



採用 NVIDIA 和 VMware Cloud Foundation™ 的人 工智慧就緒基礎架構

VMware Cloud Foundation 與 NVIDIA AI
Enterprise 軟體套件整合概觀

目錄

兼具彈性與延展性的混合雲平台，適用於人工智慧 / 機器學習工作負載	3
人工智慧 (AI) 應用程式正推動所有企業轉型	3
提供人工智慧就緒的基礎架構	4
解決方案架構	4
VMware Cloud Foundation 功能	4
著重應用程式的管理	4
人工智慧工作負載的效能、安全性和合規	5
提高生產力並減少阻礙	5
透過完整堆疊的靈活性，釋放 NVIDIA 技術的原始效能	6
提升人工智慧工作負載的原始效能，同時實現最高的利用率	8
有效率地使用並輕鬆管理人工智慧 / 機器學習工作負載的 GPU	8
基礎架構生命週期管理	8
每層堆疊都具備原生安全機制	9
延伸涵蓋私有雲和混合雲的雲端作業模式	9
VMware Cloud Foundation - 人工智慧就緒基礎架構	9

兼具彈性與延展性的混合雲平台，適用於人工智慧 / 機器學習工作負載

人工智慧 (AI) 應用程式正推動所有企業轉型

人工智慧 (AI) 與機器學習 (ML) 計畫正引領所有產業邁上數位化轉型之路，進而實現採用人工智慧支援的新穎工作負載與應用程式。在人工智慧生命週期中，這些人工智慧 / 機器學習工作負載具有特殊需求，其中包含用於訓練、推斷與執行資料分析的獨特效能特性。例如，深度學習訓練需要具備高效能，以使用私有雲基礎架構搭配 GPU 或其他特殊用途處理器和硬體元件，大規模執行平行處理作業。推斷作業則具有近乎即時的處理需求，而這需要能夠處理人工智慧訓練模型所傳送的資料，才可獲得所需成果。這類推斷工作負載可能必須位於企業內部或鄰近企業的位置，且前述工作負載的分散式本質會帶來特定的安全性需求，以確保合規。因此，企業和其 IT 部門皆希望能以安全的方式，對這類人工智慧工作負載提供卓越的效能並加快其速度。

這類人工智慧工作負載一般會與容器化和具狀態工作負載一起部署，以善用支援人工智慧 / 機器學習生命週期所需的現代化基礎架構，進而導致情況更加複雜。IT 部門發現其員工缺乏必要的技能組合，無法設定與管理容器、虛擬機和人工智慧工作負載等不同的資訊孤島。這種更為複雜的情況可能會在企業內導致成本大幅增加，並延遲這些策略性計畫開花結果的時間。為了因應這些挑戰，需要採用具有足夠彈性的混合式基礎架構以支援企業需求，同時仍可確保其兼具效能與延展性。企業則必須解決因這種人工智慧基礎架構複雜性而擴大的技能組合落差。因此，IT 團隊正著手尋找功能，以利用現有管理工具，透過自動化方式在軟體定義環境中，運用超融合式基礎架構 (HCI) 搭配加速 GPU，藉此將複雜性降至最低並簡化部署。

提供人工智慧就緒的基礎架構

VMware Cloud Foundation™ with Tanzu® 提供完整堆疊混合雲平台，可提供人工智慧就緒基礎架構，讓客戶能加快人工智慧 / 機器學習工作負載與傳統企業應用程式的速度。VMware Cloud Foundation™ with Tanzu® 以經實證的全方位軟體定義堆疊為基礎，其中包含 VMware vSphere®、VMware vSAN™、VMware NSX-T™ Data Center 和 VMware Aria Suite 等，因此可提供一組完整且安全的軟體定義服務，以因應運算、儲存、網路安全性、Kubernetes 管理與雲端管理。這樣一來，就能造就靈活、可靠且效率十足的雲端基礎架構，一舉跨越私有雲和公有雲，提供一致的作業。此外，VMware Cloud Foundation 包含內建自動化生命週期管理功能，以簡化軟體堆疊管理作業，從初始部署到修補及升級等，皆包含在內。VMware 和 NVIDIA 公司聯手提供專為人工智慧工作負載最佳化的端對端平台，可協助發揮人工智慧的強大威力。這款整合式平台提供同級最佳的人工智慧軟體，也就是 NVIDIA AI Enterprise 軟體套件，並且已針對搭配 VMware Cloud Foundation with Tanzu 運作最佳化

