 選擇具有執行任務成員能力狀態的協同合作。
3. 在分析索引標籤的資料表區段下，檢視資料表及其相關聯的分析規則類型 (自訂分析規則)。

 Note

如果您沒有在清單中看到預期的資料表，可能是因為下列原因：

- 資料表尚未[建立關聯](#)。
- 資料表未[設定分析規則](#)。

4. 在分析區段的分析模式下，選取執行分析範本。
5. 從分析範本下拉式清單中選擇 PySpark 分析範本。

來自 PySpark 分析範本的參數會自動填入定義中。

6. 如果分析範本已定義參數，請在參數下提供參數的值：
 - a. 針對每個參數，檢視參數名稱和預設值（如果已設定）。
 - b. 輸入您要覆寫的每個參數的值。

 Note

如果您不提供值，但存在預設值，則會使用預設值。

 Important

參數值最多可達 1,000 個字元，並支援 UTF-8 編碼。所有參數值都會視為字串，並透過內容物件傳遞給您的使用者指令碼。

確保您的使用者指令碼安全地驗證和處理參數值。如需安全參數處理的詳細資訊，請參閱[在 PySpark 分析範本中使用參數](#)。

7. 指定支援的工作者類型和工作者數量。

使用下表來判斷使用案例所需的類型和數量或工作者。

工作者類型	vCPU	記憶體 (GB)	儲存體 (GB)	工作者數目	無塵室處理單元總數 (CRPU)
CR.1X (預設)	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

Note

不同的工作者類型和工作者數量都有相關聯的成本。若要進一步了解定價，請參閱 [AWS Clean Rooms 定價](#)。

8. 選擇執行。

Note

如果可以接收結果的成員尚未設定任務結果設定，則無法執行任務。

9. 繼續調整參數並再次執行任務，或在新索引標籤中選擇 + 按鈕以啟動新任務。

檢視最近的任務

您可以在分析索引標籤上檢視過去 90 天內執行的任務。

Note

如果您唯一的成員能力是貢獻資料，而且您不是 [支付任務運算成本的成員](#)，則分析索引標籤不會出現在主控台上。

檢視最近的任務

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在分析索引標籤的分析下，從下拉式清單中選取所有任務，並檢視過去 90 天內已執行的任務。
5. 若要依狀態排序最近的任務，請從所有狀態下拉式清單中選取狀態。

狀態為：已提交、已開始、已取消、成功、失敗和逾時。

檢視任務詳細資訊

您可以將任務詳細資訊檢視為可執行任務的成員，或檢視為可接收結果的成員。

檢視任務的詳細資訊

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇協同合作。
4. 在分析索引標籤的分析下，從下拉式清單中選取所有任務，然後執行下列其中一項操作：
 - 選擇您要檢視之特定任務的選項按鈕，然後選擇檢視詳細資訊。
 - 選擇受保護的任務 ID。
5. 在任務詳細資訊頁面上，
 - 如果您是可執行任務的成員，請檢視任務詳細資訊、任務和結果。

您看到一則訊息，確認任務結果已交付給可接收結果的成員。

- 如果您是可以接收結果的成員，請檢視任務詳細資訊和結果。

接收和使用分析結果

[可以接收結果的成員](#)可以在 AWS Clean Rooms 主控台中或他們在加入協同合作時指定的 Amazon S3 儲存貯體中檢閱查詢結果。

Note

僅針對加密資料表，可以接收結果的成員會透過在解密模式下執行 C3R 加密用戶端來[解密](#)查詢結果。

Amazon S3 中的結果目的地不能與任何資料來源位於相同的 S3 儲存貯體中。

下列主題說明如何使用 AWS Clean Rooms 主控台接收分析結果。

主題

- [接收查詢結果](#)
- [接收任務結果](#)
- [編輯查詢結果設定的預設值](#)
- [編輯任務結果設定的預設值](#)
- [在其他 中使用查詢輸出 AWS 服務](#)

如需有關如何直接呼叫 AWS Clean Rooms API 或使用 AWS SDKs 查詢資料或檢視查詢的資訊，請參閱 [AWS Clean Rooms API 參考](#)。

如需查詢記錄的資訊，請參閱 [分析登入 AWS Clean Rooms](#)。

Note

如果您在加密資料表上執行查詢，加密資料欄的結果會加密。

接收查詢結果

Note

Amazon S3 中的結果目的地不能與任何資料來源位於相同的 S3 儲存貯體中。

查詢的結果位於 AWS Clean Rooms 主控台分析索引標籤的結果設定預設值區段。

接收查詢結果

1. 登入 AWS 管理主控台 並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您的成員能力狀態為接收結果的協同合作。
4. 若要直接從分析索引標籤 AWS Clean Rooms 的分析下接收查詢結果，請從下拉式清單中選取所有查詢，然後在受保護的查詢 ID 欄下選取查詢。
5. 在查詢詳細資訊頁面的結果下，執行下列其中一項：

如果您想要	然後選擇...
複製結果。	Copy (複製)
下載結果。	下載  Note 根據預設，下載的檔案名稱是執行查詢時Query id顯示的對應名稱 AWS Clean Rooms。
在 Amazon S3 中檢視結果。	在 Amazon S3 中檢視 Amazon S3 主控台會在單獨的索引標籤中開啟。

6. 如果您使用的是加密的資料，您現在可以[解密](#)資料表。

如需詳細資訊，請參閱[使用 C3R 加密用戶端解密資料表](#)。

接收任務結果

Note

Amazon S3 中的結果目的地不能與任何資料來源位於相同的 S3 儲存貯體中。

任務的結果位於 AWS Clean Rooms 主控台分析索引標籤的結果設定預設值區段。

接收任務結果

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您的成員能力狀態為接收結果的協同合作。
4. 若要直接從分析索引標籤 AWS Clean Rooms 的分析下接收任務結果，請從下拉式清單中選取所有任務，然後在受保護的任務 ID 欄下選取任務。
5. 在任務詳細資訊頁面的結果下，複製任務 ID。

返回分析索引標籤並展開結果設定預設值。

在結果目的地下，選取連結以在 Amazon S3 中檢視結果。

Amazon S3 主控台會在單獨的索引標籤中開啟。

在 Amazon S3 中，在搜尋列中貼上任務 ID，然後按 Enter 鍵。

包含結果的資料夾隨即出現。選取資料夾以檢視任務結果。

編輯查詢結果設定的預設值

Note

Amazon S3 中的結果目的地不能與任何資料來源位於相同的 S3 儲存貯體內。

身為可以接收結果的成員，您可以在 AWS Clean Rooms 主控台中編輯查詢結果設定的預設值。

編輯查詢結果設定的預設值

1. 登入 AWS 管理主控台 並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您的成員能力狀態為接收結果的協同合作。
4. 在分析索引標籤的結果設定預設值下，選擇編輯。
5. 在編輯結果設定預設值頁面上，視需要修改下列任何項目：
 - a. 在查詢結果下，修改 Amazon S3 中的結果目的地、結果格式或結果檔案。
 - b. （選用）對於服務存取，如果您想要將最多需要 24 小時的查詢交付到 S3 目的地，請選取新增服務角色以支援最多需要 24 小時才能完成的查詢核取方塊。

最多需要 24 小時才能完成的大型查詢將會傳送到您的 S3 目的地。

如果您未選取核取方塊，則只會將 12 小時內完成的查詢傳送到您的 S3 位置。

- 選取建立和使用新的服務角色或使用現有的服務角色，以指定服務存取許可。

Create and use a new service role

- AWS Clean Rooms 會建立具有此資料表所需政策的服務角色。
- 預設的服務角色名稱為 `cleanrooms-query-receiver-<timestamp>`
- 您必須具有建立角色和連接政策的許可。

Use an existing service role

1. 從下拉式清單中選擇現有的服務角色名稱。

如果您具有列出角色的許可，則會顯示角色清單。

如果您沒有列出角色的許可，您可以輸入要使用的角色的 Amazon Resource Name (ARN)。

2. 選擇在 IAM 外部連結中檢視，以檢視服務角色。

如果沒有現有的服務角色，則無法使用使用現有服務角色的選項。

根據預設，AWS Clean Rooms 不會嘗試更新現有的角色政策來新增必要的許可。

Note

- AWS Clean Rooms 需要根據分析規則查詢的許可。如需 許可的詳細資訊 AWS Clean Rooms，請參閱 [AWS 的 受管政策 AWS Clean Rooms](#)。
- 如果角色沒有足夠的許可 AWS Clean Rooms，您會收到錯誤訊息，指出角色沒有足夠的許可 AWS Clean Rooms。必須先新增角色政策，才能繼續。
- 如果您無法修改角色政策，您會收到錯誤訊息，指出 AWS Clean Rooms 找不到服務角色的政策。

6. 選擇儲存變更。
7. 更新的查詢結果設定會顯示在協同合作詳細資訊頁面上。

編輯任務結果設定的預設值

Note

Amazon S3 中的結果目的地不能與任何資料來源位於相同的 S3 儲存貯體中。

身為可以接收結果的成員，您可以在 AWS Clean Rooms 主控台中編輯任務結果設定的預設值。

編輯任務結果設定的預設值

1. 登入 AWS 管理主控台 並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 選擇您的成員能力狀態為接收結果的協同合作。
4. 在分析索引標籤的結果設定預設值下，選擇編輯。
5. 在編輯結果設定預設值頁面上，視需要修改下列任何項目：
 - a. 在任務結果下，修改 Amazon S3 中的結果目的地。
 - b. 在服務存取下，修改現有的服務角色名稱。
6. 選擇儲存變更。
7. 已更新的任務結果設定會顯示在協同合作詳細資訊頁面上。

在其他 中使用查詢輸出 AWS 服務

SQL 查詢輸出可用於 Clean Rooms ML 模型的種子資料。如需詳細資訊，請參閱 [AWS Clean Rooms ML](#)。

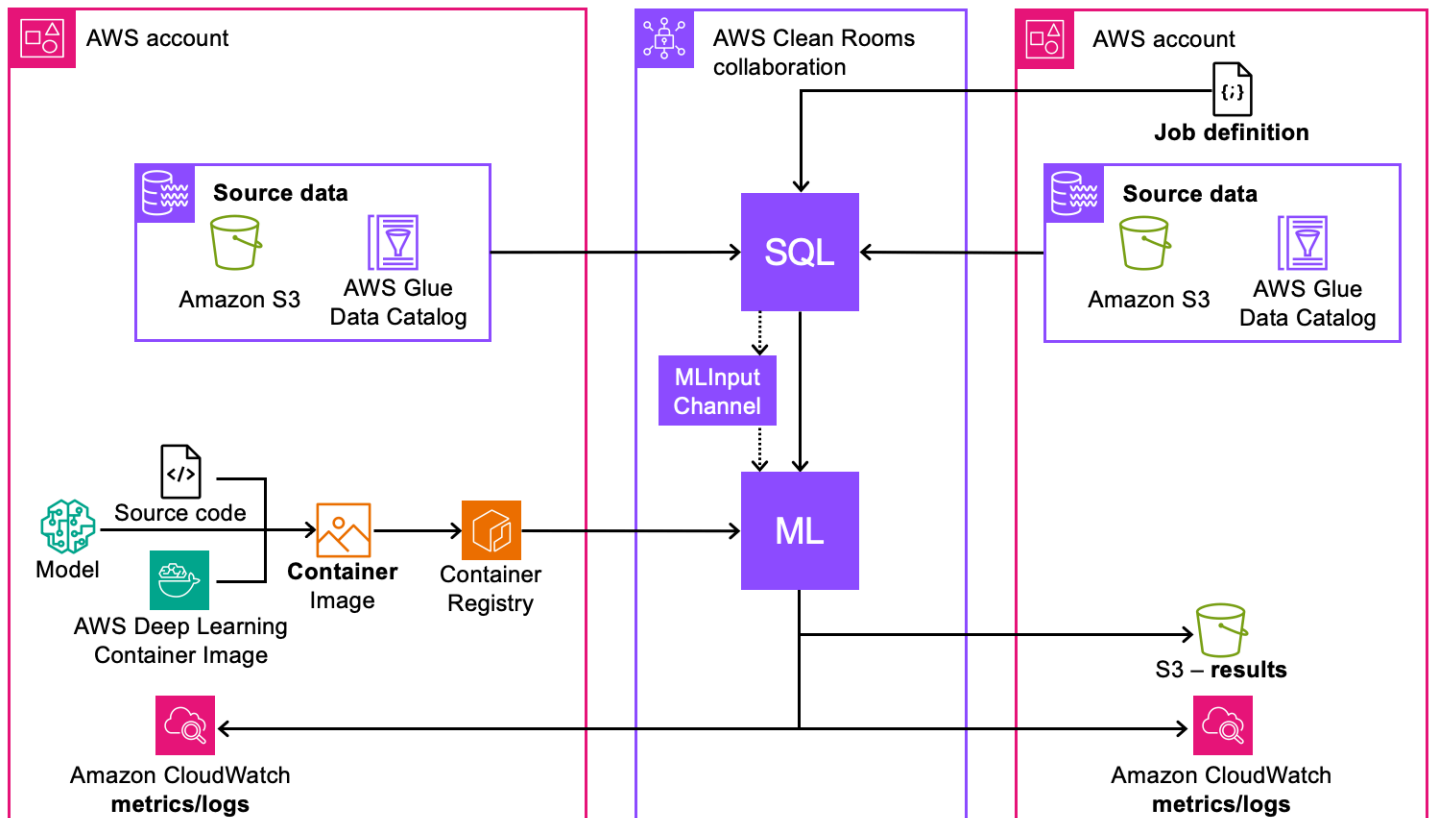
來自的查詢輸出 AWS Clean Rooms 可在主控台（如果主控台用於執行查詢）上取得，並在指定的 Amazon S3 儲存貯體中下載。您可以從該處使用其他 中的查詢輸出 AWS 服務，例如 Amazon Quick 和 Amazon SageMaker AI，取決於這些服務如何使用來自 Amazon S3 的資料。

如需 Amazon Quick 的詳細資訊，請參閱 [Amazon Quick 文件](#)。

如需 Amazon SageMaker AI 的詳細資訊，請參閱 [Amazon SageMaker AI 文件](#)。

AWS Clean Rooms ML 中的自訂建模

從技術角度來看，下圖說明自訂 ML 建模如何在 AWS Clean Rooms ML 中運作。



以下是自訂 ML 建模在 Clean Rooms ML 中的運作方式：

1. 資料來源組態

- 來源資料可以存放在 Amazon S3 目錄、AWS Glue Data Catalog 或 Snowflake 中
- AWS Glue Data Catalog 用於組織和編目
- 來自多個的資料 AWS 帳戶 可在相同的協同合作中使用

2. SQL 查詢和資料處理

- SQL 查詢用於存取和處理來源資料
- 查詢在 AWS Clean Rooms 協同合作界限內執行
- 處理的資料饋送至 ML 輸入通道以進行模型訓練

3. ML 模型開發

- 您可以使用 AWS 深度學習容器映像來開發模型的原始程式碼
- 自訂容器映像必須建立並儲存在 Amazon Elastic Container Registry 中

4. 基礎設施元件

- Amazon Elastic Container Registry 會存放和管理 ML 模型容器
- ML 處理會在安全 AWS Clean Rooms 協同合作環境中進行

5. 監控和記錄

- Amazon CloudWatch 提供兩個協作方的指標和日誌
- 監控可在 AWS 帳戶 參與協作的 之間使用
- 相關各方可存取效能指標和操作日誌

6. 結果管理

- 根據協同合作許可控制對結果的存取

開始使用之前，請參閱 [自訂 ML 建模先決條件](#) 和 [訓練容器的模型撰寫準則](#) 以取得詳細資訊。

隱私權增強的合成資料集產生

合成資料集具有與其基於的原始資料集類似的統計屬性，但不包含存在於原始資料集中的真實世界觀察。透過使用隱私權增強的合成資料集，您可以解鎖資料隱私權先前所考量的新機器學習 (ML) 模型訓練使用案例。當您建立 ML 輸入通道時，您可以產生合成資料，以在訓練 ML 模型時保護敏感資訊。

使用合成資料建立範本時，您必須：

- 要求範本輸出為合成
- 將輸出結構描述資料欄分類為數值或分類
- 根據組織需求自訂合成資料
- 調整隱私權設定：
 - 設定隱私權層級 (epsilon)
 - 設定隱私權閾值

Warning

合成資料產生可防止推斷個別屬性，無論原始資料集中是否存在特定個人，或是否存在這些個人的學習屬性。不過，它不會阻止原始資料集的常值出現，包括個人身分識別資訊 (PII) 出現在合成資料集中。

建議您避免輸入資料集中僅與一個資料主體相關聯的值，因為這些值可能會重新識別資料主體。例如，如果只有一個使用者住在郵遞區號中，則合成資料集中是否存在該郵遞區號，會確

認使用者位於原始資料集中。可以使用截斷高精確度值或將不常見目錄取代為其他等技術來降低此風險。這些轉換可以用來建立 ML 輸入通道之查詢的一部分。

如需如何為自訂模型訓練產生合成資料的詳細資訊，請參閱 [建立 SQL 分析範本](#)。

具有合成輸出的分析範本只能用來建立 ML 輸入通道。如需詳細資訊，請參閱 [在 AWS Clean Rooms ML 中建立 ML 輸入通道](#)。

合成資料產生的考量事項

使用 AWS Clean Rooms ML，協同合作成員可以建立合成資料集，以不可逆的方式從其集合資料集取消識別原始資料集的主體，以訓練自訂機器學習模型。建立協同合作時，您必須設定付款資訊，以指定誰支付合成資料產生的費用。以下是產生合成資料集和訓練自訂機器學習模型的高階步驟：

1. 協同合作成員會建立分析範本，其中包含：
 - 定義要合成的資料集所需的 SQL。
 - 隱私權相關組態，用於確保合成資料符合資料提供者的合規要求。
2. 所有資料提供者核准分析範本後，協同合作查詢執行器會使用範本建立機器學習 (ML) 輸入通道。
3. Clean Rooms ML 會產生合成資料集，並驗證是否符合分析範本中指定的隱私權閾值。
4. 如果滿足所有閾值，ML 輸入通道會填入合成資料集。
5. 然後，客戶可以使用此 ML 輸入通道來訓練與協同合作相關聯的自訂 ML 模型。

重要考量：

- 在 Clean Rooms ML 中產生的合成資料不會移除、修訂、混淆或淨化任何個別值，包括在原始資料集中找到的個人身分識別資訊 (PII)。合成資料集是由原始資料集的取樣值產生，而不是整個記錄。
- 如果原始資料集包含類似的資料列，則合成資料可能包含看起來與原始資料集中的資料列相同的資料列。

資料集準備：

- 避免類別分佈明顯不平衡的資料欄。這對預測值或「Y」資料欄特別重要。極端不平衡會降低合成資料集的整體隱私權。
- Clean Rooms ML 不支援從時間序列資料產生合成資料，其中維護序列記錄之間的關聯至關重要。
- Clean Rooms ML 不支援從文字或非結構化資料產生合成資料。

- 目前支援下列資料類型：

資
料
類
型
名
稱

BIGINT

BOOLEAN

CHAR

DATE

DECIMAL

FLOAT

INTEGER

LONG

REAL

SHORT

SMALLINT

TIME

TIMESTAMP
_LTZ

TIMESTAMP
_NTZ

TINYINT

資料
類型
名稱

VARCHAR

限制:

- 對於合成資料產生，預測資料欄的數量上限為 1。
- 如果目標欄是分類的，則原始資料集中的類別數量上限為 100。
- 在原始資料集中，資料列數目必須介於 1,500 到 250 萬之間，且資料欄數目上限為 1,000。對於目標欄中的非空值，資料列的最小數量為 1,000。

隱私權指標：

- Clean Rooms ML 提供隱私權分數，可測量產生的合成資料對成員資格推論攻擊 (MIAs) 的保護程度。此服務會從合成程序中保留 5% 的原始資料來計算此分數。
- 接近 50% 的分數被視為良好；分數越高表示對 MIAs 的保護越少。顯著低於 50% 的分數很罕見，可能是因為合成資料中原始資料的模式未呈現。

下游自訂模型：

- Clean Rooms ML 中產生的合成資料最適合訓練二進位分類模型和最多五個類別的多類別分類模型。
- 使用 Clean Rooms ML 中產生的合成資料訓練迴歸模型可能會導致模型準確性低，如根均方誤差 (RMSE) 所測量。

在 AWS Clean Rooms ML 中建立和加入協同合作

協作建立者負責建立協作、邀請成員並指派其角色。受邀成員加入協同合作並指定結果設定、訓練模型成品目的地設定，以及接受付款責任，具體取決於協同合作的設定方式。

建立機器學習的協同合作

下列程序說明如何建立機器學習的協同合作、邀請一或多個成員，以及指派可以開始模型訓練、接收結果、接收訓練過的模型結果，包括模型成品和指標，以及接收模型推論結果的成員。協作建立者也會指派將支付查詢運算、模型訓練和模型推論成本的成員。

Console

建立機器學習的協同合作（主控台）

1. [建立協同合作並邀請一或多個成員加入協同合作](#)
2. 使用查詢和任務指派下列成員分析能力：
 - 將執行查詢指派給將開始模型訓練的成員。
 - 將從分析接收結果指派給將接收查詢結果的成員。
3. 使用專用工作流程為 ML 建模指派下列成員能力：
 - 將從訓練模型接收輸出指派給將接收訓練模型結果的成員，包括模型成品和指標。
 - 將從模型推論接收輸出指派給將接收模型推論結果的成員。
4. 針對設定付款，指定將支付查詢運算、模型訓練、模型推論和合成資料集產生成本的成員。

合成資料集具有與其基於的原始資料集類似的統計屬性，但不包含存在於原始資料集中的真實世界觀察。透過使用隱私權增強的合成資料集，您可以解鎖新的 ML 模型訓練使用案例，該案例先前已防止資料隱私權問題。如需詳細資訊，請參閱[隱私權增強合成資料產生](#)。

這些成本都可以指派給相同或不同的成員。如果受邀成員是負責支付付款費用的成員，則必須在加入協同合作之前接受其付款責任。

5. 對於設定成員資格，協同合作建立者現在可以決定加入成員資格，或稍後建立成員資格。協作建立者接著必須設定 ML 組態。
 - a. 如果協同合作建立者也是結果接收者，他們也必須在結果設定預設值中指定查詢結果目的地和格式。
 - b. ML 組態為 Clean Rooms ML 提供將指標發佈至的角色 AWS 帳戶。如果協同合作建立者也收到訓練過的模型成品，他們可以指定用來接收結果的 Amazon S3 儲存貯體。
 - c. 在 ML 組態區段中，選取建立 ML 組態，然後在 Amazon S3 上指定模型輸出目的地，以及存取此位置所需的服務存取角色。
 - d. 如果協作建立者是負責支付付款成本的成員，他們必須在建立協作之前接受其付款責任。

API

建立機器學習 (API) 的協同合作

1. [建立協同合作並邀請一或多個成員加入協同合作](#)

2. 將下列角色指派給協同合作成員：

- CAN_QUERY - 指派給將開始模型訓練和推論的成員。
- CAN_RECEIVE_MODEL_OUTPUT - 指派給將收到訓練模型結果的成員。
- CAN_RECEIVE_INFERENCE_OUTPUT - 指派給將收到模型推論結果的成員。

如果協同合作建立者也是結果接收者，他們也必須在協同合作建立期間指定查詢結果目的地和格式。他們也提供服務角色 Amazon Resource Name (ARN) 將結果寫入查詢結果目的地。

- #### 3. 指定將支付查詢運算、模型訓練和模型推論成本的成員。這些成本都可以指派給相同或不同的成員。如果受邀成員是負責支付付款費用的成員，則必須在加入協同合作之前接受其付款責任。
- #### 4. 下列程式碼會建立協同合作、邀請可執行查詢和接收結果的成員，並將協同合作建立者指定為模型成品接收者。

```
import boto3
acr_client= boto3.client('cleanrooms')

collaboration = a_acr_client.create_collaboration(
    members=[
        {
            'accountId': 'invited_member_accountId',
            'memberAbilities': ["CAN_QUERY", "CAN_RECEIVE_RESULTS"],
            'displayName': 'member_display_name'
        }
    ],
    name='collaboration_name',
    description=collaboration_description,
    creatorMLMemberAbilities= {
        'customMLMemberAbilities': ["CAN_RECEIVE_MODEL_OUTPUT",
"CAN_RECEIVE_INFERENCE_OUTPUT"],
    },
    creatorDisplayName='creator_display_name',
    queryLogStatus="ENABLED",
    analyticsEngine="SPARK",
    creatorPaymentConfiguration={
        "queryCompute": {
            "isResponsible": True
```

```

    },
    "machineLearning": {
        "modelTraining": {
            "isResponsible": True
        },
        "modelInference": {
            "isResponsible": True
        }
    }
}

collaboration_id = collaboration['collaboration']['id']
print("collaborationId: {collaboration_id}")

member_membership = a_acr_client.create_membership(
    collaborationIdentifier = collaboration_id,
    queryLogStatus = 'ENABLED',
    paymentConfiguration={
        "queryCompute": {
            "isResponsible": True
        },
        "machineLearning": {
            "modelTraining": {
                "isResponsible": True
            },
            "modelInference": {
                "isResponsible": True
            }
        }
    }
)

```

5. 協作建立者接著必須設定 ML 組態。ML 組態為 Clean Rooms ML 提供將指標和日誌發佈至 的角色 AWS 帳戶。如果協同合作建立者也正在接收結果（模型成品或推論結果），他們可以指定用於接收結果的 Amazon S3 儲存貯體。

```

import boto3
acr_ml_client= boto3.client('cleanroomsmml')

acr_ml_client.put_ml_configuration(
    membershipIdentifier=membership_id,
    defaultOutputLocation={

```

```

        'roleArn': 'arn:aws:iam::account:role/roleName',
        'destination': {
            's3Destination': {
                's3Uri': "s3://bucketName/prefix"
            }
        }
    }
)

```

加入協同合作

協作建立者完成任務後，受邀的成員必須完成任務。

Console

建立成員資格並加入協同合作（主控台）

1. 受邀成員[會建立成員資格並加入協同合作](#)。
2. 如果受邀成員是負責付款的成員，包括查詢運算、模型訓練和模型推論成本，則必須在加入協同合作之前接受其付款責任。
3. 受邀成員會設定 ML 組態，為 Clean Rooms ML 提供將模型指標發佈至的角色 AWS 帳戶。如果他們也是接收訓練模型成品的成員，則必須提供存放訓練模型成品的 Amazon S3 儲存貯體。

API

建立成員資格並加入協作 (API)

1. 如果受邀成員是可以接收結果的成員，他們會指定查詢結果目的地和格式。他們也提供服務角色 ARN，允許服務寫入查詢結果目的地

如果受邀成員是負責付款的成員，包括查詢運算、模型訓練和模型推論成本，則必須在加入協同合作之前接受其付款責任。

如果受邀成員是負責支付模型訓練和模型推論以進行自訂建模的成員，則必須在加入協作之前接受其付款責任。

下列程式碼會在啟用查詢記錄的情況下建立成員資格。

```
import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_membership(
    membershipIdentifier='membership_id',
    queryLogStatus='ENABLED'
)
```

2. 受邀成員會設定 ML 組態，為 Clean Rooms ML 提供將模型指標發佈至 的角色 AWS 帳戶。如果他們也是接收訓練模型成品的成員，則必須提供存放訓練模型成品的 Amazon S3 儲存貯體。

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.put_ml_configuration(
    membershipIdentifier='membership_id',
    defaultOutputLocation={
        'roleArn':"arn:aws:iam::account:role/role_name",
        'destination':{
            's3Destination':{
                's3Uri':"s3://bucket_name/prefix"
            }
        }
    }
)
```

在 AWS Clean Rooms ML 中貢獻訓練資料

協同合作建立者建立協同合作並邀請成員加入後，您就可以將訓練資料提供給協同合作。任何成員都可以貢獻訓練資料。

Console

貢獻訓練資料（主控台）

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇 Tables (資料表)。
3. 在資料表頁面上，選擇設定新資料表。

4. 針對設定新資料表，針對資料來源選擇 Amazon S3、Amazon Athena 或 Snowflake，並根據您的資料來源完成下列步驟：

如果您使用的是	THEN
Amazon S3	1. 從下拉式清單中選擇資料庫，然後從資料庫中選取資料表。 2. 針對協同合作中允許的欄，選擇所有欄或自訂清單。 3. 如需已設定的資料表詳細資訊，請提供此資料表的名稱和選用的描述。 4. 如果您想要報告模型指標，請輸入指標的名稱和 Regex 陳述式，以搜尋輸出日誌來尋找指標。 5. 如果您想要為設定的資料表資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
Amazon Athena	1. 從下拉式清單中選擇資料庫，然後從資料庫中選取資料表。 2. 針對協同合作中允許的欄，選擇所有欄或自訂清單。 3. 如需已設定的資料表詳細資訊，請提供此資料表的名稱和選用的描述。 4. 如果您想要報告模型指標，請輸入指標的名稱和 Regex 陳述式，以搜尋輸出日誌來尋找指標。 5. 如果您想要為設定的資料表資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。

如果您使用的是	THEN
Snowflake	1. 使用現有的秘密 ARN 指定 Snowflake 登入資料，或儲存此資料表的新秘密。 2. 對於 Snowflake 資料表和結構描述詳細資訊，請手動輸入詳細資訊或自動匯入詳細資訊。 3. 針對結構描述，輸入資料欄名稱，然後從下拉式清單中選擇資料類型。 4. 如果您想要為設定的資料表資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。

5. 選擇設定新資料表。
6. 在資料表詳細資訊頁面上，選擇設定分析規則來設定此資料表的自訂分析規則。自訂分析規則會限制對資料的存取。您可以允許資料上一組特定的預先授權查詢，或允許一組特定的帳戶查詢您的資料。
 - a. 對於分析規則類型，選擇自訂，對於建立方法，選擇引導流程。
 - b. 選擇下一步。
 - c. 對於指定分析控制項，在檢閱每個新分析和允許特定協作者的任何分析之間進行選擇。
 - d. 選擇下一步。
 - e. （選用）對於指定分析結果控制項，對於輸出中不允許的資料欄，請指定是否要從輸出中排除任何資料欄。如果您選擇無，則不會從輸出中排除任何資料欄。如果您選擇自訂清單，您可以指定要從輸出中移除的特定資料欄。
 - f. 對於套用至輸出的其他分析，指定您是否要允許、拒絕或要求在產生結果之前進行其他分析。
 - g. 選擇下一步。
 - h. （選用）針對設定差異隱私權，選擇關閉。
 - i. 選擇下一步。
 - j. 檢閱檢閱和設定頁面上的資訊，然後選擇設定分析規則。
7. 在資料表詳細資訊頁面中，選擇關聯以協同合作。
8. 在關聯資料表對話方塊中，選取您要將此資料表與之建立關聯的協同合作，然後選擇選擇協同合作。

9. 在關聯資料表頁面上，檢閱並驗證資料表關聯詳細資訊、服務存取和標籤中的資訊。選擇關聯資料表。
10. 在與您相關聯的資料表中，選取您剛相關聯的資料表旁的選項按鈕。從動作功能表中，選擇協作分析規則群組中的設定。
11. 在設定協同合作分析規則頁面上，針對允許的額外分析，選擇任何協同合作成員或特定協同合作成員是否可以執行其他分析。

針對結果交付，選擇允許哪些成員從查詢輸出接收結果。

12. 選擇設定分析規則。

API

貢獻訓練資料 (API)

1. 透過 AWS Clean Rooms 提供可使用的 AWS Glue 資料表和資料欄，設定在 中使用的現有資料表。

使用特定參數執行下列程式碼。

```
import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_configured_table(
    name='configured_table_name',
    tableReference= {
        'glue': {
            'tableName': 'glue_table_name',
            'databaseName': 'glue_database_name'
        }
    },
    analysisMethod="DIRECT_QUERY",
    allowedColumns=["column1", "column2", "column3",...]
)
```

2. 設定自訂分析規則，以限制對資料的存取。您可以允許資料上一組特定的預先授權查詢，或允許一組特定的帳戶查詢您的資料。

使用特定參數執行下列程式碼。

```
import boto3
```

```
acr_client= boto3.client('cleanrooms')

acr_client.create_configured_table_analysis_rule(
    configuredTableIdentifier='configured_table_id',
    analysisRuleType='CUSTOM',
    analysisRulePolicy= {
        'v1': {
            'custom': {
                'allowedAnalyses': ['ANY_QUERY'],
                'allowedAnalysisProviders': ['query_runner_account'],
                'additionalAnalyses': "REQUIRED"
            }
        }
    }
)
```

在此範例中，允許特定帳戶對資料執行任何查詢，而且需要額外的分析。

3. 將設定的資料表與協同合作建立關聯，並提供 AWS Glue 資料表的服務存取角色。

使用特定參數執行下列程式碼。

```
import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_configured_table_association(
    name='configured_table_association_name',
    membershipIdentifier='membership_id',
    configuredTableIdentifier='configured_table_id',
    roleArn='arn:aws:iam::account:role/role_name'
)
```

Note

此服務角色具有資料表的許可。服務角色只能由 擔任 AWS Clean Rooms，以代表可查詢的成員執行允許的查詢。協作成員（資料擁有者除外）無法存取協作中的基礎資料表。資料擁有者可以關閉差異隱私權，使其資料表可供其他成員查詢。

4. 最後，將分析規則新增至設定的資料表關聯。

使用特定參數執行下列程式碼。

```

import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_configured_table_association_analysis_rule(

