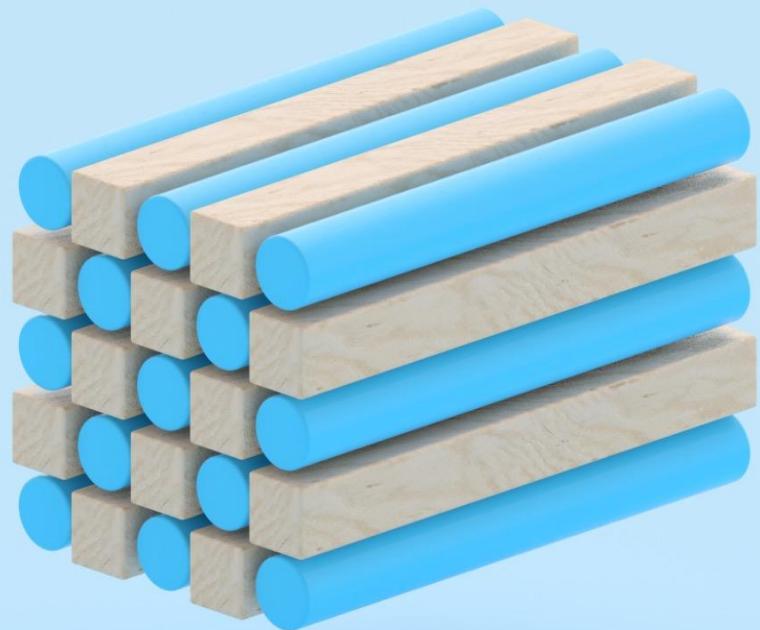


打造 AI 校園儲存平台

Henry Chen
Solutions Engineer



Agenda

- NetApp Overview
- 資料空間效率
- 資料保護
- 資料管理

#1

企業外部OEM儲存系統 台灣市佔第一

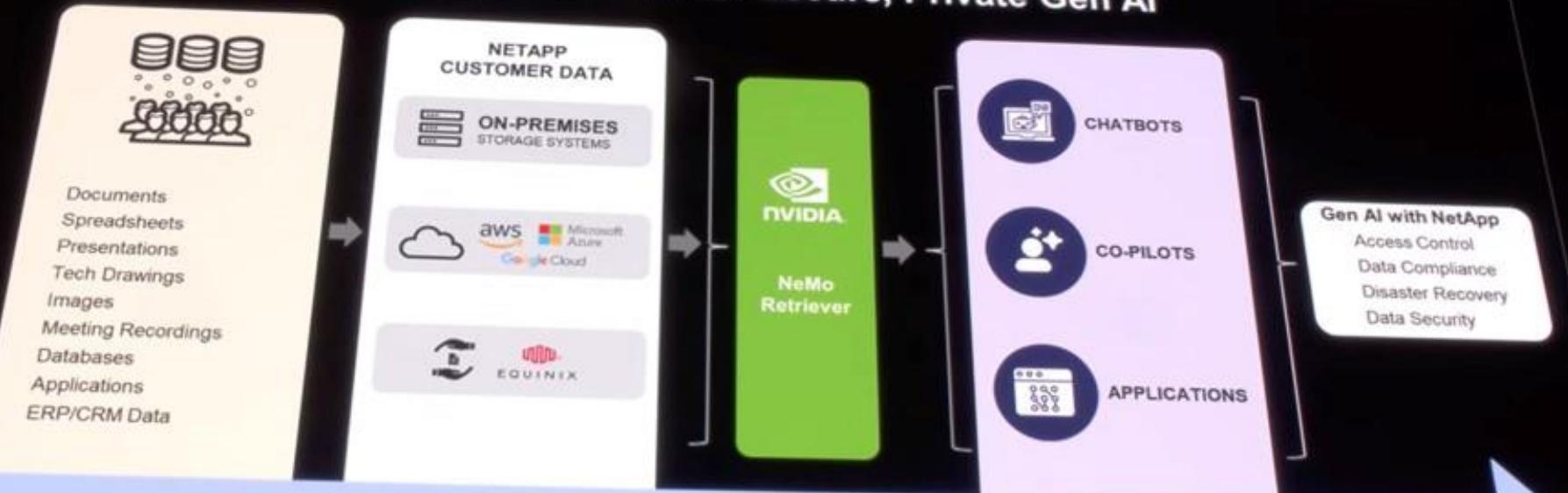
市佔率 (依廠商營收)	開放式網路儲存	全快閃陣列儲存
2024 Q1	31.6%	38.2%
Rolling 4Q	29.2%	33.3%

資料來源：IDC 全球企業儲存系統季度追蹤報告，2024年第一季（2024年6月發布）





NetApp Unlocks Exabytes of Data for Secure, Private Gen AI



NETAPP CUSTOMERS CAN NOW TALK TO THEIR DATA"

" Nearly half of the files in the world are stored on-prem on NetApp"

Jensen Huang – CEO, NVIDIA

首先, AI workload 需要 external storage?

資料規模

要讓AI模型準確度越高，訓練用的資料越多越好，在這個前提下，資料量會不斷成長

Local storage空間遲早不夠, 所以需要容易擴充的external storage. 另外, 現在的Gen AI LLM模型越來越大, 參數越來越多
GPU memory過於昂貴空間不夠的狀況下, 也需要高效能的external storage來幫忙最佳化整體成本.

LLM的典型狀況是, 一個具有1 billion (10億) 參數的model, 需要GPU memory至少80GB (32-bit full precision). Meta的LLaMA 2有約70 billions (700億) 參數,
估計需要 $80GB \times 70 = 5600GB = 5.6TB$ 的GPU memory, 而OpenAI的GPT4則是號稱有兩萬億參數, 估計需要 $80GB \times 2000 = 160TB$ 的GPU memory...?

資料分享

多個GPU node需要由統一平台提供資料給不同GPU node進行運算，避免資料孤島

所以需要分享資料用的external storage，解除local storage管理資料的限制與麻煩，提升資料科學家的工作效率

資料獨立

AI時代的應用不斷改變，但資料才是企業最寶貴的資產，**因此將計算與儲存分開, 做到計算不斷迭代, 而資料永久持續**

這也正是Microsoft, Facebook, AWS, Google, Nvidia等科技巨頭的做法，所以需要獨立運作的external storage來達到此目的.

