

Azure AutoML and Azure Databricks

Azure 雲端服務 個人體驗練習



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks

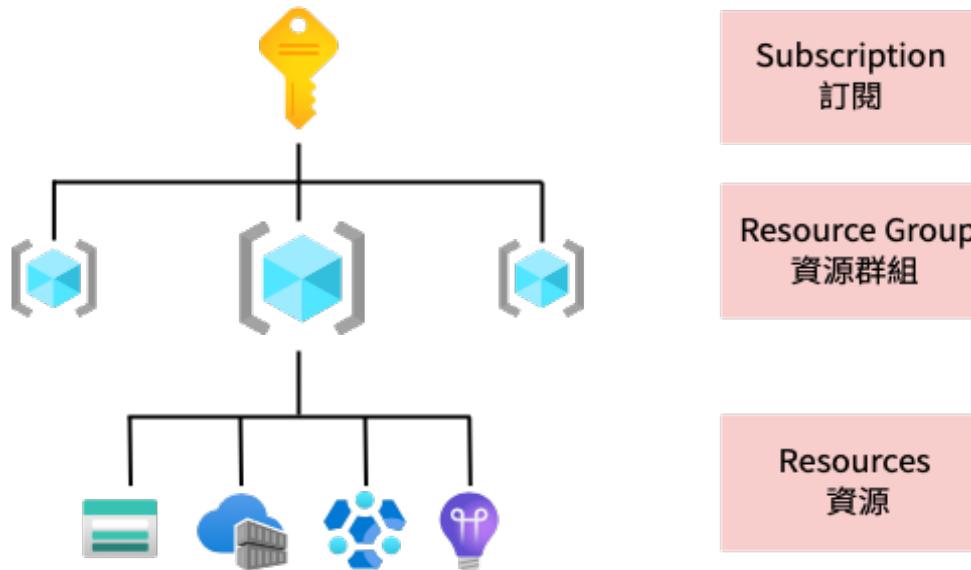
課程大綱

- 本課程的目的要體驗如何使用 Azure 雲端服務的各項工具
- 本次上課主要介紹 Azure Machine Learning 中的：
 - ◆ 自動化ML, Automated ML
由 Azure 將機器學習中的重複工作自動化，不用會寫程式也可以進行機器學習
 - ◆ Azure Databricks
由 Databricks 在 Microsoft Azure 上提供的雲端服務，分為兩種 - Azure Databricks SQL Analytics, Azure Databricks Workspace，分別對應到 SQL, Spark

Azure環境介紹 -1



- Azure環境可分為多層次的架構，可以分層進行授權管理
- 友達有一個企業戶的訂閱Subscription，其下依照單位、專案等設定不同資源群組，在每一個資源群組內按照需求建立不同的資源(AutoML, Azure Databricks等)
- 建立資源一定要有相對應的訂閱、資源群組



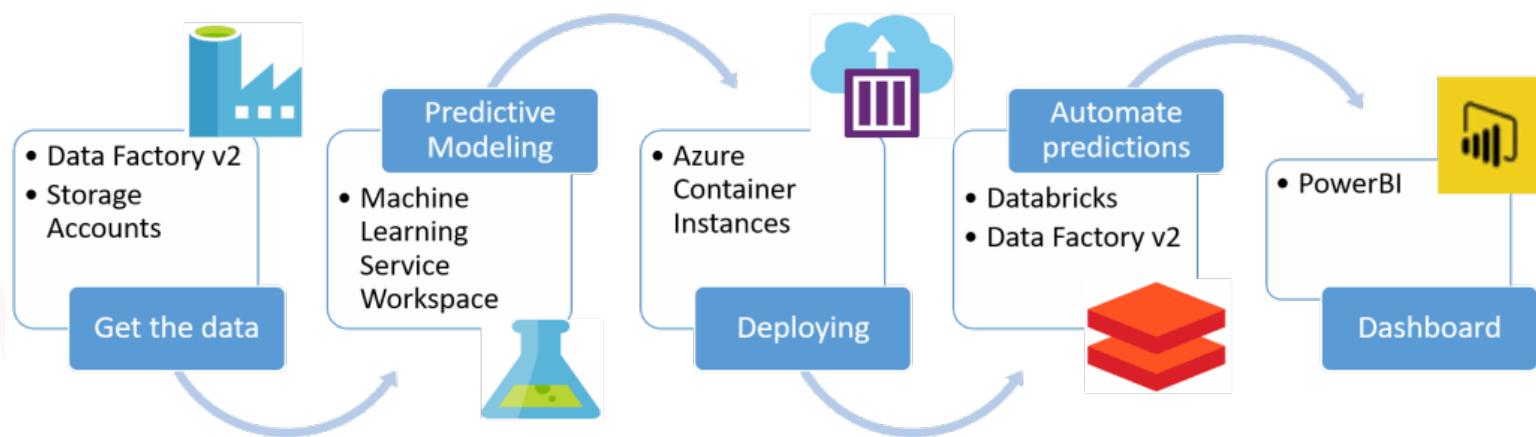
Azure環境介紹 -2

- Azure雲端上的服務資源是各自建立的，例如建立虛擬機器同時需要建立或連結到現有的儲存空間、虛擬網路等

Name	Type	Location
 MediaWebServer	Virtual machine	Japan East
 MicrosoftMonitoringAgent (Me...)	Microsoft.Compute...	Japan East
 MediaWebServer_OsDisk_1_b70...	Disk	Japan East
 mediawebserver285	Network interface	Japan East
 MediaWebServer-ip	Public IP address	Japan East
 MediaWebServer-nsg	Network security gr...	Japan East
 mslearn-vmupdate-vnet	Virtual network	Japan East

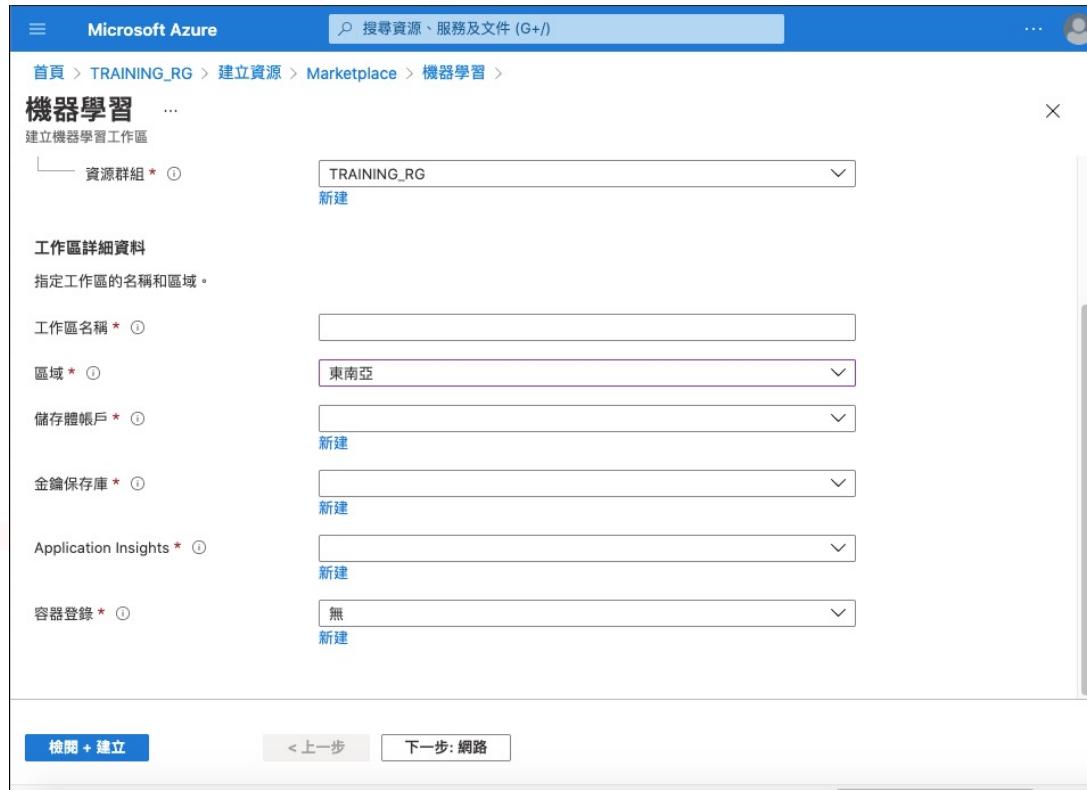
Azure雲端建置說明 -1

- Azure 雲端提供一系列的機器學習服務資源，從Raw Data的存放、資料處理、機器學習及視覺化呈現，都可以在Azure上找到相對應的資源。



Azure雲端建置說明 -2

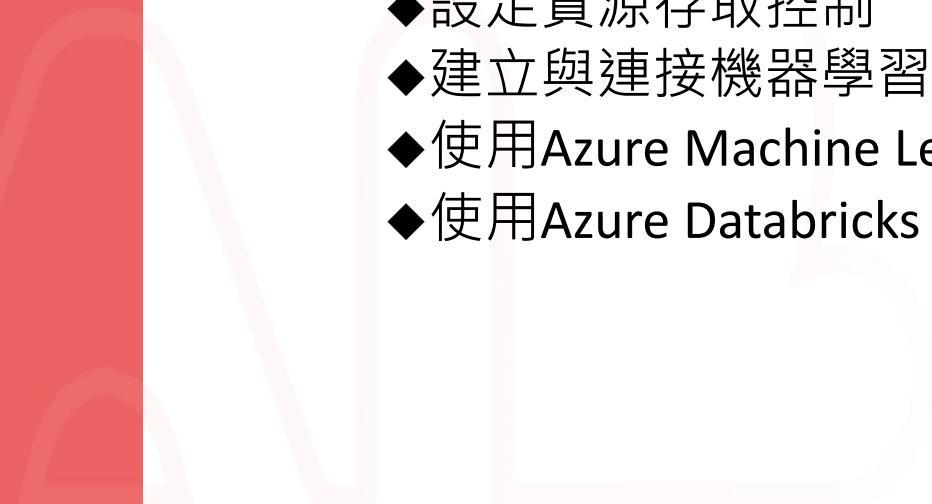
- Auto ML 是屬於 Azure Machine Learning Studio 中的一項服務
- 建立 Machine Learning Studio，會需要有相依的服務，如，存放紀錄的 Storage Account、金鑰保存庫、給予模型建議的 Application Insights、註冊機器學習產生 Docker Image 的容器登錄等資源。



Azure雲端建置說明 -3



- 透過本課程練習使用 Azure Machine Learning Studio，可以理解並熟悉 Azure 雲端各項服務的操作與使用，包括
 - ◆ 建立儲存體
 - ◆ 使用 Azure 儲存體總管傳輸資料
 - ◆ 註冊應用程式
 - ◆ 設定資源存取控制
 - ◆ 建立與連接機器學習工作區
 - ◆ 使用 Azure Machine Learning Studio 的自動化 ML
 - ◆ 使用 Azure Databricks



AUO訓練 實作環境



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks

AUO訓練實作環境介紹 -1

- 登入Azure Portal <https://portal.azure.com/>，點選Resource Group資源群組 "TRAINING_RG"

The screenshot shows the Microsoft Azure portal interface. At the top, there's a blue header bar with the Microsoft Azure logo, a search bar, and various navigation icons. On the right side of the header, the user's email (lihsiang.tu@caloudi.com) and company name (AU OPTRONICS CORP...) are displayed.

The main content area is titled "Azure 服務" (Azure Services). On the left, there's a sidebar with sections for "建立資源" (Create Resource), "最近的資源" (Recent Resources), and "瀏覽" (Browse). The "Recent Resources" section lists several items, including "TRAINING_RG".

The central part of the screen is the "資源群組" (Resource Group) blade for "TRAINING_RG". It shows a summary card with icons for SQL 資料庫, 適用於 PostgreSQL ..., Azure Cosmos DB, Kubernetes 服務, and a link to "更多服務" (More Services). Below this, there are sections for "最近的資源" (Recent Resources) and "Microsoft 提供的免費訓練" (Free training provided by Microsoft). The "Recent Resources" section shows "TRAINING_RG" was last checked 32 minutes ago. The "Free Training" section shows a course titled "使用 Azure Resource Manager 控制及組織 Az..." with 8 units and 46 minutes of content, last checked 19 hours ago.

At the bottom of the blade, there are links for "概觀" (Overview), "開始使用" (Get Started), and "文件" (Files). Navigation links at the very bottom include "訂用帳戶" (Subscription), "資源群組" (Resource Group), "所有資源" (All Resources), and "儀表板" (Dashboard).

AUO訓練實作環境介紹 -2

- 進到資源群組”TRAINING_RG”的”概觀”畫面，您可以看到在此資源群組下的所有資源，如儲存體帳戶、機器學習等等

The screenshot shows the Microsoft Azure Resource Group Overview page for the 'TRAINING_RG' resource group. The left sidebar contains navigation links for Home, TRAINING_RG (selected), Overview, Activity Log, IAM, Tags, Events, Settings (Deployments, Security, Principles, Properties, Locks), Cost Management (Cost Analysis, Budgets, Cost Alarms), and Monitoring (Insights, Alerts, Metrics). The main content area displays the 'Overview' tab with the following details:

- Subscription:** auobigdata (changed)
- Deployment Status:** 4 successful
- Deployment Identifier:** a7bdf2e3-b855-4dda-ac93-047ff722cbbd
- Location:** Southeast Asia
- Tags:** Clickable link to add new tags.
- Filtering:** Options to filter by type (All) and location (All), and a search bar.
- Resource List:** A table showing 8 resources:

Name	Type	Location	Actions
ames-ml-deploy-4iNLQWdEVEO_tnaLTv1i_A	Container Instance	Southeast Asia	...
autrainingadls	Storage Account	Southeast Asia	...
trainingdbs	Azure Databricks	Southeast Asia	...
trainingml	Machine Learning	Southeast Asia	...
trainingml3253432003	Time Series Insights	Southeast Asia	...
trainingml5524565781	Storage Account	Southeast Asia	...
trainingml6944426921	Application Insights	Southeast Asia	...
trainingmlcr	Container Registry	Southeast Asia	...

AUO訓練實作環境介紹 -3

- 由於上課學員較多，已經將各位依照工號分組，請記得您所屬的小組編號，後續使用的資源將與小組編號有關
- 上課示範用的資料集及相關資源都放在Azure Portal的儲存體帳戶**dllakedemo**
- 廠端使用WVD的學員，可以在T槽中找到一樣的資源

The screenshot shows the Azure Storage Account 'dllakedemo' interface. On the left, there's a sidebar with a list of storage accounts. The account 'dllakedemo' is selected and highlighted in grey. The main pane displays the 'Containers' blade for the 'dllakedemo' account. It includes a search bar, a toolbar with options like '+ Container', 'Change access tier', 'Restore container', 'Refresh', and 'Delete', and a table listing existing containers: 'databricks', 'designer tool', and 'sample'. The table columns are 'Name', 'Last modified', 'Access tier', and 'Status'.

名稱	上次修改	公用存取層級	租用狀況
databricks	2021/8/2 下午7:11:44	容器	可用
designer tool	2021/7/29 下午3:20:09	容器	可用
sample	2021/7/27 下午4:44:08	容器	可用

AUO訓練實作環境介紹

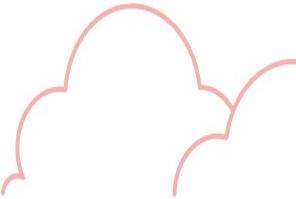
-4

資源名稱	類型	內容
建立workspace自動產生	Application Insights	Azure對資源部署的智慧建議
dllakedemo	儲存體帳戶 Storage Account	ADLS Gen2，主要放置上課使用的講義和資料集
建立workspace自動產生	儲存體帳戶 Storage Account	主要存放ML Workspace相關Log紀錄
建立模型後產生	容器執行個體 Container Instances	部署ML設計工具產生的模型
建立模型後產生	容器登錄 Container Registry	Container Registry，因Cluster產生的資源
AutoML員工編號	機器學習 Machine Learning	機器學習的工作站workspace
建立workspace自動產生	金鑰保存庫 Key Vault	儲存資源群組的金鑰

Azure儲存體介紹

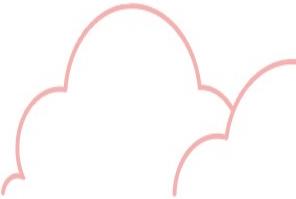


- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks



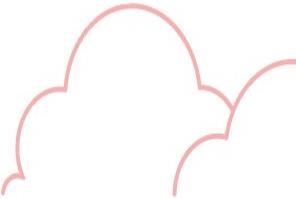
Azure Blob 儲存體 -1

- Azure Blob 儲存體是 Microsoft 針對雲端推出的物件儲存體解決方案，Blob 儲存體已針對儲存大量非結構化資料進行最佳化
- Blob 儲存體設計用來：
- 直接提供映像或文件給瀏覽器
- 儲存檔案供分散式存取
- 串流影片和音訊
- 寫入記錄檔
- 儲存備份和還原、災害復原和封存資料
- 儲存資料供內部部署或 Azure 裝載服務進行分析



