

## AI-Stack Express

# 機器學習訓練協作管理平台

AI-Stack Express 能將單台 GPU 伺服器或多台 GPU 伺服器叢集，升級為易於控管、多人共享、提升資源運用與追蹤，並賦予 GPU 伺服器橫向擴容能力的機器學習運算環境資源池，積極為 GPU 運算資源帶來彈性分配、高效協作與強化運作效益。

## 用 GPU 伺服器搞 AI，就找 AI-Stack

擁有國家高速網路與計算中心 TWGC 管理 NVIDIA DGX-1 和 HGX 叢集的成功經驗，AI-Stack Express 機器學習訓練協作管理平台是您在使用 NVIDIA DGX-1/2、DGX Station、EGX 和 DGX POD 時，不可或缺的絕佳軟體。在業界，AI-Stack Express 獲得台灣知名半導體製造大廠青睞，採用本軟體掌握高密度 GPU 伺服器叢集管理任務，並作為集團 AI 人才培訓平台。在教育研究領域中，AI-Stack Express 透過高效型管理介面，積極協助大專院校與研究單位提升 AI 研究能量。故在各項專業應用領域中，AI-Stack Express 都是您跨進 AI 世代的最佳選擇。

## 導入 AI 雲基礎架構、流程與方針

AI-Stack Express 機器學習訓練協作管理平台為 AI 工作者提供了一組新穎直觀的基礎架構、流程與方針，讓開發者、資料分析師、人工智慧訓練師或 AI 教學者專注在 AI/ML 任務，不僅節省下可觀的系統維護、調校和佈署等人力與時間，平台同時支援 GPU 運算和 AI Framework 資源調配流程自動化功能，大幅降低 AI 工程師與科學家採用 Tensorflow、Caffe 和其他深度學習工具的複雜性和學習門檻。

## 資源共享、流程簡易、快速進入機器學習任務

傳統的 GPU 伺服器只能適用於 1 人 1 機的使用情況，且僅限於單一 AI 用途，光是準備基礎建設雲端環境就需要三個星期的時間，相當於 35% 的工作量都花在環境準備上，曠日費時。現在透過 AI-Stack Express 友善的操作介面，只要幾分鐘的時間，就可以將 AI 工作所需的環境設置完畢，協助您快速開始機器學習任務。

- **簡易的 Web 介面**—提供網頁操作介面，透過點選所需的服務功能，便能快速建立容器。Web-based 網頁互動式工具，讓使用者可隨時隨地登入環境進行開發與訓練任務。
- **機器學習**—提供多種常用熱門的 AI Framework 於建立 GPU 容器時自由選擇，並自動登入常用的社群開發軟體 Jupyter Notebook 編寫演算程式。
- **資源分享**—可支配每用戶/容器配置一片或以上的實體 GPU pass-through 資源，保障運算效能充足。
- **狀態查詢**—細部查詢容器資訊，將清楚呈現容器 ID、名稱、運作狀況與連線資訊。
- **安全管理**—提供使用者透過密碼與金鑰對(key pair)方式進行容器 SSH 連線。

## 採雲端租戶機制、自助服務、管理方便

AI-Stack Express 提供您建構一個自給自足、可控可管、可共享、可橫向擴容的私有雲 AI 運算環境，提供可靠、強大、具成本競爭力、高資源效率與效益的 AI 訓練資源池。管理者可依部門層級或專案需求建立多個租戶，並將使用者加入租戶中形成不同群組，達到資源有效分配與控管目的。

- **自動化自主服務**—提供自助線上申裝模式，不需透過傳統線下作業由管理者建立容器後再後配發給使用者，平台將提供使用者自助建立屬於自己的 GPU 資源。
- **租戶管理**—租戶機制將能完整隔離每個租戶使用的資源，確保資源額度不互相佔用並提升資料隔離安全性。