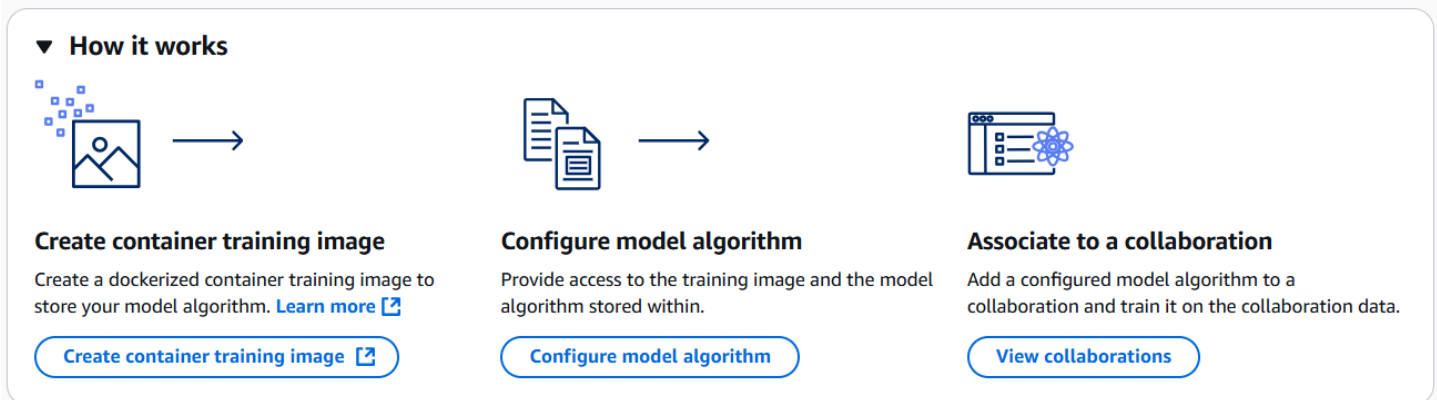
    configuredTableAssociationIdentifier='configured_table_association_identifier',
    membershipIdentifier='membership_id',
    configuredTableIdentifier='configured_table_id',
    analysisRuleType = 'CUSTOM',
    analysisRulePolicy= {
        'v1': {
            'custom': {
                'allowedAdditionalAnalyses':
                ['configured_model_algorithm_association_arns'],
                'allowedResultReceivers': ['query_runner_account']
            }
        }
    }
)

```

在 AWS Clean Rooms ML 中設定模型演算法

[建立容器訓練映像](#)之後，您必須設定模型演算法。設定模型演算法可讓您與協同合作建立關聯。

下圖顯示將模型演算法設定為建立容器訓練映像之後，以及將其與協同合作建立關聯之前發生的步驟。



Console

設定自訂 ML 模型演算法 (主控台)

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇自訂 ML 模型。
3. 在自訂 ML 模型頁面上，選擇設定模型演算法。
4. 在設定模型演算法頁面上，針對模型演算法詳細資訊，輸入名稱和選用的描述。
5. 如果您想要執行模型訓練，對於訓練映像 ECR 容器詳細資訊，
 - a. 選取指定訓練映像 URI 核取方塊。
 - b. 從下拉式清單中選取具有訓練模型、推論容器或兩者的儲存庫。
 - c. 選取映像。
 - d. (選用) 輸入進入點的值以存取訓練影像。
 - e. (選用) 輸入引數的值。
6. (選用) 如果您想要報告模型指標，請在訓練指標中輸入指標的名稱和 Regex 陳述式，以搜尋輸出日誌來尋找指標。
7. 如果您想要執行模型推論，對於推論映像 ECR 容器詳細資訊，
 - a. 選取指定推論映像 URI 核取方塊。
 - b. 從下拉式清單中選取儲存庫。
 - c. 選取映像。
8. 針對服務存取，選擇將用於存取此資料表的現有服務角色名稱。
9. 針對加密，選擇自訂加密設定以指定您自己的 KMS 金鑰和相關資訊。否則，Clean Rooms ML 會管理加密。
10. 如果您想要啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
11. 選擇設定模型演算法。

API

設定自訂 ML 模型演算法 (API)

1. 建立 SageMaker AI 相容 Docker 映像。Clean Rooms ML 僅支援 SageMaker AI 相容 Docker 映像。

2. 建立 SageMaker AI 相容 Docker 映像之後，請使用 Amazon ECR 建立訓練映像。遵循 [Amazon Elastic Container Registry 使用者指南](#) 中的指示來建立容器訓練映像。
3. 設定模型演算法以用於 Clean Rooms ML。您必須提供下列資訊：
 - Amazon ECR 儲存庫連結和其他引數來訓練模型並執行推論。Clean Rooms ML 支援在推論容器上執行批次轉換任務。
 - 允許 Clean Rooms ML 存取儲存庫的服務存取角色。
 - (選用) 推論容器。雖然您可以在個別設定的模型演算法中提供此功能，但建議您在此步驟中提供此功能，以便將訓練和推論容器作為相同資源的一部分進行管理。

使用特定參數執行下列程式碼。

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.create_configured_model_algorithm(
    name='configured_model_algorithm_name',
    trainingContainerConfig={
        'imageUri': 'account.dkr.ecr.region.amazonaws.com/image_name:tag',
        'metricDefinitions': [
            {
                'name': 'custom_metric_name_1',
                'regex': 'custom_metric_regex_1'
            }
        ]
    },
    inferenceContainerConfig={
        'imageUri': 'account.dkr.ecr.region.amazonaws.com/image_name:tag',
    },
    roleArn='arn:aws:iam::account:role/role_name'
)
```

在 AWS Clean Rooms ML 中關聯設定的模型演算法

設定模型演算法之後，您就可以將模型演算法與協同合作建立關聯。與模型演算法建立關聯可讓協同合作的所有成員使用模型演算法。

下圖顯示建立容器訓練映像和設定模型演算法之後，將設定的模型演算法關聯為最後一個步驟。

▼ How it works



Create container training image

Create a dockerized container training image to store your model algorithm. [Learn more](#)

[Create container training image](#)



Configure model algorithm

Provide access to the training image and the model algorithm stored within.

[Configure model algorithm](#)



Associate to a collaboration

Add a configured model algorithm to a collaboration and train it on the collaboration data.

[View collaborations](#)

Console

Note

模型演算法建立關聯後，就無法編輯。若要進行變更，您可以刪除相關聯的模型演算法，並關聯新的模型演算法。

建立自訂 ML 模型演算法的關聯（主控台）

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇自訂 ML 模型。
3. 在自訂 ML 模型頁面上，選擇您要與協同合作建立關聯的已設定模型演算法，然後選擇與協同合作建立關聯。
4. 在關聯設定的模型演算法視窗中，選擇您要關聯的協同合作。
5. 選擇選擇協同合作。
6. 在關聯模型演算法頁面上，針對模型演算法關聯詳細資訊，輸入名稱和選用的描述。
7. 針對模型演算法，選擇已設定的模型演算法。
8. 對於訓練過的模型匯出隱私權組態，
 - a. 若要匯出模型檔案，請選取模型檔案核取方塊。
 - b. 若要匯出輸出檔案，請選取輸出檔案核取方塊。
 - c. 輸入匯出資料的最大值。值必須介於 0.01 和 10 之間。
9. （選用）如果您想要傳送完整的錯誤日誌或較短的錯誤摘要給成員，請在經過訓練的模型推論任務隱私權組態下，

- a. 在完整日誌下，從下拉式清單中選取一或多個帳戶 IDs。
 - b. (選用) 如果您想要傳送符合篩選條件模式的日誌，請輸入篩選條件模式。
 - c. (選用) 如果您要新增另一個帳戶和選用篩選條件模式，請選擇新增日誌政策。
 - d. 在錯誤摘要下，從下拉式清單中選取一或多個帳戶 IDs。
 - e. (選用) 選取要修訂的一或多個實體，以指定要從錯誤日誌或錯誤摘要中修訂哪些實體。
 - PII – 修訂個人身分識別資訊
 - 數字 – 編輯數字
 - 自訂 – 根據自訂修訂模式進行修訂
 - i. 如果您在上一個步驟中選擇自訂，請輸入自訂修訂模式。這會記錄與此模式相符的資訊。
 - ii. (選用) 如果您想要新增另一個自訂修訂模式，請選擇新增另一個自訂模式。
10. (選用) 如果您想要設定訓練模型指標，請在訓練模型指標組態下，從下拉式清單中選取雜訊層級。
- 您可以選擇無、低、中和高。
11. (選用) 如果您想要設定成品大小上限，請在成品組態下輸入成品大小上限值。值必須介於 0.01 和 10 之間。
12. (選用) 如果您想要啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
13. 選擇關聯。

API

建立自訂 ML 模型演算法 (API) 的關聯

使用特定參數執行下列程式碼。

您也可以提供隱私權政策，定義可存取不同日誌的人員、允許客戶定義 regex，以及可以從訓練模型輸出或推論結果匯出的資料量。

Note

設定的模型演算法關聯是不可變的。

```

import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.create_configured_model_algorithm_association(
    name='configured_model_algorithm_association_name',
    description='purpose of the association',
    configuredModelAlgorithmArn='arn:aws:cleanrooms-ml:region:account:membership/
membershipIdentifier/configured-model-algorithm/identifier',
    privacyConfiguration={
        "policies": {
            "trainedModelExports": {
                "filesToExport": ['files to export'],
                "containerLogs": [
                    {
                        "allowedAccountIds": ['member_account_id'],
                        "filterPattern": ['filter pattern'],
                        "logRedactionConfiguration": {
                            "entitiesToRedact": [
                                'ALL_PERSONALLY_IDENTIFIABLE_INFORMATION',
                                'NUMBERS',
                                'CUSTOM'
                            ],
                        "customEntityConfig": {
                            "customDataIdentifiers": [
                                'custom_regex_1',
                                'custom_regex_2'
                            ]
                        }
                    }
                ]
            },
            "containerMetrics": {
                "noiseLevel": 'noise value'
            },
            "maxArtifactSize": {
                "unit": 'unit',
                "value": 'number'
            }
        },
        "trainedModelInferenceJobs": {
            "containerLogs": [
                {
                    "allowedAccountIds": ['member_account_id'],

```



```

        "filterPattern": ['filter pattern'],
        "logRedactionConfiguration": {
            "entitiesToRedact": [
                'ALL_PERSONALLY_IDENTIFIABLE_INFORMATION',
                'NUMBERS',
                'CUSTOM'
            ],
            "customEntityConfig": {
                "customDataIdentifiers": [
                    'custom_regex_1',
                    'custom_regex_2'
                ]
            }
        }
    ],
    "maxOutputSize": {
        "unit": 'unit',
        "value": 'number'
    }
}
},
tags={
    'tag': 'tag'
}
)

```

在設定的模型演算法與協同合作相關聯後，訓練資料提供者必須將協同合作分析規則新增至其資料表。此規則允許設定的模型演算法關聯存取其設定的資料表。所有貢獻的訓練資料提供者都必須執行下列程式碼：

```

import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_configured_table_association_analysis_rule(
    membershipIdentifier= 'membership_id',
    configuredTableAssociationIdentifier= 'configured_table_association_id',
    analysisRuleType= 'CUSTOM',
    analysisRulePolicy = {
        'v1': {
            'custom': {

```

```

        'allowedAdditionalAnalyses': ['arn:aws:cleanrooms-
ml:region:*:membership/*/configured-model-algorithm-association/*'],
        'allowedResultReceivers': []
    }
}
)

```

Note

由於設定的模型演算法關聯不可變，我們建議訓練希望允許列出模型的資料提供者，以便在自訂模型組態的前幾次反覆allowedAdditionalAnalyses運算期間使用中的萬用字元。這可讓模型提供者在其程式碼上反覆運算，而不需要其他訓練提供者在訓練更新後的模型程式碼與資料之前重新建立關聯。

在 AWS Clean Rooms ML 中建立 ML 輸入通道

先決條件：

- AWS 帳戶 可存取 的 AWS Clean Rooms
- 您要 AWS Clean Rooms 在其中建立 ML 輸入通道的協同合作設定
- 在協同合作中查詢資料和建立 ML 輸入通道的許可。
- (選用) 要與 ML 輸入通道建立關聯的現有模型演算法，或建立新的模型演算法的許可
- (選用) 具有分析規則的資料表，可針對您指定的模型執行。
- (選用) 用於產生資料集的現有 SQL 查詢或分析範本
- (選用) 具有適當許可的現有服務角色，或建立新服務角色的許可
- (選用) 如果您想要使用自己的加密 AWS KMS 金鑰的自訂金鑰
- 在協同合作中建立和管理 ML 模型的適當許可

ML 輸入通道是從特定資料查詢建立的資料集。能夠查詢資料的成員可以透過建立 ML 輸入通道來準備其資料以進行訓練和推論。建立 ML 輸入通道可讓該資料在相同的協同合作中用於不同的訓練模型。您應該為訓練和推論建立個別的 ML 輸入通道。

若要建立 ML 輸入通道，您必須指定用來查詢輸入資料的 SQL 查詢，並建立 ML 輸入通道。此查詢的結果絕不會與任何成員共用，並保持在 Clean Rooms ML 的界限內。參考 Amazon Resource Name (ARN) 用於後續步驟，以訓練模型或執行推論。

Console

建立 ML 輸入通道（主控台）

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 在協同合作頁面上，選擇您要建立 ML 輸入通道的協同合作。
4. 協同合作開啟後，選擇 ML 模型索引標籤。
5. 在自訂 ML 模型的 ML 輸入通道區段中，選擇建立 ML 輸入通道。
6. 在建立 ML 輸入通道頁面上，針對 ML 輸入通道詳細資訊，執行下列動作：
 - a. 在名稱中，輸入頻道的唯一名稱。
 - b. （選用）針對描述，輸入頻道的描述。
 - c. 針對關聯的模型演算法，選取要使用的演算法。

選擇關聯模型演算法以新增。

7. 針對資料集，選擇產生訓練資料集的方法：
 - 選擇 SQL 查詢以使用 SQL 查詢的結果做為訓練資料集。

如果您選擇 SQL 查詢，請在 SQL 查詢欄位中輸入您的查詢。

（選用）若要匯入您最近使用的查詢，請選擇從最近的查詢匯入。

- 選擇分析範本以使用分析範本的結果做為訓練資料集。

Warning

合成資料產生可防止推斷個別屬性，無論原始資料集中是否存在特定個人，或是否存在這些個人的學習屬性。不過，它不會阻止原始資料集的常值出現，包括個人身分識別資訊 (PII) 出現在合成資料集中。

建議您避免輸入資料集中僅與一個資料主體相關聯的值，因為這些值可能會重新識別資料主體。例如，如果只有一個使用者住在郵遞區號中，則合成資料集中是否存在該郵遞區號，會確認使用者位於原始資料集中。可以使用截斷高精確度值或將不常見目

錄取代為其他等技術來降低此風險。這些轉換可以用來建立 ML 輸入通道之查詢的一部分。

- a. 如果沒有相關聯的資料表，請選擇關聯資料表以新增具有可對指定模型執行之分析規則的資料表。
- b. 選擇建立此資料通道時要使用的工作者類型。預設工作者類型為 CR.1X。指定要使用的工作者數量。預設工作者編號為 16。若要指定 Spark 屬性：
 - i. 展開 Spark 屬性。
 - ii. 選擇新增 Spark 屬性。
 - iii. 在 Spark 屬性對話方塊中，從下拉式清單中選擇屬性名稱，然後輸入值。

下表提供每個屬性的定義。

如需 Spark 屬性的詳細資訊，請參閱 Apache Spark 文件中的 [Spark 屬性](#)。

屬性名稱	Description	預設值
spark.task.maxFailures	控制任務在任務失敗之前可以失敗的連續次數。需要大於或等於 1 的值。允許重試次數等於此值減去 1。如果任何嘗試成功，就會重設失敗計數。不同任務的故障不會累積到此限制。	4
spark.sql.files.maxPartitionBytes	設定從 Parquet、JSON 和 ORC 等檔案型來源讀取時要封裝到單一分割區的最大位元組數。	128MB
spark.hadoop.fs.s3.maxRetries	設定 Amazon S3 檔案操作的重試嘗試次數上限。	

屬性名稱	Description	預設值
spark.network.timeout	設定所有網路互動的預設逾時。如果未設定，則覆寫下列逾時設定： - spark.storage.blockManagerHeartbeatTimeoutMs - spark.shuffle.io.connectionTimeout - spark.rpc.askTimeout - spark.rpc.lookupTimeout	120 秒
spark.rdd.compress	指定是否使用 spark.io.compression.codec 壓縮序列化 RDD 分割區。適用於 Java 和 Scala 中的 StorageLevel.MEMORY_ONLY_SER，或 Python 中的 StorageLevel.MEMORY_ONLY。減少儲存空間，但需要額外的 CPU 處理時間。	FALSE
spark.shuffle.spill.compress	指定是否使用 spark.io.compression.codec 壓縮隨機溢出資料。	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	在 spark.sql.adaptive.enabled 為 true 的適應性最佳化期間，設定隨機播放分割區的目標大小，以位元組為單位。合併小型分割區或分割扭曲分割區時，控制分割區大小。	(spark.sql.adaptive.shuffle.targetPostShuffleInputSize 的值)
spark.sql.adaptive.autoBroadcastJoinThreshold	設定在聯結期間廣播至工作者節點的資料表大小上限，以位元組為單位。僅適用於自適應架構。使用與 spark.sql.autoBroadcastJoinThreshold 相同的預設值。設定為 -1 以停用廣播。	(無)

屬性名稱	Description	預設值
spark.sql.adaptive.coalescePartitions.enabled	指定是否要根據 spark.sql.adaptive.advisoryPartitionSizeInBytes 來合併連續隨機播放分割區，以最佳化任務大小。需要 spark.sql.adaptive.enabled 才能成立。	TRUE
spark.sql.adaptive.coalescePartitions.initialPartitionNum	在合併之前定義隨機播放分割區的初始數量。同時需要 spark.sql.adaptive.enabled 和 spark.sql.adaptive.coalescePartitions.enabled 為 true。預設為 spark.sql.shuffle.partitions 的值。	(無)
spark.sql.adaptive.coalescePartitions.minPartitionSize	設定合併隨機播放分割區的大小下限，以防止分割區在適應性最佳化期間變得太小。	1 MB
spark.sql.adaptive.coalescePartitions.parallelismFirst	指定是否要在分割區合併期間根據叢集平行處理而非 spark.sql.adaptive.advisoryPartitionSizeInBytes 計算分割區大小。產生的分割區大小小於設定的目標大小，以最大化平行處理。我們建議在忙碌叢集上將此設定為 false，以透過防止過多的小型任務來改善資源使用率。	TRUE
spark.sql.adaptive.enabled	指定是否啟用自適應查詢執行，以在查詢執行期間根據準確的執行時間統計資料重新最佳化查詢計劃。	TRUE
spark.sql.adaptive.forceOptimizeSkewedJoin	指定是否強制啟用 OptimizeSkewedJoin，即使它引入額外的隨機播放。	FALSE
spark.sql.adaptive.localShuffleReader.enabled	指定是否在不需隨機分割時使用本機隨機讀取器，例如從排序合併聯結轉換為廣播雜湊聯結之後。需要 spark.sql.adaptive.enabled 才能成立。	TRUE

屬性名稱	Description	預設值
spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold	設定用於建置本機雜湊映射的分割區大小上限，以位元組為單位。在下列情況下，將隨機雜湊聯結優先於排序合併聯結： - 此值等於或超過 spark.sql.adaptive.advisoryPartitionSizeInBytes - 所有分割區大小都在此限制內 覆寫 spark.sql.join.preferSortMergeJoin 設定。	0 個位元組
spark.sql.adaptive.optimizeSkewsInRebalancePartitions.enabled	指定是否根據 spark.sql.adaptive.advisoryPartitionSizeInBytes 將扭曲隨機分割區分割為較小的分割區，以最佳化扭曲隨機分割區。需要 spark.sql.adaptive.enabled 才能成立。	TRUE
spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor	定義分割期間合併分割區的大小閾值因素。小於此因素的分割區乘以 spark.sql.adaptive.advisoryPartitionSizeInBytes 會合併。	0.2
spark.sql.adaptive.skewJoin.enabled	指定是否透過分割和選擇性複寫扭曲的分割區來處理隨機聯結中的資料扭曲。適用於排序合併和隨機雜湊聯結。需要 spark.sql.adaptive.enabled 才能成立。	TRUE
spark.sql.adaptive.skewJoin.skewedPartitionFactor	決定決定分割區扭曲的大小因素。當分割區的大小超過以下兩者時，就會扭曲： - 此係數乘以中位數分割區大小 - spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes 的值	5

屬性名稱	Description	預設值
spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes	設定識別扭曲分割區的大小閾值，以位元組為單位。當分割區的大小超過以下兩者時，就會扭曲： - 此閾值 - 中位數分割區大小乘以 spark.sql.adaptive.skewJoin.skewedPartitionFactor 我們建議您設定大於 spark.sql.adaptive.advisoryPartitionSizeInBytes 的值。	256MB
spark.sql.autoBroadcastJoinThreshold	設定在聯結期間廣播至工作者節點的資料表大小上限，以位元組為單位。設定為 -1 以停用廣播。	10MB
spark.sql.broadcastTimeout	控制廣播聯結期間廣播操作的逾時期間，以秒為單位。	300 秒
spark.sql.cbo.enabled	指定是否要為計劃統計資料估算啟用成本型最佳化 (CBO)。	FALSE
spark.sql.cbo.joinReorder.dp.starfilter	指定是否要在成本型聯結列舉期間套用星狀聯結篩選條件啟發式。	FALSE
spark.sql.cbo.joinReorder.dp.threshold	設定動態程式設計演算法中允許的聯結節點數目上限。	12
spark.sql.cbo.joinReorder.enabled	指定是否要在成本型最佳化 (CBO) 中啟用聯結重新排序。	FALSE
spark.sql.cbo.planStats.enabled	指定是否要在邏輯計畫產生期間從目錄擷取資料列計數和資料欄統計資料。	FALSE

屬性名稱	Description	預設值
spark.sql.cbo.starSchemaDetection	指定是否根據星狀結構描述偵測啟用聯結重新排序。	FALSE
spark.sql.files.maxPartitionNum	設定檔案型來源 (Parquet、JSON 和 ORC) 的目標分割檔案分割區數量上限。當初始計數超過此值時，重新擴展分割區。這是建議的目標，而不是保證的限制。	(無)
spark.sql.files.maxRecordsPerFile	設定寫入單一檔案的記錄數目上限。設定為零或負值時不會套用任何限制。	0
spark.sql.files.minPartitionNum	設定檔案型來源 (Parquet、JSON 和 ORC) 的目標分割檔案分割區數量下限。預設為 spark.sql.leafNodeDefaultParallelism。這是建議的目標，而不是保證的限制。	(無)
spark.sql.inMemoryColumnarStorage.batchSize	控制單欄式快取的批次大小。增加大小可改善記憶體使用率和壓縮，但會增加out-of-memory錯誤的風險。	10000
spark.sql.inMemoryColumnarStorage.compressed	指定是否根據資料統計資料自動選取資料欄的壓縮轉碼器。	TRUE
spark.sql.inMemoryColumnarStorage.enableVectorizedReader	指定是否要為單欄式快取啟用向量化讀取。	TRUE

屬性名稱	Description	預設值
<code>spark.sql.legacy.allowHashOnMapType</code>	指定是否允許對映射類型資料結構執行雜湊操作。此舊版設定可維持與舊版 Spark 版本的映射類型處理的相容性。	
<code>spark.sql.legacy.allowNegativeScaleOfDecimal</code>	指定是否在十進位類型定義中允許負縮放值。此舊版設定可維持與支援負小數位數擴展之較舊 Spark 版本的相容性。	
<code>spark.sql.legacy.castComplexTypesToString.enabled</code>	指定是否啟用舊版行為，將複雜類型轉換為字串。維持與舊版 Spark 版本類型轉換規則的相容性。	
<code>spark.sql.legacy.charVarcharAsString</code>	指定是否將 CHAR 和 VARCHAR 類型視為 STRING 類型。此舊版設定提供與較舊 Spark 版本的字串類型處理相容性。	
<code>spark.sql.legacy.createEmptyCollectionUsingStringType</code>	指定是否使用字串類型元素建立空集合。此舊版設定可維持與較舊 Spark 版本的集合初始化行為的相容性。	
<code>spark.sql.legacy.exponentLiteralAsDecimal.enabled</code>	指定是否要將指數常值解譯為十進位類型。此舊版設定可維持與舊版 Spark 版本數值常值處理的相容性。	
<code>spark.sql.legacy.json.allowEmptyString.enabled</code>	指定是否要在 JSON 處理中允許空字串。此舊版設定可維持與舊版 Spark 版本的 JSON 剖析行為的相容性。	
<code>spark.sql.legacy.parquet.int96RebaseModelRead</code>	指定是否在讀取 Parquet 檔案時使用舊版 INT96 時間戳記重新基礎模式。此舊版設定可維持與較舊 Spark 版本時間戳記處理的相容性。	

屬性名稱	Description	預設值
spark.sql.legacy.timeParserPolicy	控制時間剖析行為以實現回溯相容性。此舊版設定會決定如何從字串剖析時間戳記和日期。	
spark.sql.legacy.typeCoercion.datetimeToString.enabled	指定是否在將日期時間值轉換為字串時啟用舊版類型強制行為。維持與較舊 Spark 版本日期時間轉換規則的相容性。	
spark.sql.maxSinglePartitionBytes	設定以位元組為單位的分割區大小上限。規劃器為較大的分割區引進隨機播放操作，以改善平行處理。	128 公尺
spark.sql.metadataCacheTTLSeconds	控制中繼資料快取的time-to-live (TTL)。適用於分割區檔案中繼資料和工作階段目錄快取。需要： - 大於零的正值 - spark.sql.catalogImplementation 設定為 hive - spark.sql.hive.filesourcePartitionFileCacheSize 大於零 - spark.sql.hive.manageFilesourcePartitions 設定為 true	-1000 毫秒
spark.sql.optimizer.collapseProjectAlwaysInline	指定是否要摺疊相鄰的投影和內嵌表達式，即使它會導致重複。	FALSE
spark.sql.optimizer.dynamicPartitionPruning.enabled	指定是否要為用作聯結索引鍵的分割區資料欄產生述詞。	TRUE

屬性名稱	Description	預設值
spark.sql.optimize r.enableC svExpress ionOptimization	指定是否要透過從_csv 操作剪除不必要的資料欄來最佳化 SQL 最佳化工具中的 CSV 表達式。	TRUE
spark.sql.optimize r.enableJ sonExpres sionOptimization	指定是否要透過下列方式最佳化 SQL 最佳化工具中的 JSON 表達式： - 從_json 操作刪除不必要的資料欄 - 簡化從_json 和 到_json 的組合 - 最佳化 named_struct 操作	TRUE
spark.sql.optimize r.excludedRules	定義要停用的最佳化工具規則，以逗號分隔的規則名稱識別。某些規則無法停用，因為其為正確性的必要條件。最佳化工具會記錄哪些規則已成功停用。	(無)
spark.sql.optimize r.runtime.bloomFil ter.applicationSid eScanSize Threshold	設定在應用程式端插入 Bloom 篩選條件所需的最小彙總掃描大小，以位元組為單位。	10GB
spark.sql.optimize r.runtime.bloomFil ter.creationSideTh reshold	定義在建立端注入 Bloom 篩選條件的大小閾值上限。	10MB
spark.sql.optimize r.runtime.bloomFil ter.enabled	指定是否要插入 Bloom 篩選條件，以便在隨機聯結的一側具有選擇性述詞時減少隨機播放資料。	TRUE
spark.sql.optimize r.runtime.bloomFil ter.expec tedNumItems	定義執行時間 Bloom 篩選條件中預期項目的預設數量。	1000000

屬性名稱	Description	預設值
spark.sql.optimize r.runtime.bloomFilter.maxNumBits	設定執行時間 Bloom 篩選條件中允許的位元數上限。	67108864
spark.sql.optimize r.runtime.bloomFilter.maxNumItems	設定執行時間 Bloom 篩選條件中允許的預期項目數量上限。	4000000
spark.sql.optimize r.runtime.bloomFilter.number.threshold	限制每個查詢允許的非 DPP 執行時間篩選條件數量上限，以防止 out-of-memory 錯誤。	10
spark.sql.optimize r.runtime.bloomFilter.numBits	定義執行時間 Bloom 篩選條件中使用的預設位元數。	8388608
spark.sql.optimize r.runtime.rowlevel OperationGroupFilter.enabled	指定是否要為資料列層級操作啟用執行時間群組篩選。允許資料來源： • 使用資料來源篩選條件刪除整個資料群組（例如檔案或分割區） • 執行執行時間查詢以識別相符的記錄 • 捨棄不必要的群組，以避免昂貴的重寫 限制： • 並非所有表達式都可以轉換為資料來源篩選條件 • 有些表達式需要 Spark 評估（例如子查詢）	TRUE

屬性名稱	Description	預設值
spark.sql.optimize r.runtimeFilter.se miJoinRed uction.enabled	指定是否要插入半聯結，以在隨機聯結的一側具有選擇性述詞時減少隨機資料。	FALSE
spark.sql.parquet. aggregate Pushdown	指定是否將彙總下推至 Parquet 以進行最佳化。支援： - 布林值、整數、浮點數和日期類型的 MIN 和 MAX - 所有資料類型的 COUNT 如果任何 Parquet 檔案頁尾缺少統計資料，則擲回例外狀況。	FALSE
spark.sql.parquet. columnarR eaderBatchSize	控制每個 Parquet 向量化讀取器批次中的資料列數。選擇平衡效能額外負荷和記憶體用量的值，以防止out-of-memory錯誤。	4096
spark.sql.session. timeZone	定義工作階段時區，以處理字串常值和 Java 物件轉換中的時間戳記。接受： - 區域/城市格式的區域型 IDs (例如 America/Los_Angeles) - 區域位移，以 (+/-)HH、(+/-)HH : mm 或 (+/-)HH : mm : ss 格式顯示 (例如 -08 或 +01 : 00) +00 : 00 的 UTC 或 Z 別名	(本機時區的值)
spark.sql.shuffle. partitions	設定聯結或彙總期間資料隨機播放的預設分割區數量。結構式串流查詢從相同檢查點位置重新啟動之間無法修改。	200

屬性名稱	Description	預設值
spark.sql.shuffledHashJoinFactor	定義用於判斷隨機雜湊聯結資格的乘數係數。當小端資料大小乘以此因素小於大型端資料大小時，會選取隨機雜湊聯結。	3
spark.sql.sources.parallelPartitionDiscovery.threshold	使用檔案型來源 (Parquet、JSON 和 ORC) 設定驅動程式端檔案清單的路徑數目上限。在分割區探索期間超過 時，會使用個別的 Spark 分散式任務列出檔案。	32
spark.sql.statistics.histogram.enabled	指定是否要在資料欄統計資料運算期間產生等高長長條圖，以改善估算準確度。需要額外的資料表掃描，超過基本資料欄統計資料所需的掃描。	FALSE

- c. 對於以天為單位的資料保留，請輸入要保留資料的天數。
- d. 針對結果格式，選擇 CSV 或 Parquet 做為 ML 輸入通道應使用的資料格式。
8. 針對服務存取，選擇將用於存取此表格的現有服務角色名稱，或選擇建立並使用新的服務角色。
9. 針對加密，選擇使用自訂 KMS 金鑰加密秘密，以指定您自己的 KMS 金鑰和相關資訊。否則，Clean Rooms ML 會管理加密。
10. 選擇建立 ML 輸入通道。

建立 ML 輸入通道需要幾分鐘的時間。您可以在 ML 模型索引標籤上查看 ML 輸入通道的清單。

Note

建立 ML 輸入通道之後，您就無法編輯它。

API

建立 ML 輸入通道 (API)

使用特定參數執行下列程式碼：

```

import boto3
acr_client = boto3.client('cleanroomsmml')

acr_client.create_ml_input_channel(
    name="ml_input_channel_name",
    membershipIdentifier='membership_id',

    configuredModelAlgorithmAssociations=[configured_model_algorithm_association_arn],
    retentionInDays=1,
    inputChannel={
        "dataSource": {
            "protectedQueryInputParameters": {
                "sqlParameters": {
                    "queryString": "select * from table",
                    "computeConfiguration": {
                        "worker": {
                            "type": "CR.1X",
                            "number": 16,
                            "properties": {
                                "spark": {
                                    "spark configuration key": "spark configuration
value",
                                }
                            }
                        }
                    }
                },
                "resultFormat": "PARQUET"
            }
        },
        "roleArn": "arn:aws:iam::111122223333:role/role_name"
    }
)
channel_arn = resp['ML Input Channel ARN']

```

在 AWS Clean Rooms ML 中建立訓練過的模型

事前準備：

- AWS 帳戶 可存取 的 AWS Clean Rooms
- 在 中設定的協同合作 AWS Clean Rooms

- 與協同合作相關聯的已設定模型演算法
- 至少一個已設定的 ML 輸入通道
- 在協同合作中建立和管理 ML 模型的適當許可

將設定的模型演算法與協同合作建立關聯後，然後建立並設定 ML 輸入通道，您就可以建立訓練過的模型。協作的成員會使用訓練過的模型來共同分析其資料。

您可以使用下列程序建立訓練過的模型。

或者，您可以使用增量訓練來改善具有新資料的現有模型，或使用分散式訓練來跨多個運算執行個體訓練模型。

主題

- [在 AWS Clean Rooms ML 中使用增量訓練](#)
- [在 AWS Clean Rooms ML 中使用分散式訓練](#)

Console

建立訓練過的模型（主控台）

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 在協同合作頁面上，選擇您要建立訓練模型的協同合作。
4. 協同合作開啟後，選擇 ML 模型索引標籤。
5. 在自訂 ML 模型下，在訓練模型區段中，選擇建立訓練模型。
6. 在建立訓練過的模型頁面上，針對關聯的模型演算法，指定演算法。
7. 如需訓練模型的詳細資訊，請輸入下列項目：
 - a. 在名稱中，輸入協同合作中模型的唯一名稱。
 - b. （選用）針對描述，輸入訓練模型的描述。
 - c. 針對訓練資料輸入模式，選擇下列其中一項：
 - 如果您有可以容納在 ML 儲存磁碟區的較小資料集，而且您偏好訓練指令碼的傳統檔案系統存取，請選取檔案。

- 選取管道讓大型資料集直接從 S3 串流資料，避免將所有資料下載到磁碟，從而提高訓練速度並減少儲存需求。
- 如果您想要將 S3 串流的優點與檔案系統存取結合，請選取 FastFile，特別是用於循序讀取資料或處理較少的檔案以縮短啟動時間時。

