



# NetApp AIPod 搭配 NVIDIA NetApp Solutions

NetApp  
August 14, 2024

# 目錄

採用 NVIDIA 的 NetApp AI .....	1
NetApp AIPod 搭配 NVIDIA DGX 系統 .....	1
NetApp ONTAP 採用NVIDIA DGX A100系統的AI .....	1
NetApp ONTAP 採用NVIDIA DGX A100系統和Mellanox Spectrum乙太網路交換器的AI .....	1
NetApp AIPod 搭配 NVIDIA DGX 系統 - 簡介 .....	1
NVA-1151-Design：採用ONTAP NVIDIA DGX A100系統設計指南的NetApp W人工 智慧 .....	10
NVA-1151-Deploy：NetApp ONTAP 支援NVIDIA DGX A100系統的AI .....	10
NVA-1153設計：採用ONTAP NVIDIA DGX A100系統和Mellanox Spectrum乙太網路交換器的NetApp W人工 智慧.....	10
NVA-1153-Deploy：採用ONTAP NVIDIA DGX A100系統和Mellanox Spectrum乙太網路交換器的NetApp W人工 智慧.....	11

# 採用 NVIDIA 的 NetApp AI

NetApp與NVIDIA提供的整合式基礎架構解決方案總覽ONTAP。

## NetApp AIPod 搭配 NVIDIA DGX 系統

- ["NetApp AI Pod 搭配 NVIDIA DGX 系統"](#)

## NetApp ONTAP 採用NVIDIA DGX A100系統的AI

- ["設計指南"](#)
- ["部署指南"](#)

## NetApp ONTAP 採用NVIDIA DGX A100系統和Mellanox Spectrum乙太網路交換器的AI

- ["設計指南"](#)
- ["部署指南"](#)

## NetApp AIPod 搭配 NVIDIA DGX 系統 - 簡介

本節將介紹 NetApp AIPod 搭配 NVIDIA DGX 系統。

NetApp 解決方案工程

NetApp ；採用 NVIDIA DGX 的 AIPod ；系統和 NetApp 雲端連線儲存系統、可消除設計複雜度和猜測、簡化機器學習（ML）和人工智慧（AI）工作負載的基礎架構部署。AIPod 採用 NVIDIA DGX BasePOD 設計、為新一代工作負載提供卓越的運算效能、搭配 NVIDIA DGX 系統的 AIPod 新增了 NetApp AFF 儲存系統、可讓客戶從小開始、不中斷地成長、同時智慧地管理從邊緣到核心、再到雲端再回來的資料。NetApp AIPod 是 NetApp AI 解決方案較大型產品組合的一部分、如下圖所示：

NetApp AI 解決方案產品組合 \_ 影像：[aipd\\_nv\\_Portfolio .png](#) []

本文件說明 AIPod 參考架構的主要元件、系統連線資訊和解決方案規模指南。本文件適用於有興趣為 ML/DL 和分析工作負載部署高性能基礎架構的 NetApp 和合作夥伴解決方案工程師及客戶策略決策者。

## NetApp AIPod 搭配 NVIDIA DGX 系統 - 簡介

本節將介紹 NetApp AIPod 搭配 NVIDIA DGX 系統。

NetApp 解決方案工程

NetApp ；採用 NVIDIA DGX 的 AIPod ；系統和 NetApp 雲端連線儲存系統、可消除設計複雜度和猜測、簡化機器學習（ML）和人工智慧（AI）工作負載的基礎架構部署。AIPod 採用 NVIDIA DGX BasePOD 設計、為新一代工作負載提供卓越的運算效能、搭配 NVIDIA DGX 系統的 AIPod 新增了 NetApp AFF 儲存系統、可讓客戶從小開始、不中斷地成長、同時智慧地管理從邊緣到核心、再到雲端再回來的資

料。NetApp AI Pod 是 NetApp AI 解決方案較大型產品組合的一部分、如下圖所示：

NetApp AI 解決方案產品組合 \_ 影像： aipd\_nv\_Portfolio .png []

本文件說明 AI Pod 參考架構的主要元件、系統連線資訊和解決方案規模指南。本文件適用於有興趣為 ML/DL 和分析工作負載部署高性能基礎架構的 NetApp 和合作夥伴解決方案工程師及客戶策略決策者。

## NetApp AI Pod 搭配 NVIDIA DGX 系統 - 硬體元件

本節重點介紹 NetApp AI Pod 搭配 NVIDIA DGX 系統的硬體元件。

### NetApp AFF 儲存系統

NetApp AFF 最先進的儲存系統、讓 IT 部門能夠以領先業界的效能、優異的靈活度、雲端整合、以及同級最佳的資料管理、滿足企業儲存需求。專為Flash而設計AFF的支援功能、可協助加速、管理及保護業務關鍵資料。