資料安全

如何**最有效率地保護資料**，備份資料、還原資料、移動資料、加密資料、防止勒索病毒、訪問控制、分析使用行為、分類識別隱私等
這些都是非常重要的安全考量，所以需要專業的external storage來提供這些企業級服務

NetApp AI Pod Reference Architecture

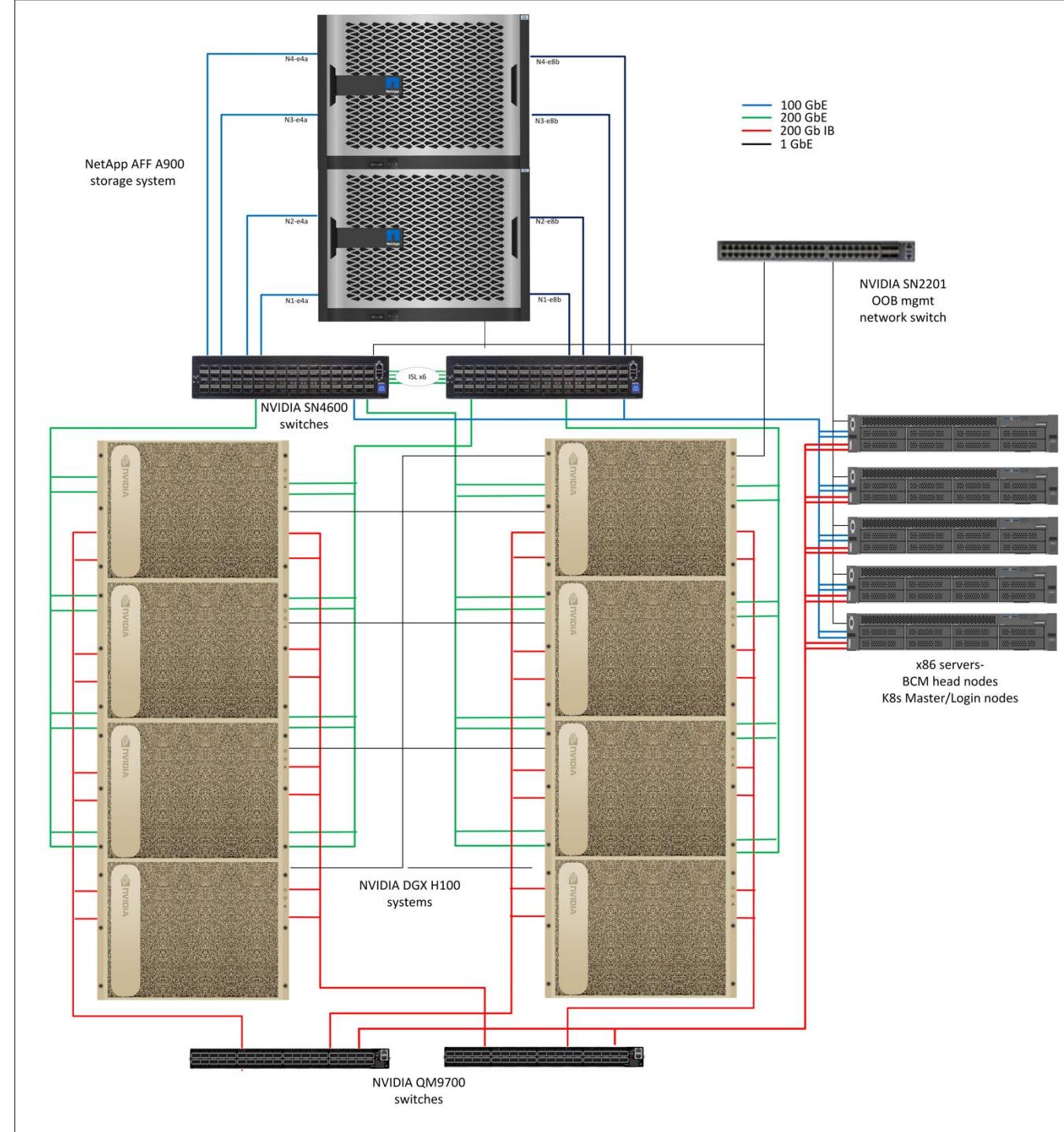
DGX H100 –
H100 GPU x8

NetApp Ontap –
A series, C series

Network –
GPU node之間溝通走
Infiniband 200Gb IB,
GPU node跟storage的溝通走
GDS over ethernet 100GbE

Nvidia認證NetApp

<https://www.nvidia.com/en-us/data-center/resources/netapp-ontap-ai-reference-architecture/>



國衛院啟用 NetApp AFF A800，為生醫 AI 平台建立快速且流暢的資料傳輸通道

以 A800 為中繼儲存，快速輸送資料至 GPU 伺服器，提升 AI 建模效率

文/ 廠商新聞稿 | 2022-10-14 發表

讚 45

分享



「我認為 NetApp AFF A800 最大優勢，在於擁有優異的資料儲存與分享效率，並擁有高穩定度，且只要我們拋出問題、NetApp都能快速回應。這些特質，足以帶給國衛院充足信心，未來不論承載多麼龐大複雜的基因資料庫，也無所畏懼。」 國衛院群體健康科學研究所所長 邱弘毅