8. 對於 ML 輸入通道詳細資訊，請執行下列動作：

- a. 針對 ML 輸入通道，指定將資料提供給模型演算法的 ML 輸入通道。

若要新增另一個頻道，請選擇新增另一個 ML 輸入頻道。您最多可以新增 19 個額外的 ML 輸入通道。

- b. 針對頻道名稱，輸入 ML 輸入頻道的名稱。

- c. 針對 Amazon S3 資料分佈類型，選擇下列其中一項：

- 選取完整複寫，為每個訓練執行個體提供完整的資料集副本。當您的資料集夠小而足以容納在記憶體中，或每個執行個體都需要存取所有資料時，這最適合使用。
- 選取依 S3 金鑰分割，根據 S3 金鑰跨訓練執行個體分割資料集。每個執行個體會收到大約總 S3 物件的 $1/n$ ，其中 'n' 是執行個體的數量。這最適合您要平行處理的大型資料集。

 Note

選取分佈類型時，請考慮您的資料集大小和訓練需求。完全複寫可提供完整的資料存取，但需要更多儲存空間，而由 S3 金鑰碎片可分散式處理大型資料集。

9. 針對訓練持續時間上限，選擇您要訓練模型的時間上限。
10. 針對超參數，指定任何演算法特定的參數及其預期值。超參數專屬於要訓練的模型，並用於微調模型訓練。
11. 針對環境變數，指定任何演算法特定的變數及其預期值。環境變數是在 Docker 容器中設定。
12. 針對加密，若要使用自訂 AWS KMS key，請選取使用自訂 KMS 金鑰加密秘密核取方塊。
13. 針對 EC2 資源組態，指定用於模型訓練之運算資源的相關資訊。
- a. 針對執行個體類型，選擇您要執行的執行個體類型。
- b. 針對執行個體計數，輸入執行個體的數量。
- c. 對於以 GB 為單位的磁碟區大小，輸入 ML 儲存磁碟區大小。
14. 選擇建立訓練過的模型。

API

建立訓練模型 (API)

能夠訓練模型的成員會選取 ML 輸入通道和模型演算法來開始訓練。

使用特定參數執行下列程式碼：

```
import boto3
acr_ml_client= boto3.client('cleanroomsmml')

acr_ml_client.create_trained_model(
    membershipIdentifier= 'membership_id',
    configuredModelAlgorithmAssociationArn = 'arn:aws:cleanrooms-
ml:region:account:membership/membershipIdentifier/configured-model-algorithm-
association/identifier',
    name='trained_model_name',
    resourceConfig={
        'instanceType': 'ml.m5.xlarge',
        'volumeSizeInGB': 1
    },
    dataChannels=[
        {
            "mlInputChannelArn": channel_arn_1,
            "channelName": "channel_name"
        },
        {
            "mlInputChannelArn": channel_arn_2,
            "channelName": "channel_name"
        }
    ]
)
```

Note

建立訓練模型後，您無法編輯模型。若要進行變更，請刪除訓練過的模型並建立新的模型。

在 AWS Clean Rooms ML 中使用增量訓練

事前準備：

- AWS 帳戶 可存取 的 AWS Clean Rooms
- 協作中的現有訓練模型
- 用於增量訓練的新資料集或更新的資料集
- 在協同合作中建立和管理 ML 模型的適當許可
- 熟悉現有模型的超參數和組態

透過增量訓練，您可以使用現有模型的成品和更新的資料集來訓練新模型。增量訓練可節省時間和資源。

使用增量訓練時，可執行下列作業：

- 使用已擴展的資料集來訓練新模型，該資料集具有先前訓練中未說明的基礎模式。
- 使用不同的超參數或使用不同的資料集來訓練模型的數個變體。

Console

執行增量訓練任務（主控台）

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 在協同合作頁面上，選擇您要用於增量訓練的模型成品存在的協同合作。
4. 協同合作開啟後，選擇 ML 模型索引標籤。
5. 在自訂 ML 模型下，在訓練模型區段中，選擇您要遞增訓練之模型旁的選項按鈕。
6. 在概觀頁面的版本下，
 - a. 選擇您要增量訓練之訓練模型旁的選項按鈕。
 - b. 選擇從版本訓練。
7. 在從版本建立訓練模型頁面上，針對訓練模型版本，選擇版本。

會自動選取基礎模型版本。如果有其他版本存在，您可以變更此版本。
8. 如需訓練模型的詳細資訊，請輸入下列項目：
 - a. 在名稱中，輸入協同合作中模型的唯一名稱。
 - b. （選用）針對描述，輸入訓練模型的描述。

- c. 針對訓練資料輸入模式，選擇下列其中一項：
- 如果您有適合 ML 儲存磁碟區的較小資料集，而且您偏好訓練指令碼的傳統檔案系統存取，請選取檔案。
 - 選取管道讓大型資料集直接從 S3 串流資料，避免將所有資料下載到磁碟，從而提高訓練速度並減少儲存需求。
 - 如果您想要將 S3 串流的優點與檔案系統存取結合，請選取 FastFile，特別是循序讀取資料或處理較少的檔案以加快啟動時間時。
- d. 針對增量訓練頻道名稱，輸入增量訓練頻道的名稱

 Note

如果您指定沒有版本 ID 的增量訓練頻道名稱，系統會使用基本模型進行增量訓練。

9. 對於 ML 輸入通道詳細資訊，請執行下列動作：

- a. 針對 ML 輸入通道，指定將資料提供給模型演算法的 ML 輸入通道。

若要新增另一個頻道，請選擇新增另一個 ML 輸入頻道。您最多可以新增 19 個額外的 ML 輸入通道。

- b. 針對頻道名稱，輸入 ML 輸入頻道的名稱。
- c. 針對 Amazon S3 資料分佈類型，選擇下列其中一項：

- 選取完整複寫，為每個訓練執行個體提供完整的資料集副本。當您的資料集夠小而足以容納在記憶體中，或每個執行個體都需要存取所有資料時，這最適合使用。
- 選取依 S3 金鑰分割，根據 S3 金鑰跨訓練執行個體分割資料集。每個執行個體會收到大約總 S3 物件的 $1/n$ ，其中 'n' 是執行個體的數量。這最適合您要平行處理的大型資料集。

 Note

選取分佈類型時，請考慮您的資料集大小和訓練需求。完全複寫可提供完整的資料存取，但需要更多儲存空間，而由 S3 金鑰碎片可分散式處理大型資料集。

10. 針對訓練持續時間上限，選擇您想要訓練模型的時間上限。

11. 對於超參數，指定任何演算法特定的參數及其預期值。超參數專屬於要訓練的模型，並用於微調模型訓練。
12. 針對環境變數，指定任何演算法特定的變數及其預期值。環境變數是在 Docker 容器中設定。
13. 針對加密，若要使用自訂 AWS KMS key，請選取使用自訂 KMS 金鑰加密秘密核取方塊。
14. 針對 EC2 資源組態，指定用於模型訓練之運算資源的相關資訊。
 - a. 針對執行個體類型，選擇您要執行的執行個體類型。
 - b. 針對執行個體計數，輸入執行個體的數量。
 - c. 對於以 GB 為單位的磁碟區大小，輸入 ML 儲存磁碟區大小。
15. 選擇從 版本建立訓練過的模型。

API

執行增量訓練任務 (API)

使用特定參數執行下列程式碼：

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.create_trained_model(
    membershipIdentifier= 'membership_id',
    configuredModelAlgorithmAssociationArn = 'arn:aws:cleanrooms-
ml:region:account:membership/membershipIdentifier/configured-model-algorithm-
association/identifier',
    name='trained_model_name',
    resourceConfig={
        'instanceType': 'ml.m5.xlarge',
        'volumeSizeInGB': 1
    },
    incrementalTrainingDataChannels=[
        {
            'trainedModelArn': trained_model_arn,
            'channelName': 'channel_name'
        },
    ],
    dataChannels=[
        {
            'mlInputChannelArn': channel_arn_1,
            'channelName': 'channel_name'
        }
    ]
)
```

```
    },  
    {  
      'mlInputChannelArn': channel_arn_2,  
      'channelName': 'channel_name'  
    }  
  ]  
)
```

 Note

限制：總計最多 20 個頻道（包括 dataChannels 和 incrementalTrainingDataChannels）。

 Note

建立訓練模型後，您無法編輯模型。若要進行變更，請刪除訓練過的模型並建立新的模型。

在 AWS Clean Rooms ML 中使用分散式訓練

事前準備：

- AWS 帳戶 可存取 的 AWS Clean Rooms
- 在 中設定的協同合作 AWS Clean Rooms
- 支援分散式訓練的已設定模型演算法
- 適合分散式處理的大型資料集
- 在協同合作中建立和管理 ML 模型的適當許可
- 足以執行多個執行個體以進行分散式訓練的 Amazon EC2 配額

分散式訓練利用許多運算節點的強大功能並行運作，以有效率地處理大量資料和更新模型參數。

如需分散式訓練的詳細資訊，請參閱《Amazon SageMaker AI 開發人員指南》中的[分散式訓練概念](#)。

Console

執行分散式訓練任務（主控台）

1. 登入 AWS 管理主控台 並開啟位於 <https://console.aws.amazon.com/cleanrooms> 的 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 在協同合作頁面上，選擇您要建立訓練模型的協同合作。
4. 協同合作開啟後，選擇 ML 模型索引標籤。
5. 在自訂 ML 模型下，在訓練模型區段中，選擇建立訓練模型。
6. 在建立訓練過的模型頁面上，針對關聯的模型演算法，指定演算法
7. 如需訓練模型的詳細資訊，請輸入下列項目：
 - a. 在名稱中，輸入協同合作中模型的唯一名稱。
 - b. （選用）針對描述，輸入訓練模型的描述。
 - c. 針對訓練資料輸入模式，選擇下列其中一項：
 - 如果您的資料集較小，可以容納 ML 儲存磁碟區，而且您偏好訓練指令碼的傳統檔案系統存取，請選取檔案。
 - 選取管道讓大型資料集直接從 S3 串流資料，避免將所有資料下載到磁碟，從而提高訓練速度並減少儲存需求。
 - 如果您想要將 S3 串流的優點與檔案系統存取結合，請選取 FastFile，特別是循序讀取資料或處理較少的檔案以加快啟動時間時。
8. 對於 ML 輸入通道詳細資訊，請執行下列動作：
 - a. 對於 ML 輸入通道，指定將資料提供給模型演算法的 ML 輸入通道。

若要新增另一個頻道，請選擇新增另一個 ML 輸入頻道。您最多可以新增 19 個額外的 ML 輸入通道。
 - b. 針對頻道名稱，輸入 ML 輸入頻道的名稱。
 - c. 針對 Amazon S3 資料分佈類型，選擇下列其中一項：
 - 選取完整複寫，為每個訓練執行個體提供完整的資料集副本。當您的資料集夠小而足以容納在記憶體中，或每個執行個體都需要存取所有資料時，這最適合使用。

- 選取依 S3 金鑰分割，根據 S3 金鑰跨訓練執行個體分割資料集。每個執行個體會收到大約總 S3 物件的 $1/n$ ，其中 'n' 是執行個體的數量。這最適合您要平行處理的大型資料集。

 Note

選取分佈類型時，請考慮您的資料集大小和訓練需求。完全複製可提供完整的資料存取，但需要更多儲存空間，而由 S3 金鑰碎片可分散式處理大型資料集。

9. 針對訓練持續時間上限，選擇您要訓練模型的時間上限。
10. 對於超參數，指定任何演算法特定的參數及其預期值。超參數專屬於要訓練的模型，並用於微調模型訓練。
11. 針對環境變數，指定任何演算法特定的變數及其預期值。環境變數是在 Docker 容器中設定。
12. 針對加密，若要使用自訂 AWS KMS key，請選取使用自訂 KMS 金鑰加密秘密核取方塊。
13. 針對 EC2 資源組態，指定用於模型訓練之運算資源的相關資訊。

- a. 針對執行個體類型，選擇您要執行的執行個體類型。

分散式訓練支援的執行個體類型如下：

- ml.m5.4xlarge
- ml.m5.12xlarge
- ml.m5.2xlarge
- ml.g5.12xlarge
- ml.g5.24xlarge

- b. 針對執行個體計數，輸入執行個體的數量。
- c. 對於以 GB 為單位的磁碟區大小，輸入 ML 儲存磁碟區大小。

14. 選擇建立訓練過的模型。

API

執行分散式訓練任務 (API)

使用特定參數執行下列程式碼：

```
import boto3
```

```

acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.create_trained_model(
    membershipIdentifier= 'membership_id',
    configuredModelAlgorithmAssociationArn = 'arn:aws:cleanrooms-
ml:region:account:membership/membershipIdentifier/configured-model-algorithm-
association/identifier',
    name='trained_model_name',
    trainingInputMode: "File",
    resourceConfig={
        'instanceCount': "3"
        'instanceType': "ml.m5.xlarge",
        'volumeSizeInGB': 3
    },
    dataChannels=[
        {
            "mlInputChannelArn": channel_arn_1,
            "channelName": "channel_name",
            "S3DataDistributionType:" "FullyReplicated"
        }
    ]
)

```

Note

建立訓練模型後，您無法編輯模型。若要進行變更，請刪除訓練過的模型並建立新的模型。

從 AWS Clean Rooms ML 匯出模型成品

此任務是選用的，應在您將CAN_RECEIVE_MODEL_OUTPUT成員能力指派給協同合作的成員時完成。

模型訓練完成後，訓練模型的成員可以啟動模型成品的匯出。訓練模型的成員會選擇誰會收到模型成品，前提是該成員可以接收結果和有效的 ML 組態。

Console

設定自訂 ML 模型演算法 (主控台)

1. 登入 AWS 管理主控台，並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 在協同合作頁面上，選擇包含您要匯出之自訂模型的協同合作。
4. 協同合作開啟後，選擇 ML 模型索引標籤，然後從自訂訓練模型資料表中選擇您的模型。
5. 在自訂訓練模型詳細資訊頁面上，按一下匯出模型輸出。
6. 針對匯出模型輸出，針對匯出模型輸出詳細資訊，輸入名稱和選用的描述。

選擇哪些成員會在匯出至協同合作成員的模型輸出中接收模型成品下拉式清單。

7. 選擇 Export (匯出)。

結果會匯出至 ML 組態中指定的 Amazon S3 位置中的下列路徑：`yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName`。只會匯出您在關聯設定的模型演算法時所選取的要匯出的檔案，最高可達指定的檔案大小上限。

API

設定自訂 ML 模型演算法 (API)

執行下列程式碼來啟動模型匯出：

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.start_trained_model_export_job(
    membershipIdentifier='membership_id',
    trainedModelArn='arn:aws:cleanrooms-ml:region:account:membership/
membershipIdentifier/trained-model/identifier',
    outputConfiguration={
        'member': {
            'accountId': 'model_output_receiver_account'
        }
    },
    name='export_job_name'
)
```

結果會匯出至 ML 組態中指定的 Amazon S3 位置中的下列路徑：`yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName`。只會匯出您在關聯設定的模型演算法時選取的 `filesToExport`，直到 `maxSize` 指定的為止。

在 AWS Clean Rooms ML 中的訓練模型上執行推論

能夠執行查詢的成員也可以在訓練任務完成後啟動推論任務。他們會挑選要執行推論的推論資料集，並參考他們想要執行推論容器的訓練模型輸出。

將接收推論輸出的成員必須獲授予成員能力 `CAN_RECEIVE_INFERENCE_OUTPUT`。

Console

建立模型推論任務（主控台）

1. 登入 AWS 管理主控台 並在 <https://console.aws.amazon.com/cleanrooms> 開啟 AWS Clean Rooms 主控台。
2. 在左側導覽窗格中，選擇協同合作。
3. 在協同合作頁面上，選擇包含您要建立推論任務之自訂模型的協同合作。
4. 協同合作開啟後，選擇 ML 模型索引標籤，然後從自訂訓練模型資料表中選擇您的模型。
5. 在自訂訓練模型詳細資訊頁面上，按一下開始推論任務。
6. 針對開始推論任務，針對推論任務詳細資訊，輸入名稱和選用的描述。

輸入下列資訊：

- 關聯的模型演算法 – 在推論任務期間使用的關聯模型演算法。
 - ML 輸入通道詳細資訊 – 將為此推論任務提供資料的 ML 輸入通道。
 - 轉換資源 – 用來執行推論任務轉換功能的運算執行個體。
 - 輸出組態 – 誰會收到推論任務輸出和輸出的 MIME 類型。
 - 加密 – 選擇自訂加密設定，以指定您自己的 KMS 金鑰和相關資訊。否則，Clean Rooms ML 會管理加密。
 - 轉換任務詳細資訊 – 推論任務的最大承載，以 MB 為單位。
 - 環境變數 – 存取推論任務容器映像所需的任何環境變數。
7. 選擇開始推論任務。

結果會匯出至 ML 組態中指定的 Amazon S3 位置中的下列路徑：`yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName`。

API

建立模型推論任務 (API)

執行下列程式碼來啟動推論任務：

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.start_trained_model_inference_job(
    name="inference_job",
    membershipIdentifier='membership_id',
    trainedModelArn='arn:aws:cleanrooms-m1:region:account:membership/
membershipIdentifier/trained-model/identifier',