#### AFF A900 儲存系統

搭載 NetApp ONTAP 資料管理軟體的 NetApp AFF A900 提供內建資料保護、選購的反勒索軟體功能、以及支援最關鍵業務工作負載所需的高效能與恢復能力。它可避免對關鍵任務作業造成中斷、將效能調校降至最低、並保護資料免受勒索軟體攻擊。它提供：

- 領先業界的效能
- 毫不妥協的資料安全性
- 簡化不中斷營運的升級

NetApp AFF A900 儲存系統 \_ 映像： aipod\_nv\_a900.png []

#### 領先業界的效能

AFF A900 可輕鬆管理深度學習、AI 和高速分析等新一代工作負載、以及 Oracle、SAP HANA、Microsoft SQL Server 和虛擬化應用程式等傳統企業資料庫。它可讓業務關鍵應用程式以最高速度執行、每個 HA 配對最多可達 240 萬 IOPS、延遲低至 100  $\mu$ s、效能比先前的 NetApp 機型高達 50%。有了 NFS over RDMA、pNFS 和工作階段 Trunking、客戶就能使用現有的資料中心網路基礎架構、達到新一代應用程式所需的高網路效能。客戶也可以透過統一化的多重傳輸協定支援來擴充 SAN、NAS 和物件儲存設備、並透過統一化的單一 ONTAP 資料管理軟體、在內部部署或雲端中提供最大的靈活度。此外、Active IQ 數位顧問（也稱為數位顧問）和 Cloud Insights 也提供 AI 型預測分析功能、可最佳化系統健全狀況。

#### 毫不妥協的資料安全性

AFF A900 系統包含一套完整的 NetApp 整合式與應用程式一致的資料保護軟體。它提供內建的資料保護功能、以及先進的反勒索軟體解決方案、可用於搶佔和攻擊後恢復。惡意檔案可能遭到封鎖、無法寫入磁碟、而且儲存異常狀況也可輕鬆監控、以獲得深入見解。

#### 簡化不中斷升級

AFF A900 可作為不中斷營運的機箱內升級、升級至現有 A700 客戶。NetApp 透過我們先進的可靠性、可用度、可維修性和管理性（RASM）功能、讓您輕鬆更新並消除對關鍵任務作業的中斷。此外、NetApp 還能進一步提升營運效率、並簡化 IT 團隊的日常活動、因為 ONTAP 軟體會自動為所有系統元件套用韌體更新。

對於規模最大的部署、AFF A900 系統提供最高的效能和容量選項、而其他 NetApp 儲存系統（例如 AFF A800、AFF C800、AFF A400、AFF C400 和 AFF A250）則提供更低成本的小型部署選項。

## NVIDIA DGX基礎POD

NVIDIA DGX BasePOD 是整合式解決方案、包含 NVIDIA 硬體和軟體元件、MLOps 解決方案和協力廠商儲存設備。客戶可運用 NVIDIA 產品和驗證的合作夥伴解決方案的橫向擴充系統設計最佳實務做法、為 AI 開發建置高效率且可管理的平台。圖 1 重點介紹 NVIDIA DGX BasePOD 的各種元件。

\_NVIDIA DGX BasePOD 解決方案\_ 影像： aipd\_nv\_basepo\_stallies.png []

## NVIDIA DGX H100 系統

NVIDIA DGX H100 系統；系統是 AI 的強大功能、可透過 NVIDIA H100 Tensor Core GPU 的突破性性能加速。

\_NVIDIA DGX H100 系統\_ 影像： aipod\_nv\_H100\_3D.png []

DGX H100 系統的主要規格如下：

- 八個 NVIDIA H100 GPU 。
- 每個 GPU 80 GB GPU 記憶體、總計 640GB 。
- 四個 NVIDIA NVSwitch™ 晶片。
- 支援 PCIe 5.0 的雙 56 核心 Intel® Xeon® Platinum 8480 處理器。
- 2 TB 的 DDR5 系統記憶體。
- 四個 OFP 連接埠、可連接八個單埠 NVIDIA ConnectX-7 ( InfiniBand / 乙太網路) 介面卡、以及兩個雙埠 NVIDIA ConnectX-7 ( InfiniBand / 乙太網路) 介面卡。
- 兩個適用於 DGX 作業系統的 1.92 TB M.2 NVMe 磁碟機、八個用於儲存 / 快取的 3.84 TB U.2 NVMe 磁碟機。
- 最大 10.2 kw 功率。

DGX H100 CPU 匣的後端連接埠如下所示。四個 OFP 連接埠可為 InfiniBand 運算架構提供八個 ConnectX-7 介面卡。每對雙連接埠 ConnectX-7 介面卡都提供平行路徑、可通往儲存和管理架構。額外連接埠用於 BMC 存取。

\_NVIDIA DGX H100 背面板\_ 影像： aipd\_nv\_H100\_REAR .png []