Azure Blob 儲存體 -2

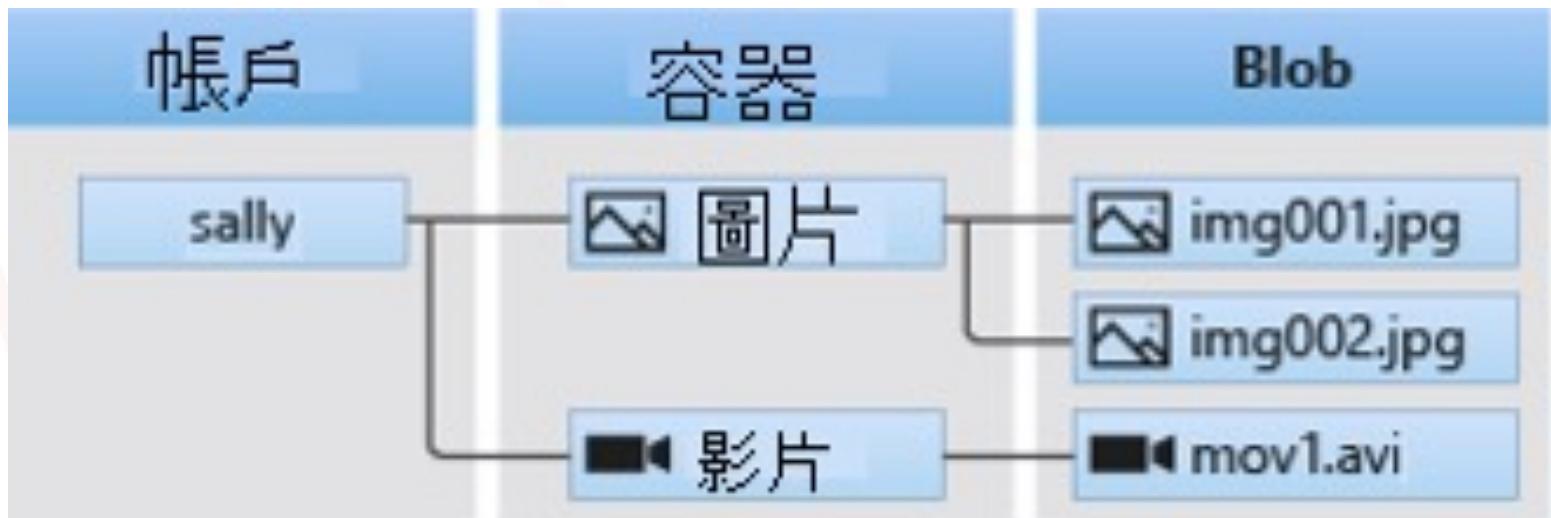
- 使用者或用戶端應用程式可以從世界各地透過 HTTP/HTTPS 存取 Blob 儲存體中的物件
- Blob 儲存體中的物件可透過 Azure 儲存體 REST API、Azure PowerShell、Azure CLI 或 Azure 儲存體用戶端程式庫存取
- 用戶端程式庫適用於不同的語言，包括：
 - ◆.NET
 - ◆Java
 - ◆Node.js
 - ◆Python
 - ◆Go
 - ◆PHP
 - ◆Ruby

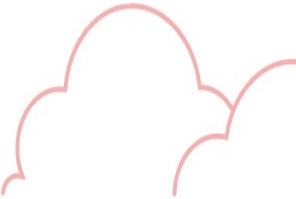


Blob 儲存體資源 -1

- Blob 儲存體提供三種類型資源：

- ◆ 儲存體帳戶
- ◆ 儲存體帳戶中的容器
- ◆ 容器中的 Blob





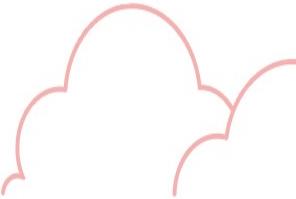
Blob 儲存體資源 -2

- 儲存體帳戶

- ◆ 儲存體帳戶會在 Azure 中為您的資料提供唯一命名空間。每個儲存在 Azure 儲存體中的物件都有一個位址，其中包含您的唯一帳戶名稱。帳戶名稱與 Azure 儲存體 Blob 端點的組合會形成儲存體帳戶中物件的基底位址
 - ◆ 例如: <http://mystorageaccount.blob.core.windows.net>

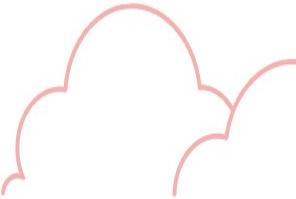
- 容器

- ◆ 容器會組織一組 Blob，類似於檔案系統中的目錄。儲存體帳戶可以包含無限數量的容器，而一個容器則可儲存無限數量的 Blob



Blob 儲存體資源 -3

- Azure 儲存體支援三種 blob 類型：
 - ◆ 區塊 Blob，可儲存文字和二進位資料。區塊 Blob 是由可個別管理的資料區塊所組成。區塊 blob 最多可以儲存約 190.7 TiB
 - ◆ 附加 Blob，和區塊 Blob 相似，由區塊所組成，但已針對附加作業最佳化。附加 Blob 很適合某些案例，例如記錄虛擬機器中的資料
 - ◆ 分頁 Blob，可儲存隨機存取檔案 (大小上限為 8 TiB)。分頁 Blob 存放區可儲存虛擬硬碟 (VHD) 檔案，以作為 Azure 虛擬機器的磁碟



Azure Data Lake Store

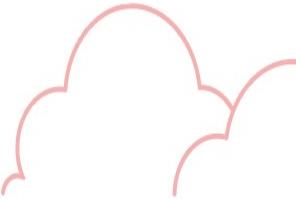
- Azure Data Lake Store提供了一個單一的存儲庫，無論大小型組織都可以在其中上傳無限大小的數據
- 用於Hadoop分佈式檔案系統工具和應用程式的高性能處理和分析，包括對低延遲工作負載的支源
- 允許使用本機格式的結構化和非結構化數據
- 允許巨大的資料流量以提高分析性能
- 提供了高可用性，耐用性和可靠性
- ADLS的類型
 - ◆ADLS Gen1
 - ◆ADLS Gen2

ADLS Gen1 -1

- 全名為 Azure Data Lake Storage Gen1
- Azure Data Lake Storage Gen1 是容納巨量資料分析工作負載的企業級超大規模存放庫。 Azure Data Lake 可讓您在單一位置擷取任何大小、類型和擷取速度的資料，以便進行運作和探究分析
- 使用與 WebHDFS 相容的 REST API，可以從 Hadoop (HDInsight 叢集所提供之) 存取 Data Lake Storage Gen1
- 為預存資料分析而設計，並針對資料分析案例而調整效能
- Data Lake Storage Gen1 包含所有企業級功能：安全性、管理性、延展性、可靠性和可用性
- 2024 年 2 月 29 日 Azure Data Lake Storage Gen1 將會淘汰

ADLS Gen1 -2

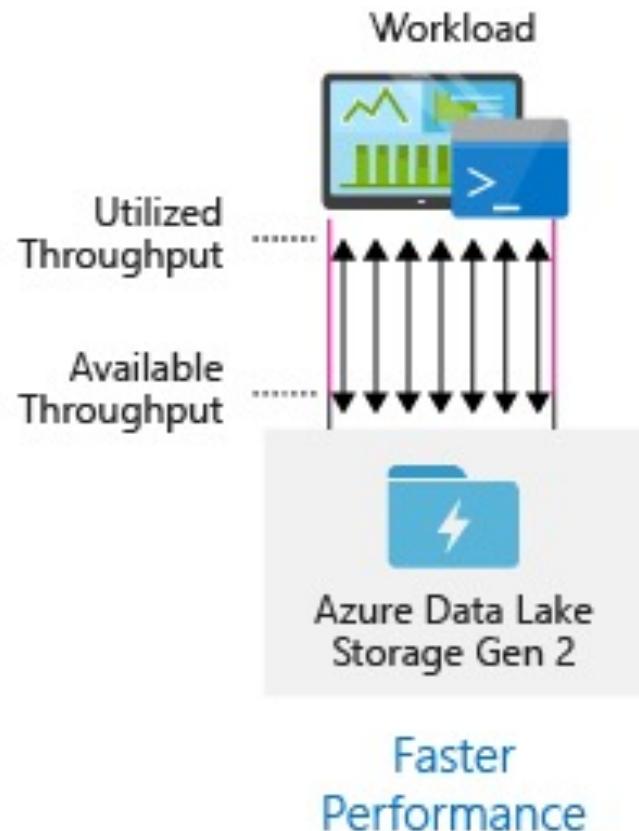
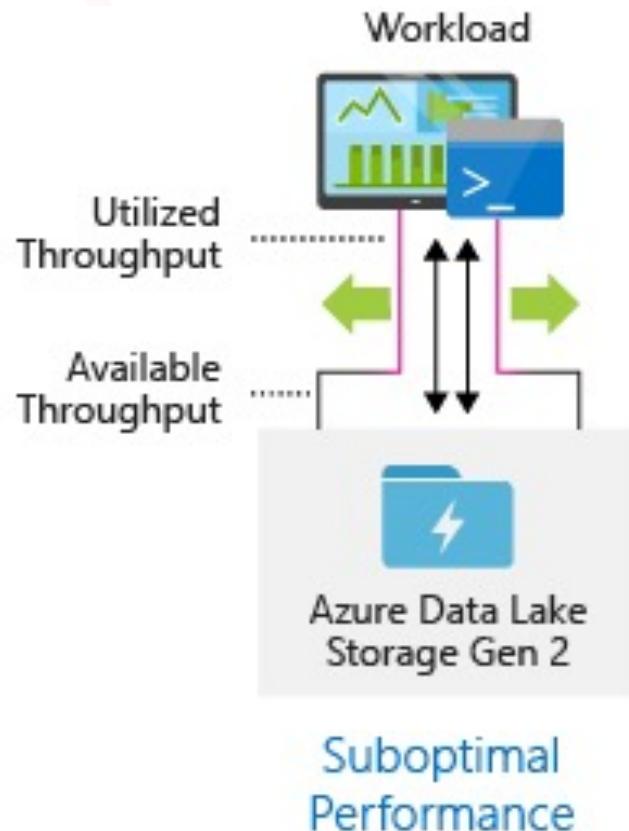




ADLS Gen2 -1

- Azure Data Lake Storage Gen2 是一組巨量資料分析的專屬功能，內建於 Azure Blob 儲存體
- Data Lake Storage Gen2 包含 Azure Data Lake Storage Gen1 的功能和 Azure Blob 儲存體
- 例如，Data Lake Storage Gen2 會提供檔案系統語法、檔案層級安全性和規模調整。因為這些功能是建置於 Blob 儲存體之上，所以您也可以取得低成本的分層式儲存體，以及高可用性/災害復原功能
- 效能經過最佳化，相較於 Blob 儲存體上的一般命名空間，階層命名空間大幅提高目錄管理作業的效能，從而提高整體作業效能。
- 管理更容易，可以透過目錄和子目錄整理和操作檔案
- 安全性是強制的，可以在目錄或個人檔案上定義 POSIX 權限

ADLS Gen2 -2



AUO 實作環境說明

- 建立ADLS Gen2 作為Big Data 儲存空間
- 建立與Machine Learning相依的Storage Account，儲存機器學習中的預設資料
- 本次課程以 **dllakedemo** 為放置公用資料集的儲存體帳戶

首頁 >

儲存體帳戶

AUO vIDE (auovide.onmicrosoft.com)

+ 建立 ⚙ 管理檢視 ▾ 重新整理 ⬇ 匯出至 CSV ⚡ 建立查詢 | 🔍 指派標籤 🗑 刪除 | 💬 意見反應

篩選任何欄位... 訂用帳戶 == **azuretrain01** 資源群組 == 全部 × 位置 == 全部 × + 加入篩選

正在顯示 1 至 87 筆記錄 (共 87 筆)。

<input type="checkbox"/> 名稱 ↑↓	類型 ↑↓	種類 ↑↓	資源群組 ↑↓
<input type="checkbox"/>  dllakedemo	儲存體帳戶	StorageV2	Training_RG
<input type="checkbox"/>  filegwadls	儲存體帳戶	StorageV2	Training_RG
<input type="checkbox"/>  m1909050	儲存體帳戶	StorageV2	M_1909050
<input type="checkbox"/>  mlstimyang4671040679	儲存體帳戶	StorageV2	ml-timyang
<input type="checkbox"/>  storageaccounttest08065	儲存體帳戶	Storage	test-0805126-RG
<input type="checkbox"/>  storageaccounttest1ace3	儲存體帳戶	Storage	test-1810052-RG
<input type="checkbox"/>  storagem00	儲存體帳戶	StorageV2	M_00_RG
<input type="checkbox"/>  test2106storage7417c2a86	儲存體帳戶	StorageV2	test-2106012-RG
<input type="checkbox"/>  test2107storage04a9b228c	儲存體帳戶	StorageV2	test-2107116-RG
<input type="checkbox"/>  testdavistoragef8fbfb1cf5	儲存體帳戶	StorageV2	test-david11-RG

< 上一頁 畫面 1 /1 下一頁 >

建立 Data Lake Storage Gen2 存放區



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks

建立資源群組 -1

- 建立新的資源群組來保存 Data Lake 儲存體
- 使用資源群組來管理相關服務和應用程式
- 資源群組也會讓清除資源的過程更為容易
- 登入 Azure 入口網站
- 選取左上方的 [+建立資源] 。
- 在搜尋方塊中，鍵入“Resource”，並從結果選取“Resource Group”
- 選取 [建立] 來新增新的資源群組
- 在 [基本] 索引標籤上，選取您想要使用的訂用帳戶，本次課程使用的是“**azuretrain01**”
- 將資源群組命名，您在此處建立的是“**automl員工編號**”
- 選擇資源群組的區域 (位置)，**小組1-10在東南亞Southeast Asia**、**小組11-20在東亞eastasia**

建立資源群組 -2

- 選取 [檢閱 + 建立] > [建立]

首頁 > 資源群組 >

建立資源群組 ...

基本 標籤 檢閱 + 建立

資源群組 - 擁有 Azure 解決方案相關資源的容器。該資源群組可包含解決方案的所有資源，或是只包含您想以群組方式管理的資源。您可以最適合您組織的方式，決定如何將資源配置到資源群組。[深入了解](#)

專案詳細資料

訂用帳戶 * ⓘ

Visual Studio Enterprise 訂閱 – MPN

資源群組 * ⓘ

mslearn-auto

資源詳細資料

區域 * ⓘ

(Asia Pacific) 東南亞

[檢閱 + 建立](#)

< 上一步

[下一步 : 標籤 >](#)

建立 Data Lake 帳戶 -1

- 在 Azure 入口網站功能表的選單，選取 [儲存體帳戶] > [建立]

The screenshot shows the Azure portal's main interface. On the left, there is a navigation sidebar with various service links like '建立資源', '首頁', '儀表板', etc., and a specific link '儲存體帳戶' which is highlighted with a red box. The main content area is titled 'Azure 服務' and contains several service icons. In the center, there is a detailed view for '儲存體帳戶' (Storage Accounts). It shows a '建立' (Create) button with a red box around it. Below it is a '描述' (Description) section with text about storage accounts. To the right, there is a table with one row showing '訂用帳戶' (Subscription) and '上次檢視時間' (Last checked time) as '10 小時之前' (10 hours ago). At the bottom of the main content area, there are sections for 'Microsoft 提供的免費訓練' (Free Microsoft training), 'Azure 監視器' (Azure Monitor), '資訊安全中心' (Information Security Center), and '成本管理' (Cost Management). The bottom of the page shows the URL 'https://portal.azure.com/#blade/HubsExtension/BrowseResourceBlade/resourceType/Microsoft.Storage%2FStorageAccounts'.

建立 Data Lake 帳戶 -2

- 選取您的“訂用帳戶azuretrain01”和“資源群組Training_RG”。
- 輸入儲存體帳戶的名稱。此名稱在整個 Azure 中必須是唯一的。您可能必須嘗試數種變化，才能找到唯一的名稱。本次課程使用“**dlakedemo**”為例，您在命名時可以使用首碼並搭配一些數字。綠色核取記號表示您輸入的名稱有效。
- 選擇資源群組的區域(位置)為東南亞
- 確認效能為標準(一般用途 V2)
- 備援更改為本地備援儲存體(LRS)
- 其餘欄位則保留預設值
- 選取上方的 [進階] 標籤

建立 Data Lake 帳戶 -3

建立儲存體帳戶 ...

基本 進階 網路 資料保護 標籤 檢閱 + 建立

選取要在其中建立新儲存體帳戶的訂閱。請選擇新的或現有的資源群組，以便一併整理和管理您的儲存體帳戶與其他資源。

訂用帳戶 *

azuretrain01

資源群組 *

autoDemo

新建

執行個體詳細資料

如果您需要建立舊版儲存體帳戶類型，請按一下[這裡](#)。

儲存體帳戶名稱 ① *

dlakecaloudi

區域 ① *

(Asia Pacific) 東南亞

效能 ① *

標準: 建議用於大多數案例 (一般用途 v2 帳戶)

進階: 建議用於需要低延遲的案例。

備援 ① *

本地備援儲存體 (LRS)

檢閱 + 建立

< 上一步

下一步 : 進階 >

建立 Data Lake 帳戶

-4

- 在 [Data Lake Storage Gen2] 區段的 [階層命名空間] 旁邊，選取 [啟用]。
- 選取 [檢閱 + 建立]
- 在驗證建立詳細資料後，選取 [建立] 來啟動部署
- 當您收到訊息，指出您的部署已完成時，請選取 [前往資源] 來確認部署

建立儲存體帳戶 ...