解決方案架構

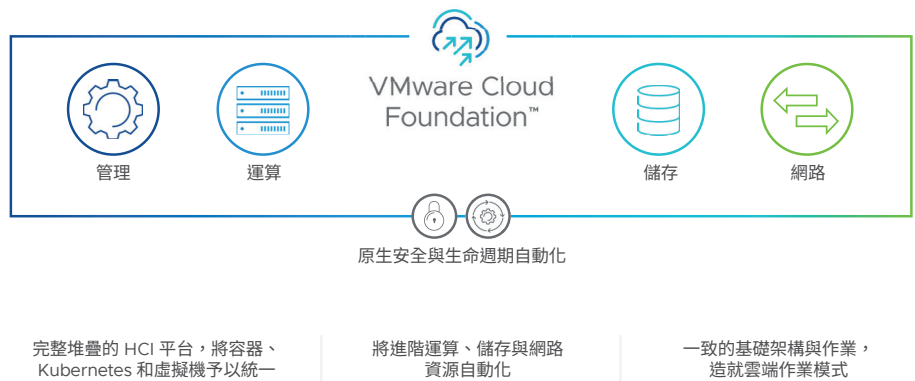


圖 1：VMware Cloud Foundation 解決方案架構



VMware Cloud Foundation 功能

著重應用程式的管理

大部分人工智慧 / 機器學習工作負載本質上皆為雲原生、部署在容器中，且已針對搭配 GPU 運作最佳化。前述人工智慧 / 機器學習工作負載必須與其他工作負載共存在共用基礎架構上，而這會造成虛擬機 (VM) 和容器混合存在。

有了 VMware Cloud Foundation，客戶即可對 vCenter 中的虛擬機、容器和 Kubernetes 叢集，取得統一的能見度。從 vCenter 的觀點來看，會將容器和 Kubernetes 叢集視為一級物件，就像虛擬機一樣。

Kubernetes 的命名空間概念已整合至 vSphere 中，並成為管理單元。透過命名空間，將虛擬機和容器等資源物件分組至邏輯應用程式後，過去需管理數千個虛擬機的虛擬基礎架構 (VI) 管理員，現在只需管理數十個應用程式命名空間，如此可大幅增加規模，並減少認知負載。這種靈活性和彈性可在人工智慧 Pipeline 中，非常有效地提升效率和利用率。



人工智慧工作負載的效能、安全性和合規

人工智慧工作負載透過使用基礎架構來運用大量資料，而前述基礎架構必須能滿足嚴格的安全性、高可用性與彈性需求。若要在訓練、推斷和分析期間，以安全且低延遲的方式存取這類資料，必須使用 GPU 加速效能，以更快速地實現業務成果，並符合所需的服務層級協定。人工智慧工作負載必須可動態使用 GPU 資源，以滿足較大型資料集與日遽增的效能需求，同時仍維持符合隱私權與安全性規範。

透過 VMware Cloud Foundation，即可集中管理安全性、效能與可用性原則，以確保在企業環境中實現一致性。管理員可定義服務品質、安全性原則、防火牆規則、加密設定、命名空間層級的可用性與存取控制規則，進而減少管理大規模人工智慧應用程式及對其進行疑難排解的所需時間。

NSX-T 經過精心設計，可做為 VMware Cloud Foundation with Tanzu 的 Pod 網路。NSX-T 可為 VMware Cloud Foundation with Tanzu 提供完整堆疊的網路與安全性，包含分散式交換器與路由、分散式防火牆、負載平衡、流入控制等。整合 Kubernetes 後，則可透過命名空間隔離來實現情境感知安全性原則。