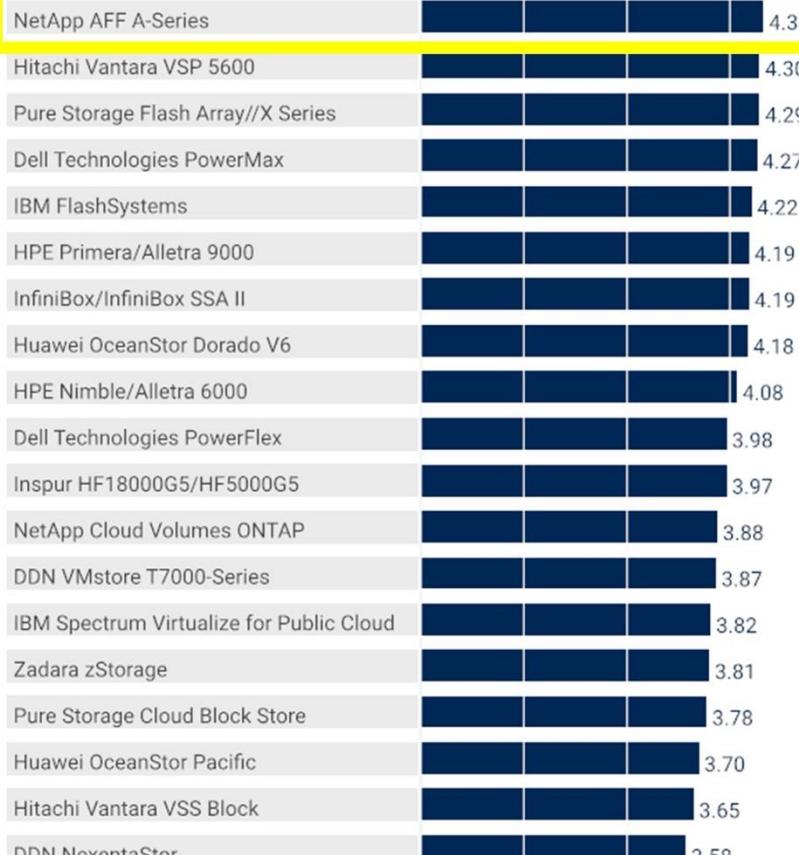


NetApp AI 方案策略

	Data Prep/Data Lake Modernization 數據處理	High Performance Model Training 從無到有建立	Model Training & Fine Tuning 套用既有model微調	RAG & Inferencing 搭配公司KB做推論
Description	<ul style="list-style-type: none">數據導入, 數據準備, 數據清洗, 數據預處理...必要的前期AI工作流程, 每間公司不論規模大小都需要這個策略.需要現代化的Data Lake平台. (如支援多協議, 集中管理)	<ul style="list-style-type: none">自己從無到有建立model做訓練.通常是非常大公司, 要有足夠資源, 人力, 技能, 數據, 才會有能力選擇這個策略. (如OpenAI, Google, Meta)需要非常高效能, 非常快速.	<ul style="list-style-type: none">套用已經預訓練好的現成model, 用自己公司的數據做微調, 產生出自己公司客製化的model.絕大多數公司都採取這個策略, 因為相對容易上手.需要處理大量參數的交換行為.數據的隱私與安全會是最重要的考量.若規模更小, 彈性利用雲端服務可以降低整體成本. (如GPU)	<ul style="list-style-type: none">RAG讓Gen AI先參考自己公司KB後, 再生成答案, 大幅提升準確度.必要的後期AI工作流程, 每間公司不論規模大小都需要這個策略.需要儲存公司KB數據和Gen AI生成的數據.數據的隱私與安全會是最重要的考量.
Storage Solutions	<ul style="list-style-type: none">NetApp AFF C series & StorageGRID Object (效率好與全面性功能)Dell EMC PowerScale & ECSPure FlashBlade	<ul style="list-style-type: none">NetApp BeeGFS, AFF A Series (高效能, 價格合理)WekaIOIBM GPFSDDN	<ul style="list-style-type: none">NetApp AFF A series (最安全, 雲端整合最完整)Dell EMC PowerScalePure FlashBlade	<ul style="list-style-type: none">Netapp AFF C series (最安全, Nvidia與Google使用案例)Dell EMC PowerScalePure FlashBlade

雲端維運

Product or Service Scores for Cloud IT Operations



As of 10 October 2022 1 2 3 4 5

Gartner

Reason:

NetApp 在 STaaS (託管儲存即服務)
展現強大 AIOps (智能維運) 支援能力

Gartner

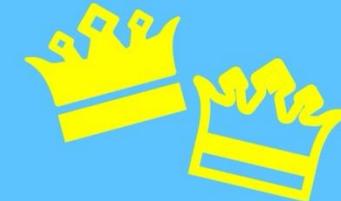
主儲存設備關鍵能力報告

NetApp AFF 系列

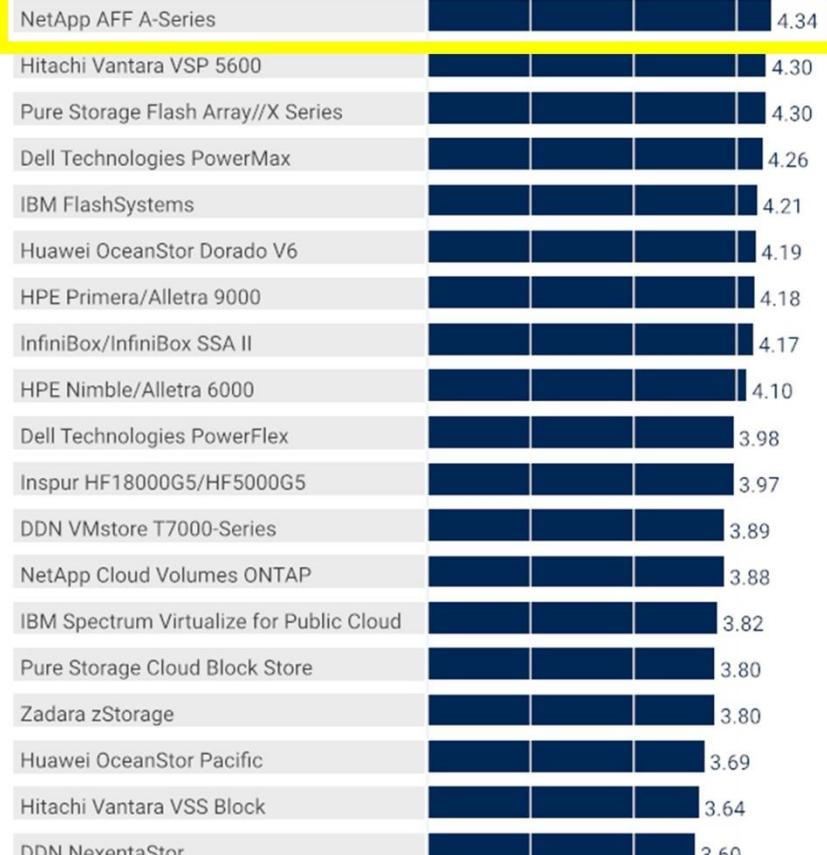
在所有關鍵能力案例中
都達到超乎水準的表現

尤其在 雲端維運 及 容器化應用

雙料第一



Product or Service Scores for Containers



As of 10 October 2022 1 2 3 4 5

Gartner

容器化應用

Reason:

NetApp 支援地端及雲端
所有主要 Kubernetes 版本



唯一通過驗證可儲存
最高機密資料的
企業級儲存供應商



美國國安局(NSA)
Commercial Solutions
for Classified (CSfC)
方案元件清單



聯邦資訊處理標準
FIPS 140-2



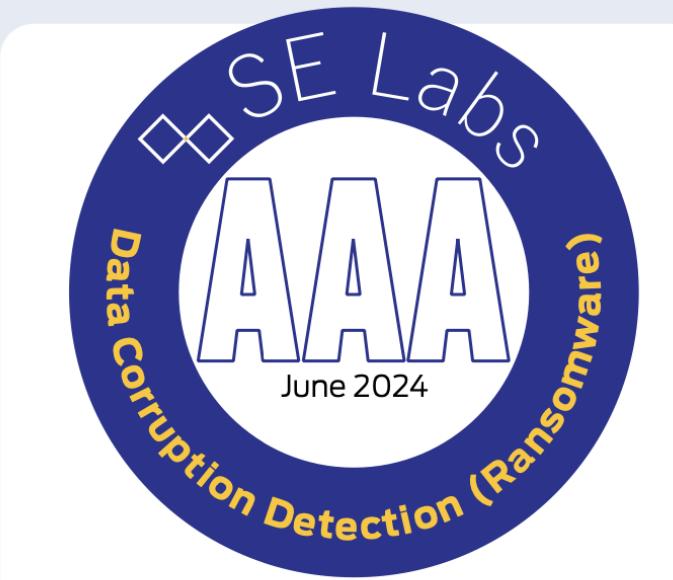
美國國防部資訊網路
(DoDIN)
核准產品清單 (APL)



Common Criteria 認證

Detection (Ransomware) Award

The following product wins the SE Labs award:



**NetApp
ONTAP Autonomous
Ransomware
Protection with AI**

We tested NetApp ONTAP Autonomous Ransomware Protection with AI against thousands of known and unknown variations of ransomware attacks. Its job was to monitor business data and detect any unwanted changes that the attacks made to those files, such as encrypting files.

We examined its abilities to detect the behaviour of:

- Known ransomware from 15 different attacker groups
- New, similar variants of ransomware from the same groups
- Legitimate operations on the network

NetApp ONTAP Autonomous Ransomware Protection with AI performed exceptionally well, providing 99% detection coverage of the ransomware's destructive behaviour. It was completely effective at handling legitimate operations, making no mistakes.

Executive Summary

Product Tested	Detection Accuracy (%)
NetApp ONTAP Autonomous Ransomware Protection with AI	99%

NetApp ONTAP 滿足各種需求的資料儲存

成本最低
次要儲存運用

混合式 Flash

統一化

FAS



最佳性價比

大容量 Flash

AFF C 系列



SAN
最佳化

ASA C 系列



最佳效能
第一級工作負載運用

高效能 Flash

AFF A 系列



ASA A 系列



ONTAP

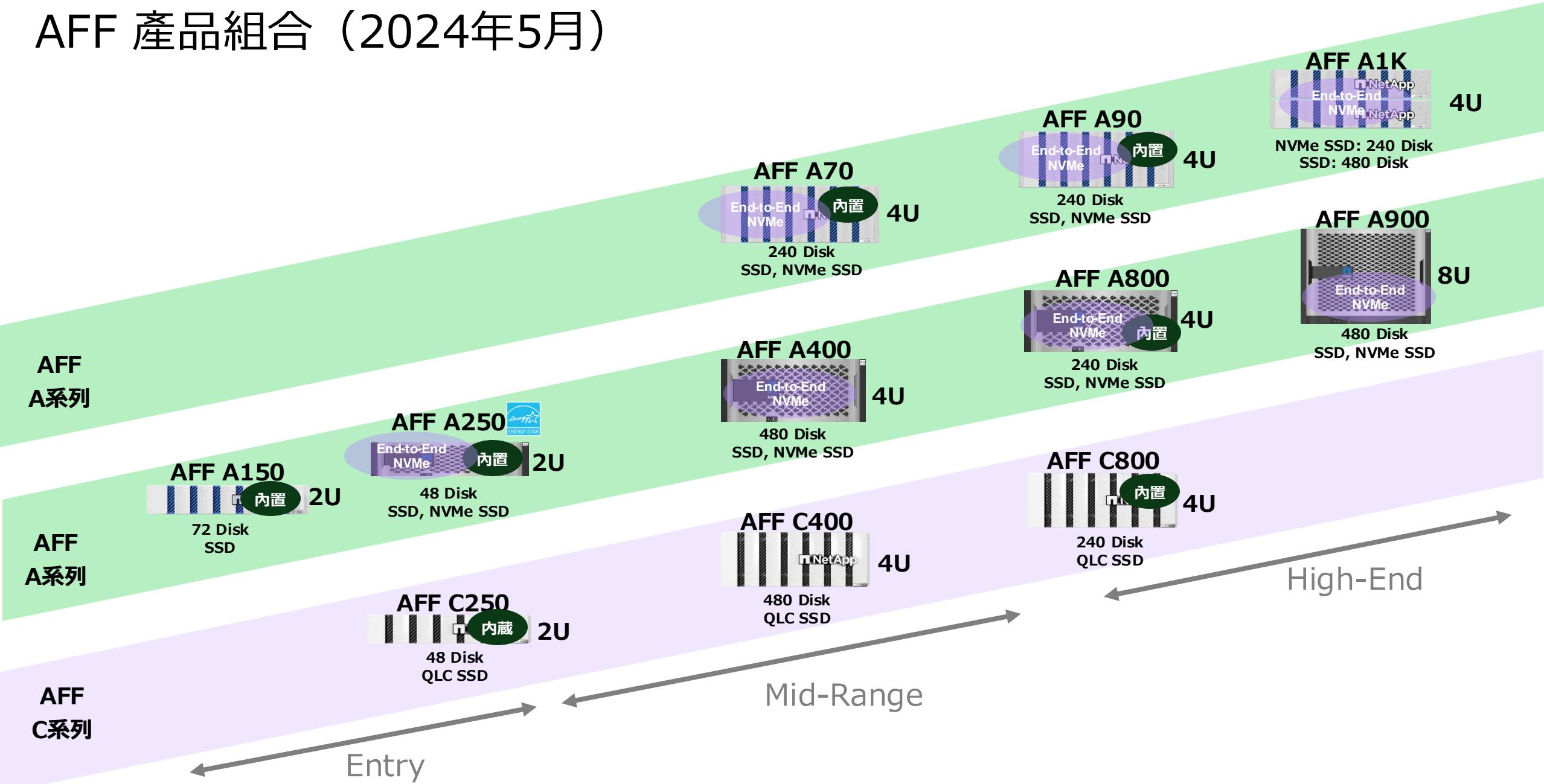
全方位管理
自動化、有效空間、File、Block、Object 資料保護與資安機能