基本 進階 網路 資料保護 標籤 檢閱 + 建立

啟用 Blob 公用存取

啟用儲存體帳戶金鑰存取

TLS 最低版本

Data Lake Storage Gen2
Data Lake Storage Gen2 階層式命名空間可加速巨量資料分析工作負載，並啟用檔案層級存取

啟用階層命名空間

Blob 儲存體

啟用網路檔案共用 v3

存取層 經常性: 經常存取的資料與例行使用案例
 非經常性: 不常存取的資料與備份案例

建立 Data Lake 帳戶 -5

首頁 >

dlakecaloudi 儲存體帳戶

搜尋 (Cmd +/)

在總管中開啟 刪除 移動 重新整理 意見反應

Microsoft 建議您升級至新的警報平台，以確保您的警報不會發生中斷。傳統警報將自 2021 年起淘汰。請升級至新的警報平台。 [深入了解](#)

JSON 檢視

概觀

活動記錄

標籤

診斷並解決問題

存取控制 (IAM)

資料移轉

事件

儲存體總管 (預覽)

資料儲存

容器

檔案共用

佇列

資料表

安全性 + 網路

網路

存取金鑰

共用存取簽章

加密

程式集

資源群組 (變更) : autoDemo

位置 : 東南亞

訂用帳戶 (變更) : azuretrain01

訂用帳戶識別碼 : a278626c-1d76-422d-a5b2-4c2cf191a0ac

磁碟狀態 : 可用

效能/存取層 : 標準/經常性

複寫 : 本地備援儲存體 (LRS)

帳戶種類 : StorageV2 (一般用途 v2)

佈建狀態 : 成功

已建立 : 2021/7/27 下午3:26:48

標籤 (變更) :

屬性 監視 功能 (5) 建議 教學課程 開發人員工具

Data Lake Storage

階層式命名空間	已啟用	需要 REST API 作業的安全傳輸	已啟用
預設存取層	Hot	儲存體帳戶金鑰存取	已啟用
Blob 公用存取	已啟用	TLS 最低版本	版本 1.2
Blob 虛刪除	已停用	基礎結構加密	已停用
容器虛刪除	已啟用 (7 天)		
版本設定	已停用		
變更摘要	已停用		

安全性

儲存體帳戶金鑰存取	已啟用
TLS 最低版本	版本 1.2
基礎結構加密	已停用

網路

允許存取自	所有網路
私人端點連線數目	0

上傳檔案



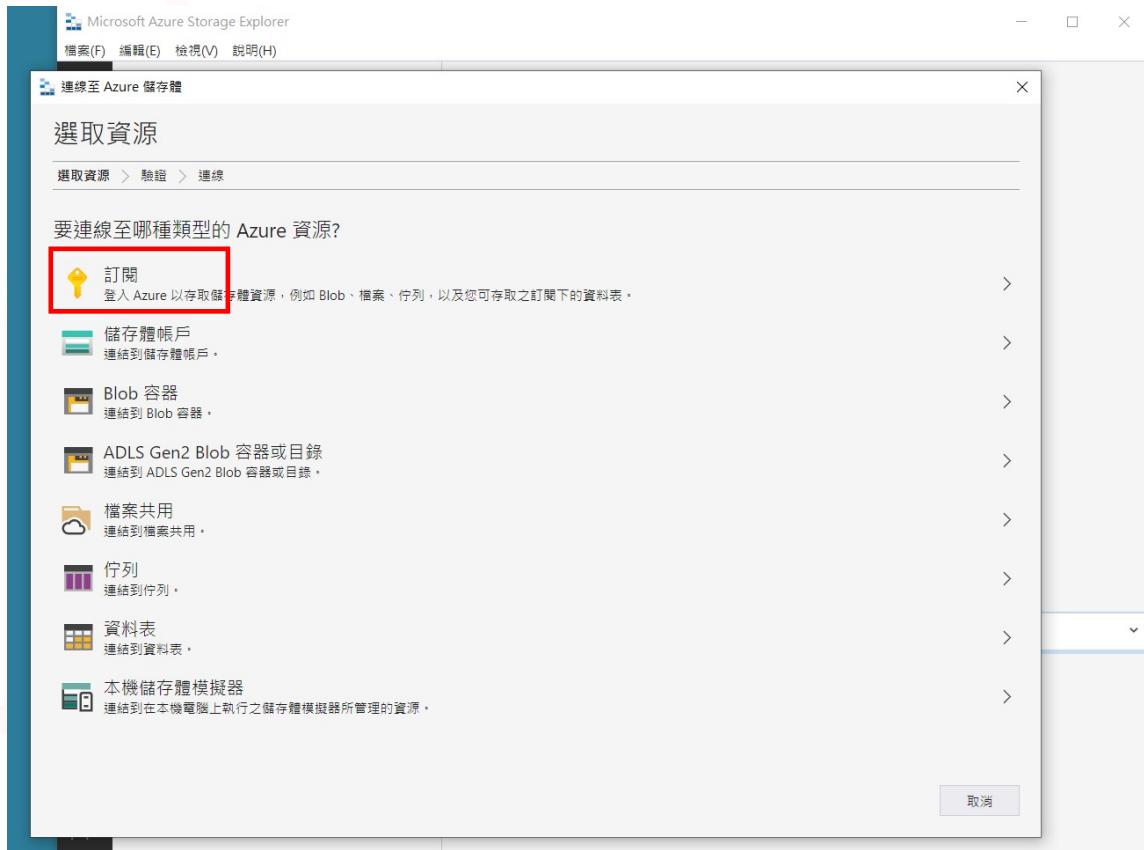
- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks

Azure 儲存體總管

- 若要對 Azure Data Lake 存放區執行臨機操作資料傳輸，請使用 Azure 儲存體總管。
- 儲存體總管是適用於 Windows、macOS 和 Linux 的免費應用程式
- 這項 Azure 應用程式旨在管理非結構化的資料，例如資料表、Blob、佇列和檔案
- 它也支援 Azure Cosmos DB 和 Azure Data Lake Storage Gen2 中的資料
- 下載及安裝儲存體總管
 - ◆ <https://azure.microsoft.com/zh-tw/features/storage-explorer/>

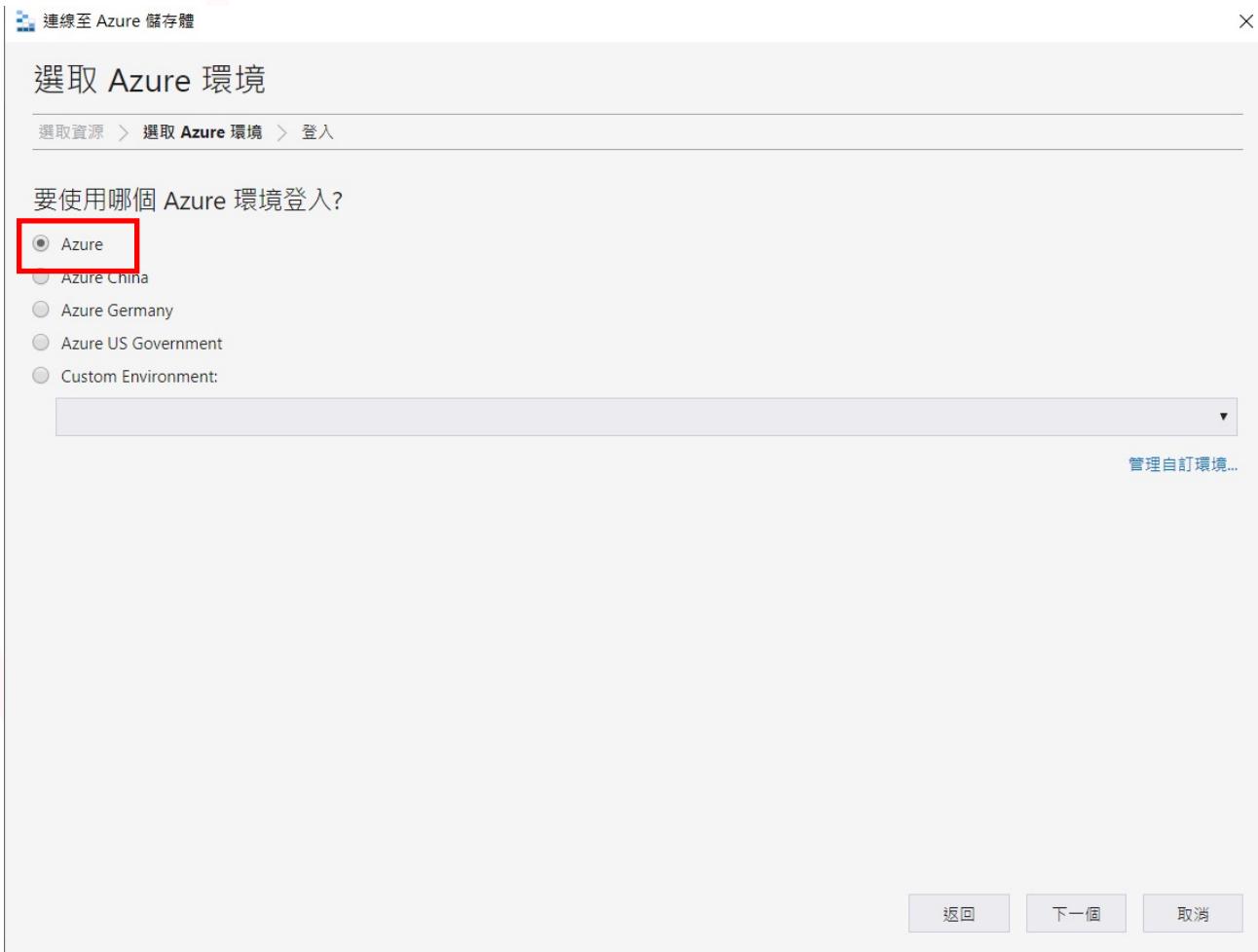
設定帳號 -1

- 若要開始使用 [儲存體總管]，請先新增您的 Azure 帳戶
- 選擇 [訂閱]



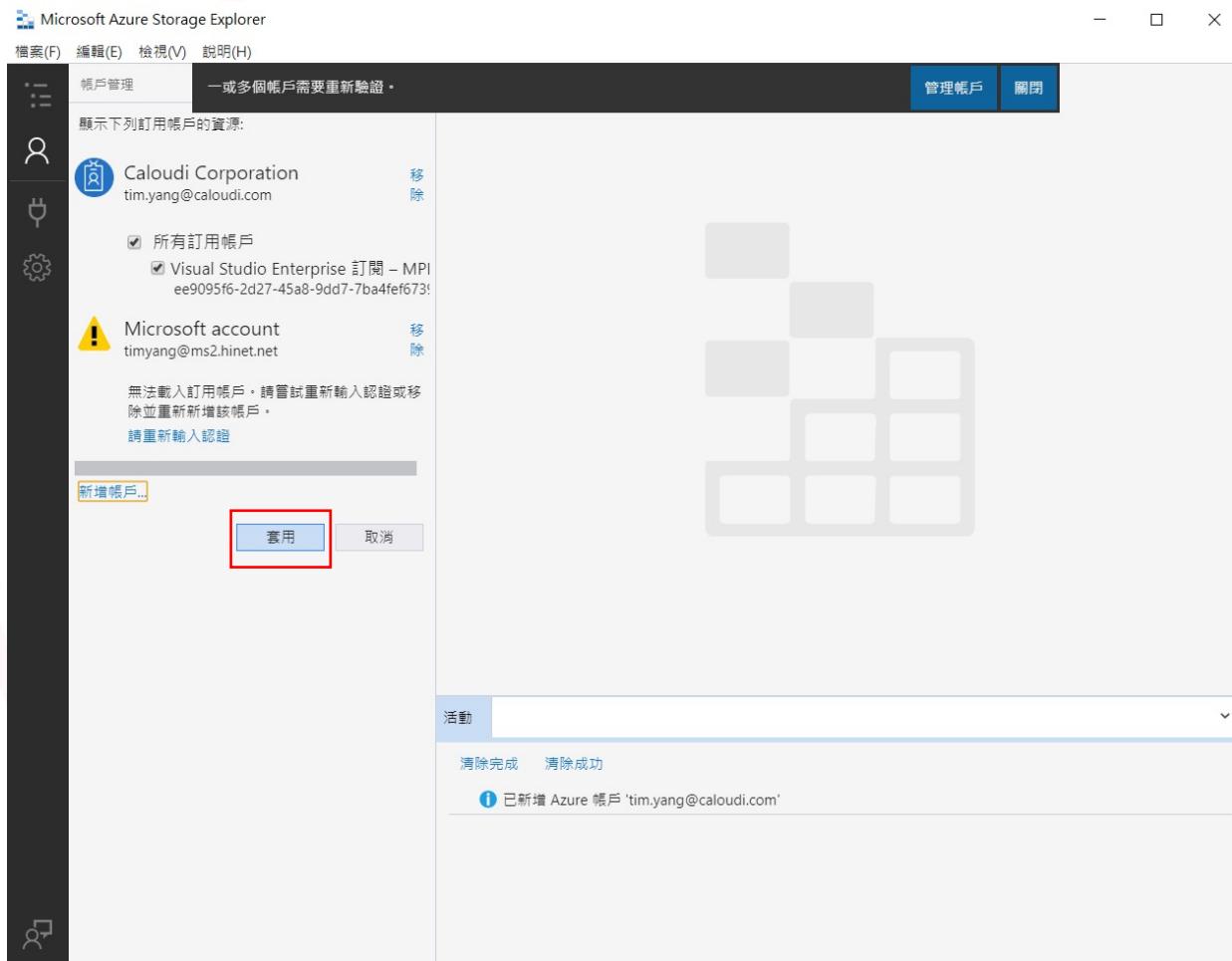
設定帳號 -2

● 選擇 [Azure] 登入



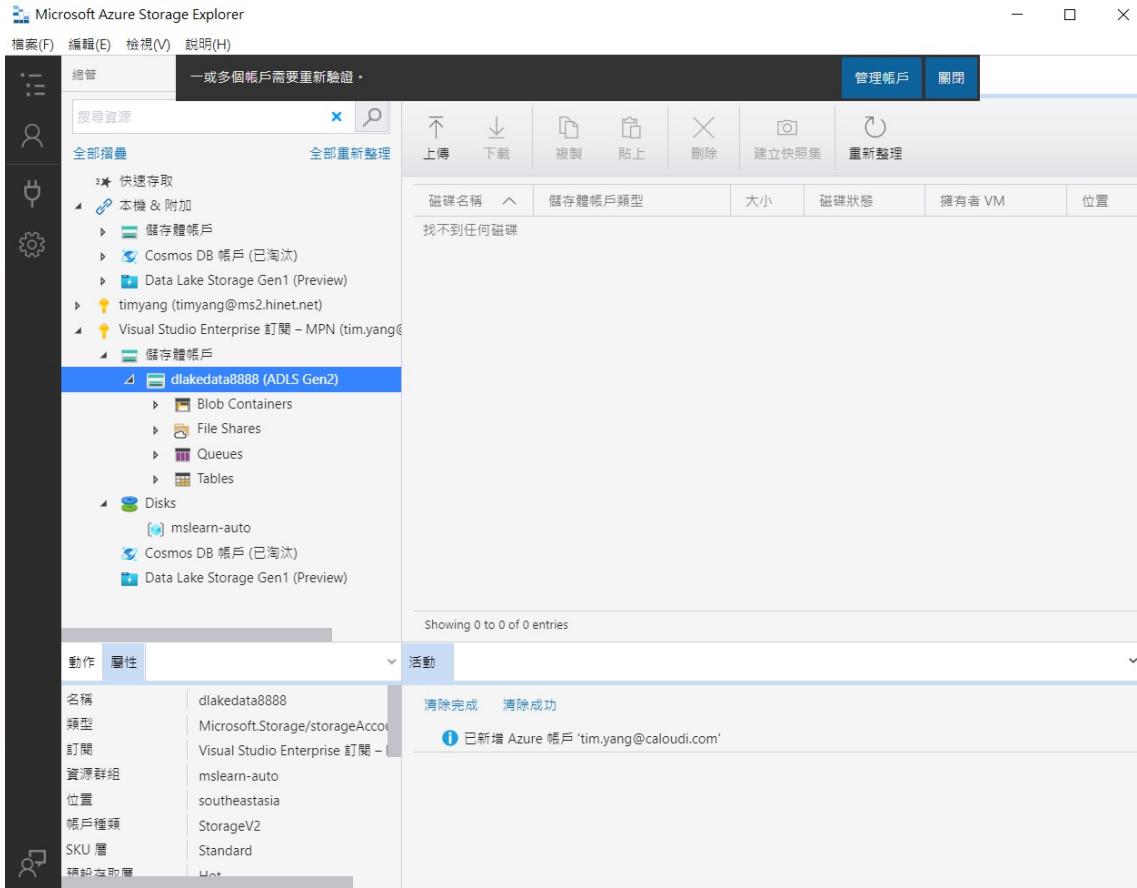
設定帳號 -3

- 在您登入後，請選取您要使用的訂用帳戶。請務必選取您已在其中建立 Azure 儲存體帳戶的訂用帳戶。



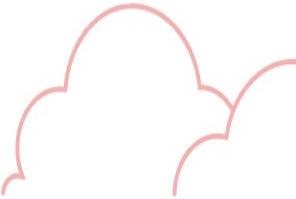
設定帳號 -4

- 應用程式會顯示您可以從訂用帳戶中使用的儲存體區域清單。您的 Azure 儲存體帳戶應該會出現在清單上



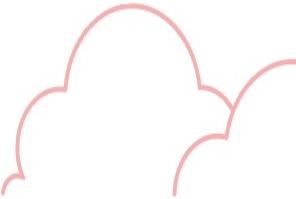
建立檔案系統

- 因為 Blob 一律會上傳至資料夾，因此您可以組織 Blob 群組，其方式就跟組織您電腦上檔案的方式雷同
- 在 Data Lake Storage Gen2 中，您會從建立「檔案系統」開始。這會定義將保存您 Data Lake 的 Blob 儲存體容器。您會在此專用區域內建立資料夾和檔案
- 在 [儲存體總管] 中，展開您的訂用帳戶，然後展開您的儲存體帳戶
- 展開您先前建立的儲存體帳戶，然後選取 [Blob 容器]
- 以滑鼠右鍵按一下 [Blob 容器]，然後選取 [建立 Blob 容器]
- 在 [Blob 容器] 下方，輸入 demodata
- 建立容器後，請選取 demodata



建立資料夾

- 資料夾可協助為您的資料提供階層結構。您可以在帳戶中建立多個層級。您只需要確保在您建立子資料夾前，主要資料夾已存在
- 在位於 [儲存體總管] 頂端的功能表上，選取 [新增資料夾]
- 將資料夾命名為 sample。然後選取 [確定] 來建立目錄。若您收到訊息，詢問您是否要重新整理您的檢視，請選取 [是]
- 按兩下新的資料夾。此動作會進入資料夾，此時資料夾應該是空白的
- 建立另一個名為 data 的資料夾。



上傳檔案

- 您可以將檔案和資料夾從本機電腦上傳至位於您 [儲存體總管] 右側檔案共用中的目錄
- 在 [儲存體總管] 中，按兩下 data 資料夾
- 在視窗頂端的功能表上，選取 [上傳]。從 [上傳] 功能表中，您可以上傳資料夾或檔案
- 選取 [上傳檔案]
- 選取您要上傳檔案
- 在 [上傳檔案] 對話方塊中，確認目的地目錄是 sample/data。然後選取 [上傳]

註冊應用程式



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks

應用程式註冊 -1

- Microsoft 身分識別平臺只會對已註冊的應用程式執行 (IAM) 的身分識別和存取管理
- 無論是用戶端應用程式 (例如 web 或行動應用程式)，或支援用戶端應用程式的 Web API，註冊會在您的應用程式與身分識別提供者 (Microsoft 身分識別平台) 之間建立信任關係
- 登入 Azure 入口網站
- 搜尋並選取 [Azure Active Directory]
- 在“管理”下選取 [應用程式註冊] > [新增註冊]
- 輸入應用程式的顯示“名稱”，例如“MyDemoApp”
- 指定可以使用應用程式的人員，預設為“僅此組織目錄中的帳戶”

應用程式註冊 -2

● 選取 [註冊] 以完成伺服器初始註冊

首頁 > 預設目錄 >
註冊應用程式 ...