    dataSource={
        "mlInputChannelArn": 'channel_arn_3'
    },
    resourceConfig={'instanceType': 'ml.m5.xlarge'},
    outputConfiguration={
        'accept': 'text/csv',
        'members': [
            {
                "accountId": 'member_account_id'
            }
        ]
    }
)
```

結果會匯出至 ML 組態中指定的 Amazon S3 位置中的下列路徑：`yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName`。

建立 AWS Clean Rooms ML 模型做為訓練資料提供者

類似模型是訓練資料提供者資料的模型，可讓種子資料提供者建立訓練資料提供者資料的類似區段，最接近其種子資料。若要建立可在協同合作中使用的類似模型，您必須匯入訓練資料、建立類似模型、設定該類似模型，然後將其與協同合作建立關聯。

使用類似模型需要兩個方，訓練資料提供者和種子資料提供者，依序在 [中工作 AWS Clean Rooms](#)，將他們的資料帶入協同合作。這是訓練資料提供者必須先完成的工作流程：

1. 訓練資料提供者的資料必須存放在使用者項目 [互動 AWS Glue](#) 的資料目錄資料表中。訓練資料至少必須包含使用者 ID 欄、互動 ID 欄和時間戳記欄。
2. 訓練資料提供者向 [註冊訓練資料 AWS Clean Rooms](#)。
3. 訓練資料提供者會建立外觀相似的模型，可與多個種子資料提供者共用。看起來像模型是一種深度神經網路，最多可能需要 24 小時才能訓練。它不會自動重新訓練，我們建議您每週重新訓練模型。
4. 訓練資料提供者會設定類似模型，包括是否共用相關性指標和輸出區段的 Amazon S3 位置。訓練資料提供者可以從單一外觀模型建立多個已設定的外觀模型。
5. 訓練資料提供者會將設定的對象模型與與種子資料提供者共用的協同合作建立關聯。

訓練資料提供者完成建立 ML 模型後，[種子資料提供者可以建立和匯出外觀客群](#)。

主題

- [匯入訓練資料](#)
- [建立外觀模型](#)
- [設定外觀模型](#)
- [關聯已設定的外觀模型](#)
- [更新已設定的類似模型](#)

匯入訓練資料

Note

您只能提供訓練資料集，以便在具有存放在 Amazon S3 中資料的 Clean Rooms ML 外觀模型中使用。不過，您可以使用 SQL 來提供類似模型的種子資料，該 SQL 會跨存放在任何支援資料來源中的資料執行。

建立類似模型之前，您必須指定包含訓練資料的 AWS Glue 資料表。Clean Rooms ML 不會儲存此資料的複本，只是中繼資料，允許它存取資料。

在中匯入訓練資料 AWS Clean Rooms

1. 使用 登入 AWS 管理主控台 並開啟[AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇 AWS ML 模型。
3. 在訓練資料集索引標籤上，選擇建立訓練資料集。
4. 在建立訓練資料集頁面上，針對訓練資料集詳細資訊，輸入名稱和選用的描述。
5. 從下拉式清單中選取您要設定的資料庫和資料表，以選擇訓練資料來源。

Note

若要驗證這是正確的資料表，請執行下列其中一項操作：

- 選擇檢視 AWS Glue。
- 開啟檢視結構描述以檢視結構描述。

6. 如需訓練詳細資訊，請從下拉式清單中選擇使用者識別碼欄、項目識別碼欄和時間戳記欄。訓練資料必須包含這三個資料欄。您也可以選取任何其他要包含在訓練資料中的資料欄。

時間戳記欄中的資料必須採用秒格式的 Unix epoch 時間。

7. （選用）如果您有任何要訓練的其他資料欄，請從下拉式清單中選擇資料欄名稱和類型。
8. 在服務存取中，您必須指定可存取資料的服務角色，並在資料加密時提供 KMS 金鑰。選擇建立並使用新的服務角色，Clean Rooms ML 會自動建立服務角色並新增必要的許可政策。如果您有要使用的特定服務角色，請選擇使用現有的服務角色，然後在服務角色名稱欄位中輸入該角色。

如果您的資料已加密，請在 AWS KMS key 欄位中輸入 KMS 金鑰，或按一下建立 AWS KMS key 以產生新的 KMS 金鑰。

9. 如果您想要為訓練資料集啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
10. 選擇建立訓練資料集。

如需對應的 API 動作，請參閱 [CreateTrainingDataset](#)。

建立外觀模型

建立訓練資料集之後，您就可以建立類似模型。您可以從單一訓練資料集建立許多外觀模型。

您必須在 中建立預設資料庫，AWS Glue Data Catalog 或在提供的角色中包含 `glue:createDatabase` 許可。

在 中建立類似模型 AWS Clean Rooms

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇 AWS ML 模型。
3. 在 Lookalike 模型索引標籤上，選擇建立外觀模型。
4. 在建立類似模型頁面上，針對 Lookalike 模型詳細資訊，輸入名稱和選用的描述。
 - a. 從下拉式清單中選擇您要建模的訓練資料集。

Note

若要驗證這是正確的訓練資料集，請開啟顯示訓練資料集詳細資訊以檢視詳細資訊。
若要建立新的訓練資料集，請選擇建立訓練資料集。

- b. （選用）輸入訓練時段。
5. 如果您想要為類似模型啟用自訂加密設定，請選擇自訂加密設定，然後輸入 KMS 金鑰。
 6. 如果您想要為類似模型啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
 7. 選擇建立外觀模型。

Note

模型訓練可能需要數小時到 2 天的時間。

如需對應的 API 動作，請參閱 [CreateAudienceModel](#)。

設定外觀模型

建立類似模型之後，您就可以將其設定為在協同合作中使用。您可以從單一外觀模型建立多個已設定的外觀模型。

在 中設定類似模型 AWS Clean Rooms

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇 AWS ML 模型。
3. 在已設定的外觀模型索引標籤上，選擇設定的外觀模型。
4. 在設定外觀模型頁面上，針對已設定的外觀模型詳細資訊，輸入名稱和選用的描述。
 - a. 從下拉式清單中選擇您要設定的 Lookalike 模型。

Note

若要驗證這是正確的外觀模型，請開啟顯示外觀模型詳細資訊以檢視詳細資訊。
若要建立新的外觀模型，請選擇建立外觀模型。

- b. 選擇您想要的最小相符種子大小。這是種子資料提供者資料中與訓練資料中的使用者重疊的最小使用者數量。此值必須大於 0。
5. 針對要與其他成員共用的指標，選擇您是否希望協同合作中的種子資料提供者接收模型指標，包括相關性分數。
 6. 對於 Lookalike 區段目的地位置，輸入匯出 lookalike 區段的 Amazon S3 儲存貯體。此儲存貯體必須位於與其他資源相同的區域。
 7. 針對服務存取，選擇將用於存取此資料表的現有服務角色名稱。
 8. 對於進階儲存貯體大小組態，請將對象大小類型指定為絕對數量或百分比。
 9. 如果您想要為設定的資料表資源啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
 10. 選擇設定外觀模型。

如需對應的 API 動作，請參閱 [CreateConfiguredAudienceModel](#)。

關聯已設定的外觀模型

設定好類似模型之後，您可以將其與協同合作建立關聯。

在 中關聯已設定的外觀模型 AWS Clean Rooms

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。

2. 在左側導覽窗格中，選擇協同合作。
3. 在具有作用中成員資格索引標籤上，選擇協作。
4. 在 ML 模型索引標籤的Ready-to-use型模型下，選擇關聯型樣模型。
5. 在關聯設定的類似模型頁面上，針對已設定的類似模型關聯詳細資訊：
 - a. 輸入相關已設定對象模型的名稱。
 - b. 輸入資料表的描述。

此描述有助於區分具有類似名稱的其他相關已設定對象模型。

6. 對於已設定的外觀模型，請從下拉式清單中選擇已設定的外觀模型。
7. 選擇關聯。

如需對應的 API 動作，請參閱 [CreateConfiguredAudienceModelAssociation](#)。

更新已設定的類似模型

將已設定的類似模型建立關聯後，您可以更新該模型以變更資訊，例如名稱、要共用的指標或輸出 Amazon S3 位置。

在 中更新相關聯的已設定外觀模型 AWS Clean Rooms

1. 登入 AWS 管理主控台 並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇 AWS ML 模型。
3. 在已設定的外觀模型索引標籤上，Ready-to-use的外觀模型下，選擇已設定的的外觀模型，然後選取編輯。
4. 在編輯頁面上，針對已設定的外觀模型關聯詳細資訊：
 - a. 更新名稱和選用的描述。
 - b. 從下拉式清單中選擇您要設定的 Lookalike 模型。
 - c. 選擇您想要的最小相符種子大小。這是種子資料提供者資料中與訓練資料中的使用者重疊的最小使用者數量。此值必須大於 0。
5. 針對要與其他成員共用的指標，選擇您是否希望協同合作中的種子資料提供者接收模型指標，包括相關性分數。

6. 對於 Lookalike 區段目的地位置，輸入匯出類似區段的 Amazon S3 儲存貯體。此儲存貯體必須位於與其他 資源相同的區域。
7. 針對服務存取，選擇將用於存取此資料表的現有服務角色名稱。
8. 針對進階儲存貯體大小組態，選擇您要如何設定對象儲存貯體大小。
9. 選擇儲存變更。

如需對應的 API 動作，請參閱 [UpdateConfiguredAudienceModel](#)。

建立 AWS Clean Rooms ML 模型做為種子資料提供者

訓練資料提供者完成建立 ML 模型後，種子資料提供者可以建立和匯出類似樣的區段。看起來像區段是訓練資料的子集，最接近種子資料。

這是種子資料提供者必須完成的工作流程：

1. 種子資料提供者的資料可以存放在 Amazon S3 儲存貯體中，也可以來自查詢結果。
2. 種子資料提供者會開啟他們與訓練資料提供者共用的協同合作。
3. 種子資料提供者會從協作頁面的 Clean Rooms ML 索引標籤建立外觀相似的區段。
4. 如果已共用關聯性指標，則種子資料提供者可以評估它們，並匯出外觀區段以供外部使用 AWS Clean Rooms。

主題

- [建立類似樣貌的客群](#)
- [匯出類似樣貌的客群](#)

建立類似樣貌的客群

Note

您只能提供訓練資料集，以便在具有存放在 Amazon S3 中資料的 Clean Rooms ML 外觀模型中使用。不過，您可以使用 SQL 來提供類似模型的種子資料，該 SQL 會跨存放在任何支援資料來源中的資料執行。

看起來像區段是訓練資料的子集，最接近種子資料。

在 中建立外觀相似的客群 AWS Clean Rooms

1. 使用 登入 AWS 管理主控台 並開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 在具有作用中成員資格索引標籤上，選擇協作。
4. 在 ML 模型索引標籤上，選擇建立外觀客群。

5. 在建立外觀客群頁面上，針對關聯的已設定外觀模型，選擇要用於此外觀客群的已設定外觀模型。
6. 對於 Lookalike 區段詳細資訊，輸入名稱和選用的描述。
7. 對於種子設定檔，選擇 選項，然後採取建議的動作，以選擇您的種子方法。

選項	建議的動作
Amazon S3 路徑	1. 選取 Amazon S3 位置。 2. (選用) 選擇在輸出中包含種子設定檔。
SQL 查詢	撰寫 SQL 查詢並使用其結果做為種子資料。
分析範本	從下拉式清單中選擇分析範本，並使用分析範本建立的結果。

8. 選擇建立此資料來源時要使用的工作者類型。預設工作者類型為 CR.1X。指定要使用的工作者數量。預設為工作者編號 16。若要指定 Spark 屬性：
 - a. 展開 Spark 屬性。
 - b. 選擇新增 Spark 屬性。
 - c. 在 Spark 屬性對話方塊中，從下拉式清單中選擇屬性名稱，然後輸入值。

下表提供每個屬性的定義。

如需 Spark 屬性的詳細資訊，請參閱 Apache Spark 文件中的 Spark [屬性](#)。

屬性名稱	Description	預設值
spark.task.maxFailures	控制任務在任務失敗之前可以失敗的連續次數。需要大於或等於 1 的值。允許重試的次數等於此值減去 1。如果任何嘗試成功，就會重設失敗計數。不同任務的故障不會累積到此限制。	4
spark.sql.files.maxPartitionBytes	設定從 Parquet、JSON 和 ORC 等檔案型來源讀取時封裝到單一分割區的最大位元組數。	128MB

屬性名稱	Description	預設值
spark.hadoop.fs.s3.maxRetries	設定 Amazon S3 檔案操作的重試嘗試次數上限。	
spark.network.timeout	設定所有網路互動的預設逾時。如果未設定，則覆寫下列逾時設定： - spark.storage.blockManagerHeartbeatTimeoutMs - spark.shuffle.io.connectionTimeout - spark.rpc.askTimeout - spark.rpc.lookupTimeout	120 秒
spark.rdd.compress	指定是否使用 spark.io.compression.codec 壓縮序列化 RDD 分割區。適用於 Java 和 Scala 中的 StorageLevel.MEMORY_ONLY_SER，或 Python 中的 StorageLevel.MEMORY_ONLY。減少儲存空間，但需要額外的 CPU 處理時間。	FALSE
spark.shuffle.spill.compress	指定是否使用 spark.io.compression.codec 壓縮隨機溢出資料。	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	在 spark.sql.adaptive.enabled 為 true 的適應性最佳化期間，設定隨機播放分割區的目標大小，以位元組為單位。合併小型分割區或分割偏斜分割區時，控制分割區大小。	(spark.sql.adaptive.shuffle.targetPostShuffleInputSize 的值)
spark.sql.adaptive.autoBroadcastJoinThreshold	設定在聯結期間廣播至工作者節點的資料表大小上限，以位元組為單位。僅適用於自適應架構。使用與 spark.sql.autoBroadcastJoinThreshold 相同的預設值。設定為 -1 以停用廣播。	(無)

屬性名稱	Description	預設值
<code>spark.sql.adaptive.coalescePartitions.enabled</code>	指定是否要根據 <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> 來合併連續隨機播放分割區，以最佳化任務大小。需要 <code>spark.sql.adaptive.enabled</code> 才能成立。	TRUE
<code>spark.sql.adaptive.coalescePartitions.initialPartitionNum</code>	在合併之前定義隨機播放分割區的初始數量。同時需要 <code>spark.sql.adaptive.enabled</code> 和 <code>spark.sql.adaptive.coalescePartitions.enabled</code> 為 true。預設為 <code>spark.sql.shuffle.partitions</code> 的值。	(無)
<code>spark.sql.adaptive.coalescePartitions.minPartitionSize</code>	設定合併隨機播放分割區的大小下限，以防止分割區在適應性最佳化期間變得太小。	1 MB
<code>spark.sql.adaptive.coalescePartitions.parallelismFirst</code>	指定是否要在分割區合併期間根據叢集平行處理而非 <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> 計算分割區大小。產生的分割區大小小於設定的目標大小，以最大化平行處理。我們建議在忙碌叢集上將此設定為 false，以透過防止過多的小型任務來改善資源使用率。	TRUE
<code>spark.sql.adaptive.enabled</code>	指定是否啟用自適應查詢執行，以在查詢執行期間根據準確的執行時間統計資料重新最佳化查詢計劃。	TRUE
<code>spark.sql.adaptive.forceOptimizeSkewedJoin</code>	指定是否強制啟用 <code>OptimizeSkewedJoin</code> ，即使它引入額外的隨機播放。	FALSE
<code>spark.sql.adaptive.localShuffleReader.enabled</code>	指定是否在不需隨機分割時使用本機隨機讀取器，例如從排序合併聯結轉換為廣播雜湊聯結之後。需要 <code>spark.sql.adaptive.enabled</code> 才能成立。	TRUE

屬性名稱	Description	預設值
<code>spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold</code>	設定用於建置本機雜湊映射的分割區大小上限，以位元組為單位。在下列情況下，優先考慮排序合併聯結的隨機雜湊聯結： - 此值等於或超過 <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> - 所有分割區大小都在此限制內 覆寫 <code>spark.sql.join.preferSortMergeJoin</code> 設定。	0 個位元組
<code>spark.sql.adaptive.optimizeSkewsInRebalancePartitions.enabled</code>	指定是否根據 <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> 將扭曲隨機播放分割區分割為較小的分割區，以最佳化扭曲隨機播放分割區。需要 <code>spark.sql.adaptive.enabled</code> 才能成立。	TRUE
<code>spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor</code>	定義分割期間合併分割區的大小閾值因素。小於此因素的分割區乘以 <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> 會合併。	0.2
<code>spark.sql.adaptive.skewJoin.enabled</code>	指定是否透過分割和選擇性複寫扭曲的分割區來處理隨機聯結中的資料扭曲。適用於排序合併和隨機雜湊聯結。需要 <code>spark.sql.adaptive.enabled</code> 才能成立。	TRUE
<code>spark.sql.adaptive.skewJoin.skewedPartitionFactor</code>	決定決定分割區扭曲的大小因素。當分割區的大小超過以下兩者時，就會扭曲： - 此係數乘以中位數分割區大小 <code>spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes</code> 的值	5

屬性名稱	Description	預設值
spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes	設定識別扭曲分割區的大小閾值，以位元組為單位。當分割區的大小超過以下兩者時，就會扭曲： - 此閾值 - 中位數分割區大小乘以 spark.sql.adaptive.skewJoin.skewedPartitionFactor 我們建議您設定大於 spark.sql.adaptive.advisoryPartitionSizeInBytes 的值。	256MB
spark.sql.autoBroadcastJoinThreshold	設定在聯結期間廣播至工作者節點的資料表大小上限，以位元組為單位。設定為 -1 以停用廣播。	10MB
spark.sql.broadcastTimeout	控制廣播聯結期間廣播操作的逾時期間，以秒為單位。	300 秒
spark.sql.cbo.enabled	指定是否要為計劃統計資料估算啟用成本型最佳化 (CBO)。	FALSE
spark.sql.cbo.joinReorder.dp.star.filter	指定是否要在成本型聯結列舉期間套用星狀聯結篩選條件啟發式。	FALSE
spark.sql.cbo.joinReorder.dp.threshold	設定動態程式設計演算法中允許的聯結節點數目上限。	12
spark.sql.cbo.joinReorder.enabled	指定是否要在成本型最佳化 (CBO) 中啟用聯結重新排序。	FALSE
spark.sql.cbo.planStats.enabled	指定是否要在邏輯計畫產生期間從目錄擷取資料列計數和資料欄統計資料。	FALSE
spark.sql.cbo.starSchemaDetection	指定是否根據星狀結構描述偵測啟用聯結重新排序。	FALSE

屬性名稱	Description	預設值
spark.sql.files.maxPartitionNum	設定檔案型來源 (Parquet、JSON 和 ORC) 的目標分割檔案分割區數量上限。當初始計數超過此值時，重新擴展分割區。這是建議的目標，而不是保證的限制。	(無)
spark.sql.files.maxRecordsPerFile	設定寫入單一檔案的記錄數目上限。設定為零或負值時不會套用任何限制。	0
spark.sql.files.minPartitionNum	設定檔案型來源 (Parquet、JSON 和 ORC) 的目標分割檔案分割區數量下限。預設為 spark.sql.leafNodeDefaultParallelism。這是建議的目標，而不是保證的限制。	(無)
spark.sql.inMemoryColumnarStorage.batchSize	控制單欄式快取的批次大小。增加大小可改善記憶體使用率和壓縮，但會增加out-of-memory錯誤的風險。	10000
spark.sql.inMemoryColumnarStorage.compressed	指定是否根據資料統計資料自動選取資料欄的壓縮轉碼器。	TRUE
spark.sql.inMemoryColumnarStorage.enableVectorizedReader	指定是否要為單欄式快取啟用向量化讀取。	TRUE
spark.sql.legacy.allowHashOnMapType	指定是否允許對映射類型資料結構執行雜湊操作。此舊版設定可維持與舊版 Spark 版本的映射類型處理的相容性。	
spark.sql.legacy.allowNegativeScaleOfDecimal	指定是否在十進位類型定義中允許負縮放值。此舊版設定可維持與支援負小數位數擴展之較舊 Spark 版本的相容性。	

屬性名稱	Description	預設值
spark.sql.legacy.castComplexTypesToString.enabled	指定是否啟用舊版行為，將複雜類型轉換為字串。維持與較舊 Spark 版本的類型轉換規則的相容性。	
spark.sql.legacy.charVarcharAsString	指定是否將 CHAR 和 VARCHAR 類型視為 STRING 類型。此舊版設定提供與舊版 Spark 版本的字串類型處理相容性。	
spark.sql.legacy.createEmptyCollectionUsingStringType	指定是否使用字串類型元素建立空集合。此舊版設定可維持與較舊 Spark 版本的集合初始化行為的相容性。	
spark.sql.legacy.exponentLiteralAsDecimal.enabled	指定是否要將指數常值解譯為十進位類型。此舊版設定可維持與舊版 Spark 版本數值常值處理的相容性。	
spark.sql.legacy.json.allowEmptyString.enabled	指定是否要在 JSON 處理中允許空字串。此舊版設定可維持與舊版 Spark 版本的 JSON 剖析行為的相容性。	
spark.sql.legacy.parquet.int96RebaseModelRead	指定是否在讀取 Parquet 檔案時使用舊版 INT96 時間戳記重新基礎模式。此舊版設定可維持與較舊 Spark 版本時間戳記處理的相容性。	
spark.sql.legacy.timeParserPolicy	控制時間剖析行為以實現回溯相容性。此舊版設定會決定如何從字串剖析時間戳記和日期。	
spark.sql.legacy.typeCoercion.datetimeToString.enabled	指定是否在將日期時間值轉換為字串時啟用舊版類型強制行為。維持與較舊 Spark 版本日期時間轉換規則的相容性。	
spark.sql.maxSinglePartitionBytes	設定以位元組為單位的分割區大小上限。規劃器為較大的分割區引進隨機播放操作，以改善平行處理。	128 公尺

屬性名稱	Description	預設值
spark.sql.metadataCacheTTLSeconds	控制中繼資料快取的time-to-live (TTL)。適用於分割區檔案中繼資料和工作階段目錄快取。需要： - 大於零的正值 - spark.sql.catalogImplementation 設定為 hive - spark.sql.hive.filesourcePartitionFileCacheSize 大於零 - spark.sql.hive.manageFilesourcePartitions 設定為 true	-1000 毫秒
spark.sql.optimizer.collapseProjectAlwaysInline	指定是否要摺疊相鄰的投影和內嵌表達式，即使它會導致重複。	FALSE
spark.sql.optimizer.dynamicPartitionPruning.enabled	指定是否要為用作聯結索引鍵的分割區資料欄產生述詞。	TRUE
spark.sql.optimizer.enableCsvExpressionOptimization	指定是否要透過從_csv 操作剪除不必要的資料欄來最佳化 SQL 最佳化工具中的 CSV 表達式。	TRUE
spark.sql.optimizer.enableJsonExpressionOptimization	指定是否要透過下列方式最佳化 SQL 最佳化工具中的 JSON 表達式： - 從_json 操作刪除不必要的資料欄 - 簡化從_json 和 到_json 的組合 - 最佳化 named_struct 操作	TRUE
spark.sql.optimizer.excludedRules	定義要停用的最佳化工具規則，以逗號分隔的規則名稱識別。某些規則無法停用，因為其為正確性的必要條件。最佳化工具會記錄哪些規則已成功停用。	(無)

屬性名稱	Description	預設值
spark.sql.optimize r.runtime.bloomFilter.applicationSideScanSizeThreshold	設定在應用程式端插入 Bloom 篩選條件所需的最小彙總掃描大小，以位元組為單位。	10GB
spark.sql.optimize r.runtime.bloomFilter.creationSideThreshold	定義在建立端注入 Bloom 篩選條件的大小閾值上限。	10MB
spark.sql.optimize r.runtime.bloomFilter.enabled	指定是否要插入 Bloom 篩選條件，以便在隨機聯結的一側具有選擇性述詞時減少隨機播放資料。	TRUE
spark.sql.optimize r.runtime.bloomFilter.expectedNumItems	定義執行時間 Bloom 篩選條件中預期項目的預設數量。	1000000
spark.sql.optimize r.runtime.bloomFilter.maxNumBits	設定執行時間 Bloom 篩選條件中允許的位元數上限。	67108864
spark.sql.optimize r.runtime.bloomFilter.maxNumItems	設定執行時間 Bloom 篩選條件中允許的預期項目數量上限。	4000000
spark.sql.optimize r.runtime.bloomFilter.number.threshold	限制每個查詢允許的非 DPP 執行時間篩選條件數量上限，以防止out-of-memory錯誤。	10
spark.sql.optimize r.runtime.bloomFilter.numBits	定義執行時間 Bloom 篩選條件中使用的預設位元數。	8388608

屬性名稱	Description	預設值
spark.sql.optimize r.runtime.rowlevel OperationGroupFilter.enabled	指定是否要為資料列層級操作啟用執行時間群組篩選。允許資料來源： - 使用資料來源篩選條件刪除整個資料群組（例如檔案或分割區） - 執行執行時間查詢以識別相符的記錄 - 捨棄不必要的群組，以避免昂貴的重寫 限制： - 並非所有表達式都可以轉換為資料來源篩選條件 - 有些表達式需要 Spark 評估（例如子查詢）	TRUE
spark.sql.optimize r.runtimeFilter.se miJoinReduction.enabled	指定是否要插入半聯結，以在隨機聯結的一側具有選擇性述詞時減少隨機資料。	FALSE
spark.sql.parquet. aggregatePushdown	指定是否將彙總下推至 Parquet 以進行最佳化。支援： - 布林值、整數、浮點數和日期類型的 MIN 和 MAX - 所有資料類型的 COUNT 如果任何 Parquet 檔案頁尾缺少統計資料，則擲回例外狀況。	FALSE
spark.sql.parquet. columnarReaderBatchSize	控制每個 Parquet 向量化讀取器批次中的資料列數。選擇平衡效能額外負荷和記憶體用量的值，以防止out-of-memory錯誤。	4096

屬性名稱	Description	預設值
spark.sql.session.timeZone	定義工作階段時區，以處理字串常值和 Java 物件轉換中的時間戳記。接受： - 區域/城市格式的區域型 IDs (例如 America/Los_Angeles) - 區域位移，以 (+/-)HH、(+/-)HH : mm 或 (+/-)HH : mm : ss 格式顯示 (例如 -08 或 +01 : 00) +00 : 00 的 UTC 或 Z 別名	(本機時區的值)
spark.sql.shuffle.partitions	設定聯結或彙總期間資料隨機播放的預設分割區數量。結構式串流查詢從相同檢查點位置重新啟動之間無法修改。	200
spark.sql.shuffledHashJoinFactor	定義用於判斷隨機雜湊聯結資格的乘數係數。當小端資料大小乘以此係數時，會選取隨機雜湊聯結，小於大端資料大小。	3
spark.sql.sources.parallelPartitionDiscovery.threshold	使用檔案型來源 (Parquet、JSON 和 ORC) 設定驅動程式端檔案清單的路徑數目上限。在分割區探索期間超過 時，會使用個別的 Spark 分散式任務列出檔案。	32
spark.sql.statistics.histogram.enabled	指定是否要在資料欄統計資料計算期間產生等高長條圖，以改善估算準確度。需要額外的資料表掃描，超過基本資料欄統計資料所需的掃描。	FALSE

- 針對服務存取，選擇將用於存取此資料表的現有服務角色名稱。
- 如果您想要為訓練資料集啟用標籤，請選擇新增標籤，然後輸入金鑰和值對。
- 選擇建立外觀客群。

如需對應的 API 動作，請參閱 [StartAudienceGenerationJob](#)。

匯出類似樣貌的客群

建立類似區段之後，您可以將該資料匯出至 Amazon S3 儲存貯體。

在 中匯出外觀相似的客群 AWS Clean Rooms

1. 登入 AWS 管理主控台 並使用 開啟 [AWS Clean Rooms 主控台](#) AWS 帳戶（如果您尚未這麼做）。
2. 在左側導覽窗格中，選擇協同合作。
3. 在具有作用中成員資格索引標籤上，選擇協作。
4. 在 ML 模型索引標籤上，選取外觀的區段，然後選擇匯出。
5. 對於匯出外觀模型，對於匯出外觀模型詳細資訊，輸入名稱和選用的描述。
6. 針對區段大小，選擇您要匯出區段的大小。
7. 選擇 Export (匯出)。

如需對應的 API 動作，請參閱 [StartAudienceExportJob](#)。

故障診斷 AWS Clean Rooms

本節說明使用時可能發生的一些常見問題，AWS Clean Rooms 以及如何修正這些問題。

問題

- [查詢參考的一或多個資料表無法由其相關聯的服務角色存取。資料表/角色擁有者必須授予 資料表的服務角色存取權。](#)
- [其中一個基礎資料集具有不支援的檔案格式。](#)
- [針對 使用密碼編譯運算時，查詢結果不如預期Clean Rooms。](#)
- [AWS Clean Rooms Spark SQL：缺少分割區資料](#)

查詢參考的一或多個資料表無法由其相關聯的服務角色存取。資料表/角色擁有者必須授予 資料表的服務角色存取權。

- 確認已視需要設定服務角色的許可。如需詳細資訊，請參閱 [設定 AWS Clean Rooms](#)。

其中一個基礎資料集具有不支援的檔案格式。

- 確保您的資料集採用其中一個支援的檔案格式：
 - Parquet
 - RCFile
 - TextFile
 - SequenceFile
 - RegexSerde
 - OpenCSV
 - AVRO
 - JSON

如需詳細資訊，請參閱[的資料格式 AWS Clean Rooms](#)。

針對 使用密碼編譯運算時，查詢結果不如預期Clean Rooms。

如果您使用加密運算 for Clean Rooms(C3R)，請確認您的查詢正確使用加密的資料欄：

- 這些sealed資料欄僅用於 SELECT子句。
- 這些fingerprint資料欄僅用於 JOIN子句（以及特定條件下的 GROUP BY子句）。
- 只有在協同合作設定需要時，您才是具有相同名稱的資料JOINingfingerprint欄。

如需詳細資訊，請參閱[the section called “密碼編譯運算”](#)及[the section called “資料欄類型”](#)。

AWS Clean Rooms Spark SQL：缺少分割區資料

中的所有分割區 AWS Glue Data Catalog 也必須有 S3 中的資料。引擎使用 Spark 設定 `spark.sql.files.ignoreMissingFiles=False`

如需詳細資訊，請參閱 <https://spark.apache.org/docs/latest/sql-data-sources-generic-options.html#ignore-missing-files>。

如果您遇到此錯誤，您會收到下列錯誤訊息："Missing partition data: One of the configured tables is partitioned and one or more of the partitions does not have data".

將 Amazon S3 中存在的資料與 AWS Glue Data Catalog 資料表的 中列出的分割區進行比較。刪除 S3 中沒有對應資料的分割區。

中的安全性 AWS Clean Rooms

的雲端安全性 AWS 是最高優先順序。身為 AWS 客戶，您可以受益於資料中心和網路架構，這些架構專為滿足最安全敏感組織的需求而建置。

安全性是 AWS 與您之間共同責任。[共同責任模型](#)將其描述為雲端的安全性和雲端中的安全性：

- 雲端的安全性 – AWS 負責保護在 AWS Cloud 中執行 AWS 服務的基礎設施。AWS 也為您提供可安全使用的服務。在[AWS 合規計劃](#)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 Clean Rooms，請參閱[合規計劃的 AWS 服務範圍](#)。
- 雲端的安全性 – 您的責任取決於您使用 AWS 的服務。您也必須對其他因素負責，包括資料的機密性、您的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 Clean Rooms。它說明如何設定 AWS Clean Rooms 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 Clean Rooms 資源。

目錄

- [中的資料保護 AWS Clean Rooms](#)
- [使用的服務連結角色 AWS Clean Rooms](#)
- [中的資料保留 AWS Clean Rooms](#)
- [中的資料協同合作最佳實務 AWS Clean Rooms](#)
- [的 Identity and Access Management AWS Clean Rooms](#)
- [的合規驗證 AWS Clean Rooms](#)
- [中的彈性 AWS Clean Rooms](#)
- [中的基礎設施安全 AWS Clean Rooms](#)
- [使用界面端點 \(AWS PrivateLink\) 存取 AWS Clean Rooms 或 AWS Clean Rooms ML](#)

中的資料保護 AWS Clean Rooms

AWS [共同責任模型](#)適用於中的資料保護 AWS Clean Rooms。如此模型所述，AWS 負責保護執行所有的全球基礎設施 AWS 雲端。您負責維護在此基礎設施上託管內容的控制權。您也同時負責所使用 AWS 服務的安全組態和管理任務。如需資料隱私權的詳細資訊，請參閱[資料隱私權常見問答集](#)。如需

有關歐洲資料保護的相關資訊，請參閱AWS 安全性部落格上的[AWS 共同責任模型和 GDPR](#) 部落格文章。

基於資料保護目的，建議您保護 AWS 帳戶 登入資料，並使用 AWS IAM Identity Center 或 AWS Identity and Access Management (IAM) 設定個別使用者。如此一來，每個使用者都只會獲得授與完成其任務所必須的許可。我們也建議您採用下列方式保護資料：

- 每個帳戶均要使用多重要素驗證 (MFA)。
- 使用 SSL/TLS 與 AWS 資源通訊。我們需要 TLS 1.2 並建議使用 TLS 1.3。
- 使用 設定 API 和使用使用者活動記錄 AWS CloudTrail。如需有關使用 CloudTrail 追蹤擷取 AWS 活動的資訊，請參閱AWS CloudTrail 《使用者指南》中的[使用 CloudTrail 追蹤](#)。
- 使用 AWS 加密解決方案，以及其中的所有預設安全控制 AWS 服務。
- 使用進階的受管安全服務 (例如 Amazon Macie)，協助探索和保護儲存在 Amazon S3 的敏感資料。
- 如果您在 AWS 透過命令列界面或 API 存取 時需要 FIPS 140-3 驗證的密碼編譯模組，請使用 FIPS 端點。如需有關 FIPS 和 FIPS 端點的更多相關資訊，請參閱[聯邦資訊處理標準 \(FIPS\) 140-3](#)。

我們強烈建議您絕對不要將客戶的電子郵件地址等機密或敏感資訊，放在標籤或自由格式的文字欄位中，例如名稱欄位。這包括當您使用 AWS Clean Rooms 或使用主控台、API AWS CLI或其他 AWS 服務 AWS SDKs 時。您在標籤或自由格式文字欄位中輸入的任何資料都可能用於計費或診斷日誌。如果您提供外部伺服器的 URL，我們強烈建議請勿在驗證您對該伺服器請求的 URL 中包含憑證資訊。

靜態加密

AWS Clean Rooms 一律會加密所有靜態服務中繼資料，而不需要任何額外的組態。當您使用 時，此加密會自動執行 AWS Clean Rooms。

Clean Rooms ML 會加密所有存放在服務中的靜態資料 AWS KMS。如果您選擇提供自己的 KMS 金鑰，外觀模型和外觀區段產生任務的內容會使用 KMS 金鑰靜態加密。

使用 AWS Clean Rooms 自訂 ML 模型時，服務會加密所有存放於 的靜態資料 AWS KMS。AWS Clean Rooms 支援使用您建立、擁有和管理的對稱客戶受管金鑰來加密靜態資料。如果未指定客戶受管金鑰，AWS 擁有的金鑰 預設會使用。

AWS Clean Rooms 使用授予和金鑰政策來存取客戶受管金鑰。您可以隨時撤銷授予的存取權，或移除服務對客戶受管金鑰的存取權。如果您這麼做，AWS Clean Rooms 則無法存取客戶受管金鑰加密的任何資料，這會影響相依於該資料的操作。例如，如果您嘗試從 AWS Clean Rooms 無法存取的加密 ML 輸入通道建立訓練模型，則操作會傳回ValidationException錯誤。

Note

您可以使用 Amazon S3 中的加密選項來保護靜態資料。

如需詳細資訊，請參閱 [《Amazon S3 使用者指南》](#) 中的指定 [Amazon S3 加密](#)。Amazon S3

在 中使用 ID 映射表時 AWS Clean Rooms，服務會加密所有以 存放的靜態資料 AWS KMS。如果您選擇提供自己的 KMS 金鑰，您的 ID 映射表的內容會透過您的 KMS 金鑰進行靜態加密 AWS Entity Resolution。如需使用 ID 映射工作流程進行加密所需許可的詳細資訊，請參閱 AWS Entity Resolution 《使用者指南》中的 [為 建立工作流程任務角色 AWS Entity Resolution](#)。

傳輸中加密

AWS Clean Rooms 使用 Transport Layer Security (TLS) 進行傳輸中的加密。與 AWS Clean Rooms 的通訊一律透過 HTTPS 完成，因此無論資料是存放在 Amazon S3、Amazon Athena 或 Snowflake 中，您的資料一律會在傳輸中加密。這包括使用 Clean Rooms ML 時傳輸中的所有資料。

加密基礎資料

如需如何加密基礎資料的詳細資訊，請參閱 [的加密運算 Clean Rooms](#)。

金鑰政策

金鑰政策會控制客戶受管金鑰的存取權限。每個客戶受管金鑰都必須只有一個金鑰政策，其中包含決定誰可以使用金鑰及其使用方式的陳述式。在建立客戶受管金鑰時，可以指定金鑰政策。如需詳細資訊，請參閱《AWS Key Management Service 開發人員指南》中的管理客戶受管金鑰的存取。

若要將客戶受管金鑰與 AWS Clean Rooms 自訂 ML 模型搭配使用，必須在金鑰政策中允許下列 API 操作：

- kms:DescribeKey – 提供客戶受管金鑰詳細資訊，AWS Clean Rooms 以允許 驗證金鑰。
- kms:Decrypt – 提供 的存取權，AWS Clean Rooms 以解密加密的資料，並將其用於相關任務。
- kms:CreateGrant - Clean Rooms ML 透過為 Amazon ECR 建立授權，來加密 Amazon ECR 中的靜態訓練和推論映像。若要進一步了解，請參閱 [Amazon ECR 中的靜態加密](#)。Clean Rooms ML 也會使用 Amazon SageMaker AI 執行訓練和推論任務，並為 SageMaker AI 建立授權，以加密連接至執行個體的 Amazon EBS 磁碟區，以及 Amazon S3 中的輸出資料。若要進一步了解，請參閱在 [Amazon SageMaker AI 中使用加密保護靜態資料](#)。

- kms:GenerateDataKey - Clean Rooms ML 使用伺服器端加密來加密存放在 Amazon S3 中的靜態資料 AWS KMS keys。若要進一步了解，請參閱[在 Amazon S3 中使用伺服器端加密搭配 AWS KMS keys \(SSE-KMS\)](#)。

以下是您可以 AWS Clean Rooms 為下列資源新增的政策陳述式範例：

ML 輸入通道與合成資料

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Allow access to principals authorized to use AWS Clean Rooms ML",
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::666666666666:role/ExampleRole"
      },
      "Action": [
        "kms:GenerateDataKey",
        "kms:CreateGrant",
        "kms:Decrypt"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "kms:ViaService": "cleanrooms-ml.us-east-1.amazonaws.com"
        },
        "ForAllValues:StringEquals": {
          "kms:GrantOperations": [
            "Decrypt",
            "Encrypt",
            "GenerateDataKeyWithoutPlaintext",
            "ReEncryptFrom",
            "ReEncryptTo",
            "CreateGrant",
            "DescribeKey",
            "RetireGrant",
            "GenerateDataKey"
          ]
        },
        "BoolIfExists": {
          "kms:GrantIsForAWSResource": true
        }
      }
    }
  ]
}
```

```

    }
  },
  {
    "Sid": "Allow describe key for principals authorized to use AWS Clean Rooms
ML",
    "Effect": "Allow",
    "Principal": {
      "AWS": "arn:aws:iam::444455556666:role/ExampleRole"
    },
    "Action": [
      "kms:DescribeKey"
    ],
    "Resource": "*",
    "Condition": {
      "StringEquals": {
        "kms:ViaService": "cleanrooms-ml.us-east-1.amazonaws.com"
      }
    }
  },
  {
    "Sid": "Allow grant operations for AWS Clean Rooms ML service principal",
    "Effect": "Allow",
    "Principal": {
      "Service": "cleanrooms-ml.amazonaws.com"
    },
    "Action": [
      "kms:GenerateDataKey",
      "kms:CreateGrant",
      "kms:Decrypt"
    ],
    "Resource": "*",
    "Condition": {
      "ForAllValues:StringEquals": {
        "kms:GrantOperations": [
          "Decrypt",
          "Encrypt",
          "GenerateDataKeyWithoutPlaintext",
          "ReEncryptFrom",
          "ReEncryptTo",
          "CreateGrant",
          "DescribeKey",
          "RetireGrant",
          "GenerateDataKey"
        ]
      }
    }
  }
]

```

```

        }
    },
    {
        "Sid": "Allow describe key for AWS Clean Rooms ML service principal",
        "Effect": "Allow",
        "Principal": {
            "Service": "cleanrooms-ml.amazonaws.com"
        },
        "Action": [
            "kms:DescribeKey"
        ],
        "Resource": "*"
    }
]
}

```

不含合成資料的 ML 輸入通道

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Allow access to principals authorized to use AWS Clean Rooms ML",
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::666666666666:role/ExampleRole"
      },
      "Action": [
        "kms:DescribeKey",
        "kms:GenerateDataKey",
        "kms:Decrypt"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "kms:ViaService": "cleanrooms-ml.us-east-1.amazonaws.com"
        }
      }
    }
  ],
  {

```



```
    "Sid": "Allow access to AWS Clean Rooms ML service principal",
    "Effect": "Allow",
    "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
    },
    "Action": [
        "kms:DescribeKey",
        "kms:GenerateDataKey",
        "kms:Decrypt"
    ],
    "Resource": "*"
  }
]
```

訓練過的模型任務或訓練過的模型推論任務

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Allow grant operations for principals authorized to use AWS Clean Rooms ML",
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::666666666666:role/ExampleRole"
      },
      "Action": [
        "kms:GenerateDataKey",
        "kms:CreateGrant",
        "kms:Decrypt"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "kms:ViaService": "cleanrooms-ml.us-east-1.amazonaws.com"
        },
        "ForAllValues:StringEquals": {
          "kms:GrantOperations": [
            "Decrypt",

```

```

        "Encrypt",
        "GenerateDataKeyWithoutPlaintext",
        "ReEncryptFrom",
        "ReEncryptTo",
        "CreateGrant",
        "DescribeKey",
        "RetireGrant",
        "GenerateDataKey"
    ]
},
    "BoolIfExists": {
        "kms:GrantIsForAWSResource": true
    }
},
{
    "Sid": "Allow describe key for principals authorized to use AWS Clean
Rooms ML",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::444455556666:role/ExampleRole"
    },
    "Action": [
        "kms:DescribeKey"
    ],
    "Resource": "*",
    "Condition": {
        "StringEquals": {
            "kms:ViaService": "cleanrooms-ml.us-east-1.amazonaws.com"
        }
    }
},
{
    "Sid": "Allow grant operations for AWS Clean Rooms ML service
principal",
    "Effect": "Allow",
    "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
    },
    "Action": [
        "kms:GenerateDataKey",
        "kms:CreateGrant",
        "kms:Decrypt"
    ],

```

```

        "Resource": "*",
        "Condition": {
            "ForAllValues:StringEquals": {
                "kms:GrantOperations": [
                    "Decrypt",
                    "Encrypt",
                    "GenerateDataKeyWithoutPlaintext",
                    "ReEncryptFrom",
                    "ReEncryptTo",
                    "CreateGrant",
                    "DescribeKey",
                    "RetireGrant",
                    "GenerateDataKey"
                ]
            }
        },
        {
            "Sid": "Allow describe key for AWS Clean Rooms ML service principal",
            "Effect": "Allow",
            "Principal": {
                "Service": "cleanrooms-ml.amazonaws.com"
            },
            "Action": [
                "kms:DescribeKey"
            ],
            "Resource": "*"
        }
    ]
}

```

Clean Rooms ML 不支援在客戶受管金鑰政策中指定服務加密內容或來源內容。CloudTrail 中的客戶可以看到服務在內部使用的加密內容。

使用的服務連結角色 AWS Clean Rooms

AWS Clean Rooms 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 Clean Rooms。服務連結角色由預先定義，AWS Clean Rooms 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 Clean Rooms 輕鬆地設定，因為您不必手動新增必要的許可。AWS Clean Rooms 會定義其服務連結角色的許可，除非另有定義，否則 AWS Clean Rooms 只能擔任其角色。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 Clean Rooms 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 Clean Rooms

AWS Clean Rooms 使用名為 `AWSServiceRoleForAWSCleanRooms` 的服務連結角色 – 將 Clean Rooms 相關 CloudWatch 指標發佈至 AWS 您的帳戶。

`AWSServiceRoleForAWSCleanRooms` 服務連結角色信任下列服務擔任該角色：

- `cleanrooms.amazonaws.com`

名為 `AWSCleanRoomsServiceRolePolicy` 的角色許可政策允許 對指定的資源 AWS Clean Rooms 完成下列動作：

- 動作：`all AWS resources, restricted to the AWS Clean Rooms namespace 上的 cloudwatch:PutMetricData`

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 建立服務連結角色 AWS Clean Rooms

您可以使用 IAM 主控台建立具有 `AWSServiceRoleForAWSCleanRooms` 使用案例的服務連結角色。在 AWS CLI 或 AWS API 中，使用服務名稱建立 `cleanrooms.amazonaws.com` 服務連結角色。如需詳細資訊，請參閱《IAM 使用者指南》中的「[建立服務連結角色](#)」。如果您刪除此服務連結角色，您可以使用此相同的程序以再次建立該角色。

編輯 的服務連結角色 AWS Clean Rooms

AWS Clean Rooms 不允許您編輯 `AWSServiceRoleForAWSCleanRooms` 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](#)。

刪除的服務連結角色 AWS Clean Rooms

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，在手動刪除服務連結角色之前，您必須先清除資源。

若要刪除 `AWSServiceRoleForAWSCleanRooms`，您必須先刪除中的所有[協同合作](#)和[成員資格](#) AWS 帳戶。

Note

如果 AWS Clean Rooms 服務在您嘗試刪除資源時使用角色，則刪除可能會失敗。若此情況發生，請等待數分鐘後並再次嘗試操作。

使用 IAM 手動刪除服務連結角色

使用 IAM 主控台 AWS CLI、或 AWS API 來刪除 `AWSServiceRoleForAWSCleanRooms` 服務連結角色。如需詳細資訊，請參閱《IAM 使用者指南》中的[刪除服務連結角色](#)。

AWS Clean Rooms 服務連結角色支援的 區域

AWS Clean Rooms 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 Regions and endpoints](#)。

中的資料保留 AWS Clean Rooms

任何暫時讀取到 AWS Clean Rooms 協同合作的資料都會在查詢完成後刪除。

當您建立外觀相似的模型時，Clean Rooms ML 會讀取您的訓練資料，將其轉換為適合我們 ML 模型的格式，並將訓練過的模型參數存放在 Clean Rooms ML 中。Clean Rooms ML 不會保留訓練資料的副本。AWS Clean Rooms SQL 查詢不會在查詢執行後保留任何資料。Clean Rooms ML 接著會使用訓練過的模型來摘要所有使用者的行為。只要您的外觀模型處於作用中狀態，Clean Rooms ML 就會將每個使用者的使用者層級資料集存放在您的資料中。

當您開始類似區段的產生任務時，Clean Rooms ML 會讀取種子資料、從相關聯的類似外觀模型讀取行為摘要，以及建立存放在 AWS Clean Rooms 服務中的類似區段。Clean Rooms ML 不會保留種子資料的副本。只要任務處於作用中狀態，Clean Rooms ML 就會儲存任務的使用者層級輸出。

如果您的種子資料來自 SQL 查詢，則該查詢的輸出只會在任務期間存放在服務中。查詢的結果會在靜態和傳輸中加密。

如果您想要移除看起來像模型或看起來像區段產生任務資料，請使用 API 將其刪除。Clean Rooms ML 會非同步刪除與模型或任務相關聯的所有資料。此程序完成後，Clean Rooms ML 會刪除模型或任務的中繼資料，而且不會再出現在 API 中。Clean Rooms ML 會保留已刪除的資料 3 天，以防止災難復原。一旦任務或模型不再顯示在 API 中且已超過 3 天，與該模型或任務相關聯的所有資料都已永久刪除。

中的資料協同合作最佳實務 AWS Clean Rooms

本主題說明在 中執行資料協作的最佳實務 AWS Clean Rooms。

AWS Clean Rooms 遵循[AWS 共同責任模型](#)。AWS Clean Rooms 提供您可以設定的[分析規則](#)，以增強您在協同合作中保護敏感資料的能力。您在 中設定的分析規則 AWS Clean Rooms 會強制執行您已設定的限制（查詢控制項和查詢輸出控制項）。您需負責判斷限制，並相應地設定分析規則。

資料協作可能不僅涉及您使用 AWS Clean Rooms。為了協助您最大化資料協同合作的優勢，我們建議您使用 執行下列最佳實務 AWS Clean Rooms，特別是分析規則。

主題

- [的最佳實務 AWS Clean Rooms](#)
- [在 中使用分析規則的最佳實務 AWS Clean Rooms](#)

的最佳實務 AWS Clean Rooms

您有責任評估每個資料協作的風險，並將其與您的隱私權要求進行比較，例如外部和內部合規計劃和政策。我們建議您在使用 時採取其他動作 AWS Clean Rooms。這些動作可能有助於進一步管理風險，並協助防止第三方嘗試重新識別您的資料（例如，差異化攻擊或側邊通道攻擊）。

例如，考慮對您的其他協作者進行盡職調查，並在參與協作之前與他們簽訂法律協議。若要監控資料的使用情況，也請考慮在您使用 時採用其他稽核機制 AWS Clean Rooms。

在 中使用分析規則的最佳實務 AWS Clean Rooms

中的分析規則 AWS Clean Rooms 可讓您透過在設定的資料表上設定查詢控制項來限制可執行的查詢。例如，您可以設定查詢控制，以了解如何聯結已設定的資料表，以及可以選取哪些資料欄。您也可以透過設定查詢結果控制項來限制查詢輸出，例如輸出列上的彙總閾值。服務會拒絕任何查詢，並移除不符合查詢中成員所設定資料表之分析規則的資料列。

我們建議您在已設定的資料表上使用分析規則的 10 個最佳實務：

- 為個別查詢使用案例建立個別設定的資料表（例如，受眾規劃或歸因）。您可以使用相同的基礎資料表建立多個已設定的 AWS Glue 資料表。
- 在分析規則中指定協同合作中查詢所需的資料欄（例如維度資料欄、列出資料欄、聯結資料欄）。這可能有助於降低不同攻擊的風險，或讓其他成員反向設計您的資料。使用允許清單資料欄功能，記下您可能希望在未來進行查詢的其他資料欄。若要自訂可用於特定協同合作的資料欄，請使用相同的基礎資料表建立其他設定的 AWS Glue 資料表。
- 在分析規則中指定在協同合作中分析所需的函數。這有助於降低罕見函數錯誤的風險，這些錯誤可能會呈現個別資料點的資訊。若要自訂可用於特定協同合作的函數，請使用相同的基礎 AWS Glue 資料表建立其他設定的資料表。
- 在資料列層級值敏感的任何資料欄上新增彙總限制。這包括已設定資料表中的資料欄，這些資料欄也存在於其他協同合作成員的資料表和分析規則中，做為彙總限制條件。這也包含已設定資料表中不可查詢的資料欄，也就是已設定資料表中但不在分析規則中的資料欄。彙總限制有助於降低將查詢結果與協同合作以外的資料建立關聯的風險。
- 建立測試協同合作和分析規則，以測試使用指定分析規則建立的限制。
- 在設定的資料表上檢閱協作者設定的資料表和成員的分析規則，以檢查它們是否符合協作商定的內容。這有助於降低其他成員自行設計資料以執行未同意之查詢的風險。
- 在您設定分析規則後，檢閱在已設定資料表上啟用的範例查詢（僅限主控台）。