## NVIDIA Networking

### NVIDIA Quantum-2 QM9700 交換器

\_NVIDIA Quantum-2 QM9700 InfiniBand 交換器\_ 影像： aipod\_nv\_QM970.png[]

NVIDIA Quantum-2 QM9700 交換器搭配 400GB / 秒 InfiniBand 連線能力、可為 NVIDIA Quantum-2 InfiniBand BasePOD 組態中的運算架構提供強大動力。ConnectX-7 單埠介面卡用於 InfiniBand 運算架構。每個 NVIDIA DGX 系統都有兩個連線至每個 QM9700 交換器、可在系統之間提供多個高頻寬、低延遲的路徑。

### NVIDIA Spectrum 3 SN4600 交換器

\_NVIDIA Spectrum 3 SN4600 交換器\_ 影像： aipd\_nv\_SN4600\_oires\_small.png []

NVIDIA Spectrum 3 SN4600 交換器總共提供 128 個連接埠（每個交換器 64 個）、可為 DGX BasePOD 的頻內管理提供備援連線能力。NVIDIA SN4600 交換器可提供介於 1 GbE 和 200 GbE 之間的速度。對於透過乙太網路連線的儲存設備、也會使用 NVIDIA SN4600 交換器。NVIDIA DGX 雙連接埠 ConnectX-7 介面卡上的連接埠可用於頻內管理和儲存連線。

### NVIDIA Spectrum SN2201 交換器

*NVIDIA Spectrum SN2201 switch* 影像： aipd\_nv\_SN2201.png[]

NVIDIA Spectrum SN2201 交換器提供 48 個連接埠、可提供額外管理的連線功能。額外管理可為 DGX BasePOD 中的所有元件提供整合式管理連線。

## NVIDIA ConnectX-7 介面卡

\_NVIDIA ConnectX-7 介面卡\_ 影像： aipd\_nv\_CX7.png[]

NVIDIA ConnectX-7 介面卡可提供 25/50/100/200/400G 的處理量。NVIDIA DGX 系統同時使用單連接埠和雙連接埠 ConnectX-7 介面卡、可在 DGX BasePOD 部署中提供高彈性、支援 400GB / 秒 InfiniBand 和 100/200Gb 乙太網路。

## NetApp AI Pod 搭配 NVIDIA DGX 系統 - 軟體元件

本節重點介紹 NetApp AI Pod 搭配 NVIDIA DGX 系統的軟體元件。

### NVIDIA 軟體

#### NVIDIA Base Command

NVIDIA Base Command<sup>®</sup> ；為每個 DGX BasePOD 提供強大動力、讓組織能夠充分發揮 NVIDIA 軟體創新的最大效益。企業可以透過備受肯定的平台、充分發揮投資潛力、包括企業級協調與叢集管理、加速運算、儲存與網路基礎架構的程式庫、以及針對 AI 工作負載最佳化的作業系統（OS）。

\_NVIDIA BaseCommand 解決方案\_ 影像： aipd\_nv\_BaseCommand\_new.png []

#### NVIDIA GPU 雲端 (NGC)

NVIDIA NGC™ 提供軟體、可滿足具有各種 AI 專業水準的資料科學家、開發人員和研究人員的需求。NGC 軟體會掃描一組彙總的常見弱點與曝險（CVE）、加密和私密金鑰。它經過測試與設計、可擴充至多個 GPU、在許多情況下、可擴充至多節點、確保使用者在 DGX 系統上的投資發揮最大效益。

\_NVIDIA GPU Cloud\_ 影像： aipd\_nv\_ngc.png []

#### NVIDIA AI Enterprise

NVIDIA AI Enterprise 是端點對端軟體平台、可為每個企業提供泛用 AI、為專為在 NVIDIA DGX 平台上執行而最佳化的泛用 AI 基礎模型提供最快且最有效率的執行時間。憑藉生產級的安全性、穩定性和管理能力、它簡化了泛型 AI 解決方案的開發。DGX BasePOD 隨附 NVIDIA AI Enterprise、可讓企業開發人員存取預先訓練的模型、最佳化架構、微服務、加速程式庫、以及企業支援。

### NetApp 軟體

#### NetApp ONTAP

NetApp 最新一代的儲存管理軟體、即支援企業將基礎架構現代化、並移轉至雲端就緒的資料中心。ONTAP 利用領先業界的資料管理功能 ONTAP、無論資料位於何處、只要使用一組工具、即可管理及保護資料。您也可以自由地將資料移至任何需要的位置：邊緣、核心或雲端。支援眾多功能、可簡化資料管理、加速及保護關鍵資料、並在混合雲架構中提供新一代基礎架構功能。ONTAP