* 名稱

此應用程式的使用者互動顯示名稱 (之後可以變更)。

MyDemoApp



支援的帳戶類型

誰可以使用此應用程式或存取此 API?

- 僅此組織目錄中的帳戶 (僅 預設目錄 - 單一租用戶)
- 任何組織目錄中的帳戶 (任何 Azure AD 目錄 - 多租用戶)
- 任何組織目錄中的帳戶 (任何 Azure AD 目錄 - 多租用戶) 和個人 Microsoft 帳戶 (例如 Skype、Xbox)
- 僅限個人 Microsoft 帳戶

[協助我選擇...](#)

重新導向 URI (選用)

我們會在成功驗證使用者之後，將驗證回應傳回給此 URI。現在已不一定需要在此時提供此 URL，可以在之後變更，但在大多數的情況下會需要值。

Web



e.g. https://example.com/auth

[繼續即表示您同意《Microsoft 平台原則》](#) ↗

[註冊](#)

應用程式註冊 -3

- 註冊完成時，Azure 入口網站會顯示應用程式註冊的 [總覽] 窗格
- 您會看到“應用程式 (用戶端) 識別碼”，此值也稱為“用戶端 識別碼”，可在 Microsoft 身分識別平臺中唯一識別您的應用程式

The screenshot shows the Azure portal interface for managing app registrations. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information (meganb@contoso.com, CONTOSO AD (DEV)). Below the navigation is a breadcrumb trail: Home > Contoso AD (dev) | App registrations > Contoso App 1. The main content area is titled "Contoso App 1". On the left, there's a sidebar with links like Overview, Quickstart, Integration assistant (preview), Manage, Branding, Authentication, Certificates & secrets, Token configuration, API permissions, and Expose an API. The "Overview" link is currently selected. The main pane displays various application details:

- Display name: Contoso App 1
- Application (client) ID: **11111111-1111-1111-1111-111111111111** (highlighted with a red box)
- Directory (tenant) ID: 00000000-0000-0000-0000-000000000000
- Object ID: 2222222-2222-2222-2222-222222222222
- Supported account types: My organization only
- Redirect URLs: Add a Redirect URI
- Application ID URI: Add an Application ID URI
- Managed application in local directory: Contoso App 1

A note at the bottom states: "Starting June 30th, 2020 we will no longer add any new features to Azure Active Directory Authentication Library (ADAL) and Azure AD Graph. We will continue to provide technical support and security updates but we will no longer provide feature updates. Applications will need to be upgraded to Microsoft Authentication Library (MSAL) and Microsoft Graph. [Learn more](#)".

新增用戶端密碼 -1

- 用戶端密碼也稱為“應用程式密碼”
- 它是一個字串值，您的應用程式可以使用此值來取代憑證本身的身分識別
- 在 Azure 入口網站的“應用程式註冊(App Registrations)”中，選取您的應用程式
- 選取 [憑證及祕密] > [新增用戶端密碼]
- 新增用戶端密碼的描述
- 選取持續時間
- 選取 [新增]。
- 記錄秘密的值，以便在您的用戶端應用程式程式碼中使用。離開此頁面後，就不會再次顯示此秘密值

新增用戶端密碼 -2

所有服務 > 應用程式註冊 > MyDemoApp

MyDemoApp | 憑證及祕密 ⚙ ...

The screenshot shows the Azure portal interface for managing application secrets. On the left, the navigation menu is visible with the '憑證及祕密' (Certificates & secrets) option selected. The main content area displays a form titled '新增用戶端密碼' (Add Client Secret). The '描述' (Description) field contains the placeholder text 'I'. The '到期' (Expires) section has three radio button options: '1 年內' (Within 1 year) is selected, while '2 年內' (Within 2 years) and '永不' (Never) are unselected. Below the form are two buttons: '新增' (Add) and '取消' (Cancel). A '新增' (Add) button is also located below the form. The '用戶端密碼' (Client Secret) section contains a note about its purpose: '應用程式在要求權杖時，用以證明其身分識別的祕密字串，也稱為應用程式密碼。' (A secret string used by the application to prove its identity when requesting an access token, also known as the application password.) Below this, a table lists the secret details:

描述	到期	值	ID
Password uploaded on Sat Apr 03 2021	3/4/2022	~_n*****	902ba91b-c42e-4757-acfd-0b308ccfe28f

The '值' (Value) column for the secret is highlighted with a red rectangle.

上課示範用密碼

- 應用程式名稱：

MyDemoApp

- Application/Client ID:

cf446eaf-3099-4179-94e3-20b51bc4a78f

- Object ID:

cece5a9f-8e44-4398-82af-4b39d6548af5

- Directory/Tenant ID:

e7c24d00-7479-4343-8247-0204699693e8

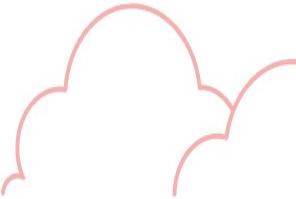
- Secrets Value:

SBoFq1TfYG2~Z1E_4N3i_ialkLs0bgN~ZU

設定資源 存取控制



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks



存取 Blob

- Azure Active Directory (Azure AD) 透過 AZURE RBAC(角色型存取控制) , 來授與受保護資源的存取權限
- Azure 儲存體會定義一組 Azure 內建角色，其中包含用來存取 blob 或佇列資料的一般許可權集
- 適用於 Blob 的 Azure 角色
 - ◆ 儲存體 Blob 資料擁有者：
用來設定擁有權和管理 Azure Data Lake Storage Gen2 的 POSIX 存取控制
 - ◆ 儲存體 Blob 資料參與者：
用來授與 Blob 儲存體資源的讀取/寫入/刪除權限
 - ◆ 儲存體 Blob 資料讀者：
用來授與 Blob 儲存體資源的唯讀權限

授與存取權 -1

- 在 [資源群組] 清單中，選擇資源群組”**automl員工編號**”
- 按一下 [存取控制 (IAM)]
 - ◆ Identity and access management
- 按一下 [角色指派] 索引標籤，以查看目前的角色指派清單
- 按一下 [新增] > [新增角色指派]，以開啟 [新增角色指派] 窗格
- 在 [角色] 下拉式清單中選取 [儲存體 Blob 資料參與者]
- 在 [選取] 文字框中，輸入選取您之前註冊的應用程式 (MyDemoApp)
- 按一下 [儲存] 以指派角色

授與存取權 -2

新增角色指派

×

角色 ①

儲存體 Blob 資料參與者 ①



存取權指派對象為 ①

使用者、群組或服務主體



選取 ①

Mydemo

找不到使用者、群組或服務主體。

選取的成員:

MyDemoAPP

移除

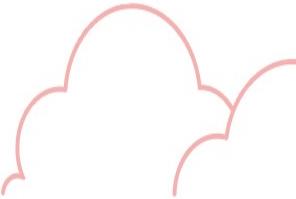
儲存

捨棄

Azure AutoML

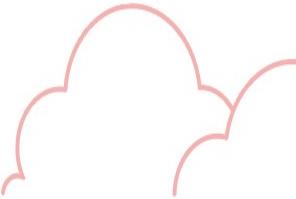


- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks



什麼是 Azure 機器學習

- Azure 機器學習是一項雲端服務
- 用來做資料建模，測試和部署預測性分析的解決方案
- 自動化機器學習(AutoML) 是將機器學習模型開發中耗時的反覆工作加以自動化的程序
- 可讓資料科學家、分析師和開發人員以高擴充性、效率和生產力來建置 ML 模型，同時維持模型品質
- 傳統機器學習模型的開發會耗用大量資源，需要大量的領域知識和時間來產生和比較數十種模型。透過自動化機器學習，您將可以更輕鬆、更高效率地縮短取得生產就緒機器學習模型所需的時間



使用AutoML的時機

- 當您想要 Azure Machine Learning 使用您指定的目標計量來為您定型和調整模型時，請套用 AutoML
- AutoML 會促進機器學習模型的開發程序大眾化，並讓其使用者 (無論其資料科學專業知識如何) 都能找出任何問題的端對端機器學習管道
- 跨產業的資料科學家、分析師和開發人員可以使用AutoML來執行下列動作：
 - ◆ 不需要豐富的程式設計知識即可實作 ML 解決方案
 - ◆ 節省時間和資源
 - ◆ 運用資料科學最佳做法
 - ◆ 提供敏捷式問題解決

建立ML工作區



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks

建立工作區 -1

- 使用您 Azure 訂閱的認證來登入 Azure 入口網站
- 在 Azure 入口網站的左上側，選取 [+建立資源]
- 使用搜尋列尋找“Machine”，選取 [Machine Learning] 或[機器學習]
- 在 [Machine Learning] 窗格中選取 [建立] 來開始操作
- 提供下列資訊來設定新的工作區
 - ◆ 工作區名稱，輸入可識別您工作區的唯一名稱，例如**“automl員工編號”**
 - ◆ 訂用帳戶，選取課程使用的訂用帳戶**azuretrain01**
 - ◆ 資源群組，選取已經建立的資源群組**“automl員工編號”**
 - ◆ 區域 (位置)，**小組1-10在東南亞Southeast Asia、小組11-20在東亞eastasia**
 - ◆ 其餘資源使用預設設定即可

建立工作區 -2

Microsoft Azure 搜尋資源、服務及文件 (G+)

首頁 > autotrain10 > 建立資源 > Marketplace > 機器學習 >

機器學習 ...

建立機器學習工作區

選取用以管理部署資源及成本的訂用帳戶。使用像資料夾這樣的資源群組來安排及管理您的所有資源。

訂用帳戶 * ⓘ

azuretrain01

資源群組 * ⓘ

autotrain10

新建

工作區詳細資料

指定工作區的名稱和區域。

工作區名稱 * ⓘ

Demo

區域 * ⓘ

東南亞

儲存體帳戶 * ⓘ

auoautotrain10

新建

金鑰保存庫 * ⓘ

auoautotrain10

Application Insights * ⓘ

auoautotrain10

新建

檢閱 + 建立 < 上一步 下一步: 網路

建立工作區 -3

- 當您完成工作區的設定時，請選取 [檢閱 + 建立]
- 檢查設定，並進行任何額外的變更或修正。當您對設定感到滿意之後，請選取 [建立]
- 程序完成後，會出現部署成功訊息，要檢視新的工作區，選取 [前往資源]

連接工作區

- 利用“工作室 Web URL”或“前往工作室”連線到工作室

The screenshot shows the Microsoft Azure portal interface for a machine learning workspace named 'automlcaloudi'. The left sidebar contains navigation links for Overview, Activity Log, Access Control (IAM), Tags, Diagnose and Solve Problems, Events, Settings, Private Endpoint, Face Recognition, Properties, Locks, Monitoring, Alerts, Metrics, Diagnostics, and Automation. The main content area displays workspace details, including a 'config.json' download link and a 'Programs' section. A red box highlights the 'Workshop Web URL' field, which contains the URL: <https://ml.azure.com/?tid=e7c24d00-7479-4343-8247-0...>. Below this, there are fields for Storage, Logging, and Application Insights, each with their respective Azure resource names. At the bottom, a section titled 'Manage your machine learning lifecycle' with a 'Start Workshop' button is also highlighted with a red box.

首頁 > Microsoft.MachineLearningServices >

自動化

automlcaloudi

機器學習

搜尋 (Cmd +/)

下載 config.json 剪切 刪除

概觀

活動記錄

存取控制 (IAM)

標籤

診斷並解決問題

事件

設定

私人端點連線

身分識別

屬性

鎖定

監視

警訊

計量

診斷設定

記錄

自動化

資源群組 : autoDemo

位置 : 東南亞

訂用帳戶 : azuretrain01

訂用帳戶識別碼 : a278626c-1d76-422d-a5b2-4c2cf191a0ac

工作室 Web URL : <https://ml.azure.com/?tid=e7c24d00-7479-4343-8247-0...>

儲存空間 : automlcaloudi3015578844

登錄 : ...

金鑰保存庫 : automlcaloudi7263158890

Application Insights : automlcaloudi0748757016

JSON 檢視

管理您的機器學習生命週期

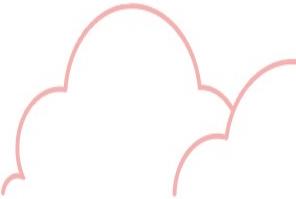
使用 Azure Machine Learning Studio (傳統版) 來建置、訓練、評估及部署機器學習模型。深入了解

啟動工作室

連接資料



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks



Azure ML 資料

- Azure Machine Learning 的資料有兩類
- 資料存放區
 - ◆ 將您的連線資訊(例如您的訂用帳戶識別碼和 token 授權)儲存在與工作區相關聯的 Key Vault，以安全地連線到您在 Azure 上的儲存體服務
- 資料集
 - ◆ 藉由建立資料集，您可以建立資料來源位置的參考，以及其中繼資料的複本
- 資料集的使用
 - ◆ 在模型定型期間存取資料
 - ◆ 共用資料並與其他使用者共同作業。
 - ◆ 利用開放原始碼程式庫(例如 pandas)來探索資料

建立資料存放區 -1

- 登入 Azure Machine Learning Studio
- 在左窗格中，選取 [管理] 底下的 [資料存放區]
- 選取 [+ 新增資料存放區]，在同一個ML Studio內，資料存放區的命名不能重複
- 完成表單，以建立並註冊新的資料存放區。此表單會根據您選取的 Azure 儲存體類型和驗證類型，以智慧方式自行更新

建立資料存放區 -2

- 資料存放區名稱:

sample_data

- 資料存放區類型:

Azure Data Lake Storage Gen2

- 訂用帳戶ID: (使用預設)

azuretrain01

- 存放區名稱:

dllakedemo (Training_RG)

- Azure Data Lake Gen2 檔案系統名稱:

sample

新增資料存放區

資料存放區名稱 *

sample_data_demo

資料存放區類型 *

Azure Data Lake Storage Gen2

選取帳戶的方法

來自 Azure 訂用帳戶

手動輸入

訂用帳戶 ID *

azuretrain01 (a278626c-1d76-422d-a5b2-4c2cf191a0ac)

存放區名稱 *

dllakedemo (Training_RG)

Azure Data Lake Gen2 檔案系統名稱 *

sample

建立資料存放區 -3

- 用戶識別碼：之前註冊的應用程式識別碼

The screenshot shows the Microsoft Azure portal interface. At the top, there's a navigation bar with 'Microsoft Azure' and a search bar. Below it, a breadcrumb trail shows 'Home > Contoso AD (dev) | App registrations >'. The main content area is titled 'Contoso App 1' with a gear icon. On the left, a sidebar has 'Overview' selected, along with 'Quickstart' and 'Integration assistant (preview)'. The main panel displays the application's details: 'Display name: Contoso App 1', 'Application (client) ID: 11111111-1111-1111-1111-111111111111', 'Supported account types: My organization only', and 'Redirect URLs' with a link to 'Add a Redirect URI'. There are also 'Delete' and 'Endpoints' buttons.

- 用戶端密碼：之前註冊的應用程式設定的密碼

The screenshot shows the 'Application secrets' section of the Azure portal. It has a header with '+ 新增用戶端密碼'. A table lists one secret: '描述: Password uploaded on Sat Apr 03 2021', '到期: 3/4/2022', '值: ~_n*****', and 'ID: 902ba91b-c42e-4757-acfd-0b308ccfe28f'. There are edit and delete icons next to the row.