Note

除了提供的範例查詢之外，也可以根據分析規則和其他協同合作成員資料表和分析規則進行其他查詢。

- 您可以在協同合作中新增或更新已設定資料表的分析規則。當您這麼做時，請檢閱已設定資料表相關聯的所有協同合作及其產生的影響。這有助於確保沒有任何協同合作使用過時的分析規則。
- 檢閱協同合作中執行的查詢，以檢查查詢是否符合協同合作商定的使用案例或查詢。（開啟查詢記錄功能時，查詢可在查詢日誌中使用。）這有助於降低執行未同意分析之成員的風險，以及潛在攻擊，例如側邊頻道攻擊。
- 檢閱協同合作成員分析規則和查詢中使用的已設定資料表資料欄，以檢查它們是否符合協同合作中同意的內容。（開啟此功能時，查詢可在查詢日誌中使用。）這有助於降低其他成員自行設計資料以進行未同意之查詢的風險。

的 Identity and Access Management AWS Clean Rooms

AWS Identity and Access Management (IAM) 是一種 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 Clean Rooms 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 Clean Rooms 如何使用 IAM](#)
- [的身分型政策範例 AWS Clean Rooms](#)
- [AWS 的 受管政策 AWS Clean Rooms](#)
- [對 AWS Clean Rooms 身分和存取進行故障診斷](#)
- [預防跨服務混淆代理人](#)
- [AWS Clean Rooms ML 的 IAM 行為](#)
- [Clean Rooms ML 自訂模型的 IAM 行為](#)

目標對象

如何使用 AWS Identity and Access Management (IAM) 會因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 [對 AWS Clean Rooms 身分和存取進行故障診斷](#))
- 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 Clean Rooms 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 Clean Rooms](#))

使用身分驗證

身分驗證是您 AWS 使用身分憑證登入的方式。您必須以 AWS 帳戶根使用者身分、IAM 使用者身分或擔任 IAM 角色來驗證（登入 AWS）。

您可以使用透過身分來源提供的憑證，以聯合身分 AWS 身分身分登入。AWS IAM Identity Center (IAM Identity Center) 使用者或您公司的單一登入身分驗證是聯合身分的範例。您以聯合身分登入時，

您的管理員先前已設定使用 IAM 角色的聯合身分。當您使用聯合 AWS 身分存取時，您會間接擔任角色。

根據您的使用者類型，您可以登入 AWS 管理主控台 或 AWS 存取入口網站。如需登入的詳細資訊 AWS，請參閱AWS 登入 《使用者指南》中的[如何登入您的 AWS 帳戶](#)。

如果您以 AWS 程式設計方式存取，AWS 會提供軟體開發套件 (SDK) 和命令列界面 (CLI)，以使用您的憑證以密碼編譯方式簽署您的請求。如果您不使用 AWS 工具，則必須自行簽署請求。如需使用建議方法來自行簽署請求的詳細資訊，請參閱 AWS 一般參考 中的[第 4 版簽署程序](#)。

無論您使用何種身分驗證方法，您可能都需要提供額外的安全性資訊。例如，AWS 建議您使用多重驗證 (MFA) 來提高帳戶的安全性。如需更多資訊，請參閱 AWS IAM Identity Center 使用者指南中的[多重要素驗證](#)和 IAM 使用者指南中的[在 AWS 中使用多重要素驗證 \(MFA\)](#)。

AWS 帳戶 根使用者

當您建立時 AWS 帳戶，您會從一個登入身分開始，該身分可完整存取帳戶中的所有 AWS 服務和資源。此身分稱為 AWS 帳戶 根使用者，使用建立帳戶時所使用的電子郵件地址和密碼即可登入並存取。強烈建議您不要以根使用者處理日常作業。保護您的根使用者憑證，並將其用來執行只能由根使用者執行的任務。如需需要您以根使用者身分登入的任務完整清單，請參閱 中的[AWS 帳戶根使用者 登入資料和 IAM 身分](#) AWS 一般參考。

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 的使用者。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。政策是中的物件，當與身分或資源建立關聯時，AWS 會定義其許可。當委託人（使用者、根使用者或角色工作階段）發出請求時，會 AWS 評估這些政策。政策中的許可決定是否允許或拒絕請求。大多數政策會以 JSON 文件 AWS 形式存放在中。如需 JSON 政策文件結構和內容的詳細資訊，請參閱《IAM 使用者指南》中的[JSON 政策概觀](#)。

管理員可以使用 AWS JSON 政策來指定誰可以存取內容。也就是說，哪個主體在什麼條件下可以對什麼資源執行哪些動作。

每個 IAM 實體（使用者或角色）在開始時都沒有許可。根據預設，使用者無法執行任何作業，甚至也無法變更他們自己的密碼。若要授予使用者執行動作的許可，管理員必須將許可政策附加到使用者。或者，管理員可以將使用者新增到具備預定許可的群組。管理員將許可給予群組時，該群組中的所有使用者都會獲得那些許可。

IAM 政策定義該動作的許可，無論您使用何種方法來執行操作。例如，假設您有一個允許 `iam:GetRole` 動作的政策。具有該政策的使用者可以從 AWS 管理主控台 AWS CLI、或 AWS API 取得角色資訊。

身分型政策

身分型政策是可以附加到身分（例如 IAM 使用者、使用者群組或角色）的 JSON 許可政策文件。這些政策可控制身分在何種條件下能對哪些資源執行哪些動作。如需了解如何建立身分型政策，請參閱《IAM 使用者指南》中的[透過客戶管理政策定義自訂 IAM 許可](#)。

身分型政策可進一步分類成內嵌政策或受管政策。內嵌政策會直接內嵌到單一使用者、群組或角色。受管政策則是獨立的政策，您可以將這些政策附加到 AWS 帳戶中的多個使用者、群組和角色。受管政策包括 AWS 受管政策和客戶受管政策。如需了解如何在受管政策及內嵌政策間選擇，請參閱 IAM 使用者指南中的[在受管政策和內嵌政策間選擇](#)。

資源型政策

資源型政策是附加到資源的 JSON 政策文件。資源型政策的最常見範例是 IAM 角色信任政策和 Amazon S3 儲存貯體政策。在支援資源型政策的服務中，服務管理員可以使用它們來控制對特定資源的存取權限。對於附加政策的資源，政策會定義指定的主體可以對該資源執行的動作以及在何種條件下執行的動作。您必須在資源型政策中[指定主體](#)。委託人可以包含帳戶、使用者、角色、聯合身分使用者或 AWS 服務。

資源型政策是位於該服務中的內嵌政策。您無法在資源型政策中使用來自 IAM 的 AWS 受管政策。

其他政策類型

AWS 支援其他較不常見的政策類型。這些政策類型可設定較常見政策類型授予您的最大許可。

- 許可界限 – 許可範圍是一種進階功能，可供您設定身分型政策能授予 IAM 實體 (IAM 使用者或角色) 的最大許可。您可以為實體設定許可界限。所產生的許可會是實體的身分型政策和其許可界限的交集。會在 Principal 欄位中指定使用者或角色的資源型政策則不會受到許可界限限制。所有這類政策中的明確拒絕都會覆寫該允許。如需許可界限的詳細資訊，請參閱《IAM 使用者指南》中的[IAM 實體許可界限](#)。
- 服務控制政策 (SCPs) – SCPs 是 JSON 政策，可指定中組織或組織單位 (OU) 的最大許可 AWS Organizations。AWS Organizations 是一種用於分組和集中管理您企業擁有 AWS 帳戶的多個的服務。若您啟用組織中的所有功能，您可以將服務控制政策 (SCP) 套用到任何或所有帳戶。SCP 會限制成員帳戶中實體的許可，包括每個實體 AWS 帳戶根使用者。如需 Organizations 和 SCP 的詳細資訊，請參閱 AWS Organizations 使用者指南中的[SCP 運作方式](#)。
- 工作階段政策 – 工作階段政策是一種進階政策，您可以在透過編寫程式的方式建立角色或聯合使用者的暫時工作階段時，作為參數傳遞。所產生工作階段的許可會是使用者或角色的身分型政策和工作階段政策的交集。許可也可以來自資源型政策。所有這類政策中的明確拒絕都會覆寫該允許。如需詳細資訊，請參《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 在涉及多種政策類型時決定是否允許請求，請參閱《IAM 使用者指南》中的[政策評估邏輯](#)。

AWS Clean Rooms 如何使用 IAM

在您使用 IAM 管理對的存取之前 AWS Clean Rooms，請先了解可使用哪些 IAM 功能 AWS Clean Rooms。

您可以搭配 使用的 IAM 功能 AWS Clean Rooms

IAM 功能	AWS Clean Rooms 支援
身分型政策	是
資源型政策	部分
政策動作	是
政策資源	是
政策條件索引鍵 (服務特定)	部分
ACL	否
ABAC (政策中的標籤)	是
臨時憑證	是
轉送存取工作階段 (FAS)	是
服務角色	是
服務連結角色	否

若要全面了解 AWS Clean Rooms 和其他 如何與大多數 IAM 功能 AWS 服務 搭配使用，請參閱 [《AWS 服務 IAM 使用者指南》中的 與 IAM 搭配使用](#)。

的身分型政策 AWS Clean Rooms

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](#)。

的身分型政策範例 AWS Clean Rooms

若要檢視 AWS Clean Rooms 身分型政策的範例，請參閱 [的身分型政策範例 AWS Clean Rooms](#)。

中的資源型政策 AWS Clean Rooms

支援資源型政策：部分

資源型政策是附加到資源的 JSON 政策文件。資源型政策的最常見範例是 IAM 角色信任政策和 Amazon S3 儲存貯體政策。在支援資源型政策的服務中，服務管理員可以使用它們來控制對特定資源的存取權限。對於附加政策的資源，政策會定義指定的主體可以對該資源執行的動作以及在何種條件下執行的動作。您必須在資源型政策中[指定主體](#)。委託人可以包含帳戶、使用者、角色、聯合身分使用者或 AWS 服務。

如需啟用跨帳戶存取權，您可以在其他帳戶內指定所有帳戶或 IAM 實體作為資源型政策的主體。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](#)。

AWS Clean Rooms 此服務僅支援一種資源類型政策，稱為已設定的外觀模型受管資源政策，該政策會連接到已設定的外觀模型。此政策定義哪些委託人可以在已設定的外觀模型上執行動作。

若要了解如何將資源型政策連接至已設定的類似模型，請參閱 [AWS Clean Rooms ML 的 IAM 行為](#)。

的政策動作 AWS Clean Rooms

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 Clean Rooms 動作清單，請參閱《服務授權參考》中的 [定義的動作 AWS Clean Rooms](#)。

中的政策動作在動作之前 AWS Clean Rooms 使用以下字首。

```
cleanrooms
```

若要在單一陳述式中指定多個動作，請用逗號分隔。

```
"Action": [  
    "cleanrooms:action1",  
    "cleanrooms:action2"  
]
```

若要檢視 AWS Clean Rooms 身分型政策的範例，請參閱 [的身分型政策範例 AWS Clean Rooms](#)。

的政策資源 AWS Clean Rooms

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 Clean Rooms 資源類型及其 ARNs，請參閱《服務授權參考》中的 [定義的資源 AWS Clean Rooms](#)。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 Clean Rooms 定義的動作](#)。

若要檢視 AWS Clean Rooms 身分型政策的範例，請參閱 [的身分型政策範例 AWS Clean Rooms](#)。

的政策條件索引鍵 AWS Clean Rooms

支援服務特定的政策條件索引鍵：部分

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](#)。

若要了解 AWS Clean Rooms ML 如何使用政策條件金鑰，請參閱 [AWS Clean Rooms ML 的 IAM 行為](#)。

中的 ACLs AWS Clean Rooms

支援 ACL：否

存取控制清單 (ACL) 可控制哪些主體 (帳戶成員、使用者或角色) 擁有存取某資源的許可。ACL 類似於資源型政策，但它們不使用 JSON 政策文件格式。

ABAC 搭配 AWS Clean Rooms

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 使用臨時登入資料 AWS Clean Rooms

支援臨時憑證：是

臨時登入資料提供 AWS 資源的短期存取權，並在您使用聯合或切換角色時自動建立。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](#)。

轉送 的存取工作階段 AWS Clean Rooms

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，並結合請求 AWS 服務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 Clean Rooms

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 Clean Rooms 功能。只有在 AWS Clean Rooms 提供指引時，才能編輯服務角色。

的服務連結角色 AWS Clean Rooms

支援服務連結角色：否

服務連結角色是連結至的一種服務角色 AWS 服務。服務可以擔任代表您執行動作的角色。服務連結角色會出現在您的 中 AWS 帳戶，並由服務擁有。IAM 管理員可以檢視，但不能編輯服務連結角色的許可。

如需建立或管理服务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 Clean Rooms

根據預設，使用者和角色不具備建立或修改 AWS Clean Rooms 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 Clean Rooms，包括每種資源類型的 ARNs 格式，請參閱《服務授權參考》中的 [的動作、資源和條件索引鍵 AWS Clean Rooms](#)。

主題

- [政策最佳實務](#)
- [使用 AWS Clean Rooms 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 Clean Rooms 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 受管政策並邁向最低權限許可 – 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，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 例如 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 Clean Rooms 主控台

若要存取 AWS Clean Rooms 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 Clean Rooms 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 Clean Rooms 主控台，請將 AWS Clean Rooms *FullAccess* 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 Clean Rooms

AWS 受管政策是由 AWS 受管政策建立和管理的獨立政策旨在為許多常用案例提供許可，以便您可以開始將許可指派給使用者、群組和角色。

請記住，AWS 受管政策可能不會授予特定使用案例的最低權限許可，因為這些許可可供所有 AWS 客戶使用。我們建議您定義特定於使用案例的[客戶管理政策](#)，以便進一步減少許可。

您無法變更 AWS 受管政策中定義的許可。如果 AWS 更新受 AWS 管政策中定義的許可，則更新會影響政策連接的所有主體身分（使用者、群組和角色）。當新的 AWS 服務 啟動或新的 API 操作可用於現有服務時，AWS 最有可能更新 AWS 受管政策。

如需詳細資訊，請參閱 IAM 使用者指南中的 [AWS 受管政策](#)。

AWS 受管政策：AWSCleanRoomsReadOnlyAccess

您可以AWSCleanRoomsReadOnlyAccess連接到您的 IAM 主體。

此政策會授予協同AWSCleanRoomsReadOnlyAccess合作中資源和中繼資料的唯讀許可。

許可詳細資訊

此政策包含以下許可：

- CleanRoomsRead - 允許主體唯讀存取 服務。
- ConsoleDisplayTables：允許主體唯讀存取所需的 AWS Glue 中繼資料，以顯示主控台上基礎 AWS Glue 資料表的資料。
- ConsoleLogSummaryQueryLogs – 允許主體查看查詢日誌。
- ConsoleLogSummaryObtainLogs – 允許主體擷取日誌結果。

如需政策詳細資訊的 JSON 清單，請參閱《AWS 受管政策參考指南》中的 [AWSCleanRoomsReadOnlyAccess](#)。

AWS 受管政策：AWSCleanRoomsFullAccess

您可以將 AWSCleanRoomsFullAccess連接到您的 IAM 主體。

此政策會授予管理許可，以允許對 AWS Clean Rooms 協同合作中的資源和中繼資料進行完整存取（讀取、寫入和更新）。此政策包含執行查詢的存取權。

許可詳細資訊

此政策包含以下許可：

- CleanRoomsAccess – 授予對所有 資源的所有動作的完整存取權 AWS Clean Rooms。
- PassServiceRole – 准許將服務角色傳遞給其名稱中具有 "cleanrooms" 的服務 (PassedToService 條件)。

- `ListRolesToPickServiceRole` – 允許主體列出其所有角色，以便在使用時選擇服務角色 AWS Clean Rooms。
- `GetRoleAndListRolePoliciesToInspectServiceRole` – 允許主體在 IAM 中查看服務角色和對應的政策。
- `ListPoliciesToInspectServiceRolePolicy` – 允許主體在 IAM 中查看服務角色和對應的政策。
- `GetPolicyToInspectServiceRolePolicy` – 允許主體在 IAM 中查看服務角色和對應的政策。
- `ConsoleDisplayTables` : 允許主體唯讀存取所需的 AWS Glue 中繼資料，以顯示主控台上基礎 AWS Glue 資料表的資料。
- `ConsolePickQueryResultsBucketListAll` – 允許主體從寫入查詢結果的所有可用 S3 儲存貯體清單中選擇 Amazon S3 儲存貯體。
- `SetQueryResultsBucket` – 允許主體選擇要寫入其查詢結果的 S3 儲存貯體。
- `ConsoleDisplayQueryResults` – 允許主體向客戶顯示查詢結果，從 S3 儲存貯體讀取。
- `WriteQueryResults` – 允許主體將查詢結果寫入客戶擁有的 S3 儲存貯體。
- `EstablishLogDeliveries` – 允許主體將查詢日誌交付到客戶的 Amazon CloudWatch Logs 日誌群組。
- `SetupLogGroupsDescribe` – 允許主體使用 Amazon CloudWatch Logs 日誌群組建立程序。
- `SetupLogGroupsCreate` – 允許主體建立 Amazon CloudWatch Logs 日誌群組。
- `SetupLogGroupsResourcePolicy` – 允許主體在 Amazon CloudWatch Logs 日誌群組上設定資源政策。
- `ConsoleLogSummaryQueryLogs` – 允許主體查看查詢日誌。
- `ConsoleLogSummaryObtainLogs` – 允許主體擷取日誌結果。

如需政策詳細資訊的 JSON 清單，請參閱《AWS 受管政策參考指南》中的 [AWSCleanRoomsFullAccess](#)。

AWS 受管政策：**AWSCleanRoomsFullAccessNoQuerying**

您可以 `AWSCleanRoomsFullAccessNoQuerying` 連接到您的 IAM principals。

此政策會授予管理許可，以允許對 AWS Clean Rooms 協同合作中的資源和中繼資料進行完整存取（讀取、寫入和更新）。此政策排除執行查詢的存取權。

許可詳細資訊

此政策包含以下許可：

- `CleanRoomsAccess` – 授予所有 資源的所有動作的完整存取權 AWS Clean Rooms，但協同合作中查詢除外。
- `CleanRoomsNoQuerying` – 明確拒絕 `StartProtectedQuery` 和 `UpdateProtectedQuery` 以防止查詢。
- `PassServiceRole` – 准許將服務角色傳遞給其名稱中具有 "cleanrooms" 的服務 (`PassedToService` 條件)。
- `ListRolesToPickServiceRole` – 允許主體列出其所有角色，以便在使用 時選擇服務角色 AWS Clean Rooms。
- `GetRoleAndListRolePoliciesToInspectServiceRole` – 允許主體在 IAM 中查看服務角色和對應的政策。
- `ListPoliciesToInspectServiceRolePolicy` – 允許主體在 IAM 中查看服務角色和對應的政策。
- `GetPolicyToInspectServiceRolePolicy` – 允許主體在 IAM 中查看服務角色和對應的政策。
- `ConsoleDisplayTables` : 允許主體唯讀存取所需的 AWS Glue 中繼資料，以顯示主控台上基礎 AWS Glue 資料表的資料。
- `EstablishLogDeliveries` – 允許主體將查詢日誌交付到客戶的 Amazon CloudWatch Logs 日誌群組。
- `SetupLogGroupsDescribe` – 允許主體使用 Amazon CloudWatch Logs 日誌群組建立程序。
- `SetupLogGroupsCreate` – 允許主體建立 Amazon CloudWatch Logs 日誌群組。
- `SetupLogGroupsResourcePolicy` – 允許主體在 Amazon CloudWatch Logs 日誌群組上設定資源政策。
- `ConsoleLogSummaryQueryLogs` – 允許主體查看查詢日誌。
- `ConsoleLogSummaryObtainLogs` – 允許主體擷取日誌結果。
- `cleanrooms` – 管理服務內的 AWS Clean Rooms 協同合作、分析範本、設定的資料表、成員資格和相關資源。執行各種操作，例如建立、更新、刪除、列出和擷取這些資源的相關資訊。
- `iam` – 將名稱包含 "cleanrooms" 的服務角色傳遞給 AWS Clean Rooms 服務。列出角色、政策，並檢查與服務相關的 AWS Clean Rooms 服務角色和政策。
- `glue` – 從中擷取資料庫、資料表、分割區和結構描述的相關資訊 AWS Glue。這是 AWS Clean Rooms 服務顯示基礎資料來源並與之互動的必要項目。
- `logs` – 管理 CloudWatch Logs 的日誌交付、日誌群組和資源政策。查詢和擷取 AWS Clean Rooms 與服務相關的日誌。這些許可對於 服務內的監控、稽核和故障診斷目的而言是必要的。

此政策也會明確拒絕動

作 `cleanrooms:UpdateProtectedQuery` , `cleanrooms:StartProtectedQuery` 並防止使用者直接執行或更新受保護的 AWS Clean Rooms 查詢，這應該透過受控制的機制完成。

如需政策詳細資訊的 JSON 清單，請參閱《AWS 受管政策參考指南》中的 [AWSCleanRoomsFullAccessNoQuerying](#)。

AWS 受管政策：**AWSCleanRoomsMLReadOnlyAccess**

您可以 `AWSCleanRoomsMLReadOnlyAccess` 連接到 IAM 主體。

此政策授予協同 `AWSCleanRoomsMLReadOnlyAccess` 合作中資源和中繼資料的唯讀許可。

此政策包含以下許可：

- `CleanRoomsConsoleNavigation` – 授予檢視 AWS Clean Rooms 主控台畫面的存取權。
- `CleanRoomsMLRead` – 允許主體唯讀存取 Clean Rooms ML 服務。
- `PassCleanRoomsResources` – 授予傳遞指定 AWS Clean Rooms 資源的存取權。

如需政策詳細資訊的 JSON 清單，請參閱《AWS 受管政策參考指南》中的 [AWSCleanRoomsMLReadOnlyAccess](#)。

AWS 受管政策：**AWSCleanRoomsMLFullAccess**

您可以將 `AWSCleanRoomsMLFullAccess` 連接到您的 IAM 主體。此政策授予管理許可，允許完整存取（讀取、寫入和更新）Clean Rooms ML 所需的資源和中繼資料。

許可詳細資訊

此政策包含以下許可：

- `CleanRoomsMLFullAccess` – 授予所有 Clean Rooms ML 動作的存取權。
- `PassServiceRole` – 准許將服務角色傳遞給其名稱中具有 "cleanrooms-ml" 的服務 (`PassedToService` 條件)。
- `CleanRoomsConsoleNavigation` – 授予檢視 AWS Clean Rooms 主控台畫面的存取權。
- `CollaborationMembershipCheck` – 當您在協同合作中開始產生受眾（看起來像是客群）任務時，Clean Rooms ML 服務會呼叫 `ListMembers` 來檢查協同合作是否有效、發起人是作用中成員，而設定的受眾模型擁有者是作用中成員。此許可一律為必要；只有主控台使用者才需要主控台導覽 SID。

- `PassCleanRoomsResources` – 授予傳遞指定 AWS Clean Rooms 資源的存取權。
- `AssociateModels` – 允許主體將 Clean Rooms ML 模型與您的協同合作建立關聯。
- `TagAssociations` – 允許主體將標籤新增至類似模型與協同合作之間的關聯。
- `ListRolesToPickServiceRole` – 允許主體列出其所有角色，以便在使用時選擇服務角色 AWS Clean Rooms。
- `GetRoleAndListRolePoliciesToInspectServiceRole` – 允許主體在 IAM 中查看服務角色和對應的政策。
- `ListPoliciesToInspectServiceRolePolicy` – 允許主體在 IAM 中查看服務角色和對應的政策。
- `GetPolicyToInspectServiceRolePolicy` – 允許主體在 IAM 中查看服務角色和對應的政策。
- `ConsoleDisplayTables` : 允許主體唯讀存取所需的 AWS Glue 中繼資料，以顯示主控台上基礎 AWS Glue 資料表的資料。
- `ConsolePickOutputBucket` – 允許主體為設定的受眾模型輸出選取 Amazon S3 儲存貯體。
- `ConsolePickS3Location` – 允許主體選取儲存貯體中已設定對象模型輸出的位置。
- `ConsoleDescribeECRRepositories` – 允許主體描述 Amazon ECR 儲存庫和映像。

如需政策詳細資訊的 JSON 清單，請參閱《AWS 受管政策參考指南》中的 [AWSCleanRoomsMLFullAccess](#)。

AWS Clean Rooms AWS 受管政策的更新

檢視自此服務開始追蹤這些變更 AWS Clean Rooms 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 Clean Rooms 文件歷史記錄頁面上的 RSS 摘要。

變更	描述	Date
AWSCleanRoomsFullAccessNoQuering – 更新現有政策	已將 <code>cleanrooms : UpdateConfiguredTableAllowedColumns</code> 和 <code>cleanrooms : UpdateConfiguredTableReference</code> 新增至 <code>CleanRoomsAccess</code> 。	2025 年 7 月 29 日
AWSCleanRoomsMLReadOnlyAccess – 更新現有政策	已新增 <code>PassCleanRoomsResources</code> 到 <code>AWSCleanRoomsMLReadOnlyAccess</code> 。新增了 <code>PassCleanRoomsResources</code> 和 <code>ConsoleDe</code>	2025 年 1 月 10 日

變更	描述	Date
	scribeECRRepositories 至 AWSCleanRoomsMLFullAccess。	
AWSCleanRoomsFullAccessNoQuering – 更新現有政策	已新增 cleanrooms:BatchGetSchemaAnalysisRule 到 CleanRoomsAccess。	2024 年 5 月 13 日
AWSCleanRoomsFullAccess – 更新現有政策	已在此政策中將 SetQueryResultsBucket 中的陳述式 ID 更新 AWSCleanRoomsFullAccessConsolePickQueryResultsBucket 為 <code>SetQueryResultsBucket</code> ，以更好地代表許可，因為需要使用和不使用 主控台來設定查詢結果儲存貯體。	2024 年 3 月 21 日
AWSCleanRoomsMLReadOnlyAccess – 新政策	新增 AWSCleanRoomsMLReadOnlyAccess 和 AWSCleanRoomsMLFullAccess 以支援 AWS Clean Rooms ML。	2023 年 11 月 29 日
AWSCleanRoomsMLFullAccess – 新政策		
AWSCleanRoomsFullAccessNoQuering – 更新現有政策	新增 cleanrooms:CreateAnalysisTemplate、cleanrooms:GetAnalysisTemplate、cleanrooms:UpdateAnalysisTemplate cleanrooms:DeleteAnalysisTemplate、cleanrooms:ListAnalysisTemplates、cleanrooms:BatchGetCollaborationAnalysisTemplate、和 cleanrooms:ListCollaborationAnalysisTemplates 至 cleanrooms:GetCollaborationAnalysisTemplateCleanRoomsAccess，以啟用新的分析範本功能。	2023 年 7 月 31 日

變更	描述	Date
AWSCleanRoomsFullAccessNoQueringing – 更新現有政策	新增 <code>cleanrooms:ListTagsForResource</code> 、 <code>cleanrooms:UntagResource</code> 和 <code>cleanrooms:TagResource</code> 至 <code>CleanRoomsAccess</code> 以啟用資源標記。	2023 年 3 月 21 日
AWS Clean Rooms 開始追蹤變更	AWS Clean Rooms 已開始追蹤其 AWS 受管政策的變更。	2023 年 1 月 12 日

對 AWS Clean Rooms 身分和存取進行故障診斷

使用以下資訊來協助您診斷和修正使用 AWS Clean Rooms 和 IAM 時可能遇到的常見問題。

主題

- [我無權在 中執行動作 AWS Clean Rooms](#)
- [我未獲得執行 `iam:PassRole` 的授權](#)
- [我想要允許 以外的人員 AWS 帳戶 存取我的 AWS Clean Rooms 資源](#)

我無權在 中執行動作 AWS Clean Rooms

如果您收到錯誤，告知您未獲授權執行動作，您的政策必須更新，允許您執行動作。

下列範例錯誤會在 `mateojackson` IAM 使用者嘗試使用主控台檢視一個虛構 `my-example-widget` 資源的詳細資訊，但卻無虛構 `cleanrooms:GetWidget` 許可時發生。

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
cleanrooms:GetWidget on resource: my-example-widget
```

在此情況下，必須更新 Mateo 政策，允許他使用 `cleanrooms:GetWidget` 動作存取 `my-example-widget` 資源。

如果您需要協助，請聯絡您的 AWS 管理員。您的管理員提供您的簽署憑證。

我未獲得執行 `iam:PassRole` 的授權

如果您收到錯誤，告知您未獲授權執行 `iam:PassRole` 動作，您的政策必須更新，允許您將角色傳遞給 AWS Clean Rooms。

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 Clean Rooms 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 以外的人員 AWS 帳戶 存取我的 AWS Clean Rooms 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。

如需進一步了解，請參閱以下內容：

- 若要了解 是否 AWS Clean Rooms 支援這些功能，請參閱 [AWS Clean Rooms 如何使用 IAM](#)。
- 若要了解如何 AWS 帳戶 在您擁有的 資源之間提供存取權，請參閱《[IAM 使用者指南](#)》中的 [在您擁有 AWS 帳戶 的另一個 中為 IAM 使用者提供存取權](#)。
- 若要了解如何將資源的存取權提供給第三方 AWS 帳戶，請參閱《[IAM 使用者指南](#)》中的 [將存取權提供給第三方 AWS 帳戶 擁有](#)。
- 如需了解如何透過聯合身分提供存取權，請參閱 IAM 使用者指南中的 [將存取權提供給在外部進行身分驗證的使用者 \(聯合身分\)](#)。
- 若要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 角色與資源型政策的差異](#)。

預防跨服務混淆代理人

混淆代理人問題屬於安全性議題，其中沒有執行動作許可的實體可以強制具有更多許可的實體執行該動作。在 AWS 中，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](#)全域條件內容索引鍵，以限制將另一個服務提供給資源的許可 AWS Clean Rooms。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`。

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。在 AWS Clean Rooms，您也必須將與 `sts:ExternalId` 條件索引鍵進行比較。

的值 `aws:SourceArn` 必須設定為擔任角色成員資格的 ARN。

下列範例示範如何在 AWS Clean Rooms 中使用 `aws:SourceArn` 全域條件內容索引鍵，以防止混淆代理人問題。

Note

此範例政策適用於服務角色的信任政策，該角色 AWS Clean Rooms 使用來存取已設定資料表的資料和中繼資料。

`<query-runner-membership-id>` 的值需要設定為查詢執行器的成員 ID。

協同合作的所有成員都可以檢視設定的資料表中繼資料，因此每個成員資格 ARN 都必須包含在成員資格 ARNs 清單中。

Note

透過 AWS Clean Rooms 主控台建立服務角色時，協作的所有目前成員預設都會包含在混淆代理人條件中。

如果您要將新成員新增至已設定與其相關聯的資料表的協同合作，請務必使用新成員的成員 ARN 更新服務角色的混淆代理人條件。

如果您在新增新成員後未更新服務角色的混淆代理人條件，則該新成員將無法存取使用該角色擷取 AWS Clean Rooms 的資訊。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowIfExternalIdMatches",
```

```

        "Effect": "Allow",
        "Principal": {
            "Service": "cleanrooms.amazonaws.com"
        },
        "Action": "sts:AssumeRole",
        "Condition": {
            "StringLike": {
                "sts:ExternalId": "arn:aws:*:us-east-1*:dbuser:*/<query-
runner-membership-id>*"
            }
        }
    },
    {
        "Sid": "AllowIfSourceArnMatches",
        "Effect": "Allow",
        "Principal": {
            "Service": "cleanrooms.amazonaws.com"
        },
        "Action": "sts:AssumeRole",
        "Condition": {
            "ForAnyValue:ArnEquals": {
                "aws:SourceArn": [
                    "arn:aws:cleanrooms:us-
east-1:111122223333:membership/<member-1-membership-id>",
                    "arn:aws:cleanrooms:us-
east-1:444455556666:membership/<member-2-membership-id>"
                ]
            }
        }
    }
]
}

```

AWS Clean Rooms ML 的 IAM 行為

跨帳戶任務

Clean Rooms ML 允許一個使用者在其帳戶中建立的特定資源由另一個使用者 AWS 帳戶 安全存取 AWS 帳戶。當 A AWS 帳戶 中的用戶端 `StartAudienceGenerationJob` 呼叫 AWS 帳戶 B 擁有 `ConfiguredAudienceModel` 的資源時，Clean Rooms ML 會為任務建立兩個 ARNs。A 中的一個 ARN，AWS 帳戶 B AWS 帳戶 中的另一個 ARN。除了 ARNs 是相同的 AWS 帳戶。

Clean Rooms ML 會為任務建立兩個 ARNs，以確保兩個帳戶都可以將自己的 IAM 政策套用至任務。例如，兩個帳戶都可以使用標籤型存取控制，並從其 AWS 組織套用政策。任務處理來自兩個帳戶的資料，因此兩個帳戶都可以刪除任務及其相關聯的資料。兩個帳戶都無法封鎖其他帳戶刪除任務。

只有一個任務執行，兩個帳戶在呼叫時都可以看到任務 `ListAudienceGenerationJobs`。兩個帳戶都可以使用 ARN Get 搭配自己的 AWS 帳戶 ID 來呼叫任務上的 `Delete`、和 `Export` APIs。

