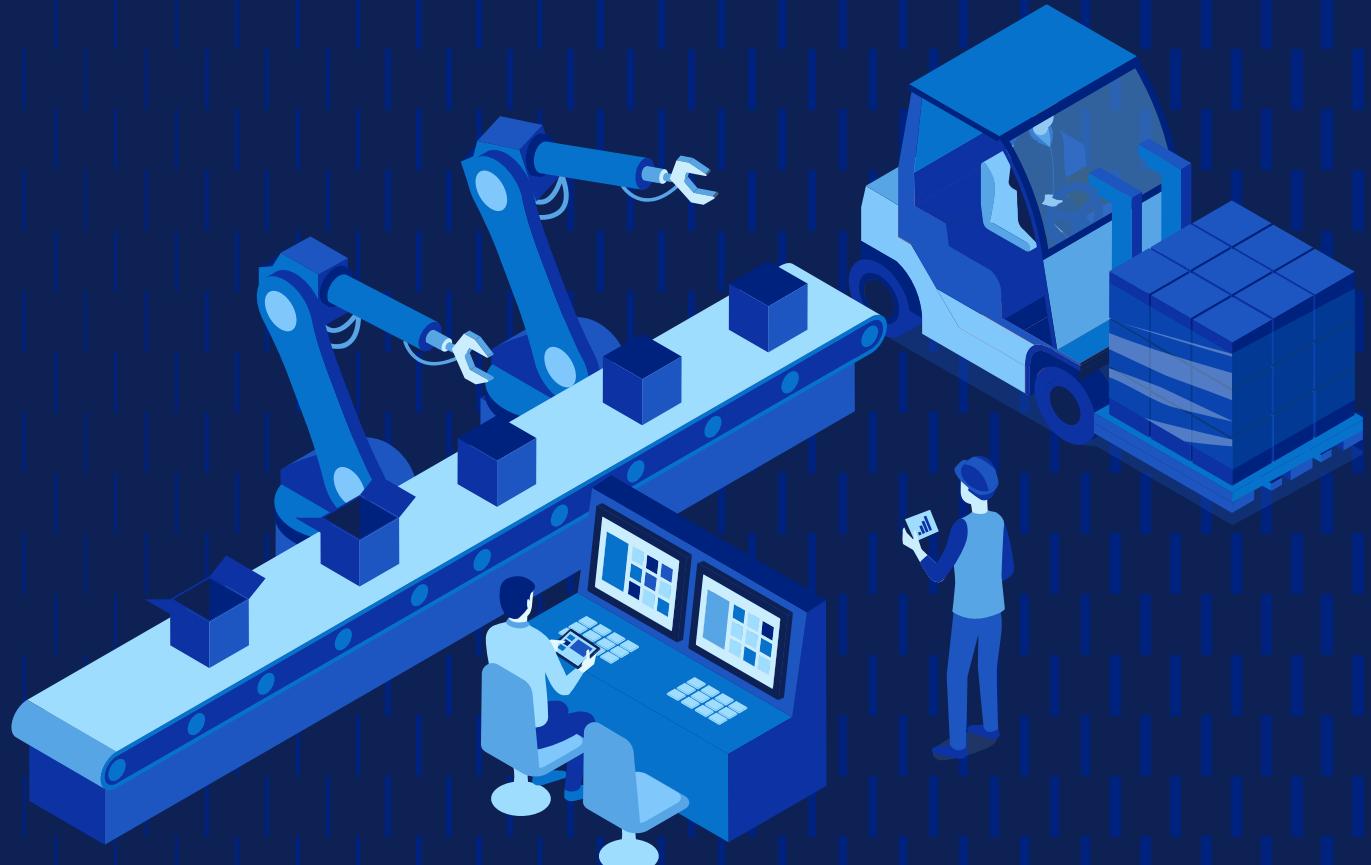


# 邁向 AI 之路

Dell AI Factory with NVIDIA：透過尖端的 AI 創新和可擴充的 AI 工作流程自動化，重新構想業務推展速度



# Dell AI Factory with NVIDIA

## 加快採用 AI 可讓您充分發揮資料的潛力

Dell AI Factory with NVIDIA 透過提供整合式 Dell 和 NVIDIA® 功能，加速採用 AI 技術的使用案例、整合資料和工作流程，並讓您設計自己的 AI 旅程，以取得可重複且可擴充的成果，藉此加快 AI 採用速度。