描述	到期	值	ID
Password uploaded on Sat Apr 03 2021	3/4/2022	~_n*****	902ba91b-c42e-4757-acfd-0b308ccfe28f

- 選取“建立”

上課示範用密碼

- Application/Client ID:

cf446eaf-3099-4179-94e3-20b51bc4a78f

- Object ID:

cece5a9f-8e44-4398-82af-4b39d6548af5

- Directory/Tenant ID:

e7c24d00-7479-4343-8247-0204699693e8

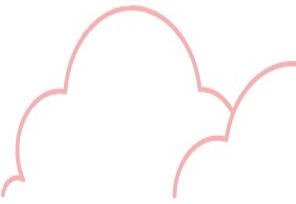
- Secrets Value:

SBoFq1TfYG2~Z1E_4N3i_ialkLs0bgN~ZU

建立AutoML



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks



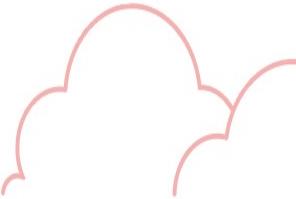
建立AutoML

- 在左側窗格中，[自動化 ML]
 - ◆ 由於這是您的第一個自動化 ML 實驗，因此您會看到一個空白清單與文件連結
- 選取 [+ 新增自動化 ML 回合]

The screenshot shows the Microsoft Azure Machine Learning studio interface. The browser title bar reads "ml.azure.com/automl/welcome?wsid=/subscriptions/ee9095f6-2d27-45a8-9dd7-7ba4fef67393/resourcegroups/mslearn-auto/workspaces/automl-demo&tid=d18376e2-1716-4..." and the URL "automl-demo". The main content area is titled "Microsoft Azure Machine Learning" and "自動化 ML". It displays a message: "無需撰寫任何程式碼，即可讓自動化 ML 依據您的資料訓練及尋找最佳模型。深入了解自動化 ML" and a button "+ 新增自動化 ML 回合". Below this, it says "沒有任何最近的自動化 ML 回合可顯示。按一下 [新增自動化 ML 回合] 建立您的第一個回合" and a link "深入了解如何建立自動化 ML 回合". At the bottom, there is a section titled "文件" with three items: "概念: 什麼是自動化 ML?", "教學課程: 使用自動化 ML 建立您的第一個分類模型", and "部落格: 使用自動化 ML 的新功能建置更精確的預測". The left sidebar has a vertical navigation menu with icons for Home, My Data, Automation, Google, What is, Hadoop, Azure, Build, and What is.

建立資料集 -1

- 在設定實驗之前，要先將資料檔案以 Azure Machine Learning 資料集的形式上傳到工作區
- 如此一來，您就可以確保資料會格式化成適合進行實驗的格式
- 透過從 [+建立資料集] 下拉式清單選取 [來自資料存放區]，以建立新的資料集
- 基本資訊
 - ◆ 名稱: wine_quality
 - ◆ 描述: Red Wine Quality
- 選取 [下一個]
- 在 [選取或建立資料存放區]，選擇“先前建立的資料存放區”



建立資料集 -2

- 選取名稱為“sample_data”的 Azure Data Lake Gen2 資料存放區
- 按一下 [選取資料存放區]
- 設定“路徑”為“winequality-red.csv”
- 選取 [下一個]
- “設定與預覽”頁面的設定是自動偵測而來。請驗證選取項目是否正確
- 正確的設定“分隔符號”、“資料行標頭”
- 選取 [下一個]

建立資料集 -3

從資料存放區建立資料集

X

基本資訊

資料存放區選取

設定與預覽

結構描述

確認詳細資料

設定與預覽

這些設定是自動偵測而來。請驗證選取項目是否正確，否則請予以更新。

檔案格式

分隔符號

分隔符號 範例
逗號 Field1,Field2,Field3

編碼

UTF-8

資料行標頭

使用第一個檔案的標頭

略過資料列

無

#	識..	0.0 fixed acidity	0.0 volatile acidity	0.0 citric acid	0.0 residual sugar
1	7.4	0.7	0	1.9	
2	7.8	0.88	0	2.6	
3	7.8	0.76	0.04	2.3	

上一步

下一個

取消

建立資料集 -4

- 進入“結構描述”頁面
 - ◆ 資料行類型會根據資料的前 200 個資料列自動偵測，可以進行必要的調整
 - ◆ 與指定資料行類型不一致的值將無法轉換，而且會填入 Null 或以錯誤值取代
- 選取 [下一個]
- 確認詳細資料後，選取“建立”來產生資料集
- 可以利用選取資料集名稱來檢視剛建立的資料集
 - ◆ 選擇要查閱的資料集(wine_quality)
 - ◆ 可瀏覽“資料預覽”，“資料統計資料”
 - ◆ 選取 [關閉]

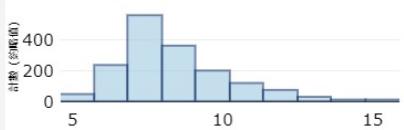
建立資料集 -5

● 資料集建立完成，選取 [下一個]

wine_quality ×

資料預覽 資料統計資料 上一頁 下一頁

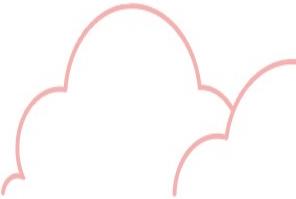
注意:統計資料的計算根據為資料的樣本

資料行	設定檔	類型	最小值	最大值	計數	遺漏... 百分比	空白... 百分比
fixed acidity		十進位	4.60	15.90	1599	0	0
volatile acidity		十進位	0.12	1.58	1599	0	0
citric acid		十進位	0.00	1.00	1599	0	0
residual sugar		十進位	0.90	15.50	1599	0	0

關閉

設定執行 -1

- 在載入和設定資料之後，您可以設定您的實驗
- 這項設定包括實驗設計工作，例如，選取計算環境的大小，以及指定您要預測的資料行
- 選取 [新建] 選項按鈕
- 新增實驗名稱: wine-automl-experiment
- 目標資料行: 選取“quality” 作為您要預測的目標資料行。此資料行會指出酒的品質
- 選取 [建立新的計算]，並設定您的計算目標。計算目標是用來執行定型指令碼或裝載服務部署的本機或雲端式資源環境。針對此實驗，我們會使用雲端式計算



設定執行 -2

- 填入“虛擬機器”表單，以設定您的計算
 - ◆ 虛擬機器優先順序: 專用
 - ◆ 虛擬機器類型: CPU
 - ◆ 虛擬機器大小:“從建議選項中選取”， Standard_DS12_V2
- 選取 [下一個]，以填入 [設定]
 - ◆ 計算名稱: **automl-員工編號**
 - ◆ 節點數目下限 : 0
 - ◆ 節點數目上限 : 4
 - ◆ 相應減少: 120 (預設值)
- 選取 [建立] 以建立您的計算目標

設定計算目標

- 建立完成後，請從下拉式清單選取新的計算目標
- 選取 [下一個]

建立計算叢集 [\(i\)](#)

設定
為您選取的虛擬機器大小設定計算叢集設定。

虛擬機器

進階設定

名稱	類別	核心數	可用的配額	RAM	儲存體	成本/節點
Standard_DS12_v2	記憶體最佳化	4	2492 個核心	28 GB	56 GB	\$0.38/小時

計算名稱 * [\(i\)](#)

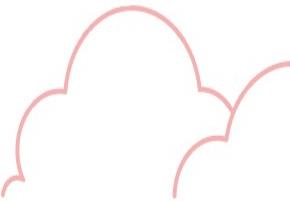
節點數目下限 * [\(i\)](#)

節點數目上限 * [\(i\)](#)

相應減少之前的閒置秒數 * [\(i\)](#)

啟用 SSH 存取 [\(i\)](#)

[上一步](#) [建立](#) [下載自動化的範本](#) [取消](#)



選取工作類型

- 在 [工作類型和設定] 表單上，藉由指定機器學習工作類型和組態設定，來完成自動化 ML 實驗的設定
 - ◆ 選取 [分類] 作為機器學習工作類型
 - ◆ 選取 [檢視其他組態設定] 並填入欄位，如下所示。這些設定可進一步控制訓練作業。否則會根據實驗選取範圍和資料來套用預設值

The screenshot shows the '建立新的自動化 ML 回合' (Create New Automation ML Run) wizard. On the left, a vertical toolbar lists various icons for different steps. The current step, '選取工作和設定' (Select Work Type and Settings), is highlighted with a blue dot. The main panel displays the 'Select Work Type' configuration screen. It includes a title, a descriptive note about selecting machine learning work types, and three options:

- 分類**: Predicts target data rows into several categories: Yes/No, Blue, Red, Green. This option is selected, indicated by a green checkmark.
- 迴歸**: Predicts continuous values.
- 時間序列預測**: Predicts values based on time.

At the bottom, there are links to '檢視其他組態設定' (View other configuration settings) and '檢視特徵化設定' (View feature engineering settings). Navigation buttons for '上一步' (Previous Step), '完成' (Finish), and '退出' (Exit) are also present.

執行實驗

- 選取 [完成] 以執行實驗。當實驗準備開始時，[回合詳細資料] 畫面隨即開啟，其頂端顯示 [執行狀態]。此狀態會隨著實驗的進行而更新。通知也會出現在 Studio 的右上角，以通知您實驗的狀態



屬性

狀態
正在執行 ▾
正在設定執行
正在將執行提交給計算

已建立
Jul 27, 2021 5:23 PM

已開始
Jul 27, 2021 5:24 PM

計算目標
[automl-caloudi](#)

執行 ID
AutoML_03505001-7e8b-47ca-9453-ea1a2a800f54

指令碼名稱
--

執行摘要

工作類型
分類 [檢視組態設定](#)

特徵化
自動

主要計量
精確度

實驗名稱
wine-auto-experiment

探索模型 -1

- 瀏覽至 [模型] 索引標籤，以查看已測試的演算法 (模型)
- 根據預設，模型會在完成時依計量分數排序，精確度最高的模型會在清單頂端
- 可選取已完成模型的演算法名稱來探索其效能詳細資料
 - ◆ 可在該演算法的 [詳細資料] 與 [計量] 索引標籤中進行瀏覽，以查看所選模型的屬性、計量與效能圖表

演算法名稱	解釋	精確度 ↓	取樣 ⓘ	已建立	持續時間
VotingEnsemble	檢視說明	0.68981	100.00 %	Apr 4, 2021 6:43 PM	1 分 37 秒
StackEnsemble		0.67855	100.00 %	Apr 4, 2021 6:43 PM	1 分 37 秒
MaxAbsScaler, LightGBM		0.66792	100.00 %	Apr 4, 2021 6:26 PM	55 秒
StandardScalerWrapper, XGBoostClassifier		0.66354	100.00 %	Apr 4, 2021 6:28 PM	2 分 3 秒
SparseNormalizer, XGBoostClassifier		0.64728	100.00 %	Apr 4, 2021 6:35 PM	58 秒
MaxAbsScaler, XGBoostClassifier		0.64103	100.00 %	Apr 4, 2021 6:26 PM	1 分 42 秒
MaxAbsScaler, LightGBM		0.62852	100.00 %	Apr 4, 2021 6:34 PM	55 秒
SparseNormalizer, XGBoostClassifier		0.62539	100.00 %	Apr 4, 2021 6:34 PM	57 秒
StandardScalerWrapper, LightGBM		0.62477	100.00 %	Apr 4, 2021 6:41 PM	52 秒

探索模型 -2

首頁 > 實驗 > wine-automl-experiment > 回合 1 > 回合 51

回合 51 已完成

(重新整理) (部署) (下載) (說明模型) (取消)

詳細資料 模型 說明 (預覽) 計量 輸出 + 記錄檔 影像 子回合 快照集

選取計量，以檢視該資料的視覺效果或資料表。

搜尋

accuracy
 accuracy_table
 AUC_macro
 AUC_micro
 AUC_weighted
 average_precision_score_macro
 average_precision_score_micro
 average_precision_score_weighted
 balanced_accuracy
 confusion_matrix
 f1_score_macro

檢視身分: (圓形) 圖表 (圓圈) 資料表

accuracy 0.69 AUC_macro 0.852 AUC_micro 0.929 AUC_weighted 0.834

Precision-Recall

The chart displays the precision-recall trade-off for four average types: Weighted Average (blue line), Macro Average (dark blue line), Micro Average (teal line), and Ideal (black dashed horizontal line at 1.0). The Y-axis represents Precision from 0.2 to 1.0, and the X-axis represents Recall. The Micro Average curve shows the highest overall precision, starting at approximately 0.95 and ending near 0.35. The Macro Average curve starts at 1.0 and drops sharply to around 0.35. The Weighted Average curve follows a similar path to the Macro Average one. The Ideal curve is a horizontal line at 1.0.

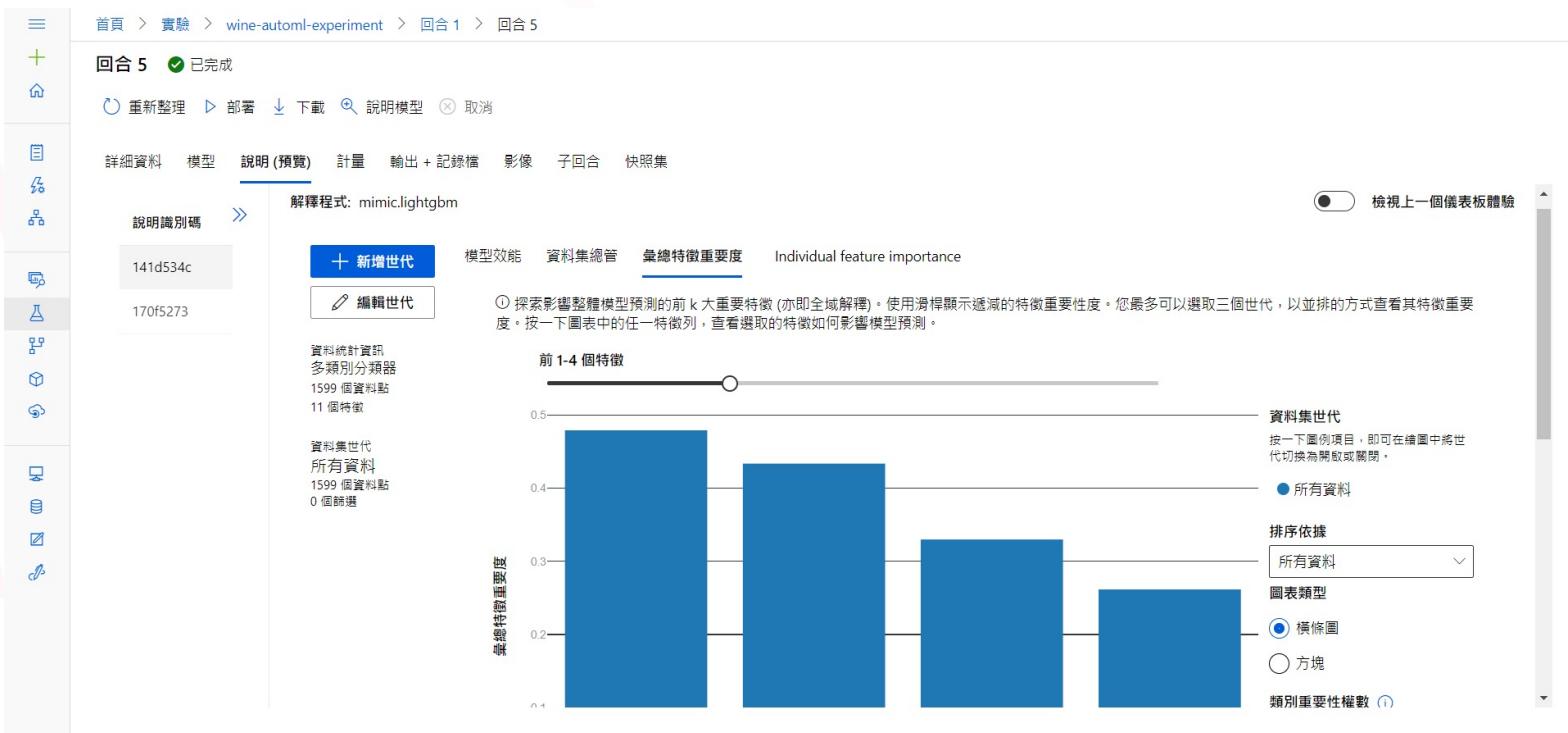
Weighted Average
Macro Average
Micro Average
Ideal
3
4
5
6
7
8

模型說明 -1

- 您也可以查看模型說明，並查看影響特定模型預測的特徵資料(原始的或是處理過的)
- 您可以視需要產生這些模型說明，並在 [模型說明] 儀表板中“說明(預覽)”標籤內檢視
- 產生模型說明
 - ◆ 選取頂端的 [回合1]
 - ◆ 選取 [模型] 索引標籤
 - ◆ 選取一個模型(例如 MaxAbsScaler、LightGBM)
 - ◆ 選取頂端的 [說明模型] 按鈕，右側的 [說明模型] 窗格隨即出現。
 - ◆ 選取先前建立的“automl-員工編號”計算，此計算叢集會產生模型說明
 - ◆ 選取底部的 [建立]，畫面頂端會出現綠色的成功訊息