#### 加速並保護資料

提供優異的效能與資料保護、並以下列方式擴充這些功能：ONTAP

- 效能與較低的延遲。ONTAP 以最低可能延遲提供最高的處理量、包括支援使用 NFS over RDMA、平行 NFS（pNFS）和 NFS 工作階段主幹的 NVIDIA GPUDirect 儲存設備（GDS）。
- 資料保護：ONTAP 提供內建的資料保護功能、以及業界最強大的反勒索軟體保證、並可在所有平台上進行通用管理。
- NetApp Volume Encryption（NVE）。支援內建和外部金鑰管理、提供原生Volume層級的加密功能。ONTAP
- 儲存設備多租戶和多因素驗證。支援以最高安全等級共享基礎架構資源。ONTAP

## 簡化資料管理

資料管理對於企業IT營運和資料科學家而言至關重要、因此可將適當的資源用於AI應用程式和訓練AI/ML資料集。下列關於NetApp技術的其他資訊超出此驗證範圍、但可能會因您的部署而有所差異。

包含下列功能的資料管理軟體、可簡化及簡化作業、並降低您的總營運成本：ONTAP

- 快照和複製功能可讓 ML/DL 工作流程進行協同作業、平行實驗和強化資料管理。
- SnapMirror 可在混合雲和多站台環境中順暢地移動資料、隨時隨地提供所需的資料。
- 即時資料精簡與擴充重複資料刪除技術。資料壓縮可減少儲存區塊內的空間浪費、重複資料刪除技術可大幅提升有效容量。這適用於本機儲存的資料、以及分層至雲端的資料。
- 最低、最大及可調適的服務品質（AQO）。精細的服務品質（QoS）控制有助於維持高共享環境中關鍵應用程式的效能等級。
- NetApp FlexGroups 可在儲存叢集中的所有節點上散佈資料、為極大型的資料集提供大容量和更高效能。
- NetApp FabricPool自動將冷資料分層至公有和私有雲端儲存選項、包括Amazon Web Services（AWS）、Azure和NetApp StorageGRID 等儲存解決方案。如需FabricPool 更多有關資訊、請參閱 "[TR-4598：FabricPool 最佳實務做法](#)"。
- NetApp FlexCache。提供遠端磁碟區快取功能、可簡化檔案發佈、減少 WAN 延遲、並降低 WAN 頻寬成本。FlexCache 可在多個站台之間進行分散式產品開發、並可從遠端位置加速存取公司資料集。

## 符合未來需求的基礎架構

下列功能可協助滿足嚴苛且不斷變化的業務需求：ONTAP

- 無縫擴充和不中斷營運。ONTAP 支援在線上新增現有控制器和橫向擴充叢集的容量。客戶可以升級至最新技術、例如NVMe和32GB FC、而不需進行昂貴的資料移轉或中斷運作。
- 雲端連線：NetApp是最具雲端連線能力的儲存管理軟體、可在所有公有雲中選擇軟體定義儲存（AI）和雲端原生執行個體（NetApp）ONTAP ONTAP Select Cloud Volumes Service。
- 與新興應用程式整合。利用支援現有企業應用程式的相同基礎架構、為新一代平台和應用程式提供企業級資料服務、例如自動駕駛車輛、智慧城市和產業4.0。ONTAP

## NetApp DataOps工具套件

NetApp DataOps Toolkit是一款以Python為基礎的工具、可簡化開發/訓練工作區和推斷伺服器的管理、這些工作區都以高效能橫向擴充的NetApp儲存設備為後盾。DataOps Toolkit 可作為獨立公用程式運作、在 Kubernetes 環境中更有效、利用 NetApp Astra Trident 來自動化儲存作業。主要功能包括：

- 快速配置以高效能橫向擴充NetApp儲存設備為後盾的新高容量JupyterLab工作區。

- 快速配置以企業級NetApp儲存設備為後盾的全新NVIDIA Triton Inference Server執行個體。
- 近乎即時的高容量 JupyterLab 工作區複製、以便進行實驗或快速迭代。
- 高容量 JupyterLab 工作區的近乎即時的快照、用於備份及 / 或可追蹤性 / 基準化。
- 近乎即時的高容量高效能資料磁碟區資源配置、複製及快照。

## NetApp Astra Trident

Astra Trident是完全受支援的開放原始碼儲存協調工具、適用於容器和Kubernetes配送、包括Anthos。Trident可搭配整個 NetApp 儲存產品組合使用、包括 NetApp ONTAP、也支援 NFS、NVMe / TCP 和 iSCSI 連線。Trident可讓終端使用者從NetApp儲存系統配置及管理儲存設備、而無需儲存管理員介入、進而加速DevOps 工作流程。

## NetApp AIPod 搭配 NVIDIA DGX 系統 - 解決方案架構

本節著重於採用 NVIDIA DGX 系統的 NetApp AIPod 架構。

### 搭載 DGX H100 系統的 NetApp AI Pod

此參考架構運用獨立的架構來進行運算叢集互連和儲存存取、並在運算節點之間提供 400GB / 秒 InfiniBand (IB) 連線能力。下圖顯示 NetApp AIPod 搭配 DGX H100 系統的整體解決方案拓撲。

