



Dell Technologies



AMD
together we advance_

透過開放式生態系統推動強大的 AI 深入見解

Dell PowerEdge XE9680 伺服器，搭載 AMD Instinct™
MI300X 加速器和 ROCm™ 6 開放式軟體平台

隨著 ChatGPT 及其他生成式 AI (GenAI) 應用程式的普遍性日益提高，人工智慧 (AI) 正迅速普及。此外，還有更多組織希望參與此行動。最近的 IDC 研究顯示：



77% 的組織

正在探索 AI 使用案例，
或已大幅投資 AI 計畫¹。

如果貴組織尚未準備好但希望利用加速的深入見解和其他 AI 優勢，則需要儘快採取行動。現在正是時機，尤其是現在有高度採用 AI 進階功能的開放式生態系統，以及準備好向前推動 AI 創新及您業務的新解決方案。

¹IDC · 《未來企業靈活度與支出問卷調查》· 第 6 波 · 2023 年 7 月 ·



前方大道正等著貴組織

隨著 GenAI 和其他進階應用程式的興起，晶片技術將變得更加強大和複雜。但是，基礎層級的專屬系統，卻為第一線的開發人員帶來了障礙。

這種情況即將改變。Dell Technologies 和 AMD 攜手合作，以奠基於開放原始碼技術的基礎結構解決方案加速您的能力。在 Dell PowerEdge XE9680 伺服器上執行 AMD Instinct™ MI300X 加速器和 AMD Radeon Open Compute (ROCm™) 開放式軟體平台，可為您提供強大的生態系統選項 (一個開放的前方路徑)，可激發商機並釋放您的潛能。

透過 Dell Technologies 和 AMD 獲得效能和彈性

Dell Technologies 和 AMD 正透過能為開發人員帶來成功的系統，協助全球企業加速取得成果。



Dell PowerEdge XE9680 伺服器是專為 AI 工作而設計的資料處理強大設備。其支援八個加速器，非常適合用於機器學習 (ML)/深度學習 (DL) 訓練和推斷工作負載，特別是訓練大型語言模型 (LLM) 的工作負載。

AMD Instinct MI300X 加速器採用第 3 代 AMD CDNA™ (Compute DNA) 架構，提供 GPU 市場最大的高頻寬記憶體容量，為訓練最大型 AI 模型提供每伺服器 1.5 TB 的一致 GPU 加速器記憶體。其亦可讓您使用較少的伺服器來增加推斷容量。

使用 ROCm 6 開啟加速

ROCm 繼而釋放 AMD Instinct 的運算能力。ROCm 是一種開放、彈性且可擴充的軟體平台，可加速運算，並配備驅動程式、API 及其他 GPU 工具，可支援 AMD Instinct 加速器的程式碼最佳化，並簡化現有 GPU 加速程式碼的可攜性。

其也可徹底簡化您的體驗。ROCm 在裝置層級中經過最佳化，並具有自由共用的程式碼基礎，可隨著系統及其零件的演進而進化。此開放的彈性也可讓您輕鬆進行移植且幾乎無須編碼。

無邊際的生態系統

AMD Instinct MI300X 加速器和 ROCm 以開放式標準和互用性為基礎，具備各種 AI 架構和程式庫對其的原生支援。這可為想要駕馭 AMD 強大功能，並仍保持與現有軟體系統相容性的使用者，提供更大的選擇和彈性。

