AI-STACK

機器學習/深度學習協作管理平台

特色

1. AI-STACK是多人共用的AI運算平台，提供多GPU伺服器的算力調度協同管理，透過統計圖表，可輕鬆掌握GPU資源的配置狀況，有效提升利用率，同時可追蹤每個使用歷程，建立費用攤提依據。
2. 透過AI-Stack可降低AI學習門檻，創造開機立即可用、硬體高度整合、軟體最佳搭配的使用經驗。
3. AI-Stack具有高度可擴張性，能協助政府、學校與企業發展不同階段的AI工作，讓想要發展AI的客戶可用最適的預算，簡單快速的開始進行AI實驗與開發，並保留日後發展擴充的彈性，從建立AI lab、資料準備、AI訓練、AI推論到資料提取均可依客戶需求進行規劃。

叢集運算

- 使用 Kubernetes 與 Docker 容器化技術管理
- 可同時管理多台 GPU 伺服器、混 GPU 卡環境
- 整合公有雲快速擴充 GPU 資源

快速建立研發環境

- 簡易步驟快速建立 (個人 / 團隊) ML環境
- 提供Jupyter Notebook、JupyterLab開發工具
- 支援NGC AI 深度學習框架，支援自製鏡像

帳號控管、群組資源配額與權限管理

- 支援Single Sign-On (SSO)，可介接LDAP/AD/OpenID/Oauth2
- 運算資源隔離，SSH安全殼層防護，可設定IP白名單
- 簽核系統可依部門層級或專案需求建群組、權限、額度

批次建立、資源共享、儲存服務

- 可批次建立容器，安排預約申請作業，批次任務定期排程
- 提供多人多容器建立在單一GPU卡上運行
- 支援儲存服務 (eg. NFS、S3)

AI-STACK 滿足IT管理者與機器學習工作者需要

1 降低使用者門檻
簡易步驟快速建立 (個人/團隊) ML環境

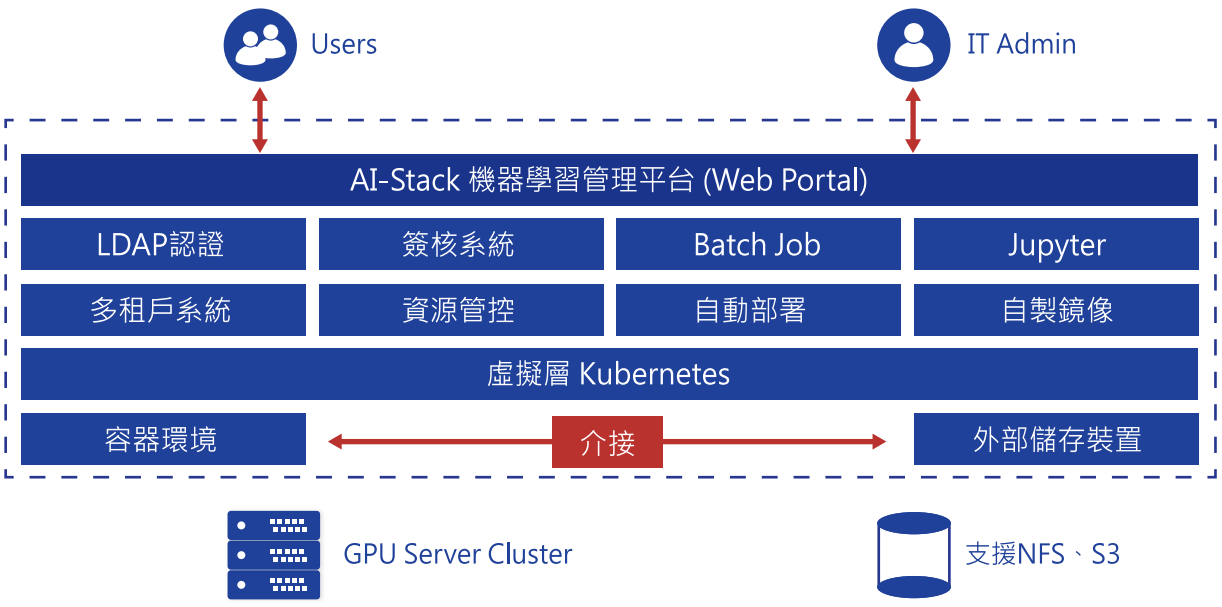
2 掌握有效資源
彈性IT資源共享、額度限制、工作排程

3 環境自主權
IT環境隔離、SSH Key或密碼登入環境

4 易於使用和共用
帳號/儲存整合,批次、預約申請作業

AI-STACK 核心架構

1. 平台提供 Web 操作介面與統一用戶入口門戶(user portal)，以圖形化方式進行深度學習容器自助式申請建立容器與日常操作。
2. 平台能對不同使用者進行資源自動分配與部署，以及可支援不同深度學習框架。
3. 平台內建 NVIDIA 優化之常用 TensorFlow 、Caffe、PyTorch、MXnet、RAPIDS 之 AI 框架，以Chainer 與 TensorRT，並具備 AI 框架擴充設計。
4. 平台流程管控與自動化內建符合 OMG 開放性業務流程三個標準 (BPMN、CMMN及DMN) 的流程引擎，以便自動化服務流程的建立、調整及管理與資源申請之簽核功能。
5. 本平台環境需求為 Ubuntu18.04, 搭配 NVIDIA Driver, NVIDIA CUDA (10.0以上), NCCL, cuDNN等技術提供AI機器學習訓練服務，並使用 NVIDIA Docker 與 Kubernetes 進行 GPU 容器管理。



AI-Stack提供企業建構自給自足、可控可管、可共享、可橫向擴容的私有雲AI運算環境，為企業提供可靠強大、具成本競爭力、高資源效率與效益的AI計算資源池，透過支援GPU和AI流程自動化，減少維護、調整與部署的時間。

INFINITIES 數位無限軟體
PRO WARE by UNIFOSA 商丞科技 系統整合

商丞科技股份有限公司
電話：(02)2914-8001 傳真：(02)2914-7975
信箱：info@proware.com.tw 網址：www.proware.com.tw