

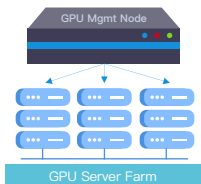
Gemini AI Console

高效能挖掘海量資料商機，優化巨資分析與AI運算流程

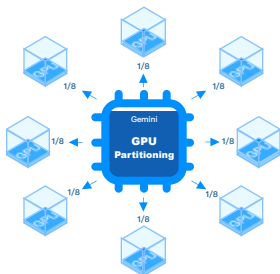
產品簡介

從大數據的資料預處理、AI 模型的訓練、一路最終至 AI 預測應用部署的複雜流程中，企業經常因不同階段所需的異質環境、多種運算架構、甚至跨部門的協作而困擾。透過 Gemini AI Console 可打造出便利企業組織進行跨單位 AI 專案協作的 GPU 管理節點，協助企業有效率地從海量資料挖掘商機。

資料科學家與開發人員可透過 Gemini AI Console 使用者服務入口，快速開啟大量預載大數據與 AI 工具的運算叢集環境，只需透過網頁表單點選即可立即開啟各種所需的 AI 運算工具服務，並掛載 GPU 伺服器本機或其他儲存空間至容器內，直接擷取資料進行運算，將底層運算架構建置化繁為簡，進而能夠將時間與人力資源專注投入在核心演算法上。



獨家技術



GPU PARTITIONING 多容器共享

✓ 最多可分成 8 x 1/8 專屬 GPU 資源

支援 1/2 1/4 1/8 分割，每個分割後的 GPU，都享有專屬運算資源，保證資源水準不受干擾

✓ 同時執行，提高 GPU 使用率

GPU 在切割後，可以在同一張 GPU 卡上同時執行多個容器或專案，漸少搶佔，資源利用率提高

✓ 縮減 GPU 使用時間，提高生產力

多個任務專案可以同時享受資源，增加公平性，因而減少任務專案的排隊時間，總任務時間下降，從而達成公司生產力提高

產品功能

功能	描述
異質/混合雲平台	本系統內建一組 Kubernetes 容器管理平台，可直接擴充您的 GPU 運算主機，並支援異質混合雲模式，可管理來自公有雲如 Google Cloud Service、Linode 等，另也完整支援 OpenStack 虛擬機管理平台，並可串接多種外部儲存設備 (如 NFS)，成為共享資源池。
GPU Partitioning 多容器共用	雙子星獨家 GPU Partitioning 多容器共享技術，支援單片 GPU 卡最多分割為 8 x 1/8 專屬 GPU 資源，可同時執行 8 組容器運算，縮短整體 GPU 使用時間，提高使用效率。
Job/Pipeline 建置 MLOps 環境	採用 Job 批次任務模式，運算完畢後即釋出資源，並透過 Pipeline 功能串接任務，讓不同類型的任務使用不同環境與資源。搭配內建模板與排程器功能，使用者不需要重複設定，即可讓 Job/Pipeline 自動化運行，以快速搭建 MLOps 環境。
AI 框架與開發工具	除了內建多種 AI Framework 之外，管理者也可透 Image Registry 自行上傳特定映像檔，並搭配軟體方案功能設定容器資源模板，快速提供使用者更多軟體開發工具。
API / Web UI 與監控	提供完整 Web UI 給管理者與使用者透過圖形化方式進行操作，同時也有相應的 API 提供使用者以程式化方式操作。每組帳號均有專屬監控儀表板，依據權限可瀏覽節點、GPU、容器等即時資源使用狀態，另有提供歷史資料報表匯出，搭配計價功能以便計費。
三層角色與多租戶管理	系統管理者 (System Admin) 可建立多個計畫 (Project)，可透過 LDAP 或內建認證機制管理所有帳號與權限，並設定計畫與帳號的資源分配。每個計畫皆有計畫管理者 (Project Admin) 與計畫使用者 (Project User)，依據不同的角色可管理與使用不同類型的功能。

產品效益

👍 簡化 IT 複雜性，優化 GPU 管理效益

管理者在單一管理節點上，管理數百至百台的 GPU 及 CPU Server，並可依據組織分工與 AI 專案管理需求優化實體與虛擬資源管理，所有用戶與專案皆可在管理者設定的 GPU、CPU、Memory 配額內共享運算資源。

👍 提升研究人員研發效率，縮短開發時間

將底層運算架構建置化繁為簡，只需透過網頁表單點選即可立即開啟各種 AI 運算工具服務，內建四種 AI 框架，包含 Tensorflow、PyTorch、Caffe2 與 CUDA，均可使用 Jupyter Notebook 與 SSH 登入使用，管理員也可透過模板自行擴充，幫助科學家與開發人員專注在分析演算法。

👍 支援不同運算架構與異質環境，完美 AI 運算體驗

單一平台同時管理大數據 + AI 機器學習運算叢集，無論底層需要的是虛擬機 (VM) 或是容器 (Docker) 搭配 GPU 都讓使用者一目了然、輕鬆管理；並支援雲端服務 (Cloud Service) 及批次工作 (Batch Job) 兩種運算模式，滿足不同的使用需求情境，提升 AI 開發流程體驗。

地址：30070 新竹市東區公道五路二段158號11樓之3

Email: contact@geminiopencloud.com Tel: (03)6590698 Fax: (03) 5753186



<https://www.geminiopencloud.com>