使用 ARN 搭配其他 AWS 帳戶 ID 時，兩者都 AWS 帳戶 無法存取任務。

任務的名稱在 中必須是唯一的 AWS 帳戶。AWS 帳戶 B 中的名稱為 `$accountA-$name`。在 AWS 帳戶 B 中檢視任務時，AWS 帳戶 A 所選擇的名稱字 AWS 帳戶 首為 A。

為了 `StartAudienceGenerationJob` 讓跨帳戶成功，AWS 帳戶 B 必須使用類似下列範例的資源政策，在 AWS 帳戶 B 中的新任務和 AWS 帳戶 B `ConfiguredAudienceModel` 中的 上允許該動作：

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Clean-Rooms-CAMA-ID",
      "Effect": "Allow",
      "Principal": {
        "AWS": [
          ""111122223333""
        ]
      },
      "Action": [
        "cleanrooms-ml:StartAudienceGenerationJob"
      ],
      "Resource": [
        "arn:aws:cleanrooms-ml:us-east-1:444455556666:configured-audience-model/id",
        "arn:aws:cleanrooms-ml:us-east-1:444455556666:audience-generation-job/*"
      ],
      "Condition": {"StringEquals": {"cleanrooms-ml:CollaborationId": "UUID"}}
    }
  ]
}
```

 Note

此 AWS Clean Rooms ML 資源政策參考兩個不同的 AWS 帳戶 IDs 以支援跨帳戶受眾產生：

- 111122223333 - 這是包含授權開始產生受眾任務之委託人（使用者、角色或服務）的帳戶。此帳戶會啟動 ML 處理工作流程。
- 444455556666 - 這是擁有 AWS Clean Rooms ML 資源的帳戶（已設定的受眾模型和受眾產生任務）。此帳戶託管 ML 模型並管理任務執行。

其他組態備註：

- 陳述式 ID (Sid)：CAMA-ID 將 AWS Clean Rooms 取代為您實際的對象模型應用程式 (CAMA) 識別符，讓政策陳述式易於識別。
- 資源 IDs：將 *ID* 取代為您設定的對象模型的實際 ID，並將 *UUID* 取代為您的特定協同合作 ID。
- 條件：cleanrooms-ml:CollaborationId 條件可確保對象產生任務只能在指定的 AWS Clean Rooms 協同合作內容中啟動，提供額外的安全界限。

此跨帳戶組態可讓一個組織管理 ML 模型和基礎設施，同時允許授權合作夥伴在其協作協議範圍內啟動受眾產生程序的案例。

如果您使用 [AWS Clean Rooms ML API](#) 來建立 `manageResourcePolicies` 設定為 `true` 的已設定外觀模型，會為您 AWS Clean Rooms 建立此政策。

此外，AWS 帳戶 A 中呼叫者的身分政策需要的 `StartAudienceGenerationJob` 許可 `arn:aws:cleanrooms-ml:us-west-1:AccountA:audience-generation-job/*`。因此，有三種適用於動作的 IAM 資源 `StartAudienceGenerationJob`：AWS 帳戶 A 任務、AWS 帳戶 B 任務和 AWS 帳戶 B `ConfiguredAudienceModel`。

 Warning

AWS 帳戶 啟動任務的 會收到有關任務的 AWS CloudTrail 稽核日誌事件。AWS 帳戶 擁有的 `ConfiguredAudienceModel` 不會收到 AWS CloudTrail 稽核日誌事件。

標記任務

當您設定 `childResourceTagOnCreatePolicy=FROM_PARENT_RESOURCE` 參數時 `CreateConfiguredAudienceModel`，您帳戶中從該已設定的外觀模型建立的所有類似區段產生任務，預設會具有與已設定的外觀模型相同的標籤。設定的類似模型是父系，而類似區段的產生任務是子系。

如果您要在自己的帳戶中建立任務，任務的請求標籤會覆寫父標籤。

其他帳戶建立的任務絕不會在您的帳戶中建立標籤。如果您設

定，`childResourceTagOnCreatePolicy=FROM_PARENT_RESOURCE` 而另一個帳戶建立任務，則有兩個任務複本。您帳戶中的複本具有父資源標籤，而任務提交者帳戶中的複本具有來自請求的標籤。

驗證協作者

將許可授予 AWS Clean Rooms 協同合作的其他成員時，資源政策應包含條件索引鍵 `cleanrooms-ml:CollaborationId`。這會強制執行 `collaborationId` 參數包含在 [StartAudienceGenerationJob](#) 請求中。當 `collaborationId` 參數包含在請求中時，Clean Rooms ML 會驗證協同合作是否存在、任務提交者是協同合作的作用中成員，而已設定的外觀模型擁有者是協同合作的作用中成員。

當 AWS Clean Rooms 管理您設定的外觀模型資源政策 (`manageResourcePolicies` 參數位於 [CreateConfiguredAudienceModelAssociation](#) 請求 `TRUE` 中) 時，會在資源政策中設定此條件金鑰。因此，您必須在 [StartAudienceGenerationJob](#) `collaborationId` 中指定。

跨帳戶存取權

`StartAudienceGenerationJob` 只能跨帳戶呼叫。所有其他 Clean Rooms ML APIs 只能與您自己的帳戶中的資源搭配使用。這可確保您的訓練資料、類似模型組態和其他資訊保持私有。

Clean Rooms ML 永遠不會顯示 Amazon S3 或跨帳戶 AWS Glue 的位置。訓練資料位置、已設定的類似模型輸出位置，以及類似區段產生任務種子位置絕不會在帳戶間顯示。除非在協同合作中啟用查詢記錄，否則種子資料是否來自 SQL 查詢，且查詢本身不會在帳戶之間顯示。如果您是另一個帳戶提交的 `Get` 對象產生任務，服務不會顯示種子位置。

Clean Rooms ML 自訂模型的 IAM 行為

跨帳戶任務

Clean Rooms ML 允許與某個建立的協同合作相關聯的特定資源，AWS 帳戶由另一個在其帳戶中安全地存取 AWS 帳戶。具有成員執行查詢能力的 AWS 帳戶 A 中的用戶端可以在協同合作中另一個成員

擁有 `ConfiguredModelAlgorithmAssociation` 的資源 `StartTrainedModelInferenceJob` 上呼叫 `CreateTrainedModelCreateMLInputChannel`、或，前提是使用建立的自訂分析規則 `ConfiguredModelAlgorithmAssociation` 允許 `CreateConfiguredTableAnalysisRule`。

此外，協同合作的任何作用中成員都可以透過 `DeleteTrainedModelOutput` 和 `DeleteMLInputChannelData` APIs 刪除與訓練模型或 ML 輸入通道相關聯的資料。

跨帳戶存取權

Clean Rooms ML 允許使用者擷取其他帳戶透過 `GetCollaboration` 和 `ListCollaboration` APIs 中繼資料。Clean Rooms ML 不會向其他帳戶顯示 KMS 金鑰 ARNs、標籤、環境變數或超參數（適用於 `TrainedModel` 動作）。

成員資格和協同合作存取權

在 Clean Rooms ML 自訂模型的內容中存取成員資格和協同合作資源時，使用者的身分政策需要動作 `cleanrooms:PassMembership`、`cleanrooms:PassCollaboration` 或兩者的許可。接受的所有 `APIs` `membershipId` 都需要 `cleanrooms:PassMembership` 許可，接受的所有 `APIs` `collaborationId` 都需要 `cleanrooms:PassCollaboration` 許可。提供可在成員 ID 內容 `createTrainedModel` 中呼叫之角色的身分政策範例，可在協作 ID 內容 `GetCollaborationTrainedModel` 中呼叫。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowCleanroomsMLActions",
      "Effect": "Allow",
      "Action": [
        "cleanrooms:PassCollaboration",
        "cleanrooms:PassMembership"
      ],
      "Resource": [
        "*"
      ]
    },
    {
```



```

        "Sid": "AllowMembershipAccess",
        "Effect": "Allow",
        "Action": [
            "cleanrooms:GetMembership"
        ],
        "Resource": [
            "arn:aws:cleanrooms:us-east-1:111122223333:membership/memberId"
        ]
    },
    {
        "Sid": "AllowCollaborationAccess",
        "Effect": "Allow",
        "Action": [
            "cleanrooms:GetCollaboration"
        ],
        "Resource": [
            "arn:aws:cleanrooms:us-east-1:111122223333:collaboration/collaborationId"
        ]
    }
]
}

```

的合規驗證 AWS Clean Rooms

若要了解 是否 AWS 服務 在特定合規計劃範圍內，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[在 中下載報告 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](#)。

中的彈性 AWS Clean Rooms

AWS 全球基礎設施是以 AWS 區域和可用區域為基礎建置。區域提供多個分開且隔離的實際可用區域，並以低延遲、高輸送量和高度備援網路連線相互連結。透過可用區域，您可以設計與操作的應用程式和資料庫，在可用區域之間自動容錯移轉而不會發生中斷。可用區域的可用性、容錯能力和擴展能力，均較單一或多個資料中心的傳統基礎設施還高。

如需 AWS 區域和可用區域的詳細資訊，請參閱 [AWS 全球基礎設施](#)。

中的基礎設施安全 AWS Clean Rooms

作為受管服務，AWS Clean Rooms 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 Clean Rooms 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。

網路安全

在查詢執行期間從 S3 儲存貯體 AWS Clean Rooms 讀取時，和 Amazon S3 之間的 AWS Clean Rooms 流量會透過 AWS 私有網路安全地路由。傳輸中的流量是使用 Amazon Signature 第 4 版通訊協定 (SIGv4) 來簽署，並使用 HTTPS 來加密。此流量是根據您已為設定資料表設定的 IAM 服務角色進行授權。

您可以透過 端點以程式設計方式連線至 AWS Clean Rooms。如需服務端點的清單，請參閱 中的[AWS Clean Rooms 端點和配額](#)AWS 一般參考。

所有服務端點都僅限 HTTPS。您可以使用 Amazon Virtual Private Cloud (VPC) 端點，以防您想要 AWS Clean Rooms 從 VPC 連線至，而且不想有網際網路連線。如需詳細資訊，請參閱《AWS PrivateLink 指南》中的[透過 存取 AWS 服務 AWS PrivateLink](#)。

您可以將 IAM 政策指派給使用 [aws : SourceVpce 內容金鑰](#)的 IAM 主體，以限制您的 IAM 主體只能 AWS Clean Rooms 透過 VPC 端點呼叫，而不是透過網際網路呼叫。

使用界面端點 (AWS PrivateLink) 存取 AWS Clean Rooms 或 AWS Clean Rooms ML

您可以使用 在虛擬私有雲端 (VPC) 和 AWS Clean Rooms 或 AWS Clean Rooms ML 之間 AWS PrivateLink 建立私有連線。您可以像在 VPC 中一樣存取 AWS Clean Rooms 或 AWS Clean Rooms

ML，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 Clean Rooms。

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 Clean Rooms 之流量的進入點。

如需詳細資訊，請參閱《AWS PrivateLink 指南》中的「[透過 AWS PrivateLink 存取 AWS 服務](#)」。

的考量事項 AWS Clean Rooms

在您設定介面端點之前 AWS Clean Rooms，請檢閱《AWS PrivateLink 指南》中的[考量事項](#)。

AWS Clean Rooms 和 AWS Clean Rooms ML 支援透過介面端點呼叫其所有 API 動作。

AWS Clean Rooms 或 AWS Clean Rooms ML 不支援 VPC 端點政策。根據預設，允許透過介面端點完整存取 AWS Clean Rooms 和 AWS Clean Rooms ML。或者，您可以將安全群組與端點網路介面建立關聯，以控制透過介面端點傳入 AWS Clean Rooms 或 AWS Clean Rooms ML 的流量。

建立的介面端點 AWS Clean Rooms

您可以使用 Amazon VPC 主控台 AWS Clean Rooms 或 AWS Command Line Interface () 為或 AWS Clean Rooms ML 建立介面端點 AWS CLI。如需詳細資訊，請參閱《AWS PrivateLink 指南》中的「[建立介面端點](#)」。

AWS Clean Rooms 使用下列服務名稱建立的介面端點。

```
com.amazonaws.region.cleanrooms
```

使用下列服務名稱建立 AWS Clean Rooms ML 的介面端點。

```
com.amazonaws.region.cleanrooms-ml
```

如果您為介面端點啟用私有 DNS，您可以使用 AWS Clean Rooms 其預設的區域 DNS 名稱向提出 API 請求。例如 `cleanrooms-ml.us-east-1.amazonaws.com`。

監控 AWS Clean Rooms

監控是維護 AWS Clean Rooms 和其他 AWS 解決方案的可靠性、可用性和效能的重要部分。AWS 提供下列監控工具，讓您監看 AWS Clean Rooms、回報錯誤，並適時採取自動動作：

- Amazon CloudWatch Logs 可讓您從 Amazon EC2 執行個體和其他來源監控 AWS CloudTrail、存放和存取您的日誌檔案。Amazon CloudWatch Logs 可監控日誌檔案中的資訊，並在達到特定閾值時通知您。您也可以將日誌資料存檔在高耐用性的儲存空間。如需詳細資訊，請參閱 [Amazon CloudWatch Logs 使用者指南](#)。

Clean Rooms ML 允許特定 API 動作的跨帳戶任務。AWS 帳戶 啟動任務的 會收到任務的 AWS CloudTrail 稽核日誌事件。如需詳細資訊，請參閱 [AWS Clean Rooms ML 的 IAM 行為](#)

- AWS CloudTrail 會擷取由 發出或代表發出的 API 呼叫和相關事件，AWS 帳戶 並將日誌檔案交付至您指定的 Amazon S3 儲存貯體。您可以識別呼叫的使用者和帳戶 AWS、進行呼叫的來源 IP 地址，以及呼叫的時間。如需詳細資訊，請參閱 [「AWS CloudTrail 使用者指南」](#)。
- Amazon EventBridge 為無伺服器事件匯流排服務，可讓您輕鬆將應用程式與來自各種來源的資料互相連線。EventBridge 可從自己的應用程式、軟體即服務 (SaaS) 應用程式和 AWS 服務提供即時資料串流，然後將該資料路由到 Lambda 等目標。這可讓您監控在服務中發生的事件，並建置事件導向的架構。如需詳細資訊，請參閱 [Amazon EventBridge 使用者指南](#) 和 [Amazon EventBridge Events 參考](#)。
- AWS 帳單與成本管理 是一項 服務，提供可協助您支付帳單和最佳化成本的功能。Amazon Web Services 會根據用量向您的帳戶收費，確保您只需為使用量付費。AWS 也可讓您將 [成本分配標籤](#) 套用至資源，以追蹤和分類您的 AWS 成本。例如，在 中 AWS Clean Rooms，您可以將標籤套用至協同合作成員資格，以追蹤每個協同合作的成本。如需詳細資訊，請參閱 [「AWS Billing 使用者指南」](#)。

分析登入 AWS Clean Rooms

分析記錄是 中的功能 AWS Clean Rooms。當您 [建立協同合作](#) 並開啟分析記錄時，成員可以在 Amazon CloudWatch Logs 中儲存來自查詢或任務日誌的相關日誌。

透過查詢日誌和任務日誌，成員可以判斷查詢是否符合分析規則，並符合協同合作協議。此外，查詢日誌有助於支援稽核。

在 AWS Clean Rooms 主控台中開啟分析記錄選項時，查詢日誌會包含下列項目：

- analysisRule – 已設定資料表的分析規則。

- `analysisTemplateArn` – 執行的分析範本（視分析規則而定）。
- `collaborationId` – 執行查詢之協同合作的唯一識別符。
- `configuredTableID` – 查詢中參考之已設定資料表的唯一識別符。
- `directQueryAnalysisRulePolicy.custom.allowedAnalysis` – 允許在設定的資料表上執行的分析範本（視分析規則而定）。
- `directQueryAnalysisRulePolicy.v1.custom.allowedAnalysisProviders` – 允許建立查詢的查詢提供者（視分析規則而定）。
- `errorCode` – 查詢無法正常執行時的錯誤代碼。
- `errorMessage` – 查詢無法正確執行時的錯誤訊息。
- `eventID` – 查詢執行的唯一識別符。2023 年 8 月 31 日之後，唯一識別符與相同 `protectedQueryID`。
- `eventTimestamp` – 查詢執行時間。
- `parameters.parametervalue` – 參數值（視查詢文字而定）。
- `queryText` – 查詢執行的 SQL 定義。如果有參數，它們會標記為 `:parametervalue`。
- `queryValidationErrors` – 查詢驗證時的查詢錯誤。
- `schemaName` – 查詢中參考的已設定資料表關聯名稱。
- `status` – 查詢的執行狀態。

接收查詢和任務日誌

您不需要在 之外執行任何動作 AWS Clean Rooms 來設定查詢日誌和任務日誌。會在每個協同合作成員 AWS Clean Rooms 建立 [成員](#) 資格後建立協同合作的日誌群組。

可以查詢的成員、可以執行查詢和任務的成員、可以接收結果的成員，以及在查詢中參考其組態表的成員，都會收到查詢日誌或任務日誌。

可以查詢的成員和可以接收結果的成員將收到查詢中參考的每個已設定資料表的查詢日誌。如果他們不擁有設定的資料表，則無法檢視設定的資料表 ID (`configuredTableID`)。

可以執行查詢和任務的成員，以及可以接收結果的成員，將會收到任務中參考之每個已設定資料表的任務日誌。如果他們不擁有設定的資料表，則無法檢視設定的資料表 ID (`configuredTableID`)。

如果成員在查詢中參考了多個已設定的資料表關聯，他們將收到每個已設定資料表的查詢日誌。

如果成員在任務中參考多個已設定的資料表關聯，他們將收到每個已設定資料表的任務日誌。

針對包含不支援和支援 SQL 的查詢建立日誌 AWS Clean Rooms。如需詳細資訊，請參閱 [AWS Clean Rooms SQL 參考](#)。

當查詢或任務參考未與協同合作相關聯的已設定資料表時，也會建立日誌。

日誌可能包含有關不正確 SQL 的資訊。

查詢和任務日誌會指出查詢的狀態，但不報告是否傳送查詢輸出。他們確認查詢或任務是由可查詢的成員提交。查詢日誌也會確認查詢包含中支援的 SQL，AWS Clean Rooms 並參考與協同合作相關聯的已設定資料表。

Example

例如，如果查詢在 AWS Clean Rooms 驗證其是否符合分析規則之後以及在查詢處理期間取消，則不會產生日誌。

如果您刪除日誌群組，則必須使用相同的日誌群組名稱（協同合作的協同合作 ID）手動重新建立日誌群組。或者，您可以在您的成員資格中關閉和開啟記錄。

如需如何開啟分析記錄的詳細資訊，請參閱 [建立協同合作](#)。

如需 Amazon CloudWatch Logs 的詳細資訊，請參閱《[Amazon CloudWatch Logs 使用者指南](#)》。

查詢和任務日誌的建議動作

我們建議成員定期採取下列動作：

- 若要驗證查詢和任務是否符合協同合作同意的使用案例或查詢，請檢閱協同合作中執行的查詢和任務。

如需如何檢視最近查詢的詳細資訊，請參閱 [檢視最近查詢](#)。

如需如何檢視最近任務的詳細資訊，請參閱 [檢視最近的任務](#)。

- 若要驗證設定的資料表資料欄是否符合協同合作的約定，請檢閱協同合作成員分析規則和查詢中使用的已設定資料表資料欄。

如需如何檢視已設定資料欄的詳細資訊，請參閱 [檢視資料表和分析規則](#)。

在 中使用 CloudWatch 進行詳細監控AWS Clean Rooms

詳細監控是 中的選擇加入功能AWS Clean Rooms。當您[建立協同合作](#)並開啟詳細監控時，可執行查詢和設定付款人的成員可以選擇在其 CloudWatch 帳戶中接收詳細的可觀測性指標。

透過詳細監控，成員可以監控查詢效能、追蹤資源使用率，以及建立儀表板以取得營運洞見。這些指標有助於容量規劃、成本分析和疑難排解效能問題。

Note

指標可用性會根據設定的資料表資料來源而有所不同：

- 對於參考以 Athena 做為資料來源的已設定資料表的分析，BytesRead和 QueryFileScanned指標一律會傳回 0。
- 對於參考以 Snowflake 做為資料來源的已設定資料表的分析，BytesRead、QueryFileScanned和 RecordsRead指標一律會傳回 0。
- 對於參考具有不同資料來源（例如 Snowflake 和 S3-backedAWS Glue資料表）的多個已設定資料表，BytesRead以及僅針對 S3-backedAWS Glue資料表QueryFileScanned傳回指標的分析。會RecordsRead傳回 S3-backedAWS Glue資料表和 Athena 資料表的指標。

指標類型

AWS Clean Rooms會將兩類指標發佈至 CloudWatch：

彙總詳細監控指標

這些指標提供跨分析的彙總層級可見性：

AnalysisRuntime

以毫秒為單位顯示分析的總持續時間。

CRPUConsumed

用於特定分析的 CRPU（乾淨房間處理單元）。如需 CRPU 計算的詳細資訊，請參閱 [AWS Clean Rooms定價](#)。

ConcurrentVCPUs

在指定時間點使用的並行 vCPU（與服務配額相關）。

AnalysesCount

指定維度在一段時間內執行的分析數目。

ConcurrentQueries

指定時間點的並行查詢計數（與服務配額相關）。

查詢層級的詳細監控指標

這些指標提供個別查詢執行的詳細洞見：

BytesRead

從來源資料表讀取以進行分析的資料總數（以位元組為單位）。

BytesWritten

為分析寫入的總資料（以位元組為單位）。

RecordsRead

從來源資料表讀取以供分析的記錄總數。

RecordsWritten

分析結果中傳回的記錄總數。

QueryFileScanned

為分析掃描的資料檔案總數。

Memory utilization

用於分析的記憶體資源。

vCPU utilization

用於分析的 vCPU 資源。

Disk utilization

用於分析的儲存資源。

指標維度

指標依下列維度組織：

- 分析狀態：已完成
- 分析類型：SparkSQL
- 成員資格：成員資格 ID

- 查詢：個別查詢 ID（適用於查詢層級指標）
- 帳戶/區域：您的 AWS 帳戶和 區域

所有指標都會發佈到 CloudWatch 中的 AWS/CleanRooms 命名空間。

誰可以存取指標

詳細的監控指標可用於：

- 查詢執行器：可在協同合作中執行查詢的成員
- 設定的付款人：被指派支付查詢執行成本的成員

僅提供資料的成員無法存取詳細的監控指標。

定價

AWS Clean Rooms 不收取發佈指標至 CloudWatch 的費用。不過，標準 CloudWatch 費用適用於指標儲存體、儀表板用量和警示。如需詳細資訊，請參閱 [Amazon CloudWatch 定價](#)。

使用 記錄 AWS Clean Rooms API 呼叫 AWS CloudTrail

AWS Clean Rooms 已與 整合 AWS CloudTrail，此服務提供使用者、角色或 AWS 服務 所採取動作的記錄 AWS Clean Rooms。CloudTrail 會將 的所有 API 呼叫擷取 AWS Clean Rooms 為事件。擷取的呼叫包括來自 AWS Clean Rooms 主控台的呼叫，以及對 AWS Clean Rooms 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 Clean Rooms。即使您未設定追蹤，依然可以透過 CloudTrail 主控台內的 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 Clean Rooms 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》](#)。

AWS Clean Rooms CloudTrail 中的資訊

當您建立帳戶 AWS 帳戶 時，您的 上會啟用 CloudTrail。當活動在 中發生時 AWS Clean Rooms，該活動會與事件歷史記錄中的其他 AWS 服務 事件一起記錄在 CloudTrail 事件中。您可以檢視、搜尋和下載 AWS 帳戶的最新事件。如需詳細資訊，請參閱 [「使用 CloudTrail 事件歷史記錄檢視事件」](#)。

若要持續記錄 中的事件 AWS 帳戶，包括 的事件 AWS Clean Rooms，請建立追蹤。線索能讓 CloudTrail 將日誌檔案交付至 Amazon S3 儲存貯體。依預設，當您在主控台中建立追蹤時，該追蹤會套用至所有的 AWS 區域。線索會記錄 AWS 分割區中所有區域的事件，並將日誌檔案交付至您指定的 Amazon S3 儲存貯體。此外，您可以設定其他 AWS 服務 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](#)
- [從多個帳戶接收 CloudTrail 日誌檔案](#)

CloudTrail 會記錄所有 AWS Clean Rooms 動作，並記錄在 [AWS Clean Rooms API 參考](#) 中。

每一筆事件或日誌專案都會包含產生請求者的資訊。身分資訊可協助您判斷下列事項：

- 是否使用根使用者或 IAM 使用者憑證提出該請求。
- 提出該請求時，是否使用了特定角色或聯合身分使用者的暫時安全憑證。
- 該請求是否由另一項 AWS 服務服務提出。

如需詳細資訊，請參閱 [CloudTrail userIdentity 元素](#)。

了解 AWS Clean Rooms 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。

Example AWS Clean Rooms CloudTrail 事件

下列範例示範 CloudTrail 事件：

主題

- [StartProtectedQuery \(成功\)](#)
- [StartProtectedQuery \(失敗\)](#)

StartProtectedQuery (成功)

```
{
  "eventVersion": "1.08",
  "userIdentity": {
    "type": "AssumedRole",
    "principalId": "EXAMPLE_PRINCIPAL_ID",
    "arn": "arn:aws:sts::123456789012:assumed-role/query-runner/jdoe",
    "accountId": "123456789012",
    "accessKeyId": "EXAMPLE_KEY_ID",
    "sessionContext": {
      "sessionIssuer": {
        "type": "Role",
        "principalId": "EXAMPLE_PRINCIPAL_ID",
        "arn": "arn:aws:iam::123456789012:role/query-runner",
        "accountId": "123456789012",
        "userName": "query-runner"
      },
      "webIdFederationData": {},
      "attributes": {
        "creationDate": "2023-04-07T19:34:32Z",
        "mfaAuthenticated": "false"
      }
    }
  },
  "eventTime": "2023-04-07T19:53:32Z",
  "eventSource": "cleanrooms.amazonaws.com",
  "eventName": "StartProtectedQuery",
  "awsRegion": "us-east-2",
  "sourceIPAddress": "203.0.113.1",
  "userAgent": "aws-internal/3",
  "requestParameters": {
    "resultConfiguration": {
      "outputConfiguration": {
        "s3": {
          "resultFormat": "CSV",
          "bucket": "cleanrooms-queryresults-jdoe-test",
          "keyPrefix": "test"
        }
      }
    }
  },
  "sqlParameters": "****",
  "membershipIdentifier": "a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
}
```

```

        "type": "SQL"
      },
      "responseElements": {
        "Access-Control-Expose-Headers": "x-amzn-RequestId,x-amzn-ErrorType,x-amzn-ErrorMessage,Date",
        "protectedQuery": {
          "createTime": 1680897212.279,
          "id": "f5988bf1-771a-4141-82a8-26fcc4e41c9f",
          "membershipArn": "arn:aws:cleanrooms:us-east-2:123456789012:membership/a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
          "membershipId": "a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
          "resultConfiguration": {
            "outputConfiguration": {
              "s3": {
                "bucket": "cleanrooms-queryresults-jdoe-test",
                "keyPrefix": "test",
                "resultFormat": "CSV"
              }
            }
          },
          "sqlParameters": "****",
          "status": "SUBMITTED"
        }
      },
      "requestID": "7464211b-2277-4b55-9723-fb4f259aefd2",
      "eventID": "f7610f5e-74b9-420f-ae43-206571ebcbf7",
      "readOnly": false,
      "eventType": "AwsApiCall",
      "managementEvent": true,
      "recipientAccountId": "123456789012",
      "eventCategory": "Management"
    }
  }
}

```

StartProtectedQuery (失敗)

```

{
  "eventVersion": "1.08",
  "userIdentity": {
    "type": "AssumedRole",
    "principalId": "EXAMPLE_PRINCIPAL_ID",
    "arn": "arn:aws:sts::123456789012:assumed-role/query-runner/jdoe",
    "accountId": "123456789012",
    "accessKeyId": "EXAMPLE_KEY_ID",

```

```

    "sessionContext": {
      "sessionIssuer": {
        "type": "Role",
        "principalId": "EXAMPLE_PRINCIPAL_ID",
        "arn": "arn:aws:iam::123456789012:role/query-runner",
        "accountId": "123456789012",
        "userName": "query-runner"
      },
      "webIdFederationData": {},
      "attributes": {
        "creationDate": "2023-04-07T19:34:32Z",
        "mfaAuthenticated": "false"
      }
    },
    "eventTime": "2023-04-07T19:47:27Z",
    "eventSource": "cleanrooms.amazonaws.com",
    "eventName": "StartProtectedQuery",
    "awsRegion": "us-east-2",
    "sourceIPAddress": "203.0.113.1",
    "userAgent": "aws-internal/3",
    "errorCode": "ValidationException",
    "requestParameters": {
      "resultConfiguration": {
        "outputConfiguration": {
          "s3": {
            "resultFormat": "CSV",
            "bucket": "cleanrooms-queryresults-jdoe-test",
            "keyPrefix": "test"
          }
        }
      },
      "sqlParameters": "****",
      "membershipIdentifier": "a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
      "type": "SQL"
    },
    "responseElements": {
      "Access-Control-Expose-Headers": "x-amzn-RequestId,x-amzn-ErrorType,x-amzn-ErrorMessage,Date",
      "message": "Column(s) [identifier] is not allowed in select"
    },
    "requestID": "e29f9f74-8299-4a83-9d18-5ddce7302f07",
    "eventID": "c8ee3498-8e4e-44b5-87e4-ab9477e56eb5",
    "readOnly": false,

```

```
"eventType": "AwsApiCall",  
"managementEvent": true,  
"recipientAccountId": "123456789012",  
"eventCategory": "Management"  
}
```

使用 Amazon EventBridge 整合 AWS Clean Rooms 到事件驅動型應用程式

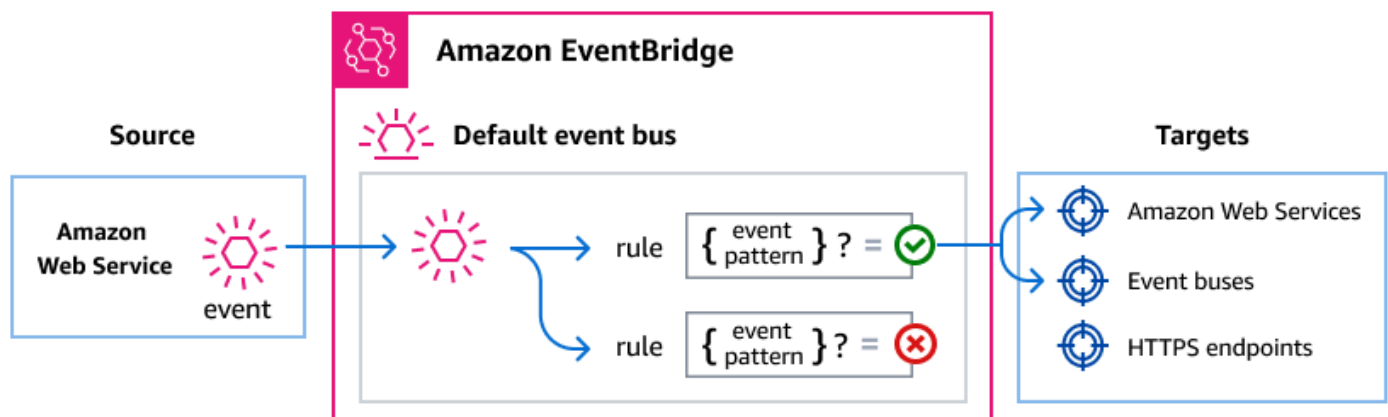
您可以將 AWS Clean Rooms 納入使用中發生的事件的事件驅動應用程式 (EDAs)，AWS Clean Rooms 可在應用程式元件之間通訊並啟動下游程序。您可以使用 Amazon EventBridge 將事件從路由 AWS Clean Rooms 到其他軟體元件來執行此操作。Amazon EventBridge 是一種無伺服器服務，使用事件將應用程式元件連接在一起，讓您更輕鬆地將等 AWS Clean Rooms 服務整合到事件驅動型架構中，而無需額外的程式碼和操作。

事件驅動架構是一種建置鬆耦合軟體系統的方式，透過發出和回應事件來協作。在此模型中，事件代表資源或環境中的變更。

以下是 EventBridge 的使用方式 AWS Clean Rooms：

如同許多 AWS 服務，AWS Clean Rooms 會產生事件並將其傳送至 EventBridge 預設事件匯流排。事件匯流排是接收事件並將其路由至您指定目的地或目標的路由器。目標可以包含其他服務 AWS、自訂應用程式和 SaaS 合作夥伴應用程式。

EventBridge 會根據您在事件匯流排上建立的規則路由事件。對於每個規則，您可以指定篩選條件或事件模式，以僅選取您想要的事件。每當事件傳送到事件匯流排時，EventBridge 都會將其與每個規則進行比較。如果事件符合規則，EventBridge 會將事件路由到指定的目標 (s)。



例如，假設您每次在帳戶中建立新的 AWS Clean Rooms 協同合作時，都想知道。您可以在預設事件匯流排上建立規則。在規則中，您會建立事件模式，從名為的 AWS Clean Rooms 指定事件 **Collaboration Created**。每次 EventBridge 收到符合這些屬性的事件時，都會將該事件路由到指定的工作流程。

AWS Clean Rooms 事件

AWS 服務可以直接將事件傳送到 EventBridge 預設事件匯流排。此外，也會 AWS CloudTrail 將來自許多 AWS 服務的事件傳送至 EventBridge。這些事件可能包括 API 呼叫、主控台登入和動作、服務事件和 CloudTrail Insights。如需詳細資訊，請參閱《EventBridge 使用者指南》中的 [AWS 透過傳遞的服務事件 AWS CloudTrail](#)。

如需傳送至 EventBridge 的完整 AWS Clean Rooms 事件清單，請參閱 [EventBridge Events Reference](#) 中的 AWS Clean Rooms 主題。