提高生產力並減少阻礙

資料科學家與開發人員需要能在測試環境中開發和測試其人工智慧模型，並快速進行迭代以取得所需成果。因此，他們需要先輕鬆建立並測試其人工智慧 / 機器學習工作負載，隨後再將其部署至生產環境中。然而，許多 IT 部門皆採用緩慢的支援工單系統來為開發人員提供基礎架構服務，因為這是唯一可治理開發人員應用程式和流程的方式。VMware Cloud Foundation 提供在命名空間層級進行管理的能力，讓管理員可一次完成命名空間設定原則、配額和角色型存取的作業。隨後開發人員即可在預先定義的邊界內，以自助方式進入命名空間。

Kubernetes 已內嵌至 vSphere 的控制平台中，因此開發人員可建立並使用雲端資源，例如 Kubernetes 叢集、磁區 (包含具狀態應用程式的持續性磁區)，以及透過宣告式模式使用 Kubernetes 和 RESTful API 的網路等。這可減少佈建與延展基礎架構的所需時間和精力，讓開發人員可聚焦於建置應用程式。同時，IT 作業人員能透過熟悉的 VMware 介面，針對這類由開發人員建立的雲端資源，保有能見度。

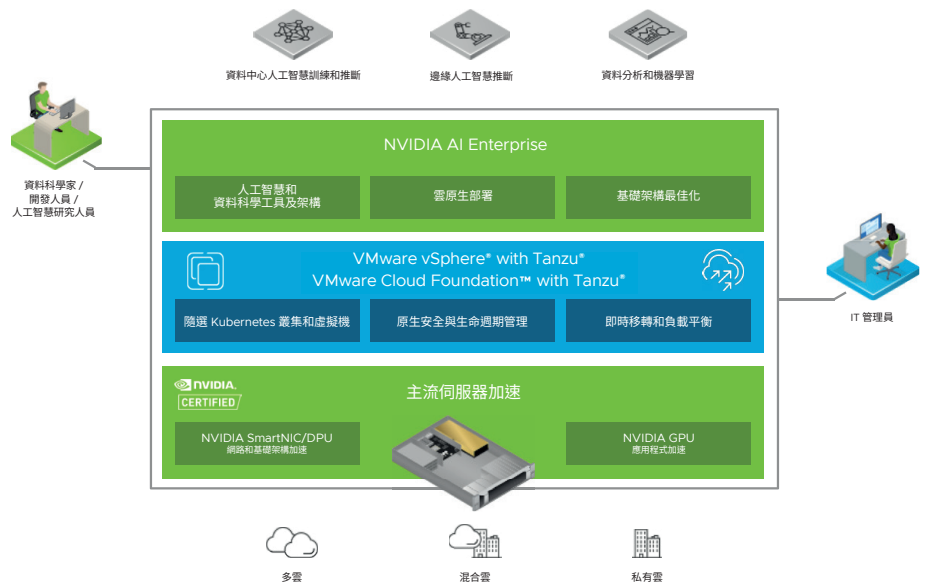


圖 2：VMware Cloud Foundation 與 NVIDIA 人工智慧就緒企業產品套件



透過完整堆疊的靈活性，釋放 NVIDIA 技術的原始效能

為了實現最佳效率與規模，應使用以容器與虛擬機組合為基礎的基礎架構，一併部署人工智慧工作負載與傳統工作負載。然而，許多 IT 部門均缺乏必要的技能組合，在部署進階人工智慧和 Kubernetes 工作負載時，無法避免建立資訊孤島。如果這些複雜的應用程式環境要以尖峰效能執行，需要一些工具與服務來維護這些環境，而這會進一步提高複雜性。

整合 NVIDIA AI Enterprise 軟體套件和 VMware Cloud Foundation 後，如 TensorFlow 與 Pytorch 等人工智慧和資料科學工具及架構，即可搭配如 NVIDIA GPU Operator 和 NVIDIA Network Operator 等雲原生部署工具使用，讓 IT 管理員能輕鬆部署這些工作負載。