NetApp AIPod 解決方案拓撲 \_ 影像：aipod\_nv\_a900top.png []

### 網路組態

在此組態中、運算叢集架構使用一對 QM9700 400GB / 秒 IB 交換器、這些交換器會連接在一起以獲得高可用度。每個 DGX H100 系統都使用八個連線連接至交換器、其中偶數連接埠連接至一台交換器、而奇數連接埠則連接至另一台交換器。

對於儲存系統存取、頻內管理和用戶端存取、會使用一對 SN4600 乙太網路交換器。交換器是透過交換器間的連結連接、並設定多個 VLAN 來隔離各種流量類型。對於較大型的部署、可將乙太網路擴充至葉脊組態、為脊椎交換器新增額外的交換器配對、並視需要新增額外的葉片。

除了運算互連和高速乙太網路之外、所有實體裝置也會連線至一或多個 SN2201 乙太網路交換器、以進行頻外管理。如需 DGX H100 系統連線的詳細資訊、請參閱 "[NVIDIA BasePOD 文件](#)"。

### 用於儲存存取的用戶端組態

每個 DGX H100 系統均配置兩個雙連接埠 ConnectX-7 介面卡、用於管理和儲存流量、而此解決方案則將每個介面卡上的兩個連接埠連接到同一個交換器。接著、每個卡的一個連接埠會設定為 LACP MLAG 連結、每個交換器都有一個連接埠、而用於頻內管理、用戶端存取和使用者層級儲存存取的 VLAN 則會裝載在此連結上。

每張卡上的另一個連接埠用於連線至 AFF A900 儲存系統、並可用於數種組態、視工作負載需求而定。對於使用 NFS over RDMA 來支援 NVIDIA Magnum IO GPUDirect 儲存設備的組態、連接埠會設定為主動 / 被動連結、因為任何其他類型的連結都不支援 RDMA。對於不需要 RDMA 的部署、儲存介面也可以設定為 LACP 繫結、以提供高可用度和額外頻寬。無論是否使用 RDMA、用戶端都可以使用 NFS v4.1 pNFS 和工作階段主幹來掛載儲存系統、以便平行存取叢集中的所有儲存節點。



## 儲存系統組態

每個 AFF A900 儲存系統都使用每個控制器的四個 100 GbE 連接埠進行連線。每個控制器的兩個連接埠用於從 DGX 系統存取工作負載資料、每個控制器的兩個連接埠則設定為 LACP 介面群組、以支援從管理平面伺服器存取叢集管理成品和使用者主目錄。儲存系統的所有資料存取都是透過 NFS 提供、其中儲存虛擬機器 (SVM) 專用於 AI 工作負載存取、另有專用於叢集管理用途的獨立 SVM。

工作負載 SVM 總共配置了八個邏輯介面 (生命)、每個實體連接埠上有兩個生命。此組態可提供最大頻寬、以及將每個 LIF 容錯移轉到同一個控制器上的另一個連接埠的方法、以便在網路故障時、兩個控制器都能保持作用中。此組態也支援 NFS over RDMA 以啟用 GPUDirect 儲存存取。儲存容量是以單一大型 FlexGroup 磁碟區的形式配置、該磁碟區橫跨叢集中的所有儲存控制器、每個控制器上有 16 個組成磁碟區。此 FlexGroup 可從 SVM 上的任何生命體存取、透過使用 NFSv4.1 搭配 pNFS 和工作階段主幹、用戶端可與 SVM 中的每個 LIF 建立連線、讓每個儲存節點的本機資料能夠平行存取、大幅提升效能。工作負載 SVM 和每個資料 LIF 也設定為用於 RDMA 傳輸協定存取。如需 ONTAP RDMA 組態的詳細資訊、請參閱 ["本文檔 ONTAP"](#)。

管理 SVM 只需要單一 LIF、裝載在每個控制器上設定的雙連接埠介面群組上。管理 SVM 上的其他 FlexGroup 磁碟區則是用來存放叢集管理產出工件、例如叢集節點映像、系統監控歷史資料、以及終端使用者主目錄。下圖顯示儲存系統的邏輯組態。

NetApp A900 儲存叢集邏輯組態 \_ 影像：aipod\_nv\_a900logical.png []

## 管理層伺服器

此參考架構也包含五部以 CPU 為基礎的伺服器、供管理層使用。其中兩個系統是 NVIDIA Base Command Manager 的主要節點、用於叢集部署和管理。其餘三個系統則用於提供額外的叢集服務、例如 Kubernetes 主節點或登入節點、以便使用 Slurm 進行工作排程。使用 Kubernetes 的部署可運用 NetApp Astra Trident CSI 驅動程式、為 AFF A900 儲存系統上的管理和 AI 工作負載提供自動化的資源配置和資料服務、