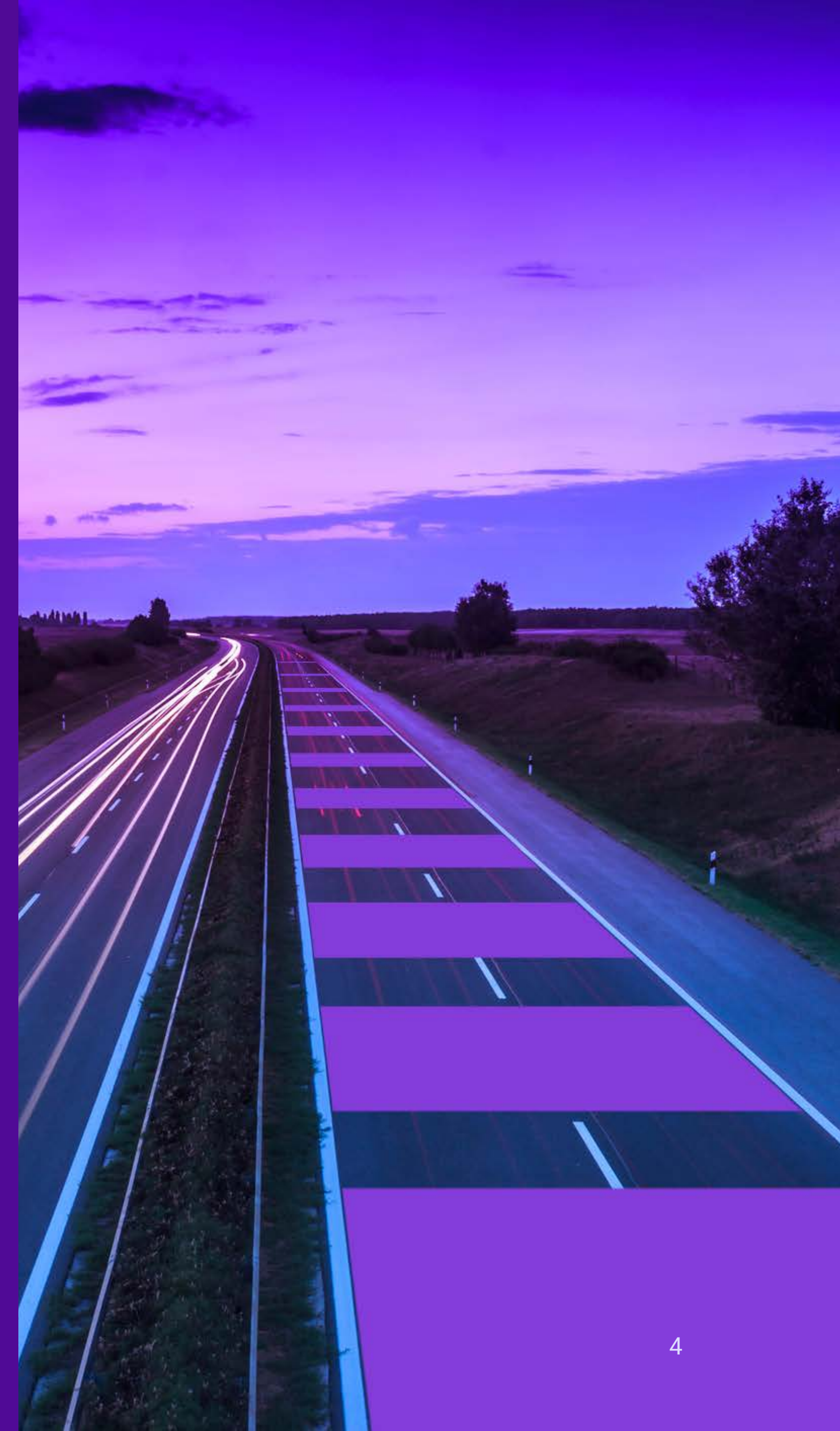


AMD 掃除障礙，讓您自由快速推動專案。

ROCm 優勢

- 開啟 GPU 強大功能，加速運算工作
- 針對大規模 ML/DL/GenAI 工作負載最佳化
- 為領先的模型和架構 (包括 PyTorch®、TensorFlow™ 及其他項目) 提供現成支援
- 能快速採用新的創新 (包括 JAX、Open Neural Network Xchange (ONNX)、OpenAI Triton 等的原生支援)
- 透過與 OSS 開發社群連線，推動創新、差異化及協同合作
- 為開放式可攜性提供彈性基礎

深入瞭解 ROCm 的新功能。



使用 ROCm 輕鬆移植並最佳化 AMD Instinct 適用的應用程式

- 實現輕鬆將軟體移植至 AMD。HIP 可讓開發人員從單一來源程式碼為 AMD Instinct MI300X 建立可攜式應用程式。
- 提供常用程式庫和運算子的所有功能
- 快速移植並最佳化大部分應用程式以實現效能