事件詳細資訊類型	Description
已建立分析範本	建立分析範本時，會通知分析範本擁有者和協同合作的所有作用中成員。
已更新分析範本	更新分析範本時，會通知分析範本擁有者和所有可查看更新之協同合作的作用中成員。
已刪除分析範本	刪除分析範本時，會通知分析範本擁有者和協同合作的所有作用中成員。
協同合作已建立	協同合作建立時，協同合作擁有者會收到通知。
協同合作已更新	協同合作擁有者和具有更新可見性的所有協同合作用中成員會在協同合作更新時收到通知。
協同合作變更請求已建立	建立協同合作變更請求時，協同合作擁有者和協同合作的所有作用中成員都會收到通知。
協作變更請求已核准	協同合作變更請求已核准時，協同合作擁有者和所有協同合作的作用中成員都會收到通知。
協作變更請求已取消	取消協同合作變更請求時，協同合作擁有者和協同合作的所有作用中成員都會收到通知。

事件詳細資訊類型	Description
已遞交協同合作變更請求	當遞交協同合作變更請求時，協同合作擁有者和協同合作的所有作用中成員都會收到通知。
已建立設定的資料表關聯	建立已設定的資料表關聯時，系統會通知已設定的資料表關聯擁有者和協同合作的所有作用中成員。
已更新設定的資料表關聯	更新已設定資料表關聯時，系統會通知已設定資料表關聯擁有者和具有更新可見性的所有協同合作用中成員。
已刪除設定的資料表關聯	刪除已設定的資料表關聯時，系統會通知已設定的資料表關聯擁有者和協同合作的所有作用中成員。
已建立設定的資料表關聯分析規則	建立已設定的資料表關聯分析規則時，系統會通知已設定的資料表關聯分析規則擁有者和協同合作的所有作用中成員。
已更新設定的資料表關聯分析規則	已設定資料表關聯分析規則擁有者和所有可查看更新之協同合作的作用中成員，會在已設定資料表關聯分析規則更新時收到通知。
已刪除設定的資料表關聯分析規則	刪除已設定的資料表關聯分析規則時，系統會通知已設定的資料表關聯分析規則擁有者和協同合作的所有作用中成員。
已建立 ID 映射表	建立 ID 映射表時，會通知 ID 映射表擁有者和協同合作的所有作用中成員。
已更新 ID 映射表	更新 ID 映射表時，將會通知 ID 映射表擁有者和具有更新可見性的所有協同合作用中成員。
已刪除 ID 映射表	刪除 ID 映射表時，會通知 ID 映射表擁有者和協同合作的所有作用中成員。
已建立 ID 命名空間關聯	建立 ID 命名空間關聯時，會通知 ID 命名空間關聯擁有者和協同合作的所有作用中成員。
ID 命名空間關聯已更新	更新 ID 命名空間關聯時，將會通知 ID 命名空間關聯擁有者和具有更新可見性的所有協同合作用中成員。

事件詳細資訊類型	Description
ID 命名空間關聯已刪除	刪除 ID 命名空間關聯時，會通知 ID 命名空間關聯擁有者和協同合作的所有作用中成員。
邀請加入協同合作	當受邀成員受邀參與協同合作時，會收到通知。
成員資格已建立	建立成員資格時，會通知成員擁有者和協同合作的所有作用中成員。
成員資格已更新	成員資格更新時會通知成員，除非成員資格已從協同合作中移除，在這種情況下，協同合作的所有作用中成員都會收到通知。
已刪除成員資格	刪除成員資格時，會通知成員擁有者和協同合作的所有作用中成員。
已提交受保護的任務	受保護任務的任務執行器、任務付款人和結果接收者會在提交受保護任務時收到通知。
受保護的任務已啟動	受保護任務的任務執行器、任務付款人和結果接收者會在受保護任務啟動時收到通知。
受保護的任務取消	當受保護任務取消時，系統會通知受保護任務的任務執行器、任務付款人和結果接收者。
已取消受保護的任務	受保護任務的任務執行器、任務付款人和結果接收者會在受保護任務取消時收到通知。
受保護任務成功	當受保護任務成功時，系統會通知受保護任務的任務執行器、任務付款人和結果接收者。
受保護任務失敗	受保護任務的任務執行器、任務付款人和結果接收者會在受保護任務失敗時收到通知。
已提交受保護的查詢	提交受保護查詢時，會通知受保護查詢的查詢執行器、查詢付款人和結果接收者（們）。
已啟動受保護的查詢	受保護查詢的查詢執行器、查詢付款人和結果接收者 (Informed Query) 會在受保護查詢啟動時收到通知。

事件詳細資訊類型	Description
受保護的查詢取消	當受保護查詢取消時，會通知受保護查詢的查詢執行器、查詢付款人和結果接收者（們）。
已取消受保護的查詢	受保護查詢的查詢執行器、查詢付款人和結果接收者（含）會在受保護查詢取消時收到通知。
受保護查詢成功	當受保護查詢成功時，會通知受保護查詢的查詢執行器、查詢付款人和結果接收者（們）。
受保護的查詢失敗	當受保護查詢失敗時，會通知受保護查詢的查詢執行器、查詢付款人和結果接收者（們）。
受保護的查詢逾時	當受保護查詢逾時時，會通知受保護查詢的查詢執行器、查詢付款人和結果接收者（們）。

使用 EventBridge 路由 AWS Clean Rooms 事件

若要讓 EventBridge 將 AWS Clean Rooms 事件路由到目標，您必須建立規則。每個規則都包含事件模式，EventBridge 會比對事件匯流排上收到的每個事件。如果事件資料符合指定的事件模式，EventBridge 會將該事件路由到規則的目標(s)。

如需建立事件匯流排規則的完整說明，請參閱EventBridge 使用者指南》中的[建立對事件做出反應的規則](#)。

建立符合事件 AWS Clean Rooms 的事件模式

每個事件模式都是 JSON 物件，它包含：

- （選用）：識別服務傳送事件的source屬性。對於 AWS Clean Rooms 事件，來源為 `aws.cleanrooms`。
- （選用）：包含要比對之事件名稱陣列的detail-type屬性。
- (選擇性)：包含要比對的任何其他事件資料的 detail 屬性。

例如，下列事件模式符合所有 ID 命名空間關聯 已更新協同合作從中刪除的事件 AWS Clean Rooms：

```
{
```

```
"source": ["aws.cleanrooms"],
"detail-type": ["Id Namespace Association Updated"],
"detail": {
  "status": ["COLLABORATION_DELETED"]
}
}
```

如需撰寫事件模式的詳細資訊，請參閱《[EventBridge 使用者指南](#)》中的[事件模式](#)。EventBridge

AWS Clean Rooms 事件詳細資訊參考

AWS 服務中的所有事件都有一組通用的欄位，其中包含事件的中繼資料，例如事件來源 AWS 的服務、事件產生的時間、事件發生的帳戶和區域，以及其他。如需這些一般欄位的定義，請參閱《Amazon EventBridge Events Reference》中的事件[結構](#)。

此外，每個事件都有一個 detail 欄位，其中包含該特定事件的特定資料。以下參考定義了各種 AWS Clean Rooms 事件的詳細資訊欄位。

使用 EventBridge 來選取和管理 AWS Clean Rooms 事件時，請謹記下列事項：

- 來自的所有事件的 source 欄位 AWS Clean Rooms 都設定為 aws.cleanrooms。
- detail-type 欄位指定事件類型。

例如 Collaboration Created。

- detail 欄位包含該特定事件的特定資料。

如需建構讓規則符合 AWS Clean Rooms 事件的事件模式的相關資訊，請參閱《Amazon EventBridge 使用者指南》中的[事件模式](#)。

分析範本建立的事件

以下是Analysis Template Created事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Analysis Template Created"]
}
```

分析範本更新事件

以下是Analysis Template Updated事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Analysis Template Updated"]
}
```

分析範本已刪除事件

以下是Analysis Template Deleted事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Analysis Template Deleted"]
}
```

協同合作建立的事件

以下是Collaboration Created事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Created"]
}
```

協同合作已更新事件

以下是Collaboration Updated事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Updated"]
}
```

協同合作變更請求建立的事件

以下是Collaboration Change Request Created事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Created"]
}
```

協同合作變更請求核准事件

以下是Collaboration Change Request Approved事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Approved"]
}
```

協同合作變更請求已取消事件

以下是Collaboration Change Request Cancelled事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Cancelled"]
}
```

協同合作變更請求遞交事件

以下是Collaboration Change Request Committed事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Committed"]
}
```

設定的資料表關聯已建立事件

以下是Configured Table Association Created事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Created"]
}
```

設定的資料表關聯已更新事件

以下是Configured Table Association Updated事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Updated"]
}
```

設定的資料表關聯已刪除事件

以下是Configured Table Association Deleted事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Deleted"]
}
```

設定的資料表關聯分析規則已建立事件

以下是Configured Table Association Analysis Rule Created事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Analysis Rule Created"]
}
```

設定的資料表關聯分析規則已更新事件

以下是Configured Table Association Analysis Rule Updated事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Analysis Rule Updated"]
}
```

已設定的資料表關聯分析規則已刪除事件

以下是Configured Table Association Analysis Rule Deleted事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Analysis Rule Deleted"]
}
```

ID 映射資料表已建立事件

以下是Id Mapping Table Created事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Mapping Table Created"]
}
```

ID 映射資料表已更新事件

以下是Id Mapping Table Updated事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Mapping Table Updated"]
}
```

ID 映射資料表已刪除事件

以下是Id Mapping Table Deleted事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Mapping Table Deleted"]
}
```

ID 命名空間關聯已建立的事件

以下是Id Namespace Association Created事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Created"]
}
```

ID 命名空間關聯已更新事件

以下是Id Namespace Association Updated事件的詳細資訊欄位。

```
{
```

```
"source": ["aws.cleanrooms"],
"detail-type": ["Id Namespace Association Updated"]
}
```

Note

刪除協同合作時，事件會在 detail.status 欄位中的 ID 命名空間關聯已更新事件中擷取。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Updated"],
  "detail": {
    "status": ["COLLABORATION_DELETED"]
  }
}
```

ID 命名空間關聯已刪除事件

以下是Id Namespace Association Deleted事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Deleted"]
}
```

受邀參與協同合作事件

以下是Invited To Collaboration事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Invited To Collaboration"]
}
```

成員資格建立的事件

以下是Membership Created事件的詳細資訊欄位。

```
{
```



```
"source": ["aws.cleanrooms"],
"detail-type": ["Membership Created"]
}
```

成員資格更新事件

以下是Membership Updated事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Membership Updated"]
}
```

已刪除成員資格事件

以下是Membership Deleted事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Membership Deleted"]
}
```

受保護的任務提交事件

以下是Protected Job Submitted事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Submitted"]
}
```

受保護的任務啟動事件

以下是Protected Job Started事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Started"]
}
```

受保護任務取消事件

以下是Protected Job Cancelling事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Cancelling"]
}
```

受保護的任務已取消事件

以下是Protected Job Cancelled事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Cancelled"]
}
```

受保護任務成功事件

以下是Protected Job Succeeded事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Succeeded"]
}
```

受保護的任務失敗事件

以下是Protected Job Failed事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Failed"]
}
```

受保護的查詢提交事件

以下是Protected Query Submitted事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Submitted"]
}
```

受保護的查詢啟動事件

以下是Protected Query Started事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Started"]
}
```

受保護的查詢取消事件

以下是Protected Query Cancelling事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Cancelling"]
}
```

受保護的查詢已取消事件

以下是Protected Query Cancelled事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Cancelled"]
}
```

受保護的查詢成功事件

以下是Protected Query Succeeded事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Succeeded"]
}
```

受保護的查詢失敗事件

以下是Protected Query Failed事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Failed"]
}
```