模型說明 -2

- 選取 [說明(預覽)] 的標籤
- 選取左側的 [說明識別碼]
- 選取右邊的 [匯總特徵重要度] 索引標籤。此圖表會顯示哪些特徵資料會影響所選模型的預測



部署最佳模型 -1

- 自動化機器學習介面可讓您透過幾個步驟，將最佳模型部署為 Web 服務
- 部署是模型的整合，因此可以根據新資料進行預測，並找出潛在的商機區域
- 查看您的實驗執行是否已完成
 - ◆ 請選取畫面頂端的 [回合1]，“已完成”狀態會顯示在畫面的左上方
- 一旦實驗執行完成後，[詳細資料] 頁面就會填入 [最佳模型摘要] 區段
- 在此實驗內容中，VotingEnsemble 會被視為最佳模型

部署最佳模型 -2

首頁 > 實驗 > wine-automl-experiment > 回合 1

回合 1 已完成

⟳ 重新整理 × 取消

詳細資料 資料護欄 模型 輸出 + 記錄檔 子回合 快照集

屬性

狀態 已完成

已建立 Apr 4, 2021 6:22 PM

已開始 Apr 4, 2021 6:23 PM

持續時間 22 分 26.73 秒

計算目標 automl-compute

執行 ID AutoML_d4ef1ca4-d94f-4e45-b4aa-20a1630d3394

指令碼名稱 --

建立者 Tim Yang

輸入資料集 輸入名稱: training data , 識別碼: aa7e7e76-f1b2-4def-b41b-21af3b40b3e8

最佳模型摘要

演算法名稱 VotingEnsemble

精確度 0.68981 ☰ 檢視所有其他計量

取樣 100.00 % ⓘ

註冊的模型 尚未註冊

部署狀態 尚未部署

執行摘要

工作類型 分類 ☰ 檢視所有回合設定

主要計量 精確度

實驗名稱

部署最佳模型 -3



- 部署程序需要幾個步驟，包括註冊模型、產生資源，以及為 Web 服務設定這些資源
- 選取 [VotingEnsemble] 以開啟模型特定頁面
- 選取左上方的 [部署] 按鈕
- 填入 [部署模型] 窗格
 - ◆ 部署名稱: wine-automl-deploy
 - ◆ 部署描述: Wine Quality Evaluate
 - ◆ 計算類型: Azure 容器執行個體
- 選取 [部署]
 - ◆ [執行] 畫面頂端會出現綠色成功訊息，而在 [模型摘要] 窗格中，狀態訊息會顯示在 [部署狀態] 底下。定期選取 [重新整理] 以檢查部署狀態
- 現在您已有可運作的 Web 服務，可用來產生預測

部署最佳模型 -4

首頁 > 實驗 > wine-automl-experiment > 回合 1 > 回合 51

回合 51 ✓ 已完成

⟳ 重新整理 ⚡ 部署 ⏪ 下載 ⓘ 說明模型 ✖ 取消

成功: 已成功觸發模型部署

詳細資料 模型 **說明 (預覽)** 計量 輸出 + 記錄檔 影像 子回合 快照集

模型摘要

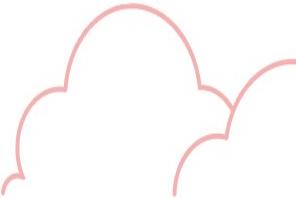
演算法名稱
VotingEnsemble

精確度
0.68981 [檢視所有其他計量](#)

取樣
100.00 % [\(i\)](#)

註冊的模型
AutoMLd4ef1ca4d46:1

部署狀態
wine-automl-deploy [正在執行](#)



測試部署 -1

- 選取 “wine-autml-deploy” ，選取 [測試] 標籤，輸入測試資料
 - ◆fixed acidity: 12.5
 - volatile acidity: 0.18
 - citric acid: 0.71
 - residual sugar: 1.7
 - chlorides: 0.065
 - free sulfur dioxide: 18.0
 - total sulfur dioxide: 62.0
 - density: 0.987
 - pH: 3.24
 - sulphates: 0.51
 - alcohol: 9.6
- 選取 [測試]，得到測試結果

測試部署 -2

首頁 > 端點 > wine-automl-deploy

wine-automl-deploy

詳細資料 測試 取用 部署記錄

輸入要測試即時端點的資料

測試

測試結果

已剖析 原始

fixed acidity
12.5

volatile acidity
0.18

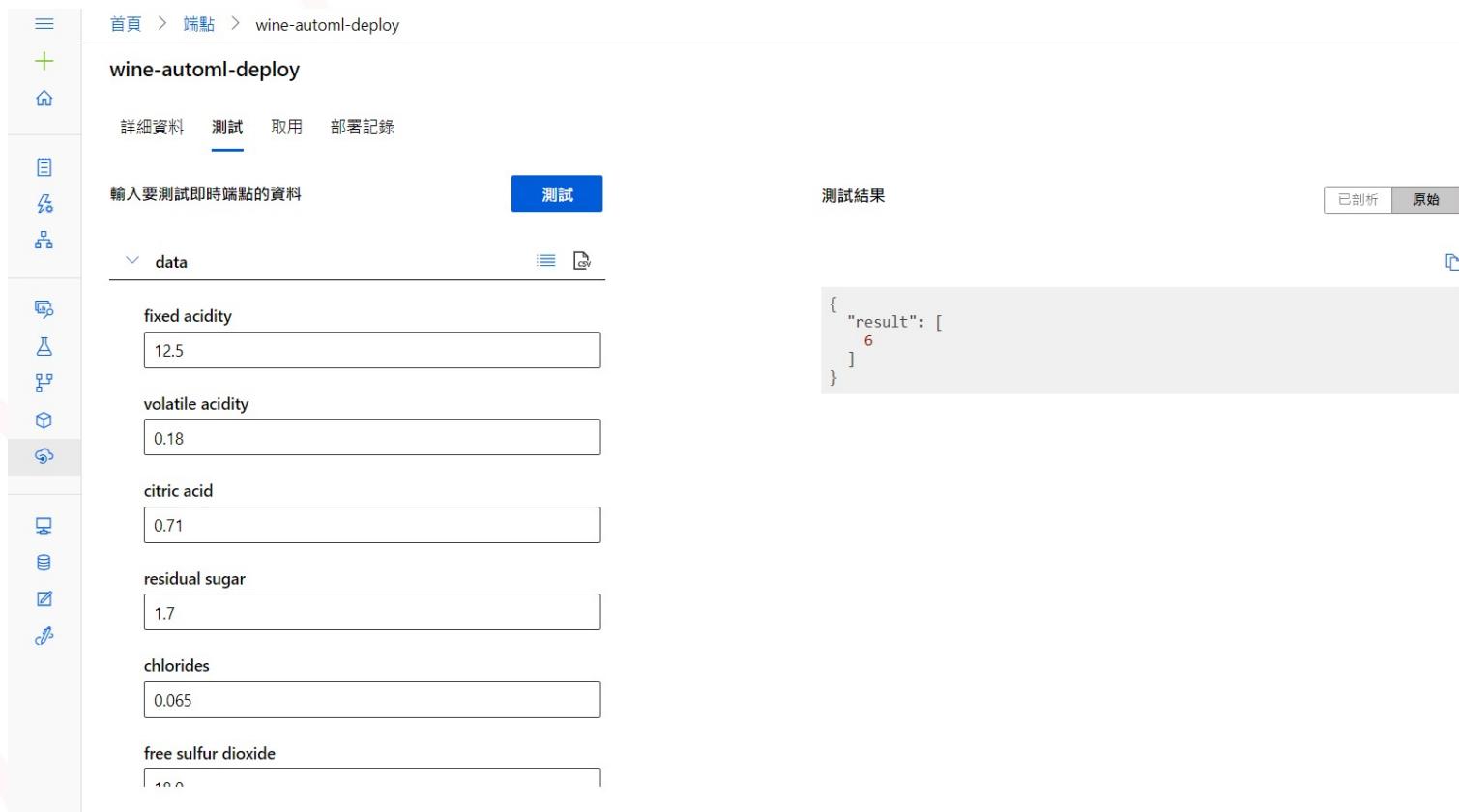
citric acid
0.71

residual sugar
1.7

chlorides
0.065

free sulfur dioxide
10.0

{ "result": [6] }



清除資源



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks

資源上限 -1



- 閒置的資源仍會被收費：
 - ◆ Azure是使用者付費 Pay-as-You-Go
 - ◆ 使用多少資源就會收多少費用，閒置的計算資源也會被收費，因為是佔用了雲端的計算資源

- 資源的使用上限：
 - ◆ 每個群組/資源都會有個別資源的使用上限
 - ◆ ML Workspace也有自己的資源使用上限
 - ◆ 虛擬機器VM的核心會有區域總數量限制及每個系列的區域總數量限制，兩個限制會同時生效

資源上限 -2

- 資源群組和資源所在的區域可以不同，在建立機器學習ML Studio時就須決定資源所在區域，ML Studio中所使用的計算叢集會和ML Studio會架設在同一個區域
- 查詢ML Studio的資源上限：
[資源群組] autml員工編號 > [機器學習] automl員工編號
> [支援與疑難排解] 使用量 + 配額
- Azure會自動帶入資源群組所在地理位置，在使用量 + 配額的欄位可以看到該區域各系列虛擬機器核心的使用量

資源上限 -3

Microsoft Azure 搜尋資源、服務及文件 (G+)

首頁 > 資源群組 > TRAINING_RG > trainingml

trainingml | 使用量 + 配額

機器學習

搜尋 (Cmd +/)

重新整理

訂用帳戶 *

auobigdata

位置 *

東南亞

訂用帳戶檢視 (selected) 工作區檢視

Azure Machine Learning 配額是預先設定的限制，指定了在任何指定時間可使用的核心數目上限。此摘要顯示訂閱中所有工作區的配額使用量與限制。深入了解

專用核心使用量 低優先順序核心使用量

專用配額: 已使用 22 個核心 | 剩餘 278 個核心
叢集配額: 已使用 14 個叢集和 CI | 剩餘 186 個

要求配額增加 設定配額

VM 系列	使用量	
標準 D 系列 vCPU	0%	已使用 0 個核心 (共 300 個)
標準 DDSv4 系列 vCPU	0%	已使用 0 個核心 (共 0 個)
> 標準 DSV2 系列 vCPU	6%	已使用 20 個核心 (共 300 個)
標準 DSV3 系列 vCPU	0%	已使用 0 個核心 (共 300 個)
> 標準 DV2 系列 vCPU	0%	已使用 2 個核心 (共 300 個)
標準 DV3 系列 vCPU	0%	已使用 0 個核心 (共 300 個)
標準 EAV4 系列 vCPU	0%	已使用 0 個核心 (共 0 個)
標準 EV3 系列 vCPU	0%	已使用 0 個核心 (共 0 個)
標準 FSv2 系列 vCPU	0%	已使用 0 個核心 (共 300 個)
標準 H 系列 vCPU	0%	已使用 0 個核心 (共 0 個)
標準 HCS 系列 vCPU	0%	已使用 0 個核心 (共 0 個)
標準 MS 系列 vCPU	0%	已使用 0 個核心 (共 0 個)
標準 NCSV2 系列 vCPU	0%	已使用 0 個核心 (共 0 個)

清除資源 -1

- Azure的雲端計算資源都有AutoScaling的功能，如果將節點數目下限設定為0時，閒置的資源就會被解除配置而不被佔用。
- 在建立AutoML時需要設定計算資源，若設定的節點nodes數目下限為1，也就是當資源計算完之後仍會保留一台節點給本次的實驗，產生閒置的資源

The screenshot shows the configuration interface for setting up a compute resource in Azure Machine Learning. It includes three main sections:

- 節點數目下限 ***: A numeric input field set to 1, with a tooltip explaining that setting it to 0 would allow resources to be deallocated when not in use.
- 節點數目上限 ***: A numeric input field set to 4, with a long horizontal slider bar extending from 1 to 100.
- 相應減少之前的閒置秒數 ***: A numeric input field set to 120.

清除資源 -2



- 從ML Studio檢查所有計算資源：

[管理] 計算 > 計算叢集

名稱	狀態	大小	建立日期 ↓	使用中的執行... 節點	閒置節點
MLCompute	成功 (1 個節點)	STANDARD_D2_V2	Apr 21, 2021 4:08 PM	0	1
testDS12	成功 (0 個節點)	STANDARD_DS12_V2	Apr 21, 2021 3:28 PM	0	0
automl-compute	成功 (1 個節點)	STANDARD_DS12_V2	Apr 20, 2021 4:35 PM	0	1

- 確認計算的狀態都已經顯示為成功，找到有閒置節點的 automl-compute，點入確認資源是閒置的狀態

automl-compute

詳細資料 節點 執行

重新整理 刪除

叢集節點狀態

1
閒置 1

■ 閑置
■ 正在離開
■ 正在準備
■ 正在執行

叢集狀態

配置狀態
成功 (1 個節點)

配置狀態轉換時間
2021/4/21 上午11:13:24

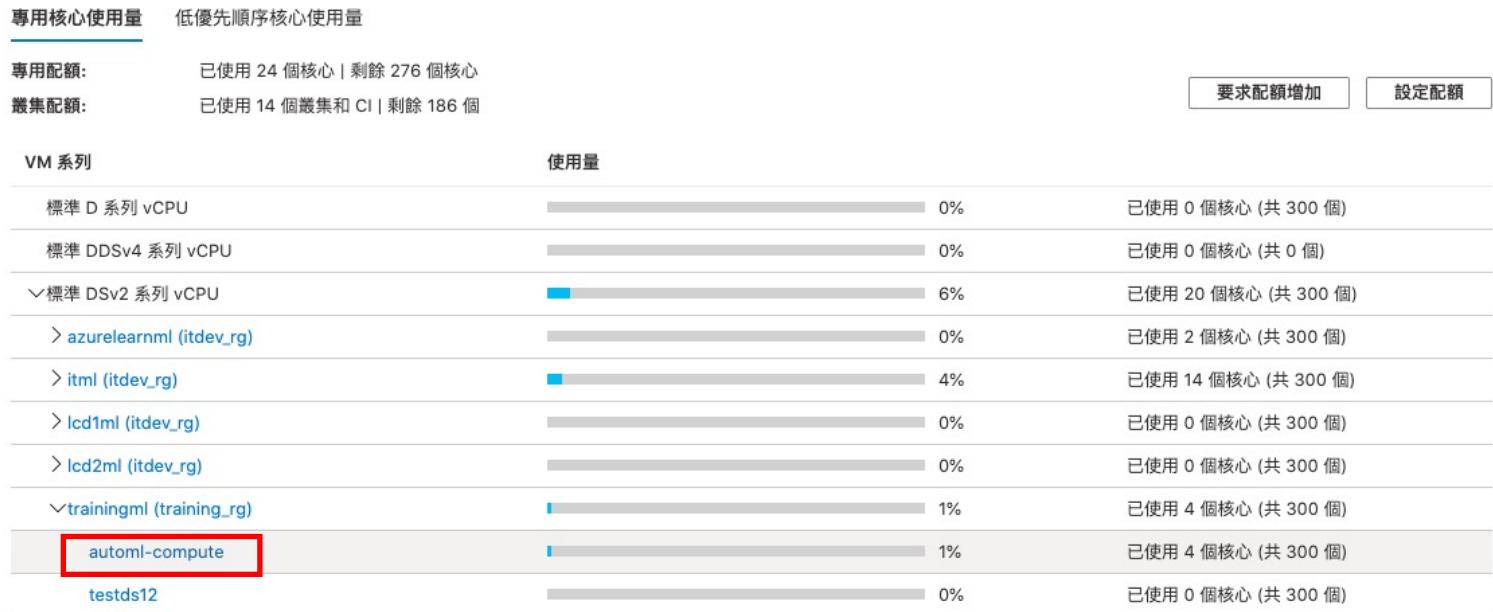
建立日期
2021/4/20 下午4:35:57

目前的節點計數
1

清除資源 -3

- 找到閒置資源後，回到 Azure Machine Learning 的頁面確認佔用的 VM vCore

[資源群組] automl員工編號 > [機器學習] automl員工編號
> [支援與疑難排解] 使用量 + 配額



清除資源 -4

- 回到ML Studio的計算叢集automl-compute頁面，點選上方的刪除，確認[刪除]，可以在該畫面重新整理確認執行狀態，直到資源被刪除出現錯誤消息(資源不存在)



清除資源 -5

- 再在[支援與疑難排解] 使用量 + 配額頁面重新整理，確認該組vCores已經不再佔用資源

設定

- 私人端點連線
- 屬性
- 鎖定

監視

- 警訊
- 計量
- 診斷設定
- 記錄

自動化

- 工作 (預覽)
- 匯出範本

支援與疑難排解

- 使用量 + 配額
- 新增支援要求

專用核心使用量 低優先順序核心使用量

專用配額: 已使用 18 個核心 | 剩餘 282 個核心
叢集配額: 已使用 13 個叢集和 CI | 剩餘 187 個

要求配額增加 設定配額

VM 系列	使用量	
標準 D 系列 vCPU	0%	已使用 0 個核心 (共 300 個)
標準 DDSv4 系列 vCPU	0%	已使用 0 個核心 (共 0 個)
標準 DSv2 系列 vCPU	5%	已使用 16 個核心 (共 300 個)
> azurelearnml (itdev_rg)	0%	已使用 2 個核心 (共 300 個)
> itml (itdev_rg)	4%	已使用 14 個核心 (共 300 個)
> lcd1ml (itdev_rg)	0%	已使用 0 個核心 (共 300 個)
> lcd2ml (itdev_rg)	0%	已使用 0 個核心 (共 300 個)
> trainingml (training_rg)	0%	已使用 0 個核心 (共 300 個)
testds12	0%	已使用 0 個核心 (共 300 個)
標準 DSv3 系列 vCPU	0%	已使用 0 個核心 (共 300 個)
> 標準 Dv2 系列 vCPU	0%	已使用 2 個核心 (共 300 個)
標準 Dv3 系列 vCPU	0%	已使用 0 個核心 (共 300 個)