不斷增加的上游支援

ROCm 為現今最熱門的架構提供上游支援，包括但不限於：

- PyTorch
- TensorFlow
- Open Neural Network Xchange (ONNX)
- JAX



開放式可攜性，節省寶貴時間

對於可能會擔心新生態系統及其對現有程式碼影響的開發人員，ROCm 可讓其完全安心。ROCm 平台的主要優點是，您能以對現有程式碼進行最低限度修改的方式移植應用程式。移植非常簡單，且可節省大量的時間。

以四種不同的方式，將現有應用程式移植至 AMD Instinct MI300X 加速器：

1 利用現有的 AI 架構，為 ROCm 提供上游支援。

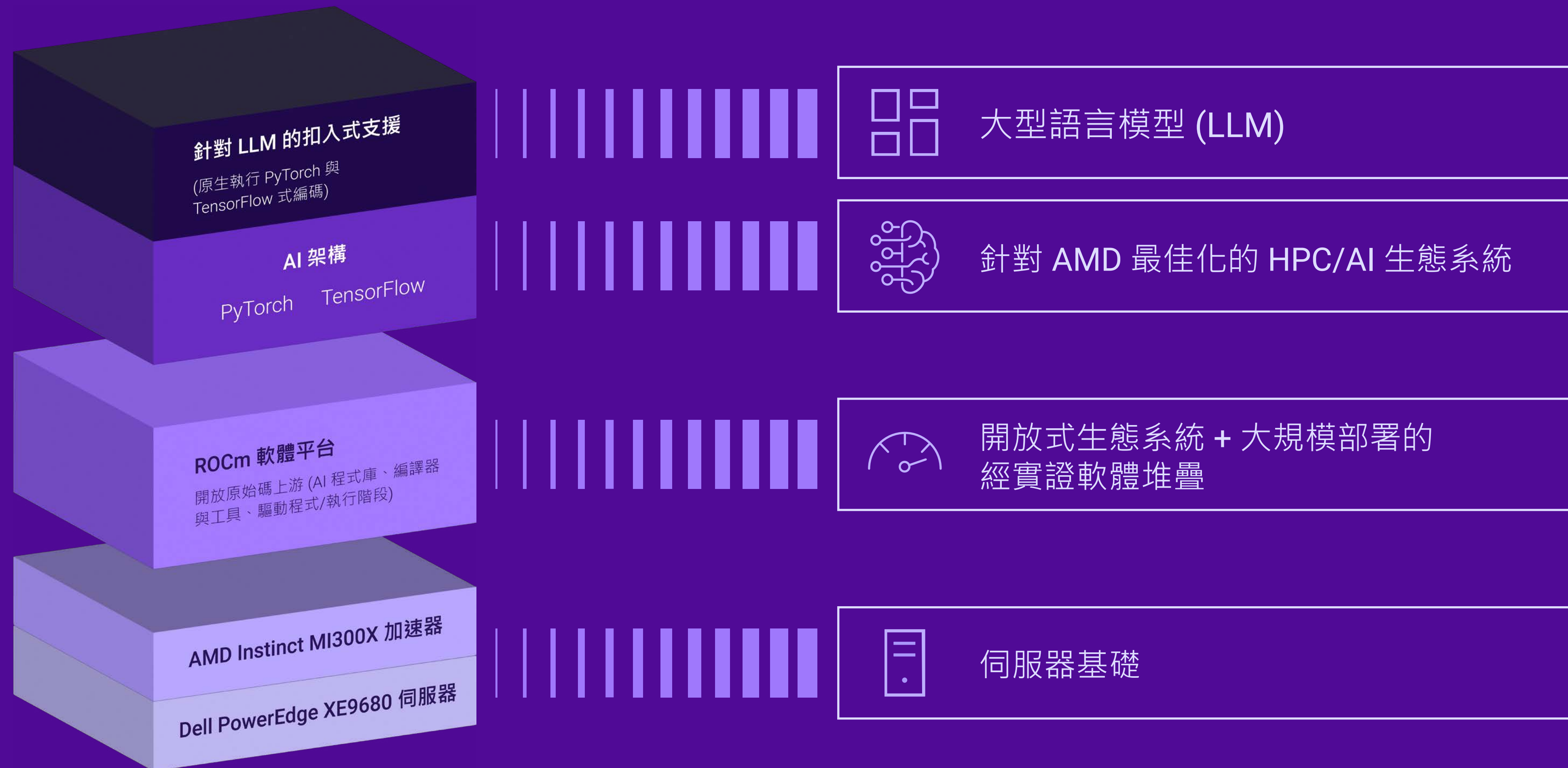
2 使用針對 AMD Instinct MI300X 最佳化的扣入式 ML 核心。或善用 AMD HIP 轉換工具，讓您在 ROCm 型 GPU 上最佳化 CUDA 應用程式。