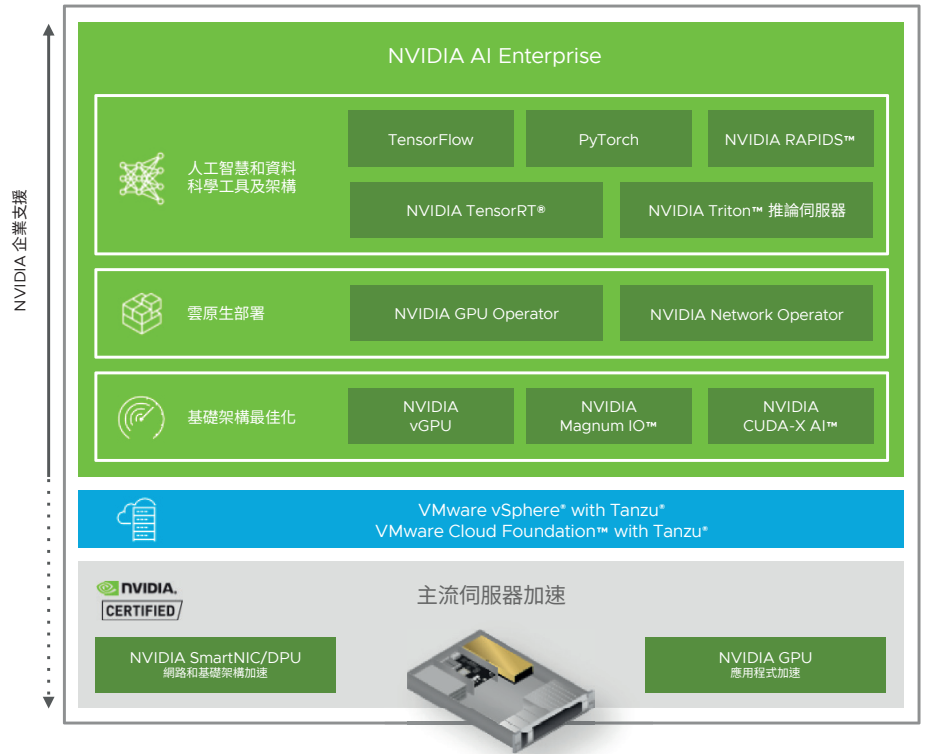


圖 3：NVIDIA AI Enterprise 軟體套件

有鑑於人工智慧 / 機器學習工作負載的分散式本質，因此需要可輕鬆並迅速於地端、邊緣或雲端中部署工作負載。VMware Cloud Foundation 可透過分散式工作負載網域，因應這項需求。工作負載網域是原則式資源架構，具備特定的可用性與效能屬性。其將運算 (vSphere)、儲存 (vSAN)、網路 (NSX-T) 和雲端管理 (VMware Aria) 結合為可供使用的單一實體。工作負載網域會以自動化方式部署底層基礎架構和 Kubernetes 元件，可大幅加快 Kubernetes 的具體化作業。工作負載網域也讓 IT 作業人員和開發人員能安全地進行沙箱作業，並為容器與虛擬機配置合適的基礎架構。



提升人工智慧工作負載的原始效能，同時實現最高的利用率

VMware Cloud Foundation with Tanzu 現在透過整合 NVIDIA AI Enterprise 軟體套件，支援 NVIDIA Ampere A30 和 A100 GPU。如此一來，即可在多個執行實例間共用 GPU，進而提升資源的利用率和最佳化成果，同時享有減少成本的優勢。

人工智慧訓練與推斷工作負載需要高效能的基礎架構，而 GPU 和網路加速器也可為其帶來助益。NVIDIA AI Enterprise 整合 VMware Cloud Foundation，以在簡化的完整堆疊 HCI 環境中，佈建和管理多個 GPU。其包含可讓加速虛擬機和容器達到接近裸機效能的最佳化功能。NVIDIA GPU 可針對時間切割或多執行實例 GPU (MIG) 進行設定，以提供可預測的效能，同時提高這些關鍵資源的利用率。