Microsoft

Google Cloud

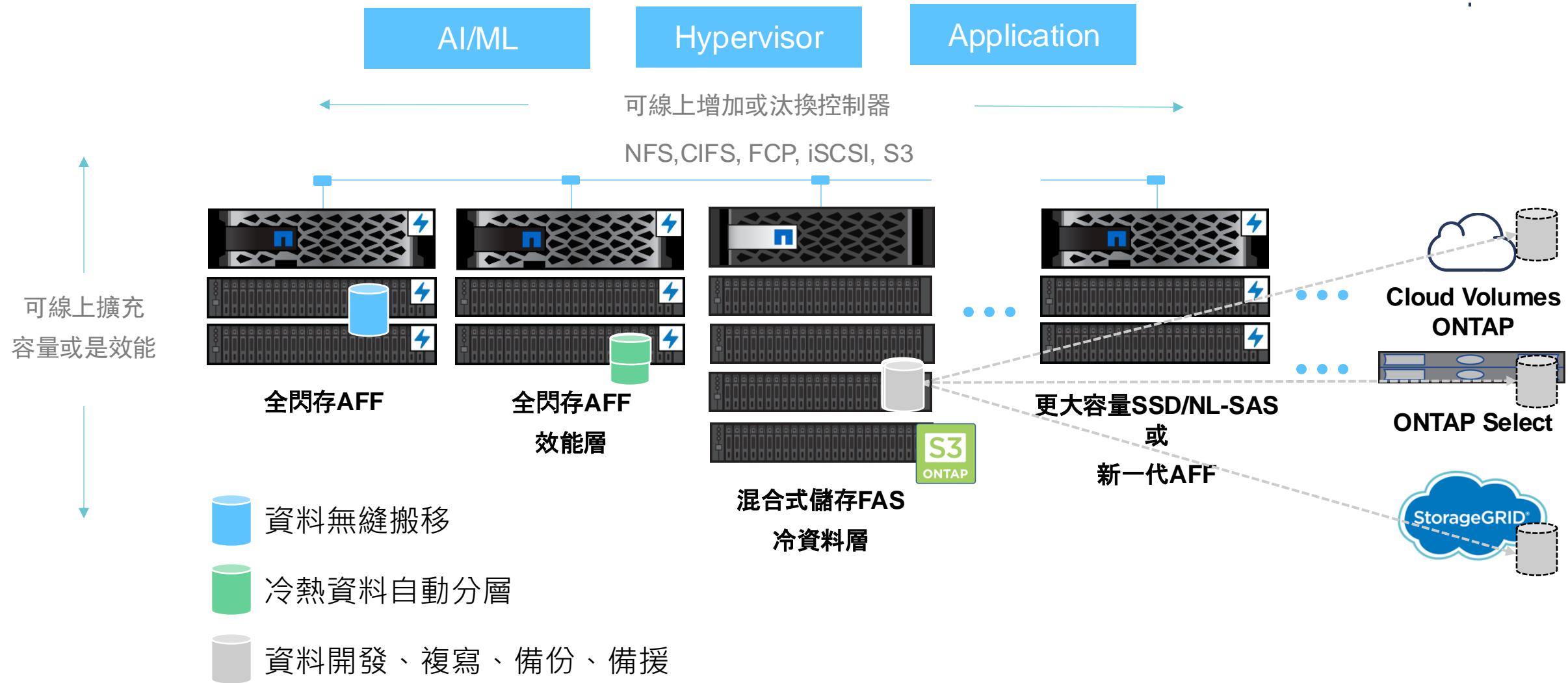
AFF 產品組合 (2024年5月)