受保護的查詢逾時事件

以下是Protected Query Timed Out事件的詳細資訊欄位。

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Timed Out"]
}
```

成本分配標記

AWS Clean Rooms 支援使用成本分配標籤來追蹤您的 AWS 成本。您可以在 AWS 帳單與成本管理 儀表板上啟用這些標籤。AWS 會使用標籤來分類您的成本，並傳送每月成本分配報告給您。使用者定義的成本分配標籤可以套用至 AWS Clean Rooms 資源，以協助追蹤和分配協同合作的成本。

成本分配的可標記資源

下表列出產生費用 AWS Clean Rooms 的資源，並且可以使用使用者定義的成本分配標籤進行標記，以追蹤和組織成本。

計費資源	已標記的資源
SQL 查詢	成員資格
PySpark 任務	成員資格
ML 訓練任務	成員資格
ML 推論任務	成員資格

計費資源	已標記的資源
合成資料產生	成員資格
Lookalike 模型訓練	Lookalike 模型
Lookalike 區段匯出	Lookalike 區段

使用 建立 AWS Clean Rooms 資源 AWS CloudFormation

AWS Clean Rooms 已與 整合 AWS CloudFormation，此服務可協助您模型化和設定 AWS 資源。由於此整合，您可以減少建立和管理資源和基礎設施的時間。您可以建立範本來描述您想要的所有 AWS 資源，並為您 CloudFormation 佈建和設定這些資源。資源的範例包括協同合作、設定的資料表、設定的資料表關聯和成員資格。

使用 時 CloudFormation，您可以重複使用範本來一致且重複地設定 AWS Clean Rooms 資源。描述您的資源一次，然後在多個 和 中逐一佈建相同的資源 AWS 帳戶 AWS 區域。

AWS Clean Rooms 和 CloudFormation 範本

若要佈建和設定 AWS Clean Rooms 和相關服務的資源，您必須了解 [CloudFormation 範本](#)。範本是以 JSON 或 YAML 格式化的文本檔案。這些範本說明您要在 CloudFormation 堆疊中佈建的資源。如果您不熟悉 JSON 或 YAML，您可以使用 CloudFormation 設計工具來協助您開始使用 CloudFormation 範本。如需詳細資訊，請參閱《AWS CloudFormation 使用者指南》中的 [什麼是 CloudFormation 設計器？](#)。

AWS Clean Rooms 支援在其中建立協同合作、設定的資料表、設定的資料表關聯和成員資格 CloudFormation。如需詳細資訊，包括協同合作的 JSON 和 YAML 範本範例、設定的資料表、設定的資料表關聯和成員資格，請參閱AWS CloudFormation 《使用者指南》中的 [AWS Clean Rooms](#)和 [AWS Clean Rooms ML](#) 資源類型參考。

進一步了解 CloudFormation

若要進一步了解 CloudFormation，請參閱下列資源：

- [AWS CloudFormation](#)
- [AWS CloudFormation 使用者指南](#)
- [CloudFormation API 參考](#)
- [AWS CloudFormation 命令列介面使用者指南](#)

的配額 AWS Clean Rooms

您的 AWS 帳戶 具有每個 的預設配額，先前稱為限制 AWS 服務。除非另有說明，否則每個配額都專屬於 AWS 區域。您可以要求提高某些配額，而其他配額無法提高。

若要檢視 的配額 AWS Clean Rooms，請開啟 [Service Quotas 主控台](#)。在導覽窗格中，選擇 AWS 服務，然後選取 AWS Clean Rooms。

若要請求提升配額，請參閱《[Service Quotas 使用者指南](#)》中的請求提升配額。如果 Service Quotas 中尚未提供配額，請使用[服務限制增加表單](#)。

主題

- [AWS Clean Rooms 配額](#)
- [AWS Clean Rooms ML 配額](#)

AWS Clean Rooms 配額

您的 AWS 帳戶 具有下列與 相關的配額 AWS Clean Rooms。

名稱	預設	可調整	Description
分析規則大小	每個支援的區域： 100 KB	否	分析規則的 JSON 大小上限
每個成員資格的分析範本	每個受支援的區域：25	是	每個成員資格的分析範本數目上限
每個帳戶建立的協同合作	每個受支援的區域：10	是	每個帳戶建立的協同合作數目上限
每個已設定資料表允許清單的資料欄	每個受支援的區域：100	是	每個已設定資料表可允許列出的資料欄數目上限
每個帳戶的並行 PySpark 任務 vCPU 用量	us-east-1： 1,024	是	每個帳戶所有同時執行的 PySpark 任務的總 vCPU 用量上限

名稱	預設	可調整	Description
	每個其他支援的區域：512		
每個帳戶的並行 PySpark 任務	us-east-1：5 每個其他支援的區域：2	是	每個帳戶同時執行的 PySpark 任務數量上限
每個帳戶的並行 SQL 查詢	us-east-1：5 每個其他支援的區域：2	是	每個帳戶同時執行 SQL 查詢的數量上限
每個帳戶的並行 SQL 查詢 vCPU 用量	每個支援的區域：512	是	每個帳戶所有同時執行 SQL 查詢的 vCPU 總用量上限
每個成員資格的並行進行中任務	每個受支援的區域：1	否	每個成員資格的並行進行中任務數目上限
每個成員資格的並行持續查詢	每個受支援的區域：5	否	每個成員資格的並行進行中查詢數目上限
每個成員資格的設定受眾模型關聯	每個受支援的區域：5	是	每個成員資格的已設定對象模型關聯數目上限
每個帳戶的已設定資料表	每個受支援的區域：250	是	每個帳戶建立的已設定資料表數目上限
每個受保護查詢的設定資料表	每個受支援的區域：15	是	受保護查詢中設定資料表的數量上限
每個成員資格的 ID 映射表	每個受支援的區域：5	是	每個成員資格的 ID 映射資料表數目上限

名稱	預設	可調整	Description
每個成員資格的 ID 命名空間關聯	每個受支援的區域：10	是	每個成員資格的 ID 命名空間關聯數目上限
每個協同合作邀請的成員	每個受支援的區域：5	是	每個協同合作邀請的成員數量上限
每個帳戶的成員資格	每個受支援的區域：100	是	每個帳戶的成員資格數目上限
每次協同合作存取預算的隱私權預算範本	每個受支援的區域：25	是	每次協同合作存取預算的隱私權預算範本數量上限
查詢文字長度（使用差異隱私權）	每個支援的區域：8 KB	否	使用差異隱私權的 SQL 查詢陳述式文字長度上限
Clean Rooms SQL 分析引擎上的查詢文字長度	每個支援的區域：90 KB	否	Clean Rooms SQL 分析引擎上 SQL 查詢陳述式的文字長度上限
Spark 分析引擎上的查詢文字長度	每個支援的區域：500 KB	否	Spark 分析引擎上 SQL 查詢陳述式的文字長度上限
BatchGetCollaborationAnalysisTemplate 請求的速率	每個支援的區域：每秒 5 個	是	每秒 BatchGetCollaborationAnalysisTemplate API 呼叫的數量上限
BatchGetSchema 請求的速率	每個支援的區域：每秒 5 個	是	每秒 BatchGetSchema API 呼叫的數量上限
BatchGetSchemaAnalysisRule 請求的速率	每個支援的區域：每秒 5 個	是	每秒 BatchGetSchemaAnalysisRule API 呼叫的數量上限
CreateAnalysisTemplate 請求的速率	每個支援的區域：每秒 5 個	是	每秒 CreateAnalysisTemplate API 呼叫的數量上限

名稱	預設	可調整	Description
CreateCollaboration 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreateCollaboration API 呼叫的數量上限
CreateCollaborationChangeRequest 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreateCollaborationChangeRequest API 呼叫的數量上限
CreateConfiguredAudienceModelAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreateConfiguredAudienceModelAssociation API 呼叫的數量上限
CreateConfiguredTable 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreateConfiguredTable API 呼叫的數量上限
CreateConfiguredTableAnalysisRule 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreateConfiguredTableAnalysisRule API 呼叫的數量上限
CreateConfiguredTableAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreateConfiguredTableAssociation API 呼叫的數量上限
CreateConfiguredTableAssociationAnalysisRule 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreateConfiguredTableAssociationAnalysisRule API 呼叫的數量上限
CreateIdMappingTable 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreateIdMappingTable API 呼叫的數量上限
CreateIdNamespaceAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreateIdNamespaceAssociation API 呼叫的數量上限

名稱	預設	可調整	Description
CreateMembership 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreateMembership API 呼叫的數量上限
CreatePrivacyBudgetTemplate 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 CreatePrivacyBudgetTemplate API 呼叫的數量上限
DeleteAnalysisTemplate 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteAnalysisTemplate API 呼叫的數量上限
DeleteCollaboration 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteCollaboration API 呼叫的數量上限
DeleteConfiguredAudienceModelAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteConfiguredAudienceModelAssociation API 呼叫的數量上限
DeleteConfiguredTable 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteConfiguredTable API 呼叫的數量上限
DeleteConfiguredTableAnalysisRule 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteConfiguredTableAnalysisRule API 呼叫的數量上限
DeleteConfiguredTableAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteConfiguredTableAssociation API 呼叫的數量上限
DeleteConfiguredTableAssociationAnalysisRule 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteConfiguredTableAssociationAnalysisRule API 呼叫數目上限
DeleteldMappingTable 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteldMappingTable API 呼叫的數量上限

名稱	預設	可調整	Description
DeleteIdNamespaceAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteIdNamespaceAssociation API 呼叫的數量上限
DeleteMember 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteMember API 呼叫的數量上限
DeleteMembership 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeleteMembership API 呼叫的數量上限
DeletePrivacyBudgetTemplate 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 DeletePrivacyBudgetTemplate API 呼叫的數量上限
GetAnalysisTemplate 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetAnalysisTemplate API 呼叫的數量上限
GetCollaboration 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetCollaboration API 呼叫的數量上限
GetCollaborationAnalysisTemplate 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetCollaborationAnalysisTemplate API 呼叫的數量上限
GetCollaborationChangeRequest 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetCollaborationChangeRequest API 呼叫的數量上限
GetCollaborationConfiguredAudienceModelAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetCollaborationConfiguredAudienceModelAssociation API 呼叫的數量上限

名稱	預設	可調整	Description
GetCollaborationIdNamespaceAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetCollaborationIdNamespaceAssociation API 呼叫的數量上限
GetCollaborationPrivacyBudgetTemplate 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetCollaborationPrivacyBudgetTemplate API 呼叫的數量上限
GetConfiguredAudienceModelAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetConfiguredAudienceModelAssociation API 呼叫的數量上限
GetConfiguredTable 請求的速率	每個受支援的區域： 每秒 20 個	是	每秒 GetConfiguredTable API 呼叫的數量上限
GetConfiguredTableAnalysisRule 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetConfiguredTableAnalysisRule API 呼叫的數量上限
GetConfiguredTableAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetConfiguredTableAssociation API 呼叫的數量上限
GetConfiguredTableAssociationAnalysisRule 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetConfiguredTableAssociationAnalysisRule API 呼叫的數量上限
GetIdMappingTable 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetIdMappingTable API 呼叫的數量上限
GetIdNamespaceAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetIdNamespaceAssociation API 呼叫的數量上限

名稱	預設	可調整	Description
GetMembership 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetMembership API 呼叫的數量上限
GetPrivacyBudgetTemplate 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetPrivacyBudgetTemplate API 呼叫的數量上限
GetProtectedJob 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetProtectedJob API 呼叫的數量上限
GetProtectedQuery 請求的速率	每個受支援的區域： 每秒 20 個	是	每秒 GetProtectedQuery API 呼叫的數量上限
GetSchema 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetSchema API 呼叫的數量上限
GetSchemaAnalysisRule 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 GetSchemaAnalysisRule API 呼叫的數量上限
ListAnalysisTemplates 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListAnalysisTemplates API 呼叫的數量上限
ListCollaborationAnalysisTemplates 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListCollaborationAnalysisTemplates API 呼叫的數量上限
ListCollaborationChangeRequests 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListCollaborationChangeRequests API 呼叫的數量上限
ListCollaborationConfiguredAudienceModelAssociations 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListCollaborationConfiguredAudienceModelAssociations API 呼叫數目上限

名稱	預設	可調整	Description
ListCollaborationIdNamespaceAssociations 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListCollaborationIdNamespaceAssociations API 呼叫的數量上限
ListCollaborationPrivacyBudgetTemplates 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListCollaborationPrivacyBudgetTemplates API 呼叫的數量上限
ListCollaborationPrivacyBudgets 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListCollaborationPrivacyBudgets API 呼叫的數量上限
ListCollaborations 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListCollaborations API 呼叫數目上限
ListConfiguredAudienceModelAssociations 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListConfiguredAudienceModelAssociations API 呼叫的數量上限
ListConfiguredTableAssociations 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListConfiguredTableAssociations API 呼叫的數量上限
ListConfiguredTables 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListConfiguredTables API 呼叫的數量上限
ListIdMappingTables 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListIdMappingTables API 呼叫的數量上限
ListIdNamespaceAssociations 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListIdNamespaceAssociations API 呼叫的數量上限
ListMembers 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListMembers API 呼叫數目上限

名稱	預設	可調整	Description
ListMemberships 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListMemberships API 呼叫的數量上限
ListPrivacyBudgetTemplates 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListPrivacyBudgetTemplates API 呼叫的數量上限
ListPrivacyBudgets 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListPrivacyBudgets API 呼叫的數量上限
ListProtectedJobs 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListProtectedJobs API 呼叫數目上限
ListProtectedQueries 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 ListProtectedQueries API 呼叫的數量上限
ListSchemas 請求的速率	每個受支援的區域：5	是	每秒 ListSchemas API 呼叫的數量上限
ListTagsForResource 請求速率	每個支援的區域： 每秒 5 個	是	每秒 ListTagsForResource API 呼叫的數量上限
PopulateIdMappingTable 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 PopulateIdMappingTable API 呼叫的數量上限
PreviewPrivacyImpact 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 PreviewPrivacyImpact API 呼叫的數量上限
StartProtectedJob 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 StartProtectedJob API 呼叫的數量上限
StartProtectedQuery 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 StartProtectedQuery API 呼叫的數量上限

名稱	預設	可調整	Description
TagResource 請求速率	每個支援的區域： 每秒 5 個	是	每秒 TagResource API 呼叫的數量上限
UntagResource 請求速率	每個支援的區域： 每秒 5 個	是	每秒 UntagResource API 呼叫的數量上限
UpdateAnalysisTemplate 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateAnalysisTemplate API 呼叫的數量上限
UpdateCollaboration 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateCollaboration API 呼叫數上限
UpdateCollaborationChangeRequest 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateCollaborationChangeRequest API 呼叫的數量上限
UpdateConfiguredAudienceModelAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateConfiguredAudienceModelAssociation API 呼叫的數量上限
UpdateConfiguredTable 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateConfiguredTable API 呼叫的數量上限
UpdateConfiguredTableAnalysisRule 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateConfiguredTableAnalysisRule API 呼叫的數量上限
UpdateConfiguredTableAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateConfiguredTableAssociation API 呼叫的數量上限

名稱	預設	可調整	Description
UpdateConfiguredTableAssociationAnalysisRule 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateConfiguredTableAssociationAnalysisRule API 呼叫的數量上限
UpdateIdMappingTable 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateIdMappingTable API 呼叫的數量上限
UpdateIdNamespaceAssociation 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateIdNamespaceAssociation API 呼叫的數量上限
UpdatePrivacyBudgetTemplate 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdatePrivacyBudgetTemplate API 呼叫的數量上限
UpdateProtectedJob 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateProtectedJob API 呼叫的數量上限
UpdateProtectedQuery 請求的速率	每個支援的區域： 每秒 5 個	是	每秒 UpdateProtectedQuery API 呼叫的數量上限
每個成員資格的資料表關聯	每個受支援的區域：25	是	每個成員資格的資料表關聯數目上限

AWS Clean Rooms 資源參數限制

資源	預設	Description
查詢文字長度	90 KB	SQL 查詢陳述式的文字長度上限

資源	預設	Description
查詢文字長度（使用差異隱私權）	90 KB	使用差異隱私權的 SQL 查詢陳述式文字長度上限
查詢執行時間	24 小時	在逾時之前執行查詢的持續時間上限

AWS Clean Rooms ML 配額

您的 AWS 帳戶 具有 AWS Clean Rooms ML 的預設配額，先前稱為限制。

若要檢視 AWS Clean Rooms ML 的服務配額，請執行下列其中一項操作：

- 遵循 Service Quotas [使用者指南中檢視服務配額](#) 的步驟，然後選取 AWS Clean Rooms ML 做為服務。
- 請參閱 中的 [AWS Clean Rooms ML 服務配額](#) Amazon Web Services 一般參考。

為了維持服務的效能並確保適當使用 AWS Clean Rooms ML，指派給帳戶的預設配額可能會根據區域因素、付款歷史記錄、詐騙使用量和/或配額增加請求的核准而更新。

名稱	預設	可調整	Description
每個對象產生任務的作用中對象匯出任務	每個受支援的區域：25	否	對象產生任務的作用中對象匯出任務數量上限
每個成員資格的作用中設定模型演算法關聯	每個受支援的區域：1,000	是	每個成員資格的作用中設定模型演算法關聯數目上限
每個成員資格的作用中設定模型演算法	每個受支援的區域：1,000	是	每個成員資格的作用中設定模型演算法數目上限
每個成員資格的作用中自訂模型輸入通道	每個受支援的區域：100	是	每個成員資格的作用中自訂模型輸入通道數量上限

名稱	預設	可調整	Description
每個帳戶作用中訓練執行個體的數量上限	每個受支援的區域：10	是	每個帳戶可用於建立訓練模型的作用中訓練執行個體數目上限。
每個訓練模型的active/pending/in進行中訓練模型版本數量上限	每個受支援的區域：100	是	每個訓練模型可建立的active/pending/in進行中訓練模型版本數量上限。
產生合成資料的輸入資料欄數目上限	每個支援的區域：1,000	否	產生合成資料的輸入資料欄數目上限
產生合成資料的輸入列數目上限	每個支援的區域：2, 500, 000	否	產生合成資料的輸入列數目上限
每個帳戶的 ml.c4.2xlarge 訓練執行個體數目上限	每個支援的區域：0	是	每個帳戶可用於建立訓練模型的 ml.c4.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c4.4xlarge 訓練執行個體數目上限	每個支援的區域：0	是	每個帳戶可用於建立訓練模型的 ml.c4.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c4.8xlarge 訓練執行個體數目上限	每個支援的區域：0	是	每個帳戶可用於建立訓練模型的 ml.c4.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c4.xlarge 訓練執行個體數目上限	每個支援的區域：0	是	每個帳戶可用於建立訓練模型的 ml.c4.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c5.18xlarge 訓練執行個體數目上限	每個支援的區域：0	是	每個帳戶可用於建立訓練模型的 ml.c5.18xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.c5.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c5.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c5.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c5.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c5.9xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c5.9xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c5.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c5.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c5n.18xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c5n.18xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c5n.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c5n.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c5n.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c5n.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c5n.9xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c5n.9xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c5n.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c5n.xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.c6i.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c6i.12xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c6i.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c6i.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c6i.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c6i.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c6i.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c6i.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c6i.32xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c6i.32xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c6i.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c6i.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c6i.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c6i.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c6i.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c6i.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c7i.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c7i.12xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.c7i.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c7i.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c7i.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c7i.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c7i.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c7i.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c7i.48xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c7i.48xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c7i.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c7i.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c7i.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c7i.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.c7i.large 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c7i.large 訓練執行個體數目上限。
每個帳戶的 ml.c7i.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.c7i.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g4dn.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g4dn.12xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.g4dn.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g4dn.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g4dn.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g4dn.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g4dn.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g4dn.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g4dn.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g4dn.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g4dn.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g4dn.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g5.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g5.12xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g5.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g5.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g5.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g5.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g5.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g5.2xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.g5.48xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g5.48xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g5.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g5.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g5.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g5.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g5.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g5.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6.12xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6.48xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6.48xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.g6.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6e.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6e.12xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6e.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6e.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6e.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6e.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6e.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6e.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6e.48xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6e.48xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6e.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6e.4xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.g6e.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6e.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.g6e.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.g6e.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m4.10xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m4.10xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m4.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m4.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m4.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m4.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m4.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m4.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m4.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m4.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m5.12xlarge 訓練執行個體數目上限	每個支援的區域： 3	是	每個帳戶可用於建立訓練模型的 ml.m5.12xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m5.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m5.24xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.m5.2xlarge 訓練執行個體數目上限	每個支援的區域： 3	是	每個帳戶可用於建立訓練模型的 ml.m5.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m5.4xlarge 訓練執行個體數目上限	每個支援的區域： 3	是	每個帳戶可用於建立訓練模型的 ml.m5.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m5.large 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m5.large 訓練執行個體數目上限。
每個帳戶的 ml.m5.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m5.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m6i.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m6i.12xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m6i.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m6i.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m6i.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m6i.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m6i.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m6i.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m6i.32xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m6i.32xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.m6i.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m6i.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m6i.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m6i.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m6i.large 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m6i.large 訓練執行個體數目上限。
每個帳戶的 ml.m6i.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m6i.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m7i.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m7i.12xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m7i.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m7i.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m7i.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m7i.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m7i.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m7i.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m7i.48xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m7i.48xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.m7i.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m7i.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m7i.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m7i.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.m7i.large 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m7i.large 訓練執行個體數目上限。
每個帳戶的 ml.m7i.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.m7i.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.p2.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.p2.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.p2.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.p2.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.p2.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.p2.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.p4d.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.p4d.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.p4de.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.p4de.24xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.p5.48xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.p5.48xlarge 訓練執行個體數目上限。
每個帳戶的 ml.p5en.48xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.p5en.48xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5.12xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5.large 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5.large 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.r5.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5d.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5d.12xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5d.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5d.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5d.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5d.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5d.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5d.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5d.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5d.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5d.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5d.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r5d.large 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5d.large 訓練執行個體數目上限。
每個帳戶的 ml.r5d.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r5d.xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.r7i.12xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r7i.12xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r7i.16xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r7i.16xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r7i.24xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r7i.24xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r7i.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r7i.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r7i.48xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r7i.48xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r7i.4xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r7i.4xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r7i.8xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r7i.8xlarge 訓練執行個體數目上限。
每個帳戶的 ml.r7i.large 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r7i.large 訓練執行個體數目上限。
每個帳戶的 ml.r7i.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.r7i.xlarge 訓練執行個體數目上限。

名稱	預設	可調整	Description
每個帳戶的 ml.t3.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.t3.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.t3.large 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.t3.large 訓練執行個體數目上限。
每個帳戶的 ml.t3.medium 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.t3.medium 訓練執行個體數目上限。
每個帳戶的 ml.t3.xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.t3.xlarge 訓練執行個體數目上限。
每個帳戶的 ml.trn1.2xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.trn1.2xlarge 訓練執行個體數目上限。
每個帳戶的 ml.trn1.32xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.trn1.32xlarge 訓練執行個體數目上限。
每個帳戶的 ml.trn1n.32xlarge 訓練執行個體數目上限	每個支援的區域： 0	是	每個帳戶可用於建立訓練模型的 ml.trn1n.32xlarge 訓練執行個體數目上限。
分類預測欄中的唯一值上限	每個受支援的區域：100	否	分類預測欄中的最大唯一值
產生合成資料所需的輸入資料欄數目下限	每個受支援的區域：5	否	產生合成資料所需的輸入資料欄數目下限

名稱	預設	可調整	Description
產生合成資料所需的輸入列數目下限	每個支援的區域：1,500	否	產生合成資料所需的輸入列數下限
每個客戶的待定/進行中對象匯出任務	每個受支援的區域：20	否	每個客戶待處理/進行中對象匯出任務的數量上限
每個客戶的待定/進行中對象產生任務	每個受支援的區域：10	<u>是</u>	每個客戶待定/進行中對象產生任務的數量上限
每個客戶的待定/進行中對象模型	每個受支援的區域：2	<u>是</u>	每個客戶待處理/進行中對象模型訓練任務的數量上限
每個帳戶的待定/進行中自訂模型推論任務	每個受支援的區域：10	<u>是</u>	每個帳戶的待處理/進行中自訂模型推論任務數量上限
每個成員資格的待定/進行中自訂模型推論任務	每個受支援的區域：5	<u>是</u>	每個成員資格的待處理/進行中自訂模型推論任務數量上限
每個帳戶的待定/進行中自訂模型訓練任務	每個受支援的區域：10	<u>是</u>	每個帳戶待處理/進行中自訂模型訓練任務的數量上限
每個成員資格的待定/進行中自訂模型訓練任務	每個受支援的區域：5	<u>是</u>	每個成員資格的待處理/進行中自訂模型訓練任務數量上限
每個帳戶的待處理/進行中合成 MLIC 產生任務	每個受支援的區域：2	<u>是</u>	每個帳戶的待處理/進行中合成 MLIC 產生任務數量上限

AWS Clean Rooms ML 具有下列其他配額

Clean Rooms ML 配額

資源	預設	Description
自訂建模推論任務的最長持續時間	25 小時	
互動數量上限	200 億	訓練資料中允許的互動數量上限。較大的輸入會進行取樣。
互動次數下限	100 萬	
類似模型訓練的不同使用者數目上限	1 億	如果包含更多，則只會使用前 1 億個，並依互動次數排名。
類似模型訓練的不同使用者人數下限	100,000	
匯出類似客群（對象）任務的使用者人數下限	10,000	
用於模型訓練的不同項目數量上限。	100 萬	您最多可以包含 5,000 萬個項目，但只會使用最熱門的 100 萬個項目。
訓練資料集中的特徵資料欄數量上限。	10	
每個使用者的不同項目數量下限	2	AWS Clean Rooms ML 要求每一列或使用者有兩個或更多項目，包括重複項目。
種子受眾的大小上限	500,000	
種子受眾的大小下限	500	訓練資料提供者可以將此值設定為低至 25。
API	每位客戶	
作用中訓練資料集的總數	500	

資源	預設	Description
作用中外觀模型的總數（對象模型）	500	
主動設定的外觀模型總數（對象模型）	10,000	
已完成的類似客群（對象）產生任務總數	沒有限制	
已完成的匯出類似區段（對象）任務總數	沒有限制	
產生類似模型（對象模型）任務的最長持續時間	1 天 (24 小時)	
產生類似客群（對象）任務的最長持續時間	10 小時	提供種子後，Clean Rooms ML 最多需要 10 小時才能產生外觀的區段。如果您使用 SQL 查詢做為種子資料，除了 10 小時之外，執行查詢最多可能需要 12 小時才能產生類似樣的區段。
區段（對象）大小儲存貯體的最小百分比	1%	
區段（對象）大小儲存貯體的百分比上限	20%	
區段（對象）大小儲存貯體的最小絕對大小	1% 的不同使用者數量	
區段大小（對象）大小儲存貯體的絕對大小上限	20% 的不同使用者數量	

Clean Rooms ML API 限流配額

您的 AWS 帳戶 具有每個 帳戶每個端點配額的每秒以下 API 交易 (TPS)。

資源	速率限制	Description
CreateAudienceModel 請求率	1 TPS 速率、3 TPS 爆量	每秒 CreateAudienceModel API 呼叫的數量上限
CreateConfiguredAudienceModel 請求率	10 TPS	每秒 CreateConfiguredAudienceModel API 呼叫數目上限
CreateTrainingDataset 請求率	10 TPS	每秒 CreateTrainingDataset API 呼叫的數量上限
DeleteAudienceGenerationJob 請求率	2 TPS 速率、10 TPS 爆量	每秒 DeleteAudienceGenerationJob API 呼叫的數量上限
DeleteAudienceModel 請求率	2 TPS 速率、10 TPS 爆量	每秒 DeleteAudienceModel API 呼叫的數量上限
DeleteConfiguredAudienceModel 請求率	10 TPS	每秒 DeleteConfiguredAudienceModel API 呼叫的數量上限
DeleteConfiguredAudienceModelPolicy 請求率	25 TPS	每秒 DeleteConfiguredAudienceModelPolicy API 呼叫數上限
DeleteTrainingDataset 請求率	10 TPS	每秒 DeleteTrainingDataset API 呼叫數上限
GetAudienceGenerationJob 請求率	50 TPS	每秒 GetAudienceGenerationJob API 呼叫數上限
GetAudienceModel 請求率	50 TPS	每秒 GetAudienceModel API 呼叫數上限

資源	速率限制	Description
GetConfiguredAudienceModel 請求率	50 TPS	每秒 GetConfiguredAudienceModel API 呼叫數上限
GetConfiguredAudienceModelPolicy 請求率	50 TPS	每秒 GetConfiguredAudienceModelPolicy API 呼叫的數量上限
GetTrainingDataset 請求率	50 TPS	每秒 GetTrainingDataset API 呼叫數上限
ListAudienceExportJobs 請求率	50 TPS	每秒 ListAudienceExportJobs API 呼叫數上限
ListAudienceGenerationJobs 請求率	50 TPS	每秒 ListAudienceGenerationJobs API 呼叫數上限
ListAudienceModels 請求率	50 TPS	每秒 ListAudienceModels API 呼叫數上限
ListConfiguredAudienceModels 請求率	50 TPS	每秒 ListConfiguredAudienceModels API 呼叫數上限
ListTagsForResource 請求率	50 TPS	每秒 ListTagsForResource API 呼叫數上限
ListTrainingDatasets 請求率	50 TPS	每秒 ListTrainingDatasets API 呼叫的數量上限
PutConfiguredAudienceModelPolicy 請求率	25 TPS	每秒 PutConfiguredAudienceModelPolicy API 呼叫數上限
StartAudienceExportJob 請求率	1 TPS 速率、3 TPS 爆量	每秒 StartAudienceExportJob API 呼叫的數量上限

資源	速率限制	Description
StartAudienceGenerationJob 請求率	1 TPS 速率、5 TPS 爆量	每秒 StartAudienceGenerationJob API 呼叫數上限
TagResource 請求率	10 TPS	每秒 TagResource API 呼叫數目上限
UntagResource 請求率	50 TPS	每秒 UntagResource API 呼叫數上限
UpdateConfiguredAudienceModel 請求率	10 TPS	每秒 UpdateConfiguredAudienceModel API 呼叫數量上限
CreateConfiguredModelAlgorithm 請求率	10 TPS	每秒 CreateConfiguredModelAlgorithm API 呼叫的數量上限。
CreateConfiguredModelAlgorithmAssociation 請求率	10 TPS	每秒 CreateConfiguredModelAlgorithmAssociation API 呼叫的數量上限。
PutMLConfiguration 請求率	10 TPS	每秒 PutMLConfiguration API 呼叫的數量上限。
CreateTrainedModel 請求率	1 TPS 速率、3 TPS 爆量	每秒 CreateTrainedModel API 呼叫的數量上限。
StartTrainedModelExportJob 請求率	10 TPS	每秒 StartTrainedModelExportJob API 呼叫的數量上限。
StartTrainedModelInferenceJob 請求率	1 TPS 速率、3 TPS 速率	每秒 StartTrainedModelInferenceJob API 呼叫的數量上限。

資源	速率限制	Description
GetConfiguredModelAlgorithm 請求率	50 TPS	每秒 GetConfiguredModelAlgorithm API 呼叫的數量上限。
GetConfiguredModelAlgorithmAssociation 請求率	50 TPS	每秒 GetConfiguredModelAlgorithmAssociation API 呼叫的數量上限。
GetTrainedModel 請求率	50 TPS	每秒 GetTrainedModel API 呼叫的數量上限。
GetMLConfiguration 請求率	50 TPS	每秒 GetMLConfiguration API 呼叫的數量上限。
GetTrainedModelInferenceJob 請求率	50 TPS	每秒 GetTrainedModelInferenceJob API 呼叫的數量上限。
ListConfiguredModelAlgorithm 請求率	50 TPS	每秒 ListConfiguredModelAlgorithm API 呼叫的數量上限。
ListConfiguredModelAlgorithmAssociations 請求率	50 TPS	每秒 ListConfiguredModelAlgorithmAssociations API 呼叫的數量上限。
ListTrainedModels 請求率	50 TPS	每秒 ListTrainedModels API 呼叫的數量上限。
ListCollaborationTrainedModelExportJobs 請求率	50 TPS	每秒 ListCollaborationTrainedModelExportJobs API 呼叫的數量上限。
ListCollaborationTrainedModelInferenceJobs 請求率	50 TPS	每秒 ListCollaborationTrainedModelInferenceJobs API 呼叫的數量上限。

資源	速率限制	Description
DeleteConfiguredModelAlgorithm 請求率	2 TPS 速率、10 TPS 爆量	每秒 DeleteConfiguredModelAlgorithm API 呼叫的數量上限。
DeleteConfiguredModelAlgorithmAssociation 請求率	2 TPS 速率、10 TPS 爆量	每秒 DeleteConfiguredModelAlgorithmAssociation API 請求的數量上限。
DeleteMLConfiguration 請求率	2 TPS 速率、10 TPS 爆量	每秒 DeleteMLConfiguration API 請求的數量上限。
DeleteTrainedModelOutput 請求率	2 TPS 速率、10 TPS 爆量	每秒 DeleteTrainedModelOutput API 請求的數量上限。

AWS Clean Rooms 使用者指南的文件歷史記錄

下表說明 文件的發行版本 AWS Clean Rooms。

如需有關此文件更新的通知，您可以訂閱 RSS 摘要。若要訂閱 RSS 更新，必須為所使用的瀏覽器啟用 RSS 外掛程式。

變更	描述	日期
已更新成本分配標記區段的文件	新增 的文件 AWS 帳單與成本管理，並建立與成本分配相關的可標記資源區段。	2026 年 2 月 20 日
支援 cleanrooms-ml 火花運算組態	客戶現在可以透過為 SQL 查詢設定支援的 Spark 屬性，自訂 Spark 執行時間行為並最佳化 SQL 效能。此功能現在可用於 受眾產生任務建立 和 MLInputChannel 建立 。	2026 年 1 月 5 日
支援使用 CloudWatch 進行詳細監控	客戶現在可以選擇使用 CloudWatch 指標進行詳細監控，以操作監控協同合作，包括查詢效能和資源使用率。	2026 年 1 月 2 日
支援協同合作變更請求 EventBridge 通知事件	新增協同合作變更請求生命週期事件的 EventBridge 事件。	2025 年 12 月 19 日
支援管理協同合作成員資格和設定的變更請求	協作建立者現在可以提交變更請求，以新增新成員、更新現有成員功能，以及修改自動核准設定。所有協同合作成員都必須在變更生效之前核准變更請求。	2025 年 12 月 18 日
AWS Clean Rooms SQL 分析引擎的終止支援時間表	AWS Clean Rooms SQL 分析引擎的支援已於 2025 年 12 月	2025 年 12 月 17 日

	17 日結束。客戶現在可以僅使用 Spark SQL 分析引擎。	
支援其他 EventBridge 通知事件	新增分析範本、已設定資料表關聯、已設定資料表關聯分析規則、Id 映射資料表、Id 命名空間關聯、受保護任務和受保護查詢的新 EventBridge 事件。	2025 年 12 月 15 日
支援用於訓練自訂機器學習模型的隱私權增強合成資料集產生	客戶現在可以產生隱私權增強的合成資料集，以訓練自訂機器學習模型，同時保護敏感資料。	2025 年 11 月 30 日
支援 SQL 查詢中的可設定 Spark 屬性	客戶現在可以透過設定 SQL 查詢支援的 Spark 屬性來自訂 Spark 執行時間行為。	2025 年 10 月 30 日
支援跨區域協同合作	AWS Clean Rooms 現在支援 AWS Glue 和 Athena 資料表的跨區域協同合作，讓客戶能夠處理存放在單一 AWS 區域中的資料，並將結果交付至所有協同合作成員核准的區域。	2025 年 10 月 3 日
支援資料存取預算	客戶現在可以設定和管理資料存取預算，以控制資料表用量。	2025 年 10 月 1 日
填入 ID 映射表 – 更新	客戶現在可以在規則型 ID 映射工作流程中使用增量處理，更有效地處理大型資料集。客戶也可以從 ID 映射工作流程中刪除記錄，以協助遵守資料管理法規。	2025 年 9 月 23 日

支援 PySpark 任務的可變工作負載大小	客戶現在可以選擇和調整其運算大小，以更有效地執行 PySpark 任務。	2025 年 9 月 4 日
支援將新成員新增至現有的協同合作	客戶現在可以將新成員新增至現有的協同合作。此功能包含設定自動核准設定和封鎖成員資格請求的選項。	2025 年 9 月 3 日
新增設定自訂模型日誌錯誤摘要和修訂的支援	客戶現在可以在設定的模型演算法關聯中，為自訂模型日誌設定錯誤摘要和修訂。	2025 年 9 月 3 日
支援 Snowflake 身分驗證選項	客戶現在可以在 中存放 Snowflake 憑證 AWS Secrets Manager 以進行 AWS Clean Rooms 連線時，搭配 Snowflake PEM 私有金鑰使用金鑰對身分驗證。	2025 年 8 月 29 日
新增 PySpark 分析範本的錯誤組態	客戶現在可以在 PySpark 分析範本中設定基本或詳細的錯誤訊息層級，以管理錯誤可見性和故障診斷功能。	2025 年 8 月 19 日
支援 EventBridge	客戶現在可以使用 Amazon EventBridge 將事件從 路由 AWS Clean Rooms 到其他軟體元件。	2025 年 7 月 31 日

更新現有政策	下列新許可已新增至 AWSCleanRoomsFullAccessNoQuerying 受管政策：cleanrooms:UpdateConfiguredTableAllowedColumns 和 cleanrooms:UpdateConfiguredTableReference。	2025 年 7 月 29 日
增加已設定資料表的編輯功能	客戶現在可以修改所設定資料表的允許資料欄和資料表參考欄位。	2025 年 7 月 29 日
ML 輸入通道支援 Parquet 檔案類型	客戶現在可以選擇 Parquet 檔案類型，做為 ML 輸入通道可以使用的資料格式。	2025 年 7 月 17 日
支援 Spark SQL 分析引擎上的差異隱私權	客戶現在可以搭配 Spark SQL 分析引擎使用差異隱私權。此外，使用差異隱私權的 SQL 查詢陳述式文字長度上限現在是 90 KB。	2025 年 7 月 16 日
AWS Clean Rooms SQL 分析引擎的終止支援時間表	使用 Clean Rooms SQL 分析引擎的新協同合作支援將於 2025 年 7 月 16 日結束。自 2025 年 7 月 17 日起，您無法與舊版 Clean Rooms SQL 分析引擎建立新的協同合作。將於 2025 年 12 月 17 日 AWS Clean Rooms 終止對 Clean Rooms SQL 分析引擎的支援。	2025 年 7 月 16 日
支援增量和分散式訓練	客戶現在可以為其自訂模型使用增量訓練和分散式訓練。	2025 年 7 月 1 日

[Clean Rooms SQL 分析引擎的 終止支援時間表](#)

從現在開始到 2025 年 7 月 16 日，使用舊版 Clean Rooms SQL 分析引擎的新協作需要提高限制。從 2025 年 7 月 17 日開始，您無法與此引擎建立新的協同合作。在 2025 年 12 月 17 日 AWS Clean Rooms 結束對 Clean Rooms SQL 分析引擎的支援。

2025 年 5 月 22 日

[更新至 PySpark 分析範本](#)

客戶現在可以使用任何有效的 Python 檔案名稱來命名使用者指令碼檔案。

2025 年 5 月 15 日

[支援多個結果接收者](#)

客戶現在可以選取多個成員，從單一查詢接收查詢結果。

2025 年 4 月 30 日

[支援將協同合作遷移至 Spark SQL](#)

AWS Clean Rooms 除了自訂分析規則之外，SQL 現在還支援彙總和列出分析規則。此外，客戶可以更新現有的協同合作，以使用支援 Spark SQL 的 Spark 分析引擎。

2025 年 4 月 2 日

[支援 PySpark 任務](#)

客戶現在可以使用核准的 PySpark 分析範本執行任務來分析資料。

2025 年 3 月 18 日

現有政策的更新	下列新許可已新增至 AWSCleanRoomsMLReadOnlyAccess 受管政策：PassCleanRoomsResources。下列新許可已新增至 AWSCleanRoomsMLFullAccess 受管政策：PassCleanRoomsResources 和 ConsoleDescribeECRRepositories。	2025 年 1 月 10 日
支援多個運算工作者	客戶現在可以指定運算工作者的類型，以及在建立類似客群時要佈建的數量。	2024 年 12 月 17 日
支援多個資料來源和雲端	客戶現在可以使用多個資料來源和雲端，例如 Amazon Athena 和 Snowflake，與其合作夥伴的資料集協作。	2024 年 12 月 1 日
Clean Rooms ML Custom Modeling 現已推出	客戶現在可以在協同合作中使用自己的自訂 ML 模型。	2024 年 11 月 7 日
新的分析引擎	擁有大型資料集的客戶現在可以使用 Spark SQL 分析引擎支援的 SQL 函數來執行複雜的查詢。	2024 年 10 月 29 日
增強的隱私權保護、建置看起來像受眾的對象、選擇多個結果接收者	您可以保護資料，同時允許使用其他分析和協同合作分析規則進行複雜的啟用查詢。您可以從 SQL 查詢或分析範本建立外觀相似的對象模型。您可以選擇多個成員來接收結果。	2024 年 7 月 24 日

[中的實體解析 AWS Clean Rooms](#)

在 AWS Entity Resolution 中 AWS Clean Rooms，您可以在兩個 ID 命名空間之間建立 ID 映射表，以跨不同的身分空間查詢事件資料。

2024 年 7 月 23 日

[更新現有政策](#)

下列新許可已新增至 `AWSCleanRoomsFullAccessNoQuerying` 受管政策：`cleanrooms:BatchGetSchemaAnalysisRule`。

2024 年 5 月 13 日

[AWS Clean Rooms ML 現已全面推出](#)

AWS Clean Rooms ML 為雙方提供隱私權增強方法，以識別其資料中的類似使用者，而無需彼此共用其資料。

2024 年 4 月 3 日

[更新現有政策](#)

`AWSCleanRoomsFullAccess` 受管政策中的陳述式 ID 已從 `更新ConsolePickQueryResultsBucket` 為 `SetQueryResultsBucket`，以更好地代表自許可可以來的許可。

2024 年 3 月 21 日

[AWS Clean Rooms ML 的新受管政策](#)

已新增兩個新的 受管政策：`AWSCleanRoomsMLReadOnlyAccess` 和 `AWSCleanRoomsMLFullAccess`。

2023 年 11 月 29 日

[AWS Clean Rooms ML \(預覽版\)](#)

AWS Clean Rooms ML 為雙方提供隱私權增強方法，以識別其資料中的類似使用者，而無需彼此共用其資料。

2023 年 11 月 29 日

AWS Clean Rooms 差異隱私權 (預覽版)	客戶現在可以使用 AWS Clean Rooms 差異隱私權來協助保護使用者的隱私權。	2023 年 11 月 29 日
付款組態	協同合作建立者現在可以設定可執行查詢的成員或協同合作中的不同成員，以支付查詢運算成本。	2023 年 11 月 14 日
查詢執行時間 - 更新	在逾時從 4 小時更新為 12 小時之前執行查詢的持續時間上限。	2023 年 10 月 6 日
CloudFormation 資源 - 更新	AWS Clean Rooms 已新增下列新資源：AWS::CleanRooms::MembershipProtectedQueryOutputConfiguration、AWS::CleanRooms::MembershipProtectedQueryResultConfiguration 和 AWS::CleanRooms::MembershipProtectedQueryS3OutputConfiguration。	2023 年 9 月 7 日
CloudFormation 資源 - 更新	AWS Clean Rooms 已新增下列新資源：AWS::CleanRooms::AnalysisTemplate 和 AWS::CleanRooms::ConfiguredTableAnalysisRuleCustom。	2023 年 8 月 31 日

[分隔成員能力](#)

協作建立者現在可以將一個成員指定為可查詢的成員，將另一個成員指定為可接收結果的成員。這可讓協同合作建立者確保可查詢的成員無法存取查詢結果。

2023 年 8 月 30 日

[AWS Clean Rooms 詞彙表](#)

僅文件更新，以新增 AWS Clean Rooms 詞彙詞彙表。

2023 年 8 月 30 日

[支援Apache Iceberg資料表 \(預覽\)](#)

AWS Clean Rooms 現在支援 Apache Iceberg資料表 (預覽)。

2023 年 8 月 25 日

[配額更新](#)

[配額區段](#)已更新，以反映每個帳戶成員資格的新預設配額。

2023 年 8 月 9 日

[更新現有政策](#)

下列新許可已新增至
 AWSCleanRoomsFullAccessNoQuerying 受
 管政策：cleanroom
 s:CreateAnalysisTem
 plate、cleanroom
 s:GetAnalysisTempl
 ate、cleanroom
 s:UpdateAnalysisTe
 mplate、cleanroom
 s>DeleteAnalysisTe
 mplate、cleanroom
 s>ListAnalysisTemp
 lates、cleanroom
 s:GetCollaboration
 AnalysisTemplate、
 cleanrooms:BatchGe
 tCollaborationAnal
 ysisTemplate 和
 cleanrooms:ListCol
 laborationAnalysis
 Templates。

2023 年 7 月 31 日

[分析範本和自訂分析規則](#)

AWS Clean Rooms 現在支援
 分析範本和自訂分析規則。分
 析範本可讓協作者建置或匯入
 自己的自訂 SQL 查詢，以便在
 協作中使用。使用自訂分析規
 則，資料表擁有者可以在其設
 定的資料表上核准自訂 SQL 查
 詢。

2023 年 7 月 31 日

[分析規則支援OR邏輯條件](#)

AWS Clean Rooms 分析規則
 現在支援 OR JOIN子句中的邏
 輯條件。

2023 年 6 月 29 日

CloudFormation 整合	AWS Clean Rooms 現在與 整合 CloudFormation。	2023 年 6 月 15 日
分析建置器	可以查詢和接收結果的成員現在能夠對某些資料表執行查詢，而無需使用分析建置器 UI 撰寫 SQL 程式碼。	2023 年 6 月 15 日
SQL 函數	僅文件更新，以釐清支援的 SQL 函數。	2023 年 5 月 5 日
疑難排解	針對常見問題新增故障診斷區段的僅限文件更新。	2023 年 4 月 27 日
支援的 資料類型 AWS Clean Rooms	僅文件更新，以新增列出支援 AWS Glue Data Catalog 資料類型的新區段。	2023 年 4 月 26 日
AWS CloudTrail 事件範例	僅文件更新，以新增 StartProtectedQuery (成功) 和 StartProtectedQuery (失敗) 的 CloudTrail 事件範例。	2023 年 4 月 20 日
更新現有政策	下列新許可已新增至 AWSCleanRoomsFullAccessNoQuerying 受管政策：cleanrooms:UntagResource、cleanrooms:ListTagsForResource 和 cleanrooms:TagResource。如需詳細資訊，請參閱 AWS 受管政策 。	2023 年 3 月 21 日
一般可用性	AWS Clean Rooms 現已正式推出。	2023 年 3 月 21 日

[預覽版](#)

AWS Clean Rooms 使用者指南的預覽版本

2023 年 1 月 12 日

AWS Clean Rooms 詞彙表

請參閱此詞彙表，熟悉所使用的術語 AWS Clean Rooms。

彙總分析規則

允許查詢沿著選用維度使用 COUNT、SUM 或 AVG 函數彙總分析的查詢限制。這些查詢不會顯示資料列層級資訊。

支援行銷活動規劃、媒體觸及、頻率和轉換測量等使用案例。

其他類型的分析規則是 [自訂](#) 和 [清單](#)。

分析規則

授權特定查詢類型的查詢限制。

分析規則類型會決定可以在設定的資料表上執行哪種分析。每種類型都有預先定義的查詢結構。您可以透過查詢控制項來控制資料表資料欄在結構中的使用方式。

分析規則的類型包括 [彙總](#)、[清單](#) 和 [自訂](#)。

分析範本

可重複使用的協同合作特定預先核准查詢。

支援的格式：SQL 程式碼或適用於 Spark 的 Python 程式碼。

如果使用 SQL，分析範本可以包含常值通常會出現在 SQL 查詢中的參數。如需支援參數類型的詳細資訊，請參閱 AWS Clean Rooms SQL 參考中的 [資料類型](#)。

分析範本僅適用於 [自訂分析規則](#)。

C3R 加密用戶端

適用於 Clean Rooms(C3R) 加密用戶端的加密運算。

C3R 是具有命令列界面的用戶端加密 SDK，用於加密和解密資料。

純文字欄

未以密碼編譯方式保護 JOIN 或 SELECT SQL 建構的欄。

純文字資料欄可用於 SQL 查詢的任何部分。

協作

安全邏輯界限 AWS Clean Rooms，其中成員可以在設定的資料表上執行 SQL 查詢。

協同合作是由[協同合作建立者](#)建立。

只有受邀參與協同合作的成員才能加入協同合作。

協同合作只能有一個[成員可以查詢](#)資料，或一個[成員可以執行查詢和任務](#)。

協同合作只能有一個[成員可以接收結果](#)。

協同合作只能有一個[成員支付查詢運算成本](#)，或一個[成員支付查詢和任務運算成本](#)。

所有成員都可以在加入協同合作之前，查看協同合作中受邀參與者的清單。

協作建立者

建立協同合作的成員。

每個協同合作只有一個協同合作建立者。

只有協同合作建立者可以從協同合作中移除成員或刪除協同合作。

設定的資料表

每個設定的資料表代表 中已設定用於 AWS Glue Data Catalog 的現有資料表參考 AWS Clean Rooms。設定的資料表包含分析規則，可決定如何使用資料。

目前，AWS Clean Rooms 支援關聯存放在 Amazon Simple Storage Service (Amazon S3) 中的資料，而這些資料是透過 編製目錄 AWS Glue。

如需 的詳細資訊 AWS Glue，請參閱 [AWS Glue 開發人員指南](#)。

設定的資料表可以與一或多個協同合作相關聯。

 Note

AWS Clean Rooms 目前不支援向 註冊的 Amazon S3 儲存貯體位置 AWS Lake Formation。

自訂分析規則

允許一組特定預先核准的查詢 ([分析範本](#)) 或允許一組特定帳戶，可提供使用您資料的查詢或任務的查詢限制。

支援使用案例，例如初次觸控歸因、增量分析和受眾探索分析。

支援差異隱私權。

其他類型的分析規則是[彙總](#)和[清單](#)。

解密

將加密資料轉換回原始格式的程序。只有在您可以存取私密金鑰時，才能執行解密。

差異隱私權

一種數學上嚴格的技術，可保護成員的協同合作資料，這些成員可以接收了解特定個人的結果。

加密

將資料編碼為使用稱為金鑰之秘密值隨機顯示的形式的程序。無法在無法存取金鑰的情況下判斷原始純文字。

指紋資料欄

以密碼編譯方式保護 JOIN SQL 建構的欄。

ID 映射工作流程方法

您希望 ID 映射如何執行。

有兩種 ID 映射工作流程方法：

- 規則型 ID 映射 – 使用相符規則將來源的第一方資料轉譯為 ID 映射工作流程中目標的方法。
- 提供者服務 ID 映射 – 您使用提供者服務將第三方編碼資料從來源轉譯到 ID 映射工作流程中目標的方法。

AWS Clean Rooms 目前支援 LiveRamp 做為提供者服務型 ID 映射工作流程方法。您必須透過 訂閱 LiveRamp，AWS Data Exchange 才能使用此方法。如需詳細資訊，請參閱《使用者指南》中的[在上訂閱供應商服務 AWS Data Exchange](#)。AWS Entity Resolution

ID 映射表

中的資源 AWS Clean Rooms，可在協同合作中啟用第一方比對規則或多方身分轉碼。

ID 映射資料表是 中現有資料表的參考 AWS Glue Data Catalog。它包含 [ID 映射表分析規則](#)，可決定如何查詢資料 AWS Clean Rooms。ID 映射表可以與一或多個協同合作相關聯。

ID 映射表分析規則

一種由 AWS Clean Rooms 管理的分析規則類型，用於聯結不同的身分資料，以方便查詢。它會自動新增至 [ID 映射表](#)，而且無法編輯。只要這些分析規則是同質的，就會繼承協同合作中其他分析規則的行為。

ID 映射工作流程

根據指定的 [ID 映射工作流程方法](#)，將資料從來源映射到目標的資料處理任務。它會產生 [ID 映射表](#)。

ID 命名空間

中的資源 AWS Clean Rooms，其中包含說明多個資料集的中繼資料，AWS 帳戶 以及如何在 [ID 映射工作流程](#)中使用這些資料集。

ID 命名空間關聯

ID 命名空間資源的關聯，可協助您探索其 [ID 映射工作流程](#)中的輸入。

任務

一種方法，可使用支援的一組函數、類別和變數，在協同合作中存取和分析設定的資料表。

AWS Clean Rooms 目前支援 PySpark 任務類型。

AWS Clean Rooms 目前支援使用 PySpark 分析範本執行中的任務。

列出分析規則

允許輸出此資料表與可查詢之成員資料表之間重疊之資料列層級屬性分析的查詢限制。

支援擴充和受眾建置或禁止等使用案例。

其他類型的分析規則是[彙總](#)和[自訂](#)的。

Lookalike 模型

訓練資料提供者資料的模型，可讓種子資料提供者建立[訓練資料提供者資料的外觀區段](#)，最接近其[種子資料](#)。

Lookalike 區段

訓練資料的子集，最接近[種子資料](#)。

成員

身為[協同合作](#)參與者 AWS 的客戶。

使用成員的 來識別成員 AWS 帳戶。

所有成員都可以貢獻資料。

可查詢的成員

可在[協同合作](#)中查詢資料的成員。

每個協同合作只能有一個成員可以查詢，而且該成員不可變。

管理使用者可以使用 AWS Identity and Access Management (IAM) 許可來控制哪些 IAM 主體（例如使用者或角色）可以在協同合作中查詢資料。如需詳細資訊，請參閱[建立服務角色以從 Amazon S3 讀取資料](#)。

可執行查詢和任務的成員

可以對[協同合作](#)中的資料執行查詢和任務的成員。

每個協同合作只有一個成員可以執行查詢和任務，而且該成員不可變。

管理使用者可以使用 AWS Identity and Access Management (IAM) 許可來控制哪些 IAM 主體（例如使用者或角色）可以在協同合作中執行查詢和任務。如需詳細資訊，請參閱[建立服務角色以從 Amazon S3 讀取資料](#)。

可接收結果的成員

可以接收查詢結果的成員。可以接收結果的成員會指定 Amazon S3 目的地和查詢結果格式 (CSV 或 Parquet) 的查詢結果設定。

對於使用 Spark 分析引擎的分析，可以接收結果的成員也會指定檔案輸出為單一檔案或多個檔案。

可以有一個以上的成員可以在協同合作中收到結果。

成員支付查詢運算成本

負責支付查詢運算成本的成員。

只有一位成員負責支付每次協同合作的查詢運算成本，而且該成員不可變。

如果協同合作建立者尚未將任何人指定為支付查詢運算成本的成員，則[可以查詢的成員](#)即為預設付款人。

支付查詢運算成本的成員會收到已在協同合作中執行之查詢的帳單。

成員支付查詢和任務運算成本

負責支付查詢和任務運算成本的成員。

只有一位成員負責支付每次協同合作的查詢和任務運算成本，而且該成員不可變。

如果協作建立者尚未將任何人指定為支付查詢和任務運算成本的成員，則[可以查詢的成員](#)是預設付款人。

支付查詢和任務運算成本的成員會收到已在協同合作中執行之查詢的帳單。

成員資格

當[成員](#)加入[協同](#)合作時建立的資源。

成員與協同合作相關聯的所有資源都是成員資格的一部分，或與成員資格相關聯。

只有擁有成員資格的成員才能新增、移除或編輯該成員資格中的資源。

密封資料欄

以密碼編譯方式保護 SELECT SQL 建構的欄。

種子資料

種子資料提供者的資料，用於建立[看起來像區段](#)。種子資料可以直接提供，也可以來自 AWS Clean Rooms 查詢的結果。看起來像區段輸出是一組來自訓練資料的 使用者，最接近種子使用者。

Spark 分析引擎

中的分析選項 AWS Clean Rooms，可讓客戶使用 Apache Spark SQL 函數，對存放在 Amazon S3、Amazon Athena 或 Snowflake 中的大型資料集執行複雜的查詢。它也支援 中的 PySpark 分析 AWS Clean Rooms。

當您使用 [CreateCollaboration API](#) 建立協同合作時，Spark 分析引擎的值為 SPARK。

Query

一種方法，可使用支援的一組函數、類別和變數，存取和分析協同合作中設定的資料表。

AWS Clean Rooms 目前支援 SQL 查詢語言。

AWS Clean Rooms 目前支援執行直接 SQL 查詢或使用 SQL 分析範本執行查詢。

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。