更快實現價值



轉換深入見解



最佳化生產力



提供可信賴的結果

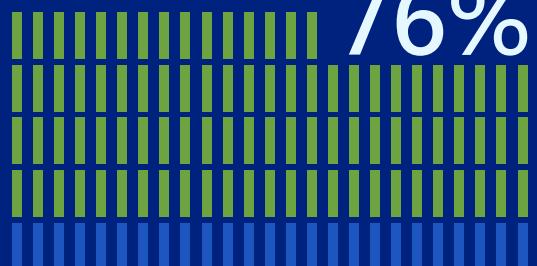
全世界最廣泛的 GenAI 解決方案產品組合，從桌上型電腦到資料中心再到雲端，一應俱全<sup>2</sup>

75%

相較於公有雲 IaaS，推斷成本效益更高。<sup>3</sup>

86%

透過一站式的完整堆疊解決方案與服務，將時間轉換為價值的速度提升的比例。<sup>4</sup>



的 IT 和業務主管認為，GenAI 可為組織創造改變現狀的價值。<sup>1</sup>

Dell Technologies 與 NVIDIA 之間有著長期的合作關係，擁有超過 25 年的聯合創新經驗，著重於加速創新，並提供尖端平台、解決方案和軟體，為我們的共同客戶帶來轉型成果。

<sup>1</sup>《Dell Technologies Innovation Catalyst Study》(Dell Technologies 創新催化劑研究) · 2024 年 2 月。

<sup>2</sup>根據 2023 年 8 月的 Dell 分析，Dell Technologies 提供專為支援 AI 工作負載所打造的解決方案，包括工作站電腦 (行動或固定) 及高效能運算伺服器、資料儲存、雲端原生軟體定義式基礎結構、網路交換器、資料保護、HCI 與各項服務。

<sup>3</sup>根據 2024 年 4 月 Dell 委託撰寫的企業策略集團研究，其中比較了內部部署 Dell 基礎結構與原生公有雲基礎結構即服務。分析的模型顯示，對於擁有 5 千名使用者的組織，運用 RAG 的 7B 參數 LLM 的成本效益高出 38%，而對於擁有 5 萬名使用者的組織，運用 RAG 的 70B 參數 LLM 的成本效益高出 75%。實際結果可能有所差異。

<sup>4</sup>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，與手動部署通用設計的時間。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。

# Dell AI Factory with NVIDIA

## 業界首款端對端企業 AI 解決方案

Dell AI Factory with NVIDIA 提供全方位的 AI 技術產品組合、經過驗證的一站式解決方案，以及專家服務，協助您更快實現 AI 成果。

透過廣泛的 Dell AI 基礎結構和服務產品組合，搭配 NVIDIA 領先業界的加速運算完整堆疊，其中包含 GPU、網路功能，以及 NVIDIA AI Enterprise 軟體、NVIDIA Inference Microservice (NIM)、模型和代理程式藍圖，藉此運用 AI 和 GenAI 大規模擴展您的企業。



# 您的資料就是差異化優勢

AI 功能從更完善的資料管理和準備工作開始。透過運用強大、準確且全面的資料集，加速業務決策的能力，將使業界佼佼者脫穎而出。

## 資料驅動 AI Factory

最重要的資料通常位於內部部署或邊緣，亦即您擁有獨佔擁有權之處。這可以協助將延遲問題降至最低，並確保所有敏感資訊安全無虞。



## 透過儲存、保護及管理資料的業界領先廠商，運用 AI 處理最重要的資料

您的資料管理旅程不是一次性任務，而是一個持續的過程，利用一系列實務和工具，幫助您的組織持續發揮資料的價值。有效的資料管理，對於任何成功的生成式 AI 計畫都至關重要。

