Dell AI Factory

Dell 生成式 AI 解決方案搭配 AMD

透過適用於複雜 GenAI 的可擴充式模組化架構，
加速創新、降低成本並保護資料。

75%
相較於公有雲 IaaS 推斷成本效益提高比率[1]

86%
價值實現時間縮短比率[2]

關鍵使用案例需要效能、彈性及擴充能力

助理、聊天機器人和內容創作

加速器即服務

多模檢索增強生成 (RAG)

簡化

透過備受肯定且經驗證的解決方案，以及超過 34 萬小時的工程時間，簡化 GenAI 部署。

最佳化效能
高效能加速器、開放式架構及 AI 最佳化網狀架構

隨時隨地運用 AI
利用多雲儲存彈性，隨時隨地取得資料

一站式多節點
經實證的全堆疊 AI 基礎，可加快交付成果

客製化

AMD ROCm™ 開放原始碼軟體和開放式生態系統可提升 AI 開發及作業。

加快創新速度
使用開放原始碼軟體和生態系統來開發獨特的應用程式。

加速開發
利用業界標準架構與彈性技術堆疊。

啟用您的資料
有效率地同時執行多個 AI 使用案例。

值得信賴

82% 的 ITDM 偏好採用內部部署或混合模式。[3]
資料決定您的成果。務必周全保護。

快速入門
具有信任根安全性和完整控制能力的內部部署基礎

簡化連線
安全且功能豐富的網狀架構，具備擴充能力及最佳化的流量

自動佈建
部署和管理高效能叢集的開放原始碼基礎

Dell GenAI 解決方案搭配 AMD

推論

在單一 AMD Instinct™ MI300X 加速器上執行 70B 參數模型。[4]

自訂

在單一 Dell PowerEdge XE9680 上部署並微調 8 個並行 70B 模型。[4]

增強

將您的資料整合至生成式程序。

透過經實證的開放式解決方案，大規模提供安全的內部部署 AI 應用程式，實現競爭優勢。

深入瞭解

[1]企業策略集團，最大化 AI 設資報酬率：通過 Dell Technologies 在內部部署進行推斷，成本效益比公有雲高出 75%，2024 年 4 月。
[2]估計值根據 2024 年 5 月 Dell 分析結果而得，比較使用自動化指令檔為一般用途 LLM 設定 2 個節點 Kubernetes 叢集、與手動部署適用設計的時間，設定時間僅包括基本安裝，實際設定時間將因解決方案組態而有所不同。
[3]Dell Technologies，生成式 AI 脈動調查，2023 年 9 月與 9 月。
[4]Dell Technologies 部落格，好多元性：在搭載 AMD Instinct MI300X 加速器的 PowerEdge XE9680 上部署 GenAI，2024 年 5 月。

版權所有 © 2024 Dell Inc. 或其子公司。保留所有權利。Dell 與其他商標均為 Dell Inc. 或其子公司的商標。AMD、AMD Arrow 標誌、ROCm™、Instinct™ 與其相關組合為 Advanced Micro Devices, Inc 的商標。PyTorch® 為 PyTorch 或 PyTorch 授權人的商標或註冊商標。TensorFlow™ 和任何相關標誌均為 Google Inc. 的商標。其他商標是屬於其各自擁有者之財產。資訊圖表在 2024 年 10 月於美國出版。

Dell Technologies 確信本文件在出版日期之時，相關的資訊正確無誤。以上資訊若有變更恕不另行通知。