

啟用 GPU 即服務

GPU 基礎架構使用容器的雲端型體驗

在現在運算密集型的業務環境中，利用資料的能力對企業的成功非常重要。隨著採用 AI 的企業增加，Hewlett Packard Enterprise 為了迎接機器學習 (ML)、深度學習 (DL)，以及進階資料分析所帶來的挑戰，提供所需的運算及儲存功能。現在，HPE 提供可實現內部部署的全新 GPU 即服務 (GPUaaS) 解決方案。

可供企業進行 AI 部署的 GPU 加速工作負載

開發 ML 和 DL 預測模型需要密集運算。相較於只使用 CPU 的系統，使用圖形處理單元 (GPU) 等加速器可提升效能，使開發的速度大幅增加，因此 GPU 成為 ML 和 DL 的通用基礎架構選擇。

不過，現在大部分企業的 IT 團隊發現，若有多種不同的 ML/DL 應用和使用案例，將會面臨要滿足多個資料科學團隊對於 GPU 的需求持續增加的挑戰。

此外，處理複雜的適當軟體元件和相關基礎架構非常耗時，而且每當要求新的 ML/DL 應用程式時都必須重複相同的程序。

一旦佈建了基礎架構，IT 就幾乎無法查看使用率，因此無法為不同的應用程式重新配置基礎架構。缺乏這種可見性，也讓 IT 難以實作更穩健的成本分攤模型。

有的公用雲端提供隨需部署虛擬化 GPU 資源 (也就是 GPU 即服務) 的能力，但公用雲端並非唯一的解決方案，而且在某些情況下可能無法選擇。許多企業組織基於安全性、效能或資料引力的考量，有需要內部部署的工作負載需求。

隨需和彈性佈建 GPU 資源

現在有一種 GPUaaS 解決方案，結合同級最佳的 HPE 基礎架構和 HPE Ezmeral 容器平台，以及 HPE Pointnext Services，可確保成功部署。這個全新的 HPE 解決方案，可以讓 IT 企業組織提供內部部署的 GPUaaS，以提高業務敏捷性、優化 GPU 使用率，以及減少 GPU 的整體 TCO。

使用容器型 HPE Ezmeral 容器平台，即可將來自多個異質伺服器的 GPU 合併和跨多個應用程式共用 - 只要點幾下滑鼠，即可隨需和彈性佈建容器化的 GPU 資源。若要啟用 GPUaaS，可以將 HPE Ezmeral 容器平台部署在支援 GPU 的伺服器內，這些伺服器包括搭載 NVIDIA® Tesla 或 Quadro GPU 的 HPE Apollo 和 HPE ProLiant 伺服器。

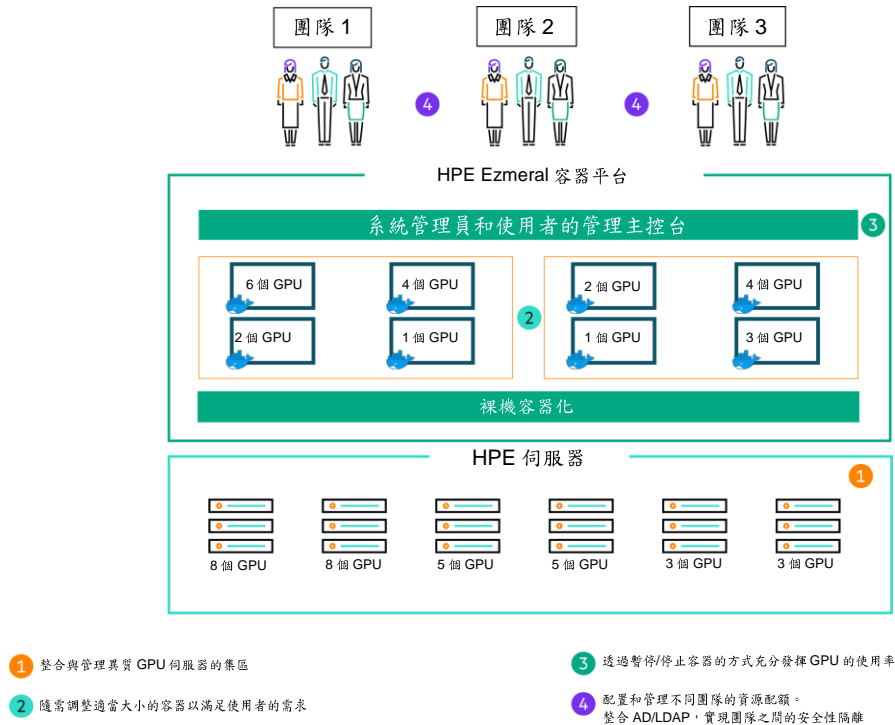
此外，利用暫停容器的獨特能力 (釋放其中的 GPU、CPU 和記憶體資源，並保留整體應用程式狀態)，資料科學家團隊即可在共用的 GPU 基礎架構上執行多個不同的 ML/DL 應用程式，而不需要重新建立或重新安裝其應用程式和程式集。