每部伺服器都會實體連接至 IB 交換器和乙太網路交換器、以啟用叢集部署和管理、並透過管理 SVM 將 NFS 裝載至儲存系統、以儲存叢集管理產出工件、如前所述。

## NetApp AIPod 搭配 NVIDIA DGX 系統 - 解決方案驗證與規模調整指南

本節著重於 NetApp AIPod 搭配 NVIDIA DGX 系統的解決方案驗證與規模調整指南。

### 解決方案驗證

此解決方案中的儲存組態已使用一系列使用開放原始碼工具 FIO 的綜合工作負載進行驗證。這些測試包括讀寫 I/O 模式、用於模擬 DGX 系統執行深度學習訓練工作所產生的儲存工作負載。儲存組態已通過驗證、使用雙插槽 CPU 伺服器叢集同時執行 FIO 工作負載、以模擬 DGX 系統叢集。每個用戶端都設定了先前所述的相同網路組態、並加入下列詳細資料。

此驗證使用下列掛載選項：

ves=4.1	啟用 pNFS 以平行存取多個儲存節點
proto=RDMA	將傳輸通訊協定設定為 RDMA、而非預設 TCP
連接埠 = 20049	為 RDMA NFS 服務指定正確的連接埠
max_connect = 16	啟用 NFS 工作階段主幹以彙總儲存連接埠頻寬
寫入 = 渴望	改善緩衝寫入的寫入效能

rsz=262144、wsz=262144	將 I/O 傳輸大小設為 256k
-----------------------	-------------------

此外、用戶端設定的 NFS max\_Session\_插 槽值為 1024。在透過 RDMA 使用 NFS 測試解決方案時、儲存網路連接埠已設定為主動 / 被動連結。此驗證使用下列連結參數：

mode=active-backup	將連結設定為主動 / 被動模式
primary = <interface name>	所有用戶端的主要介面都分散在交換器上
Mll-monitor-interval = 100	指定 100ms 的監控時間間隔
容錯移轉 -Mac-policy=active	指定主動鏈的 MAC 地址是綁定的 MAC 地址。這是在連結介面上正確操作 RDMA 所需的。

儲存系統的設定方式如前所述、為兩對 A900 HA（4 個控制器）搭配兩個 NS224 磁碟櫃（每對 HA 連接 24 個 1.9TB NVMe 磁碟機）。如架構一節所述、所有控制器的儲存容量都是使用 FlexGroup 磁碟區進行組合、而來自所有用戶端的資料則分散在叢集中的所有控制器上。

儲存系統規模調整指南

NetApp 已成功完成 DGX BasePOD 認證、兩對通過測試的 A900 HA 可輕鬆支援八個 DGX H100 系統的叢集。對於儲存效能需求較高的大型部署、可在單一叢集中新增最多 12 個 HA 配對（24 個節點）的額外 AFF 系統至 NetApp ONTAP 叢集。使用本解決方案所述的 FlexGroup 技術、24 節點叢集可在單一命名空間中提供 40 PB 以上的資料傳輸量、以及高達 300 Gbps 的傳輸量。其他 NetApp 儲存系統（例如 AFF A400、A250 和 C800）則提供較低的效能和 / 或較高的容量選項、以較低的成本進行較小型的部署。由於 ONTAP 9 支援混合模式叢集、因此客戶可以從較小的初始佔用空間開始、並在容量和效能需求增加時、將更多或更大的儲存系統新增至叢集。下表顯示每個 AFF 機型所支援的 A100 和 H100 GPU 數量的粗略估計值。

NetApp 儲存系統規模調整指南

		Throughput <sup>2</sup>	Raw capacity (typical / max)	Connectivity	# NVIDIA A100 GPUs supported <sup>3</sup>	# NVIDIA H100 GPUs supported <sup>4</sup>
NetApp® AFF A900	1 HA pair <sup>1</sup>	28GB/s	182TB / 14.7PB	100 GbE	1 - 64	1-32
	12 HA pairs	336GB/s	2.1PB / 176.4PB		768	384
AFF A800	1 HA pair	25GB/s	368TB / 3.6PB	100 GbE	1 - 64	1-32
	12 HA pairs	300GB/s	4.4PB / 43.2PB		768	384
AFF C800	1 HA pair	21GB/s	368TB / 3.6PB	100 GbE	1-48	1-24
	12 HA pairs	252GB/s	4.4PB / 43.2PB		576	288
AFF A400	1 HA pair	11GB/s	182TB / 14.7PB	40/100 GbE	1 - 32	1-16
	12 HA pairs	132GB/s	2.1PB / 176.4PB		384	192
AFF C400	1 HA pair	8GB/s	182TB / 14.7PB	40/100 GbE	1 - 16	1-8
	12 HA pairs	128GB/s	2.1PB / 176.4PB		192	96
AFF A250	1 HA pair	7.4GB/s	91.2TB / 4.4PB	25 GbE 40/100GbE	1 - 16	1-8
	4 HA pairs	29.6GB/s	364.8TB / 17.6PB		64	32
AFF C250	1 HA pair	5 GB/s	91.2TB / 4.4PB	25 GbE 40/100GbE	1-8	1-4
	4 HA pairs	20 GB/s	364.8TB / 17.6PB		32	8