3 使用針對 ROCm 最佳化和維護的現有 CUDA 程式庫軟體儲存庫鏡像版本。

4 存取 AMD 預先封裝的 AI 應用程式和容器。

立即開始使用

透過 Dell Technologies 和 AMD 的開放架構，加速現有 AI 和新 GenAI 專案。在經實證的 Dell PowerEdge 基礎上驅動您的工作負載，該基礎支援最新的 AMD 加速器，包括 AMD Instinct MI300X 加速器。造訪 [ROCm 開發人員中心](#)。

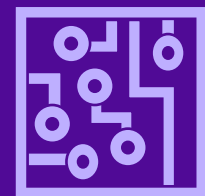


AMD Instinct MI300X 加速器已準備好部署

AMD Instinct MI300X 加速器提供突破性效能，可提供現今所需的速度、擴充能力和彈性，以及開放式生態系統，尤其是當企業在因應資源、複雜性、延遲和架構挑戰時。AMD Instinct MI300X 加速器在以下方面可提供獨特的優勢：



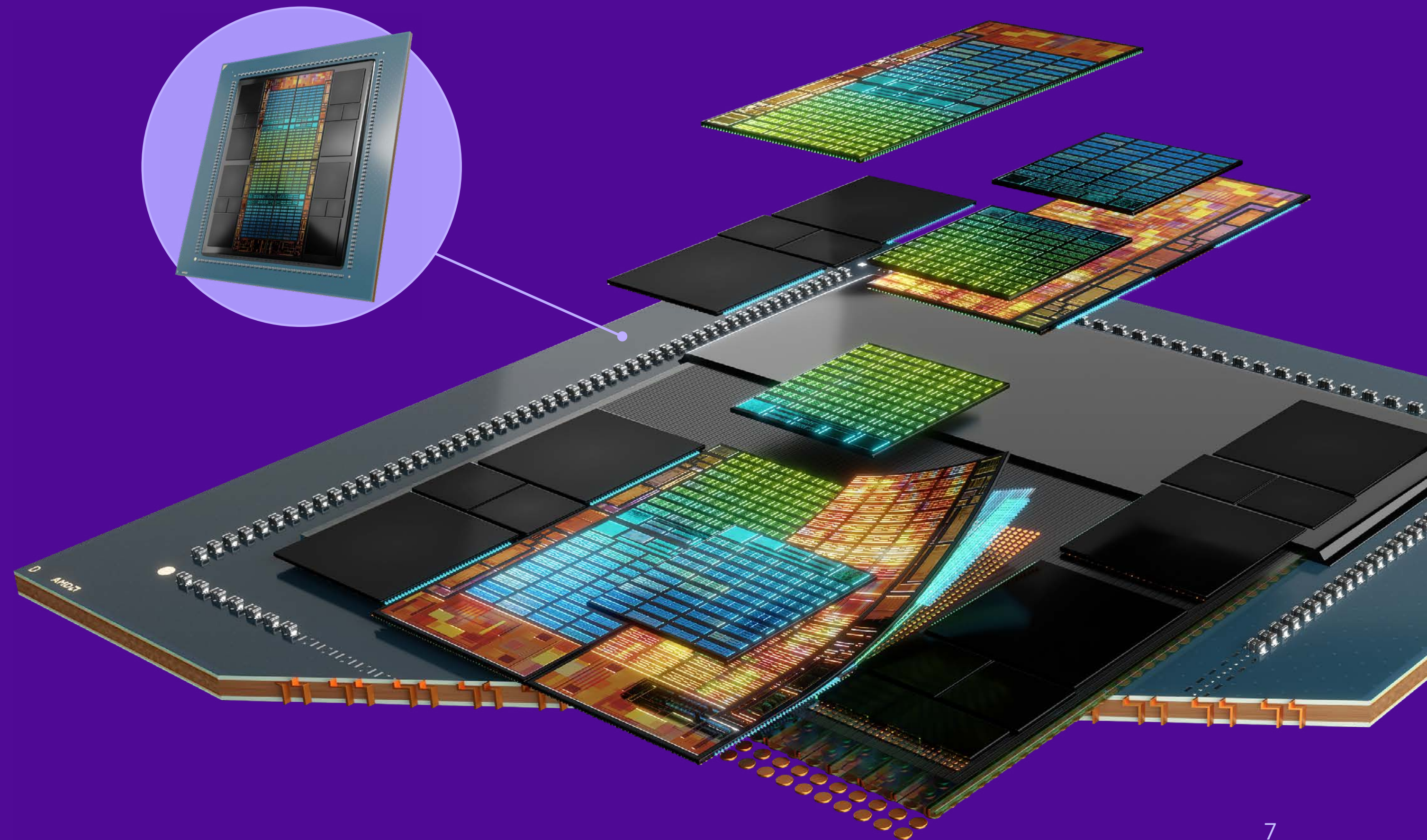
效能：以前所未有的每 GPU (以每秒 5.3 TB 運作) 192 GB 的堆疊式 HBM3 記憶體，實現全新的 AI 工作流程。