Dell Technologies 與 NVIDIA 合作，協助像您這樣的組織輕鬆完成資料管理旅程。在 AI 導向的資料工作負載領域中，過程和目的地同樣重要。

Dell AI Factory with NVIDIA 可搭配 Dell Data Lakehouse 策略，協助您管理及視覺化資料，並準備好透過 Dell Services 及合作夥伴消化 GenAI。透過支援 AI 的 Dell PowerScale 進一步強化資料儲存方式，為 AI Factory 以及 NVIDIA DGX Pod、DGX Superpod 和 OVX 系統提供高效能且經認證的平台。

# 透過使用案例促進成果

透過經驗證的解決方案和量身打造的服務，協助您簡單部署最重要的AI 使用案例。



## 內容與程式碼產生

- 策劃品牌內容。
- 增強內容個人化。
- 更新市場報告。
- 彙總產業新聞。
- 監控競爭活動。
- 整合企業系統。
- 自動化測試流程。
- 促進程式碼重構。
- 強化安全通訊協定。
- 簡化資料庫查詢。



## 數位助理

- 提供面向客戶的支援。
- 利用訓練和教育。
- 使用助理最佳化排程。
- 追蹤專案截止日期。
- 自動處理例行作業。



## 資料建立

- 強化商業智慧。
- 簡化專案研究。
- 促進法律查詢。
- 最佳化採購搜尋。
- 改善財務分析。



## 電腦視覺

- 偵測失竊。
- 目視檢查設備和產品。
- 自動化機械作業。
- 分析工作者安全。
- 監控路況。



## 數位分身

- 客戶體驗。
- 產品開發和原型設計。
- 維護和營運。
- 物流和供應鏈。
- 訓練和指引。



# AI 最佳化基礎結構

基礎結構是 AI Factory 的基礎。透過從桌上型電腦到資料中心的廣泛 Dell AI 產品組合，在任何地方彈性執行工作負載，讓您將 AI 投資調整為合適的規模。

## 確認使用案例

您期望的業務成果有哪些基礎結構需求？

## 決定您需要哪種 AI 模型/類型

瞭解您的業務目標和使用案例後，您就可以決定效能和成本的適當平衡。

## 調整投資規模至適當程度

有些模型可以部署在筆記型電腦上，有些則需要大量 GPU，例如 Dell PowerEdge XE9680 搭載的 8 個 GPU。這真的取決於您想達成的目標。這就是 AI Factory 框架有用之處。

## 產品特點

Dell NativeEdge 是第一個可在邊緣自動化交付 NVIDIA AI Enterprise 的邊緣協調平台。<sup>5</sup>

Dell PowerEdge XE9680L 伺服器以精巧的 4U 外型規格，提供直接液體冷卻和 8 個新一代 NVIDIA Blackwell Tensor 核心 GPU。