資料空間效率

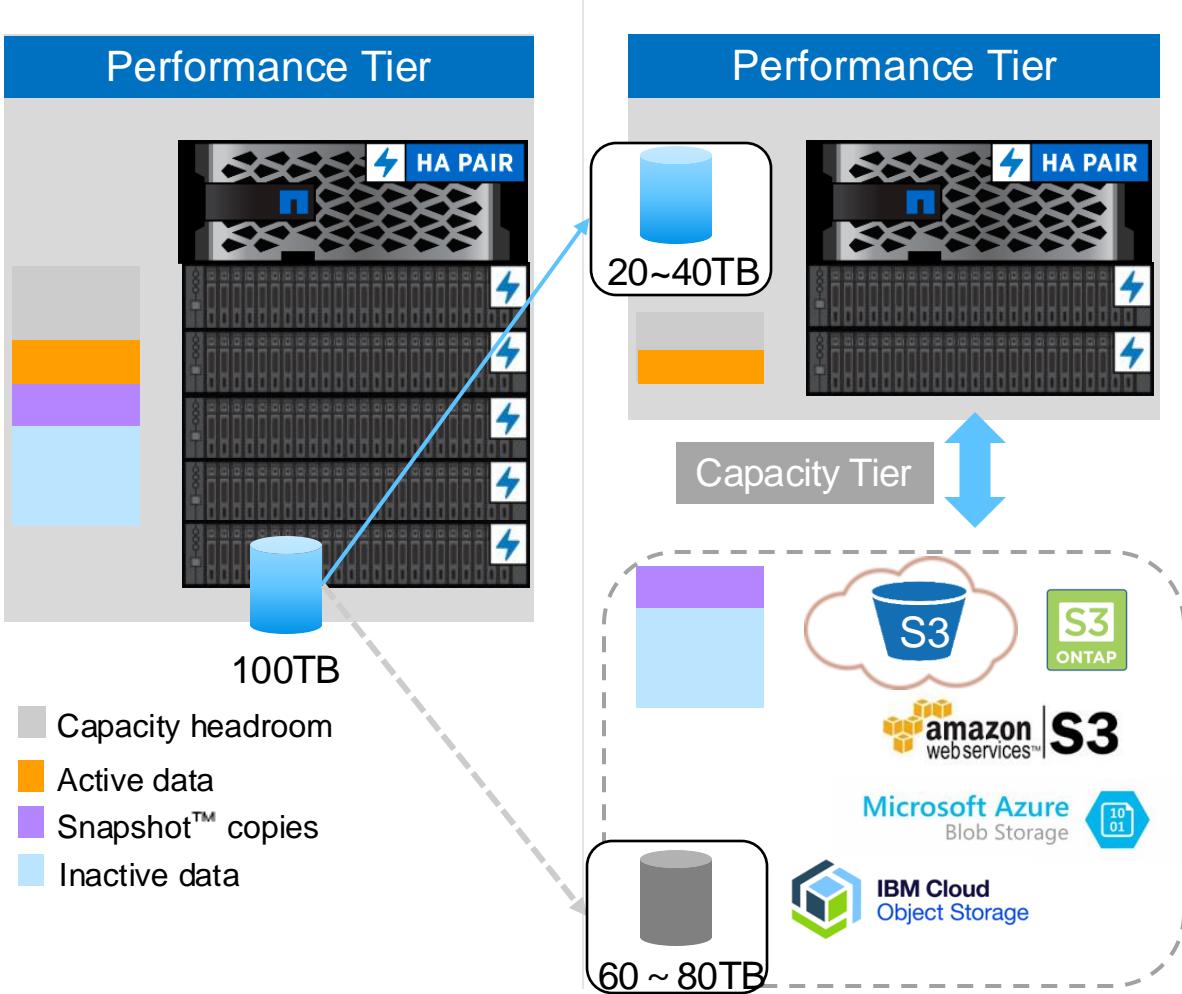
企業級資源池分享

因應需求，更新和擴展叢集群，而不會中斷業務



FabricPool 技術優勢

智能的資料自動分層技術-快照備份、冷資料、長期保存



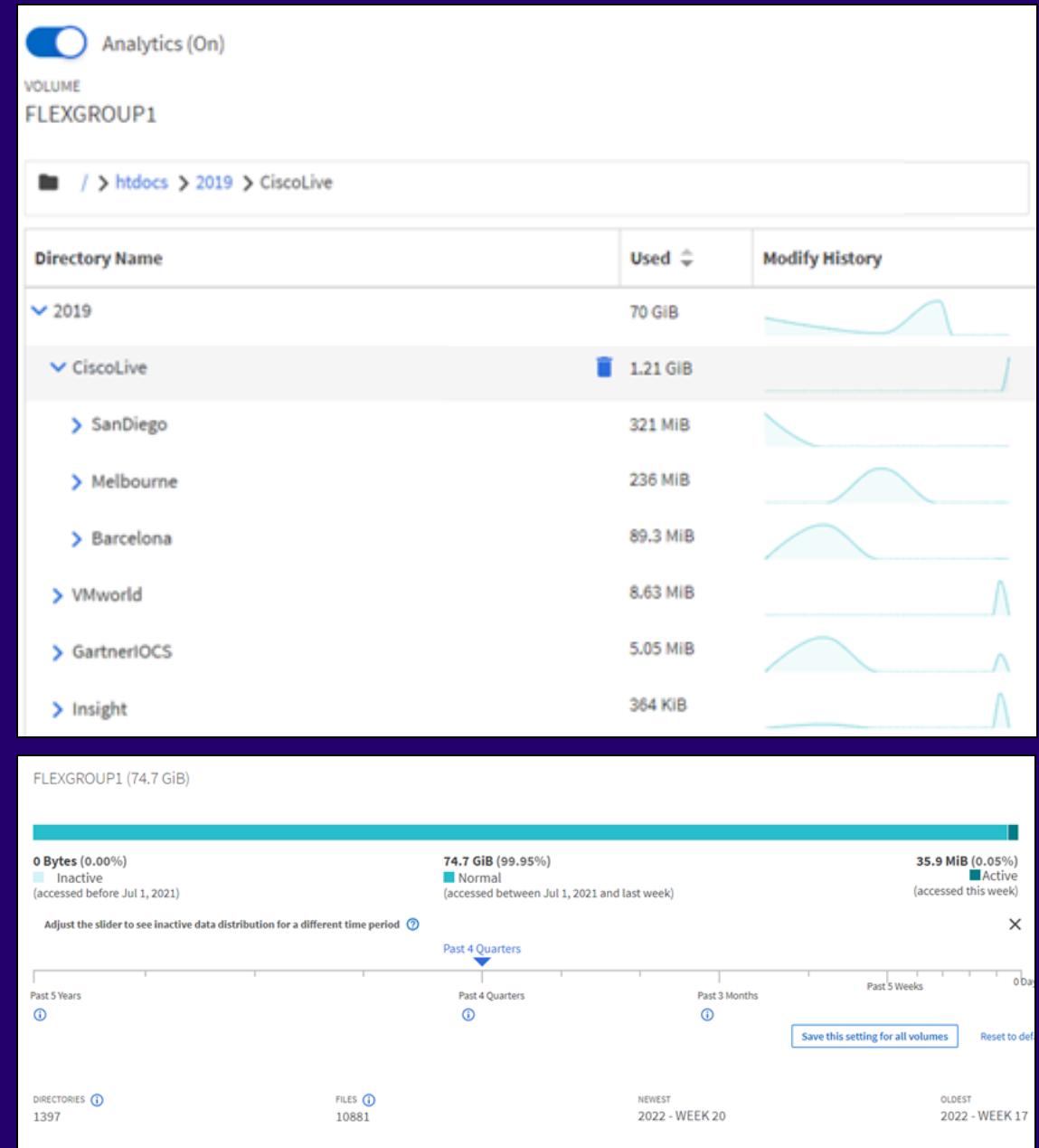
- NetApp高效率的空間節省技術會繼承到冷資料層
- 可以套用在 FCP,iSCSI,NFS,CIFS 協定
- 檔案系統及使用者行為不需要改變
- 自動化冷熱資料分層
- 降低整體營運成本
- 設定即遺忘管理

NetApp ONTAP file-system analytics

Real-time collection and insights into NAS data

Hierarchical structure of directories in NAS

- Storage VM (SVM)
- Volume
- Subdirectories count and performance
- File size, count, location, and performance
- File age histograms for modification and access history



NetApp Efficiency Guarantee Program

業界最有效的空間效率保證，提供企業滿滿的資料大平台 !!

空間效率保證，免簽署文件，一次性申請，簡單零風險 !!

簡單

- 滿足現今企業多元的應用負載，提供對應的效率保障
- **Virtual Servers 保證3:1, SAN 保證 4:1, NAS 1.5**
- **Containerized/Kubernetes 1.5:1**
- 透過NetApp® Snapshot™ 技術，提供更多效率保障