清除資源 -6

- 提醒您，Azure雲端服務是依照使用的資源計費，閒置未使用的資源亦可能因相依的服務被收費(例如儲存體帳戶)，除了產生不必要的費用之外，佔用各項資源的配額也可能導致他人無法建立所需資源。
- 建議您，在確認無需使用該項服務時，解除配置或刪除閒置資源

Azure Databricks



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks

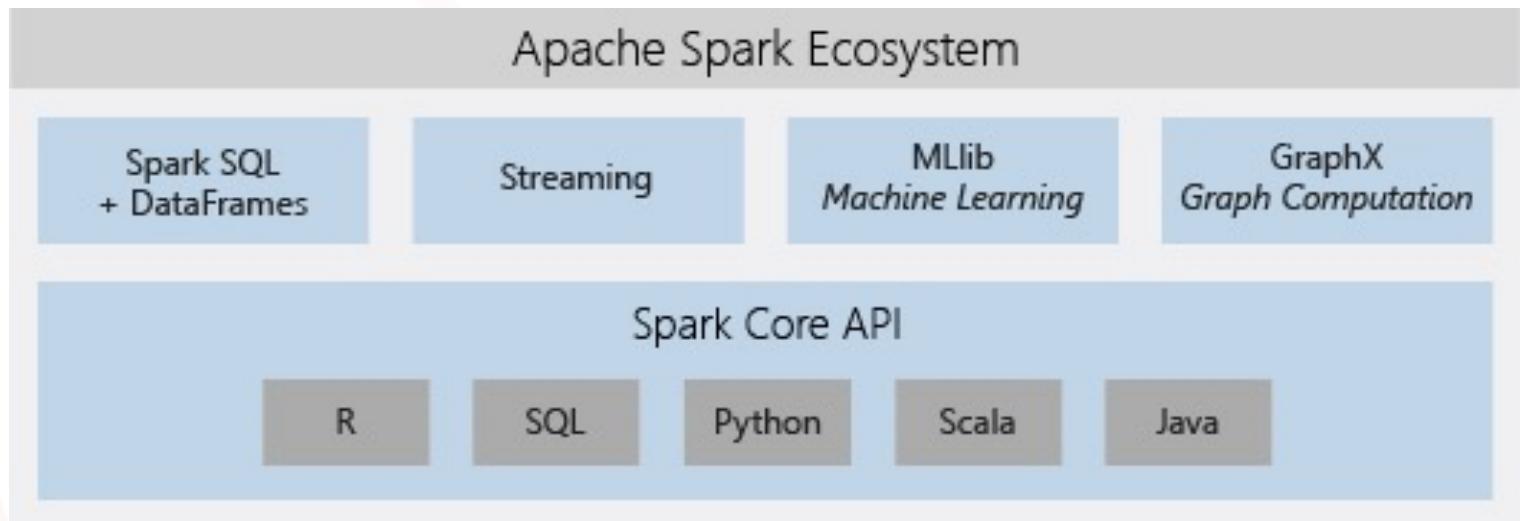
Apache Spark種類

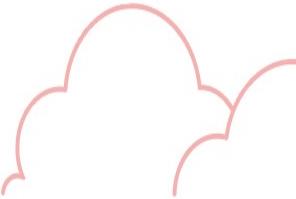
Apache Spark		HDInsight	Azure Databricks	Synapse Spark
WHAT	Is an Open Source memory optimized system for managing big data workloads	Microsoft implementation of Open Source Spark managed within the realms of Azure	A managed Spark as a Service solution	Embedded Spark capability within Azure Synapse Analytics
WHEN	When you want to benefits of spark for big data processing and/or data science work without the Service Level Agreements of a provider	When you want to benefits of OSS spark with the Service Level Agreement of a provider	Provides end to end data engineering and data science solution and management platform	Enables organizations without existing Spark implementations to fire up a Spark cluster to meet data engineering needs without the overheads of the other Spark platforms listed
WHO	Open Source Professionals	Open Source Professionals wanting SLA's and Microsoft Data Platform experts	Data Engineers and Data Scientists working on big data projects every day	Data Engineers, Data Scientists, Data Platform experts and Data Analysts
WHY	To overcome the limitations of SMP systems imposed on big data workloads	To take advantage of the OSS Big Data Analytics platform with SLA's in place to ensure business continuity	It provides the ability to create and manage an end to end big data/data science project using one platform	It provides the ability to scale efficiently with spark clusters within a one stop shop DataWarehousing platform of Synapse.



Apache Spark分析平台

- Azure Databricks 工作區包含完整的開放原始碼 Apache Spark 叢集技術和功能。 Azure Databricks 工作區中的 Spark 包括下列元件：





Azure Databricks

- Azure Databricks 是一個針對 Microsoft Azure 雲端服務平台進行最佳化的資料分析平台
- Azure Databricks 提供兩個環境來開發資料密集的應用程式
 - ◆ Azure Databricks SQL 分析
 - ◆ Azure Databricks 工作區
- Azure Databricks 工作區 提供互動式工作區，可讓資料工程師、資料科學家和機器學習工程師共同作業
- Azure Databricks 的平台可以與 Spark 配合使用，該平台提供了自動化的集群管理和 IPython 樣式的筆記本

使用Azure Databricks



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks

建立 Azure Databricks 服務 -1

- 在 Azure 入口網站中，搜尋欄搜尋 “Databricks” 就可以找到 Azure Databricks 的服務，選取 [Azure Databricks] 建立

The screenshot shows the Azure portal's search interface. A search bar at the top contains the text "databricks". Below the search bar, there are two main sections: "服務" (Services) and "資源" (Resources). In the "服務" section, the "Azure Databricks" service is highlighted with a grey background. Other listed services include "Azure Database for MySQL flexible servers" and "Azure Database for PostgreSQL server groups". In the "Marketplace" section, there are four recommended add-ons: "Infoworks for Databricks", "Unravel for Azure Databricks and Azure HDInsight", "Trifacta Enterprise with Databricks (BYOL)", and "StreamSets for Databricks". A "查看全部" (View All) link is located at the top right of the Marketplace section.

- 在 [Azure Databricks 服務] 下方提供資訊，以建立 Databricks 服務

建立Azure Databricks服務 -2

- ◆ 訂用帳戶
- ◆ 資源群組: 使用現有資源群組(本次上課使用**Training_RG**)
- ◆ 工作區名稱 (本次使用**DatabricksDemo**)
- ◆ 位置
- ◆ 定價層: 選取 [標準]
- 選取 [審核+建立]
- 選取 [建立]
- 程序完成後，會出現部署成功訊息，選取 [前往資源]

首頁 > Azure Databricks > 建立 Azure Databricks 工作區 ...

基本 網路功能 進階 標籤 審核+建立

專案詳細資料

選取用以管理部署資源及成本的訂用帳戶。使用像資料夾這樣的資源群組來安排及管理您的所有資源。

訂用帳戶 * ① Visual Studio Enterprise 訂閱 – MPN

資源群組 * ① mslearn-auto 新建

執行個體詳細資料

工作區名稱 * mydatabricks

區域 * 東南亞

定價層 * ① 標準 (Apache Spark，使用 Azure AD 保護)

審核+建立 <上一步 下一步：網路功能 >

建立Azure Databricks服務 -3

● 上課操作使用DatabricksDemo(Training_RG)

The screenshot shows the Microsoft Azure portal interface for managing a Databricks service. The main title bar reads "Microsoft Azure" and "搜尋資源、服務及文件 (G+/-)". The top navigation bar includes icons for search, refresh, notifications (8), help, and user profile.

The left sidebar shows the navigation path: 首頁 > Training_RG_DatabricksDemo > DatabricksDemo. The main content area displays the "概觀" (Overview) tab for the "DatabricksDemo" service. Key details include:

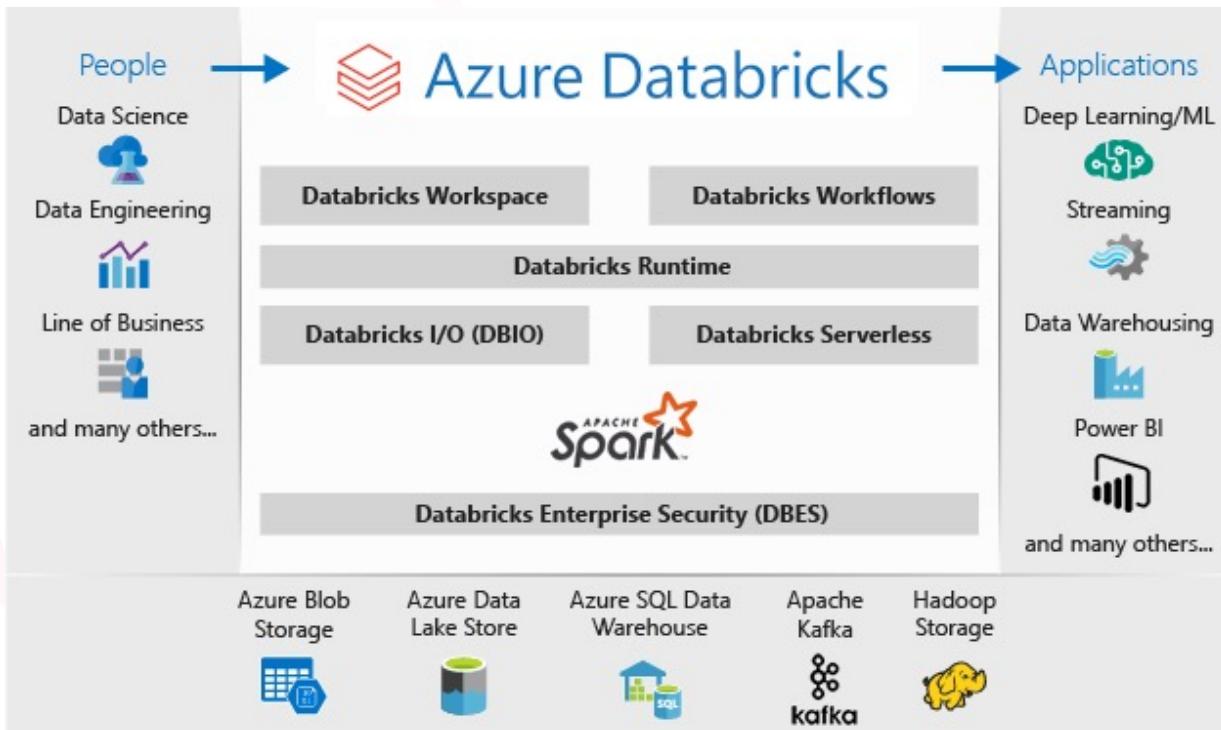
- 狀態 (Status): Active
- 資源群組 (Resource Group): Training_RG
- 位置 (Location): 東南亞 (Southeast Asia)
- 訂用帳戶 (Billing Account): azuretrain01
- 訂用帳戶識別碼 (Billing Account ID): a278626c-1d76-422d-a5b2-4c2cf191a0ac
- 受控資源群組 (Controlled Resource Group): databricks-rg-DatabricksDemo-jbnxxxiy5vf2y
- URL: <https://adb-8297892229005301.azuredatabricks.net>
- 定價層 (Pricing Tier): standard

Below the overview, there is a large red Databricks logo icon and a blue "啟動工作區" (Launch Workspace) button. At the bottom, there are several quick access links:

- 文件 (File)
- 開始使用 (Get Started)
- 從檔案匯入資料 (Import from File)
- 從 Azure 儲存體匯入資料 (Import from Azure Storage)
- 筆記本 (Notebook)
- 系統管理指南 (System Management Guide)
- 連結 Azure ML 工作區 (Connect to Azure ML workspace)

Azure Databricks工作區

- Azure Databricks 工作區是 Apache Spark 型分析平台
- 可與 Azure 整合，提供一鍵式設定、順暢的工作流程以及互動式的工作區，可讓資料工程師、資料科學家及機器學習工程師共同作業



啟動Databricks工作區

- 在 Azure 入口網站中，移至您所建立的 Databricks 服務，然後選取 [啟動工作區]
- 系統會將您重新導向至 Azure Databricks 入口網站

The screenshot shows the Azure Databricks landing page. On the left is a dark sidebar with icons for Home, Workspace, Recents, Data, Clusters, Jobs, Models, and Search. The main area has a header "Azure Databricks". Below it are three main sections: "Explore the Quickstart Tutorial" (with a "New Notebook" icon), "Import & Explore Data" (with a "Drop files or click to browse" button and a "Cloud" icon), and "Create a Blank Notebook" (with a "New Notebook" icon). At the bottom, there are three tabs: "Common Tasks", "Recents" (which is highlighted in grey), and "Documentation". Under "Common Tasks", there are links for New Notebook, Create Table, New Cluster, New Job, New MLflow Experiment, Import Library, and Read Documentation. Under "Recents", it says "Recent files appear here as you work.". Under "Documentation", there are links for Documentation, Release Notes, and Getting Started.

Explore the Quickstart Tutorial

Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.

Import & Explore Data

Quickly import data, preview its schema, create a table, and query it in a notebook.

Create a Blank Notebook

Create a notebook to start querying, visualizing, and modeling your data.

Common Tasks

- New Notebook
- Create Table
- New Cluster
- New Job
- New MLflow Experiment
- Import Library
- Read Documentation

Recents

Recent files appear here as you work.

Documentation

- Documentation
- Release Notes
- Getting Started

建立 Spark 叢集 -1

- 在 Azure Databricks 入口網站中選取 [New Cluster]
- 在 [新增叢集] 頁面上，提供資料以建立叢集
- 填寫下列欄位的值，然後接受其他欄位的預設值：
 - ◆ 輸入叢集的名稱(例如 my-spark-cluster)，此名稱需唯一
 - ◆ 務必要選取 [在停止活動 120 分鐘後終止] 核取方塊。請提供用來終止叢集的叢集未使用持續時間 (以分鐘為單位)
- 選取 [建立叢集]
- 叢集執行後，您就可以將 Notebook 連結至叢集，並執行 Spark 作業
- 本次上課請使用 "**SparkCluster小組編號**" 之計算叢集

建立 Spark 叢集 -2

Create Cluster

New Cluster

[Cancel](#) [Create Cluster](#)

2-8 Workers: 28.0-112.0 GB Memory, 8-32 Cores, 1.5-6 DBU
1 Driver: 14.0 GB Memory, 4 Cores, 0.75 DBU [?](#)

Cluster Name: my-spark-cluster

Cluster Mode: Standard

Pool: None

Databricks Runtime Version: Runtime: 8.0 (Scala 2.12, Spark 3.1.1) [Learn more](#)

Note: Databricks Runtime 8.x uses Delta Lake as the default table format. [Learn more](#)

Autopilot Options:

Enable autoscaling [?](#)

Terminate after minutes of inactivity [?](#)

Worker Type: Standard_DS3_v2 (14.0 GB Memory, 4 Cores, 0.75 DBU) [Min Workers](#) [Max Workers](#)

Min Workers: 2 [8](#) [A](#) Spot instances [?](#)

Driver Type: Same as worker (14.0 GB Memory, 4 Cores, 0.75 DBU) [Advanced Options](#)

Navigation sidebar:

- Home
- Workspace
- Recents
- Clusters (selected)
- Jobs
- Models
- Search

擷取資料 -1

- 您可以使用 OAuth 2.0 搭配 Azure Active Directory(Azure AD) 應用程式服務主體進行驗證，安全地存取 Azure Data Lake Storage Gen2(ADLS Gen2) 帳戶中的資料
- 使用服務主體進行驗證時，會提供兩個選項來存取儲存體帳戶中的資料
 - ◆ 特定檔案或路徑的掛接點
 - ◆ 直接存取資料
- 若要將特定路徑或檔案的存取權提供給多個工作區使用者，請對所需的儲存體資源和路徑建立掛接點
- 若要為具有不同許可權的多個工作區使用者提供存取權，請直接透過 Azure Blob File System (ABFS) 驅動程式存取資料

擷取資料 -2

- 在先前註冊的 [應用程式註冊] 總覽中，複製並儲存 應用程式(用戶端) 識別碼 和目錄(租使用者) 識別碼

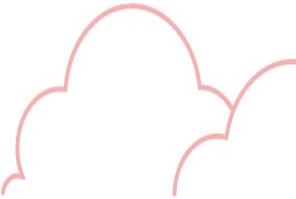
The screenshot shows the Microsoft Azure portal's App registrations section. A specific application named "Contoso App 1" is selected. The "Overview" tab is active. Key details visible include:

- Display name: Contoso App 1
- Application (client) ID: 11111111-1111-1111-1111-111111111111
- Supported account types: My organization only
- Redirect URIs: Add a Redirect URI

- 用戶端密碼：之前註冊的應用程式設定的密碼

The screenshot shows the password upload history for the "Contoso App 1" application. A single entry is listed:

描述	到期	值	ID
Password uploaded on Sat Apr 03 2021	3/4/2022	~_n*****	902ba91b-c42e-4757-acfd-0b308ccfe28f



掛接 ADLS Gen2 儲存體 -1

- 要掛接 ADLS Gen2 儲存體
- 使用服務主體作為認證，將 OAuth 2.0 驗證設定至 ADLS Gen2 儲存體帳戶
- 透過 Databricks API 建立掛接點
- 進入 Azure Databricks工作區，在“Common Tasks”下，選取 [NewNotebook]
- 在 [Create Notebook] 對話方塊中，輸入 Notebook 的名稱為“**Databricks員工編號**”。選取 [Python] 作為語言，然後選取您的 Spark 叢集。
- 選取 [Create] 進入 Notebook

掛接 ADLS Gen2 儲存體 -2

The screenshot shows the Databricks workspace interface. On the left is a vertical sidebar with icons for Home, Workspace, Recents, Data, Clusters, Jobs, Models, and Search. The main area features the 'Azure Database' logo and a 'Quickstart Tutorial' section. A central modal window titled 'Create Notebook' is open, prompting for a Name (mylab), Default Language (Python), and Cluster (my-spark-cluster). Below the modal are three cards: 'Explore the Quickstart Tutorial', 'Create a Blank Notebook', and 'Documentation'. The 'Common Tasks' and 'Recents' tabs are visible at the bottom.