Dell PowerScale 是全球第一個乙太網路型 NVIDIA DGX SuperPOD 認證儲存解決方案。

## 基礎結構



筆記型電腦



工作站



運算



網路功能和資料保護



儲存裝置

## 基礎結構

筆記型電腦和工作站資料

Dell 伺服器

Dell Storage

NVIDIA GPU

NVIDIA 網路

<sup>5</sup> 根據 2024 年 5 月的 Dell 分析結果而得。

# NVIDIA AI Enterprise - AI 作業系統

## 彈性且健全的企業 AI 端對端平台



雲端原生軟體平台，旨在加速資料科學流水線以及簡化生產級 AI 應用的開發與部署。

易於使用的微服務提供最佳化的模型效能和業界標準 API，以確保從原型到生產的順暢轉換。

安全的軟體生命週期、端對端可管理性和企業支援，讓組織能放心建構和執行 AI 解決方案。

# 從桌上型電腦到資料中心，加速您的 AI 採用

快速完成部署並上線運作，為您的團隊提供本地原型設計環境的實作經驗。透過清楚掌握資料管道和業務流程，開始建立組織技能，使您能夠隨著需求的增長順暢擴充。



## 簡單的開發控制板

在安全可靠的環境中進行本地原型設計。



## 開發並準備擴充

使用容器式微服務，快速建立 AI 和資料工作流程。



## 在企業中部署

透過可擴充的基礎結構框架，進行協調與最佳化。



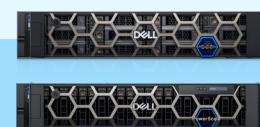
### AI 工工作站

Precision



### 運算

PowerEdge 與 NVIDIA GPU



### 儲存裝置

PowerScale/PowerStore



### 網路功能

NVIDIA Spectrum-X /  
NVIDIA Quantum InfiniBand

運用專為各式各樣工作負載所設計的廣泛產品組合，隨成長擴充。

# 可擴充的合併式基礎結構和 微服務架構



## AI 工工作站

Precision

## 運算

PowerEdge 與 NVIDIA GPU

## 儲存裝置

PowerScale/PowerStore

## 網路功能

NVIDIA Spectrum-X /  
NVIDIA Quantum InfiniBand

評估模型

NVIDIA API 目錄

將模型匯出為  
NVIDIA NIM

NVIDIA NGC 目錄

測試和自訂  
AI 工作流程

NVIDIA NIM  
代理程式藍圖

在生產環境中執行  
AI 工作流程

NVIDIA NIM 和  
NeMo 微服務

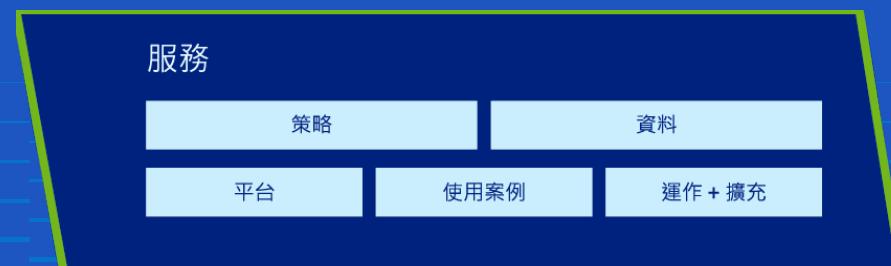
自訂模型

NVIDIA NeMo  
微服務

使用開放原始碼組建自己的 GenAI 工具時，無需猜測。相反地，您可以整合與連結這些不同層面並加以善用，透過適用於大規模部署 GenAI 應用程式的全方位解決方案，讓您更輕易地獲得成功。

# 有效採用生成式 AI 需要一段發展過程

Dell Technologies 在引導客戶完成 AI 歷程方面具備豐富經驗，因此您可以藉由大規模調整業務目標，與正確的技術解決方案達成一致，加速實現成果。



# 立即開始使用的兩種方式

透過 Dell AI Factory with NVIDIA，快速開始提供組織價值。

## 策略規劃

### 生成式 AI Accelerator Workshop

- 透過免費的半天研討會，制定制勝策略，展開發展之旅。
- 解決挑戰和差距，確定目標的優先順序並找出機會。
- 此外，我們還提供了更完整的準備程度評估，深入瞭解基礎結構需求、AI 模型、營運整合等。



## 技術準備

透過簡單易用的行動實驗室，包括 Dell Mobile Precision Workstation 和兩天的諮詢服務，協助您順利上手，快速展開您的成功之旅。

- 可攜式 GenAI 測試和示範。
- 在預先驗證的環境中快速製作原型。
- 以符合成本效益且低風險的方式，探索 GenAI 使用案例。
- 縮短 GenAI 技能差距。