重要功能

- **統一的管理主控台**：系統管理員的圖形化使用者介面 (GUI) 提供監控和管理共用 GPU 資源集區的能力，以及跨多部伺服器和多個使用者群組的完整 GPU 使用率可見性和使用情況報告。
- **隨需的彈性佈建 GPU 資源**：存取一或多個 GPU，即可快速且輕鬆部署應用程式。可以隨需佈建配備 GPU 的全新容器化環境，然後當不再需要時取消佈建 (釋放 GPU)。
- **暫停和重新啟動應用程式**：提供可在保留應用程式目前狀態的同時暫停應用程式和釋放已連接 GPU 的能力。這可讓 IT 系統管理員監控使用情況，並且在執行 GPU 專屬的程式碼時重新指派 GPU。
- **企業級安全性和多組織用戶管理**：提供可在共用相同基礎架構 (包含 GPU) 的多個使用者和專案團隊之間進行的多組織用戶管理和資料隔離。這包括與安全性和驗證的整合，例如 LDAP、Active Directory，以及 Kerberos。
- **現成支援 GPU 的應用程式映像**：包括預先整合的容器映像，可供支援 GPU 的通用應用程式和 ML/DL 工具使用，例如 TensorFlow、H2O、Caffe2 和 JupyterHub - 以及 Ubuntu 和 CentOS 的公用程式映像 - 包含 NVIDIA CUDA 驅動程式。IT 可以將這些映像快速升級到新版，並且視需要新增新的工具。

解決方案簡介

- **容器的裸機效能**：專利的創新功能，可以為 ML/DL 工作負載提供容器的敏捷性優勢，同時確保與其裸機伺服器相同的效能。
- **外部儲存連線與資料存取控制**：將運算和資料儲存分隔的能力，可免除複製或移動資料的需求。敏感性資料可以停留在您配備企業級資料控管的安全儲存系統，而沒有建立和維護多重副本或移動大規模資料的成本和風險。



主要優勢

- **降低成本**：透過改善 GPU 使用率、控制使用情況、消除叢集蔓延，以及將重複資料刪減到最少來實現節省成本。
- **簡化部署**：只要按幾下滑鼠，即可為每個工作負載提供適當大小的 GPU 環境，以及為您的資料科學家提供所需適當數量的 GPU。
- **加速 ML/DL 部署**：在數分鐘內佈建和取消佈建 GPU 資源(而不是數天)。透過可用於一般 ML/DL 應用程式、資料科學工具，以及資料架構的預先整合容器映像，可以進行快速原型設計和更多測試。
- **保持安全性和控制**：整合貴企業的安全性和驗證系統，提供可用於 GPU 和其他資源的內建控管與微調存取控制。

透過 HPE GPUaaS 解決方案，企業可以打造其資料科學團隊，並且在先進 HPE 基礎架構的多節點容器化環境中利用 GPU 加速的應用程式快速執行。可以在數分鐘內佈建配備 GPU 的完整設定環境，讓資料科學團隊可以快速建立原型、反覆查看，以及利用其慣用的 ML/DL 工具和架構進行測試。企業 IT 團隊可以在多組織用戶共享的架構中確保企業級的安全性、資料保護，以及效能，同時提高投資報酬率(ROI)並降低其 GPU 基礎架構的整體 TCO。

了解更多訊息

hpe.com/info/container-platform

做出正確的購買決定。
與我們的售前規劃專家聯繫。



交談



電子郵件



撥打



獲取更新

**Hewlett Packard
Enterprise**

© 2019-2020 版權所有，Hewlett Packard Enterprise Development LP。本文所含資訊如有更改，恕不另行通知。Hewlett Packard Enterprise 產品與服務的保固載於產品與服務隨附的明示保固聲明中。本文中的任何資訊均不構成額外保固。Hewlett Packard Enterprise 對於本文所含的技術、編輯錯誤或遺漏概不負責。

NVIDIA 是 NVIDIA Corporation 在美國及其他國家/地區的商標和/或註冊商標。Active Directory 是 Microsoft Corporation 在美國和/或其他國家/地區的註冊商標或商標。Docker 標誌是 Docker, Inc. 在美國和/或其他國家/地區的商標或註冊商標。所有第三方商標均為其各自所有者的財產。

a00075067ENW，2020 年 9 月，第 2 修訂版