有效率地使用並輕鬆管理人工智慧 / 機器學習工作負載的 GPU

透過 VMware Cloud Foundation 和 NVIDIA AI Enterprise 軟體套件，在建置使用這類 vGPU 的人工智慧 / 機器學習資料 Pipeline 時，IT 管理員就可快速且輕鬆地向資料科學家與開發營運團隊佈建自助功能。開發人員可透過選擇預先設定的 vGPU 設定檔，以在 IT 管理員團隊中的資料科學家獨立提出需求時，使用能獨立擴充或縮減的 GPU 資源。



基礎架構生命週期管理

VMware Cloud Foundation 提供自動化生命週期管理，以簡化每個工作負載網域上的基礎架構堆疊管理作業。所有元件的可用更新都會針對互通性進行測試，並且會搭配所需邏輯，以透過適當的順序安裝。隨後會為每個工作負載網域排定更新套件的自動安裝排程。這讓管理員可透過獨立於環境其他部分的方式，針對特定工作負載或環境 (例如開發與生產) 進行更新。

資源

- 深入瞭解 [VMware Cloud Foundation](#)
- 查看 Cloud Foundation [部落格](#)、[Twitter](#) 和 [YouTube](#)，瞭解 Cloud Foundation 的最新資訊
- 想要親自試用 Cloud Foundation 嗎？請造訪 [VMware Cloud Foundation Hands-On Lab](#)



每層堆疊都具備原生安全機制

在容器映像層，Tanzu Kubernetes Grid 包含同級最佳容器倉管中心，並且內建弱點掃描、映像簽名和稽核功能。

在運算層，vSphere 提供全方位的內建安全性機制，可保護資料、基礎架構，以及操作簡易的存取功能。原則導向安全性可提供虛擬機層級或 Pod 層級加密，以防範未經授權存取靜態與動態資料。

在網路層，NSX-T 可為個別虛擬機或 Pod 工作負載提供微分段與精密的安全性，從根本實現更安全的資料中心。無論工作負載位於網路拓撲的哪個位置，安全性原則都會如影隨形地伴隨工作負載。

在儲存層，vSAN 提供可在叢集層級加密靜態資料和傳輸中資料的功能。建置儲存加密功能旨在滿足合規需求，並透過支援所有符合金鑰管理互通性通訊協定 (KMIP) 標準的金鑰管理工具，提供簡易的金鑰管理作業。

在管理層級，VMware Aria 解決方案可將手動作業自動化，以減少人為錯誤、提供監控和稽核完整堆疊的功能，以及提供自行驅動作業，以在識別出問題時快速加以修復。



延伸涵蓋私有雲和混合雲的雲端作業模式

在 VMware Cloud Foundation 私有雲部署中所運用的相同核心軟體定義的基礎架構堆疊，也是 VMware Cloud on AWS 與其他 VMware Cloud Provider™ 計畫合作夥伴等 VMware 架構公有雲採用的基礎技術。擁有 VMware Cloud Foundation 技術支援的雲端可提供一致的基礎架構和作業，讓客戶可轉為採用不同的 IT 作業方式，如此一來，就能更妥善地配合企業的服務使用需求進行服務交付作業。採用雲端作業模式，代表轉為推動應用程式現代化改造，並朝可實現數位計畫的全新應用程式架構邁進。

VMware Cloud Foundation - 人工智慧就緒基礎架構

- **可輕鬆部署**與執行整合式人工智慧就緒基礎架構，包含適用於現代化應用程式的運算、儲存、網路、安全性和雲端管理服務，且這些現代化應用程式均位於與傳統應用程式相同的平台上。
- **提升開發人員生產力**，讓應用程式和資料科學團隊可透過業界標準 API，存取他們早已熟悉的雲端資源。
- **易於執行混合雲策略並讓其迎合未來需求**，前述混合雲策略在地端和外部部署中皆一致且相容，並且具備部署虛擬機、容器和任何新一代應用程式需求的能力。