可靠

- 提供更多儲存效率，更高效率與性價比
- 一旦未達保證於出貨180天內，提出申請
- 我們說到做到！未達空間效率保證立即免費補償

「應用容納更多，空間佔用更少，效能更勝以往」

適用於NetApp多元全閃存解決方案：AFF * + ASA* + FAS500f + ONTAP® Select (SSD) + SolidFire®

* AFF = All Flash FAS
ASA = All SAN Array

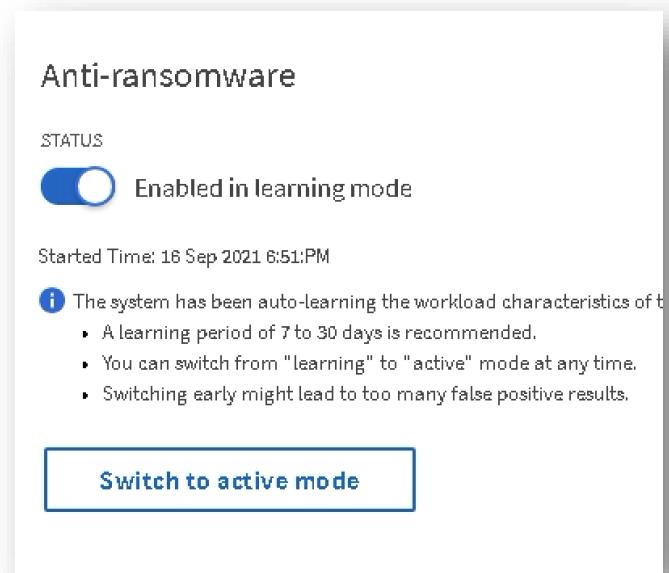
資料保護

NetApp Solution for Ransomware

自動觸發快照保護

ONTAP Anti-ransomware 儲存設備自動偵測與保護

- ONTAP 9.10.1 On-Box Anti-Ransomware
- 儲存設備底層出發
- Volume 第一層級揭露潛在感染
- 7 至 30 天機器學習模式
- 自動產出受感染檔案報告



The screenshot shows the 'Snapshot Copies' tab of the storage management interface. It displays a list of snapshots with the following details:

- Name: Anti_ransomware_backup.2021-11-18_1506
- Type: hourly.2021-11-18_1505

The screenshot shows the 'Suspected File Types' section. It includes:

- A header: 'Suspected File Types'.
- A note: 'You can mark the suspected file types as false positive or potential ransomware attack. The system will make necessary adjustments to the way it detects abnormal activity in the future.'
- A search bar and filter icon.
- A table with columns 'Type' and 'File'.
- Under 'Type': 'lckd'.
- Under 'File': A list of 20 files starting with '/Top_Dir_1/Sub_Dir_11/test_file...'.

Demo – Anti-Ransomware + Cloud Secure



NetApp 小學堂 勒索病毒的預防及保護

蘇子婷 Ting Su
NetApp 產品技術顧問

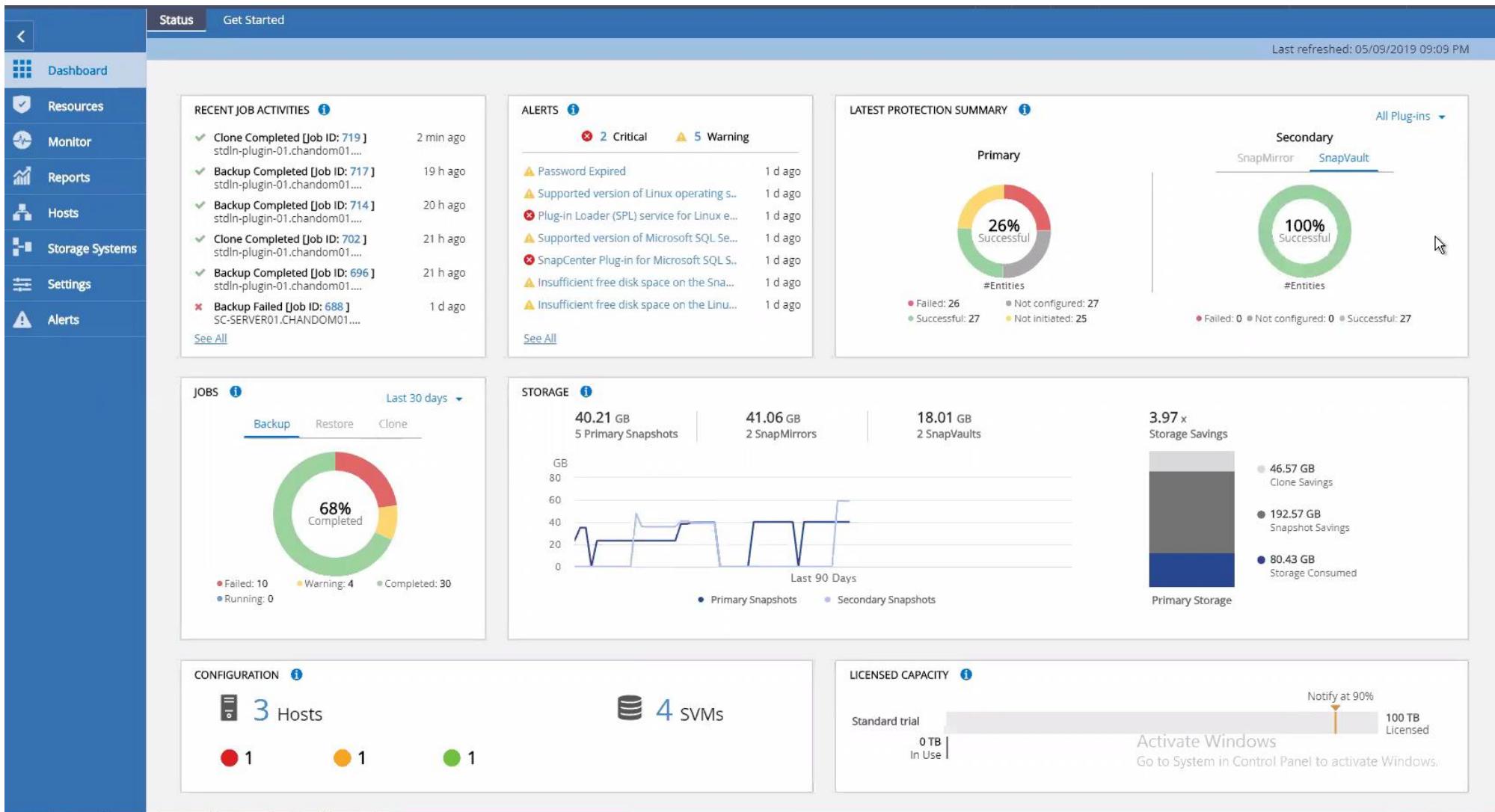
© 2021 NetApp, Inc. All rights reserved. — NETAPP CONFIDENTIAL —