密度：運用最先進的晶粒堆疊、小晶片技術，為每部伺服器提供更多運算和記憶體容量，進而減少資料移動需求。



開放式：透過專為存取和適應性所設計的生態系統，輕鬆實現結果。



運用 Dell PowerEdge XE9680 伺服器釋放您的 AI 優勢

您可以使用專為 AI 設計的伺服器平台，加速邁向更快速、更智慧化的成果。藉由可極速分析精闢洞見、提供受信任的 AI 並簡化作業流程的解決方案，強化資料導向業務、提供視覺成果，並在整個企業加速 AI 作業。



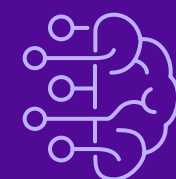
受信任的 AI

透過值得信賴的高品質解決方案基礎結構，降低風險並節省時間，同時加速您的 AI 生命週期。PowerEdge XE9680 從晶片信任根一路到資產淘汰，均提供內建保護功能。



簡化作業流程

透過 AI 基礎結構自動化，隨時隨地有效控制和管理您的 AI 基礎結構和效能工作負載。Dell Technologies 提供共同運作及獨立運作的智慧型系統，可加速工作並簡化您的體驗。



極速分析精闢洞見

驅動改善整個 AI 生命週期的效能，並以符合業務需求的速度交付 AI 作業。Dell Technologies 提供創新的運算能力，協助您以最大效能駕馭 AI。

推動創新

Dell Technologies 和 AMD 正努力為我們寶貴的客戶推動創新。我們攜手合作，以可行的解決方案為市場帶來矽多元性，讓開發人員自如運用。我們的聯合解決方案可讓您：



- 以更多選項和更少的障礙**更快進入市場**。

- 利用不會阻礙您的開放式生態系統**更快獲得深入見解**。

- **選擇最適合您的選擇**。仰賴 Dell 供應鏈優勢，在最需要時持續將創新技術推向市場。

透過 Dell Technologies 和 AMD 全速領先

AI 現已推出，但其影響力只會持續擴展並加速向前邁進。對於希望善用其功能的企業來說，其可能性似乎無窮無盡。

這就是為何您不會希望受到有時可能會妨礙您的技術障礙限制。改用專屬於 AI 工作負載的強大解決方案堆疊，讓您的選項保持多元。

Dell PowerEdge XE9680 伺服器上搭載 AMD Instinct MI300X 加速器和 ROCm 軟體就是解決辦法。加快新應用程式的開發速度、更快上市，並開始推動對您業務和整個世界都至關重要的成果。

如需深入瞭解，請造訪 Dell.com/GPU

Copyright © 2023 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、Instinct™、ROCm™ 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™ 及任何相關標誌均為 Google Inc. 的商標。CUDA® 是 NVIDIA Corporation 在美國和其他國家/地區的商標和/或註冊商標。其他商標是屬於其各自擁有者之財產。電子書 12 月 23 日於美國發行。

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。