1 – 1 AFF = 1 HA pair = 2 Nodes. 12 HA pairs = 24 nodes  
2 – 100% sequential read

3 – Based on workload testing in NVA-1153  
4 – Based on BasePOD validation test results

## NetApp AIPOd 搭配 NVIDIA DGX 系統 - 結論與其他資訊

本節包含 NetApp AIPOd 搭配 NVIDIA DGX 系統的其他資訊參考資料。

### 結論

DGX BasePOD 架構是新一代的深度學習平台、需要同樣進階的儲存與資料管理功能。透過將 DGX BasePOD 與 NetApp AFF 系統結合、NetApp AIPOd 與 DGX 系統架構幾乎可在 24 節點 AFF A900 叢集上以任何規模實作、最多可擴充至 48 個 DGX H100 系統。AFF 結合 NetApp ONTAP 優異的雲端整合功能和軟體定義功能、提供跨越邊緣、核心和雲端的完整資料傳輸管道、讓 DL 專案成功完成。

### 其他資訊

若要深入瞭解本文件所述資訊、請參閱下列文件和 / 或網站：

- NetApp ONTAP 數據管理軟體ONTAP —資訊庫  
["https://docs.netapp.com/us-en/ontap-family/"](https://docs.netapp.com/us-en/ontap-family/)
- NetApp AFF A900 儲存系統 -  
["https://www.netapp.com/data-storage/aff-a-series/aff-a900/"](https://www.netapp.com/data-storage/aff-a-series/aff-a900/)
- NetApp ONTAP RDMA 資訊 -  
["https://docs.netapp.com/us-en/ontap/nfs-rdma/index.html"](https://docs.netapp.com/us-en/ontap/nfs-rdma/index.html)
- NetApp DataOps工具套件  
["https://github.com/NetApp/netapp-dataops-toolkit"](https://github.com/NetApp/netapp-dataops-toolkit)
- NetApp Astra Trident  
["總覽"](#)
- NetApp GPUDirect 儲存部落格 -  
["https://www.netapp.com/blog/ontap-reaches-171-gpudirect-storage/"](https://www.netapp.com/blog/ontap-reaches-171-gpudirect-storage/)
- NVIDIA DGX基礎POD  
["https://www.nvidia.com/en-us/data-center/dgx-basepod/"](https://www.nvidia.com/en-us/data-center/dgx-basepod/)
- NVIDIA DGX H100 系統  
["https://www.nvidia.com/en-us/data-center/dgx-h100/"](https://www.nvidia.com/en-us/data-center/dgx-h100/)
- NVIDIA Networking  
["https://www.nvidia.com/en-us/networking/"](https://www.nvidia.com/en-us/networking/)
- NVIDIA Magnum IO GPUDirect 儲存設備  
["https://docs.nvidia.com/gpudirect-storage"](https://docs.nvidia.com/gpudirect-storage)

- NVIDIA Base Command

["https://www.nvidia.com/en-us/data-center/base-command/"](https://www.nvidia.com/en-us/data-center/base-command/)

- NVIDIA Base Command Manager

["https://www.nvidia.com/en-us/data-center/base-command/manager/"](https://www.nvidia.com/en-us/data-center/base-command/manager/)

- NVIDIA AI Enterprise

["https://www.nvidia.com/en-us/data-center/products/ai-enterprise/"](https://www.nvidia.com/en-us/data-center/products/ai-enterprise/)

感謝

本文件為 NetApp 解決方案與 ONTAP 工程團隊的工作成果：David Arnette、Olga Kornievskaia、Dustin Fischer、Srikanth Kaligotla、Mohit Kumar 和 Rajeev Badrinath。作者也想感謝 NVIDIA 和 NVIDIA DGX BasePOD 工程團隊持續提供支援。

## **NVA-1151-Design：採用ONTAP NVIDIA DGX A100系統設計指南的NetApp W人工 智慧**

NetApp公司David Arette和Sung-HAN LIN

NVA-1151-Design說明NetApp認證的機器學習架構、以及使用NetApp AFF ®XA800儲存系統、NVIDIA DGX A100系統和NVIDIA Mellanox網路交換器的人工智慧工作負載。其中也包含已實作架構的基準測試結果。

["NVA-1151-Design：採用ONTAP NVIDIA DGX A100系統設計指南的NetApp W人工 智慧"](#)