Create Notebook

Name mylab

Default Language Python

Cluster my-spark-cluster

Cancel Create

Explore the Quickstart Tutorial

Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.

Quickly import data, preview its schema, create a table, and query it in a notebook.

Create a Blank Notebook

Create a notebook to start querying, visualizing, and modeling your data.

Documentation

New Notebook

Create Table

New Cluster

New Job

New MLflow Experiment

Import Library

Read Documentation

test

Documentation

Release Notes

Getting Started

掛接 ADLS Gen2 儲存體 -3

- 將下列程式碼寫入第一個程式格中

```
◆ configs = {"fs.azure.account.auth.type": "OAuth",
    "fs.azure.account.oauth.provider.type":
"org.apache.hadoop.fs.azurebfs.oauth2.ClientCredsTokenPro
vider",
    "fs.azure.account.oauth2.client.id": "<appId>",
    "fs.azure.account.oauth2.client.secret": "<clientSecret>",
    "fs.azure.account.oauth2.client.endpoint":
"https://login.microsoftonline.com/<tenant>/oauth2/token",
    "fs.azure.createRemoteFileSystemDuringInitialization":
"true"}
```

掛接 ADLS Gen2 儲存體 -4

- ◆dbutils.fs.mount(
source = "abfss://<container-name>@<storage-account-name>.dfs.core.windows.net/<folder1>",
mount_point = "/mnt/**samples員工編號**",extra_configs = configs)
- <appId> : 應用程式(用戶端) 識別碼
- <clientSecret> : 用戶端密碼
- <tenantID> : 應用程式目錄 (租用戶) 識別碼(2fa430c4-fd7e-4dfe-996c-46fc2cda44f8)
- <container-name> : ADLS Gen2 storage 帳號的 Container 名稱 (databricks)
- <storage-account-name> : ADLS Gen2 storage 帳號名稱 (dlakedemo)
- <folder1> : ADLS Gen2 的 Container 目錄(sample/data)
- Mount_point請改成**samples員工編號**

開始使用資料 -1

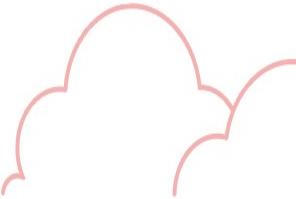
- 汇入相關套件
 - ◆ from sklearn import preprocessing
 - ◆ from sklearn.model_selection import train_test_split
 - ◆ from sklearn.linear_model import LogisticRegression
 - ◆ from matplotlib import pyplot
 - ◆ from pandas.plotting import scatter_matrix
 - ◆ import pandas as pd
- 汇入資料到 Pandas dataframe
 - ◆ df=pd.read_csv("/dbfs/mnt/**samples員工編號**/iris.csv")
- 顯示資料
 - ◆ df.shape
 - ◆ df.head()

開始使用資料 -2

- ◆ df.tail()
- ◆ df.dtypes
- ◆ df.info()
- ◆ df.describe()
- ◆ df['species'].unique()
- 資料作圖
 - ◆ scatter_matrix(df, figsize=(8,6))
 - ◆ pyplot.show()
- 建立 species 顏色
 - ◆ le = preprocessing.LabelEncoder()
 - ◆ df['species_le'] = le.fit_transform(df['species'])
 - ◆ df['color'] = df['species_le'].map({0:'r',1:'g',2:'b'})

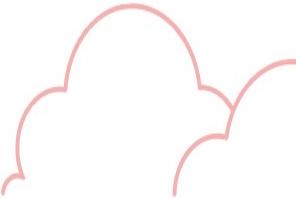
開始使用資料 -3

- 繪製散佈圖
 - ◆ `df.plot.scatter(x='sepal_length',y='sepal_width',c=df['color'])`
- 分割資料
 - ◆ `array = df.values`
 - ◆ `X = array[:,0:4] # 特徵值`
 - ◆ `y = array[:,4] # 標籤值`
- 將資料分為訓練與測試兩組
 - ◆ `X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=1)`



機器學習

- 建立模型
 - ◆ `lrm = LogisticRegression(solver='lbfgs')`
- 訓練模型
 - ◆ `lrm.fit(X_train, y_train)`
- 測試模型
 - ◆ `y_pred = lrm.predict(X_test)`
- 比較結果
 - ◆ `print(y_pred) # 預測值`
 - ◆ `print(y_test) # 實際值`
- 顯示精確度
 - ◆ `lrm.score(X_test,y_test)`



使用PySpark

- 使用 PySpark DataFrame

- ◆`sdf = spark.read.csv("/mnt/samples員工編號/iris.csv", sep=",", inferSchema="true", header="true")`

- 建立 View

- ◆`sdf.createOrReplaceTempView('iris')`

- 執行 SQL 查詢

- ◆`spark.sql('select count(species) from iris group by species').show()`

- 使用 sql magic

- ◆`%sql`
`select count(species) from iris group by species`

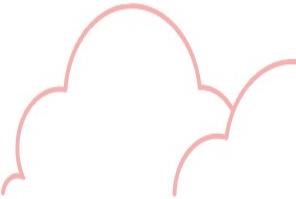
- 卸載掛接點

- ◆`dbutils.fs.unmount("/mnt/samples員工編號")`

清除資源



- Azure 雲端介紹
- Azure各項資源介紹
- AutoML介紹
- Lab – AutoML
- Azure Databrick介紹
- Lab – Azure Databricks



資源上限

- 閒置的資源仍會被收費：
 - ◆ Azure是使用者付費 Pay-as-You-Go
 - ◆ 使用多少資源就會收多少費用，閒置的計算資源也可能會被收費，因為是佔用了雲端的計算資源

- 資源的使用上限：
 - ◆ 每個群組/資源都會有個別資源的使用上限
 - ◆ Azure Databricks的叢集Cluster也有自己的資源使用上限
 - ◆ 叢集在未使用時可以手動終止，或是設定閒置一段時間後自動終止，叢集可以重新啟動使用。系統管理員可以訂選叢集，未被釘選的叢集會在終止後的30天由系統永久刪除，最多可以訂選40個叢集，當終止的叢集超過100個時會由最舊的叢集開始刪除

清除資源 -1

- 在您的Notebook中將叢集解除配置

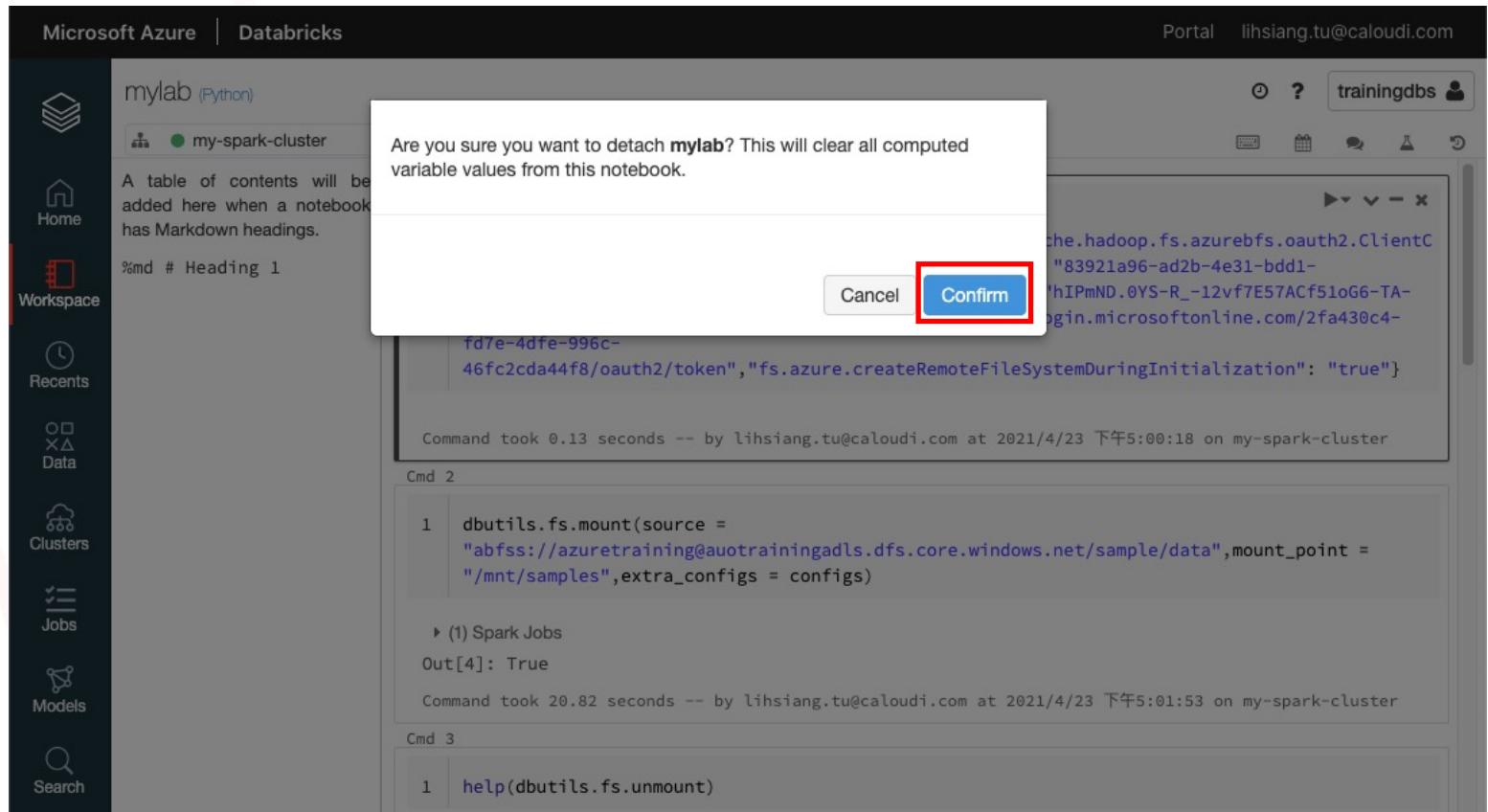
[Cluster]my-spark-cluster > Detach

The screenshot shows the Microsoft Azure Databricks workspace interface. On the left, there is a sidebar with icons for Home, Workspace (selected), Recents, Data, Clusters (highlighted with a red box), and Jobs/Models. The main area displays the attached cluster 'my-spark-cluster' (14.00 GB | 4 Cores | DBR 8.1 | Spark 3.1.1 | Scala 2.12). A red box highlights the 'Detach' button in the cluster menu. Below the cluster information, a command history notebook is open, showing the following content:

```
d326615af9a8", "fs.azure.account.oauth2.client.secret": "hIPmND.0YS-R_-12vf7E57ACf51oG6-TA-",  
", "fs.azure.account.oauth2.client.endpoint": "https://login.microsoftonline.com/2fa430c4-fd7e-4dfe-996c-  
46fc2cda44f8/oauth2/token", "fs.azure.createRemoteFileSystemDuringInitialization": "true"}  
  
Command took 0.13 seconds -- by lihsiang.tu@caloudi.com at 2021/4/23 下午5:00:18 on my-spark-cluster  
  
Cmd 2  
1 dbutils.fs.mount(source =  
"abfss://azureretraining@auotrainingadls.dfs.core.windows.net/sample/data",mount_point =  
"/mnt/samples",extra_configs = configs)  
  
▶ (1) Spark Jobs  
Out[4]: True  
  
Command took 20.82 seconds -- by lihsiang.tu@caloudi.com at 2021/4/23 下午5:01:53 on my-spark-cluster
```

清除資源 -2

● 確認刪除 Confirm



清除資源 -3

- 進入Clusters頁面，找到已經解除配置的Cluster

The screenshot shows the Microsoft Azure Databricks interface. The top navigation bar includes 'Microsoft Azure | Databricks' on the left and 'Portal lihsiang.tu@caloudi.com' on the right. The left sidebar has icons for Home, Workspace, Recents, Data, Clusters (which is highlighted with a red box), Jobs, Models, and Search. The main content area is titled 'Clusters' and shows a table of clusters. The table has columns: Name, State, Nodes, Runtime, Driver, Worker, Creator, and Actions. One row is selected, showing 'my-spark-cluster' as the name, 'Running' as the state, '1' as the number of nodes, '8.1 (includes Apache Spark 3.1.1,...)' as the runtime, and 'Standa...' and 'tim.yang...' as the driver and worker respectively. A red box highlights the 'my-spark-cluster' row. At the bottom of the table, it says '1 - 1 of 1'.

Name	State	Nodes	Runtime	Driver	Worker	Creator	Actions
my-spark-cluster	Running	1	8.1 (includes Apache Spark 3.1.1,...)	Standa...	Standa...	tim.yang...	1

清除資源 -4

- 確認沒有掛接在這個叢集下的Notebook，就可以選擇終止

The screenshot shows the Databricks Cluster configuration interface for a cluster named 'my-spark-cluster'. The left sidebar includes icons for Home, Workspace, Recents, Data, Clusters (selected), Jobs, Models, and Search. The main panel displays the cluster details: Policy (Unrestricted), Cluster Mode (Single Node), Databricks Runtime Version (8.1 (includes Apache Spark 3.1.1, Scala 2.12)), Autopilot Options (Terminate after 120 minutes of inactivity checked), and Node Type (Standard_DS3_v2). The 'Notebooks (0)' tab is currently active, and the 'Terminate' button in the top navigation bar is highlighted with a red box.

清除資源 -5

- 待叢集終止後就可以選擇刪除

The screenshot shows the Microsoft Azure Databricks Cluster configuration interface. The cluster name 'my-spark-cluster' is highlighted with a red box. The 'Delete' button in the top right corner of the cluster card is also highlighted with a red box. The left sidebar shows navigation options: Home, Workspace, Recents, Data, Clusters (selected), Jobs, Models, and Search.

Microsoft Azure | Databricks

Clusters / my-spark-cluster

my-spark-cluster

Edit Permissions Start Clone Delete

Configuration Notebooks (0) Libraries Event Log Spark UI Driver Logs Metrics Apps Spark Cluster UI - Master

Policy ? Unrestricted

Cluster Mode ? Single Node

Databricks Runtime Version 8.1 (includes Apache Spark 3.1.1, Scala 2.12)

Autopilot Options Terminate after 120 minutes of inactivity ?

Node Type ? Standard_DS3_v2 14.0 GB Memory, 4 Cores, 0.75 DBU

Advanced Options

清除資源 -6

● 確認刪除

The screenshot shows the Microsoft Azure Databricks Cluster UI. On the left, there's a sidebar with icons for Home, Workspace, Recents, Data, Clusters (which is selected and highlighted in red), Jobs, Models, and Search. The main area shows a cluster named "my-spark-cluster". A modal dialog box is open, prompting the user to confirm the deletion of the cluster. The dialog contains the text: "Deleting 'my-spark-cluster'." and "Deleting a cluster terminates the cluster and removes its configuration. You cannot undo this action." Below this, it asks "Are you sure you want to delete 'my-spark-cluster'?" with "Cancel" and "Delete" buttons. The "Delete" button is highlighted with a red box.

Microsoft Azure | Databricks

Clusters / my-spark-cluster

?

trainingdbs

my-spark-cluster

Configuration Notebooks (0)

Policy ?

Unrestricted

Cluster Mode ?

Single Node

Deleting "my-spark-cluster".

Deleting a cluster terminates the cluster and removes its configuration. You cannot undo this action.

Are you sure you want to delete "my-spark-cluster"?

Cancel Delete

Cluster UI - Master

Databricks Runtime Version

8.1 (includes Apache Spark 3.1.1, Scala 2.12)

Autopilot Options

Terminate after 120 minutes of inactivity ?

Node Type ?

Standard_DS3_v2 14.0 GB Memory, 4 Cores, 0.75 DBU

Advanced Options

清除資源 -7

- 回到叢集列表，確認叢集已經被刪除

The screenshot shows the Microsoft Azure Databricks interface. The top navigation bar includes 'Microsoft Azure' and 'Databricks'. On the right, it shows 'Portal lihsiang.tu@caloudi.com'. The left sidebar has icons for Home, Workspace, Recents, Data, Clusters (which is selected and highlighted in red), Jobs, Models, and Search. The main content area is titled 'Clusters' and shows tabs for 'All-Purpose Clusters' (selected), 'Job Clusters', 'Pools', and 'Cluster Policies'. A blue button labeled '+ Create Cluster' is visible. Below the tabs is a search bar and a table header with columns: Name, State, Nodes, Runtime, Driver, Worker, Creator, and Actions. The main message in the center says 'No Clusters'.

清除資源 -8

- 提醒您，Azure雲端服務是依照使用的資源計費，閒置未使用的資源亦可能因相依的服務被收費(例如儲存體帳戶)，除了產生不必要的費用之外，佔用各項資源的配額也可能導致他人無法建立所需資源。
- 建議您，在確認無需使用該項服務時，解除配置或刪除閒置資源