搭載 NVIDIA GPU 的 Dell Mobile Precision Workstation 5690/7780

適用於 GenAI 實驗的沙箱環境  
已部署與配置的 RAG 架構

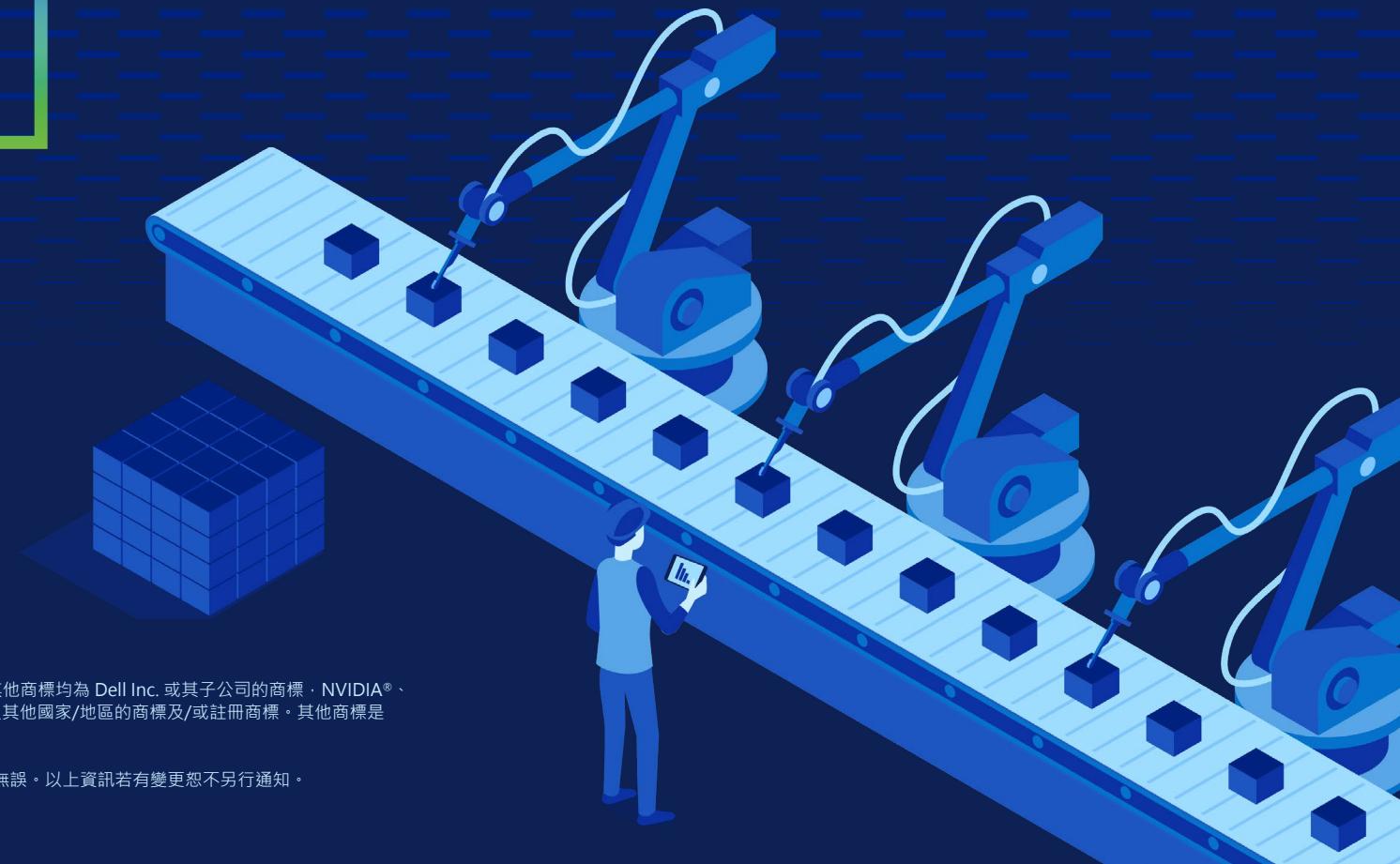
可立即供開發人員使用的 NVIDIA AI Workbench 平台  
已預先配置為可執行軟體，並已驗證效能

支援進一步探索的基礎使用案例  
以您的資料實作的初始聊天機器人使用案例

# 與 Dell Technologies 和 NVIDIA 合作， 充分發揮 GenAI 潛能

造訪我們的網站 [Dell.com/en-us/dt/nvidia-ai.htm](https://Dell.com/en-us/dt/nvidia-ai.htm)

瞭解更多資訊，立即開始您的 AI 之旅。



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。NVIDIA®、Omniverse™ 和 Spectrum™-X 是 NVIDIA Corporation 在美國及其他國家/地區的商標及/或註冊商標。其他商標是屬於其各自擁有者之財產。電子書 2024 年 9 月於美國發行

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。