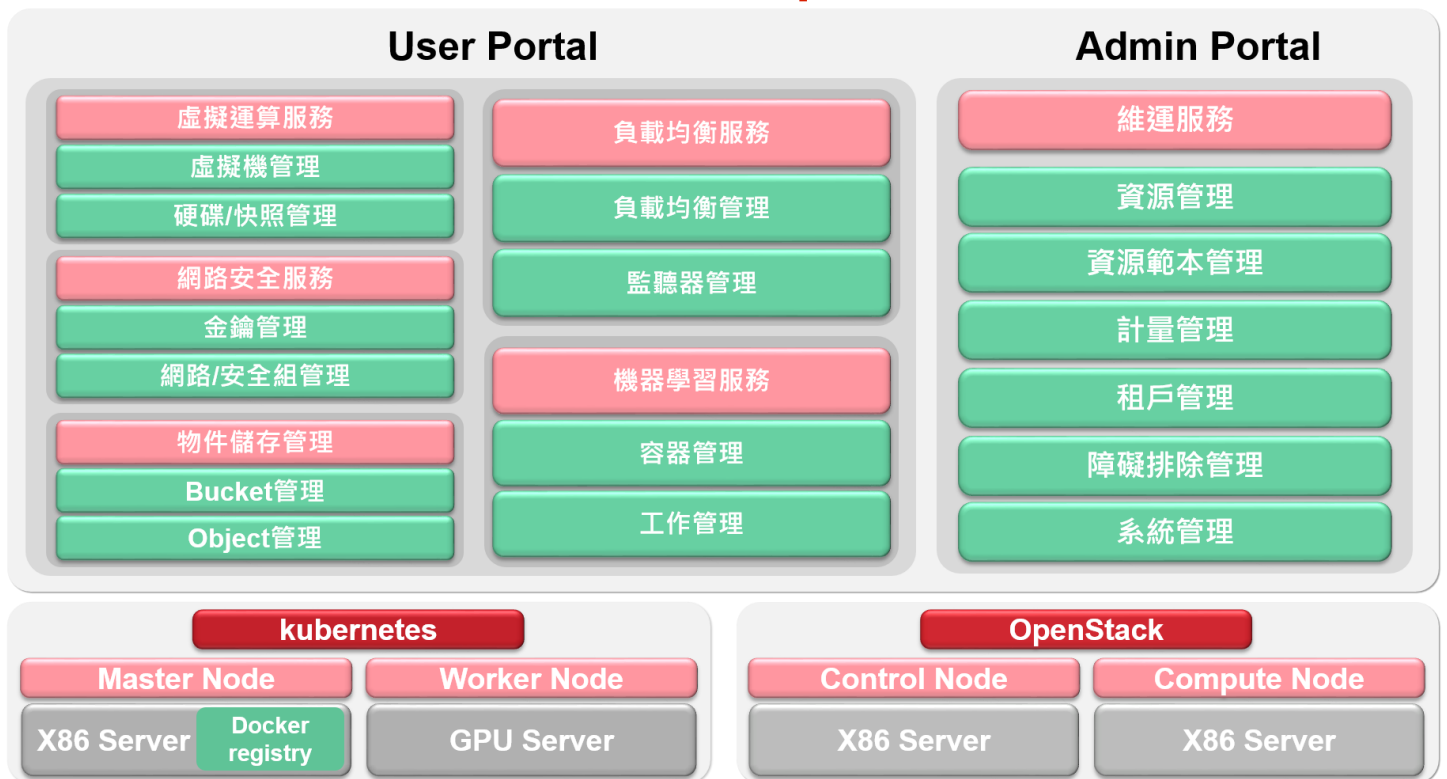
- **額度控管**—每個用戶能使用的 CPU、Memory、GPU 額度將妥善受到平台監控與管理，提升資源運用效率。
- **Image 範本/規格管理**—具備 AI 容器規格及 Image 範本的管理功能，同時可指定 Image 範本之適用對象，達到開發環境與研究成效的權限管控。
- **日誌分析**—紀錄使用者登入系統後的操作行為，提供後續操作行為分析與解讀。

## 進階功能、有效控管、提升資源使用效率

AI-Stack Express 提供進階管理功能，讓租戶管理者除了一般容器建立權限外，可透過進階功能的批次建立容器操作、容器使用時長管理以及租戶成員申請容器的審核功能，提升整體 AI 資源使用效率與管控能力。

- **批次建立**—使用同一個容器樣板一次為多位使用者建立容器，減少使用者個別建立與操作時間。
- **使用時長管理**—預約容器的啟用與刪除時間，於特定時間自動建立並於結束時自動釋放容器資源功能，以達到有效控管資源使用。
- **流程服務**—提供租戶使用資源申請機制，透過租戶管理者的審核機制來防止資源濫用的可能。

# AI-Stack Express



➤ 了解更多資訊  
[infinitiesoft.com](http://infinitiesoft.com)