SnapCenter – Supported Plugins

Filesystems, Hypervisors, Databases and Storage Systems

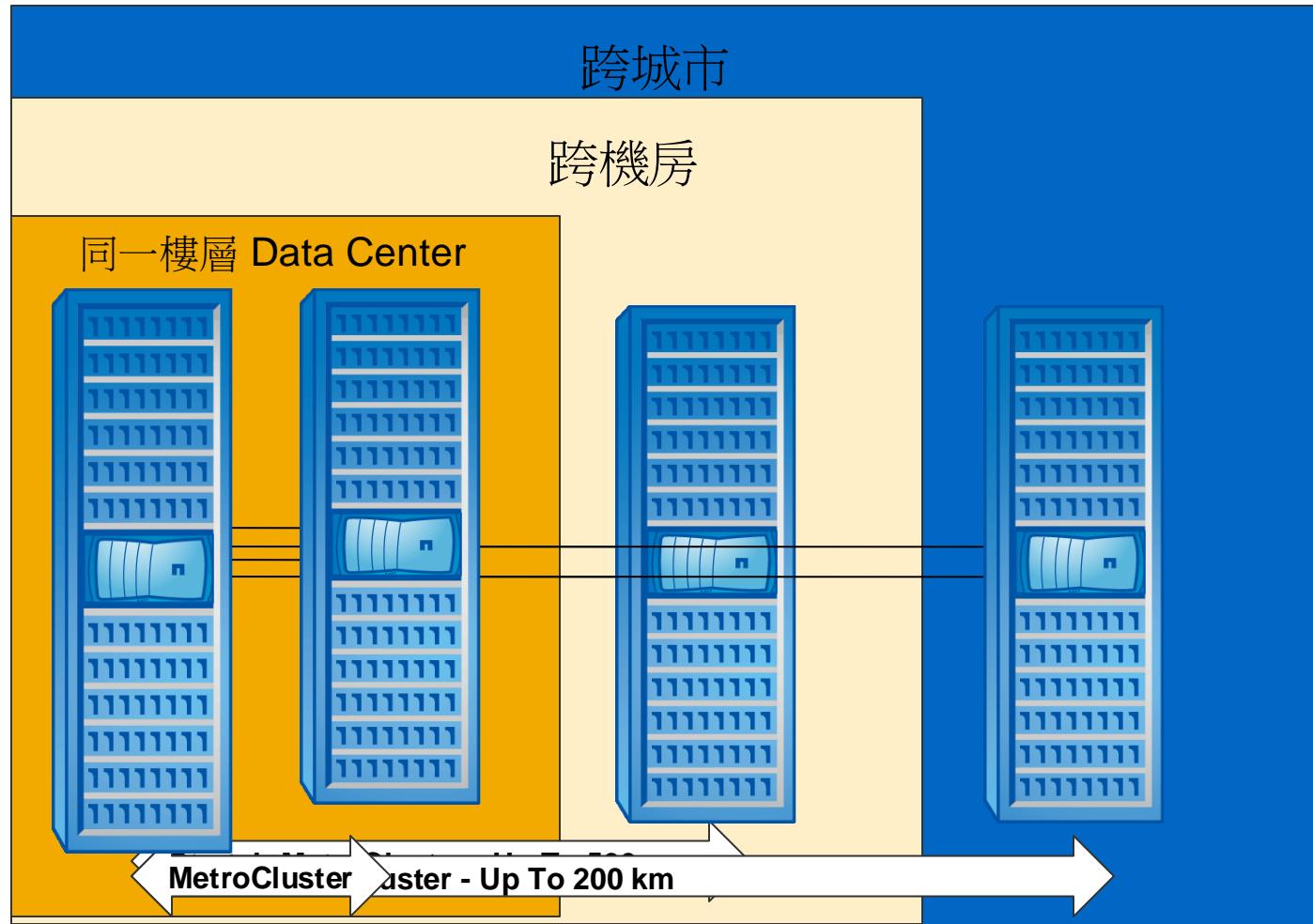
Applications/ databases	 Exchange	 Microsoft SQL Server	 ORACLE DATABASE	 SAP HANA	
Filesystems				 Windows Server	
Hypervisors				 vmware®	
Custom plug-in	 mongoDB®	 SAP MaxDB The SAP Database	 IBM DB2 Sybase	 MySQL®	 PostgreSQL
Storage systems			 ONTAP 9	 ONTAP	 ONTAP

SnapCenter Dashboard – Monitoring workflows from a single screen



資料管理

NetApp 優勢：完整的HA方案 (MetroCluster)



- 無單點故障架構
- 控制器可置於不同機櫃、跨機房甚至不同城市
- 可達到雙資料主中心，並可加上異地備援方案

MetroCluster 可達到完全的無單點故障

IT 理想的一天 - Active IQ 視覺化管理儀錶板

The screenshot shows the NetApp Active IQ Storage Admin dashboard. The left sidebar includes a 'Customized Directory' section for roles and units, a 'Performance' section, and a 'Assets Inventory' section. The main dashboard features a 'Wellness Score' (Average), several performance and availability cards, an 'Inventory' pie chart, a 'Planning' section, an 'Upgrades' section, and a 'Valuable Insights' section. A red bracket on the left indicates the 'Customized Directory' and 'Assets Inventory' sections. Red arrows point from the 'Performance' section to the 'Upgrades' section, from the 'Assets Inventory' section to the 'Valuable Insights' section, and from the 'Valuable Insights' section to the 'Systems Sending AutoSupport' section. Labels on the right side categorize these sections: 'REST APIs 自助服務區' (top right), '升級計劃協助' (Upgrades section), '精準預算評估' (Valuable Insights section), and 'AutoSupport 年度報告' (Systems Sending AutoSupport section).

客製化目錄
(角色、單位等)

執行建議

資產設備一覽
(System 層級連結)

REST APIs
自助服務區

升級計劃協助

精準預算評估

AutoSupport
年度報告

Storage Admin

Wellness Score: Average

Performance & Efficiency: 8

Availability & Protection: 9

Capacity: 11

Configuration: 3

Security: 16

Renewals: 16

Inventory:
Sites: 6 of 6
Clusters: 32 of 36
Systems: 136 of 145

Planning:
Hardware Add-ons: 2
Software Add-ons: 5
Renewals: 23
Refresh: 3

Upgrades (12 Clusters):
Interoperability Summary: 2
Unsupported: 11
Supported: 4
Upgrades: 2
Issues fixed: 86

Valuable Insights:
Insights in the last year:
136 Systems, 197 Security Risks Recognized, 472 Critical Risks Recognized, 963 Total Risks Recognized, 963 Risks Mitigated, 103 Automated Cases.
Systems not under Service Contracts:
4 Systems, 13 Security Risks, 19 Critical Risks, 34 Recognized Risks.

Systems Sending AutoSupport: 136

多

服務協定多,彈性佳
功能最多,防勒索病毒,橫向擴充
雲端地管理工具,趨勢分析
支援&資源最多, AI, K8S, DB, VM, Cloud

穩

台灣市占率第一名
大型客戶使用經驗
橫向擴充或資料移轉不需要停機
可靠度最高 99.99993 IDC認證

快

NVMe效能最快
備份還原最快
上三大雲端最快

省

空間效率保證
冷熱資料分離