## **NVA-1151-Deploy：NetApp ONTAP 支援NVIDIA DGX A100系統的AI**

NetApp公司David Arignette

NVA-1151-Deploy內含NetApp驗證架構（NVA）的儲存系統部署指示、適用於使用NetApp AFF ArfoA800儲存系統、NVIDIA DGX A100系統及NVIDIA Mellanox網路交換器進行機器學習（ML）和人工智慧（AI）工作負載。其中也包含在部署完成後執行驗證基準測試的指示。

["NVA-1151-Deploy：NetApp ONTAP 支援NVIDIA DGX A100系統的AI"](#)

## **NVA-1153設計：採用ONTAP NVIDIA DGX A100系統和Mellanox Spectrum乙太網路交換器的NetApp W人工 智慧**

NetApp公司David Arette和Sung-HAN LIN

NVA-1153設計說明NetApp驗證的機器學習（ML）架構與人工智慧（AI）工作負載、使

用NetApp AFF 的S4A800儲存系統、NVIDIA DGX A100系統、以及NVIDIA Mellanox Spectrum SN3700V 200GB乙太網路交換器。這項設計採用融合式乙太網路（RoCE）的RDMA、適用於運算叢集互連架構、為客戶提供完全乙太網路架構、以滿足高效能工作負載的需求。本文件也包含已實作架構的基準測試結果。

["NVA-1153設計：採用ONTAP NVIDIA DGX A100系統和Mellanox Spectrum乙太網路交換器的NetApp W人工智慧"](#)

## **NVA-1153-Deploy：採用ONTAP NVIDIA DGX A100系統和Mellanox Spectrum乙太網路交換器的NetApp W人工智慧**

NetApp公司David Arignette

NVA-1153-Deploy內含儲存系統部署說明、適用於NetApp驗證的機器學習（ML）架構與人工智慧（AI）工作負載、使用NetApp AFF ArfoA800儲存系統、NVIDIA DGX A100系統、以及NVIDIA Mellanox Spectrum SN3700V 200GB乙太網路交換器。其中也包含在部署完成後執行驗證基準測試的指示。

["NVA-1153-Deploy：採用ONTAP NVIDIA DGX A100系統和Mellanox Spectrum乙太網路交換器的NetApp W人工智慧"](#)

## 版權資訊

Copyright © 2024 NetApp, Inc. 版權所有。台灣印製。非經版權所有人事先書面同意，不得將本受版權保護文件的任何部分以任何形式或任何方法（圖形、電子或機械）重製，包括影印、錄影、錄音或儲存至電子檢索系統中。

由 NetApp 版權資料衍伸之軟體必須遵守下列授權和免責聲明：

此軟體以 NETAPP「原樣」提供，不含任何明示或暗示的擔保，包括但不限於有關適售性或特定目的適用性之擔保，特此聲明。於任何情況下，就任何已造成或基於任何理論上責任之直接性、間接性、附隨性、特殊性、懲罰性或衍生性損害（包括但不限於替代商品或服務之採購；使用、資料或利潤上的損失；或企業營運中斷），無論是在使用此軟體時以任何方式所產生的契約、嚴格責任或侵權行為（包括疏忽或其他）等方面，NetApp 概不負責，即使已被告知有前述損害存在之可能性亦然。

NetApp 保留隨時變更本文所述之任何產品的權利，恕不另行通知。NetApp 不承擔因使用本文所述之產品而產生的責任或義務，除非明確經過 NetApp 書面同意。使用或購買此產品並不會在依據任何專利權、商標權或任何其他 NetApp 智慧財產權的情況下轉讓授權。

本手冊所述之產品受到一項（含）以上的美國專利、國外專利或申請中專利所保障。

有限權利說明：政府機關的使用、複製或公開揭露須受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中的「技術資料權利 - 非商業項目」條款 (b)(3) 小段所述之限制。

此處所含屬於商業產品和 / 或商業服務（如 FAR 2.101 所定義）的資料均為 NetApp, Inc. 所有。根據本協議提供的所有 NetApp 技術資料和電腦軟體皆屬於商業性質，並且完全由私人出資開發。美國政府對於該資料具有非專屬、非轉讓、非轉授權、全球性、有限且不可撤銷的使用權限，僅限於美國政府為傳輸此資料所訂合約所允許之範圍，並基於履行該合約之目的方可使用。除非本文另有規定，否則未經 NetApp Inc. 事前書面許可，不得逕行使用、揭露、重製、修改、履行或展示該資料。美國政府授予國防部之許可權利，僅適用於 DFARS 條款 252.227-7015(b)（2014 年 2 月）所述權利。

## 商標資訊

NETAPP、NETAPP 標誌及 <http://www.netapp.com/TM> 所列之標章均為 NetApp, Inc. 的商標。文中所涉及的所有其他公司或產品名稱，均為其各自所有者的商標，不得侵犯。