Bigstack

Agentic Infrastructure 時代的基石

Cube 虛擬化與 Dell 伺服器驅動 AI 賦能校園服務

劉又豪 Howard Liu

堆疊股份有限公司 資深客戶經理

從虛擬化到自主化的演進旅程

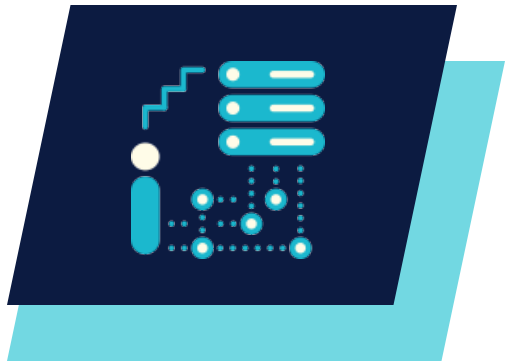
校園 IT 建設正經歷關鍵轉型：從傳統基礎設施邁向智慧自主管理，透過 CubeCOS 虛擬化平台與 Dell 伺服器效能優勢，為教育單位打造新世代數位學習環境。

虛擬化

統一資源管理
建立穩定基礎設施

智慧化

AI 驅動監控
與資源調度優化



自主化

自我優化決策
實現真正智慧校園

校園基礎設施面臨的核心挑戰

資源分散與管理困難

各廠牌硬體、軟體環境差異大，維護成本持續攀升，IT 團隊疲於奔命。

彈性不足

應用需求與使用量變化劇烈，傳統硬體架構無法快速調整資源配置。

安全與穩定性壓力

多用戶、多部門共用環境容易出現資源競爭，系統穩定性難以保證。

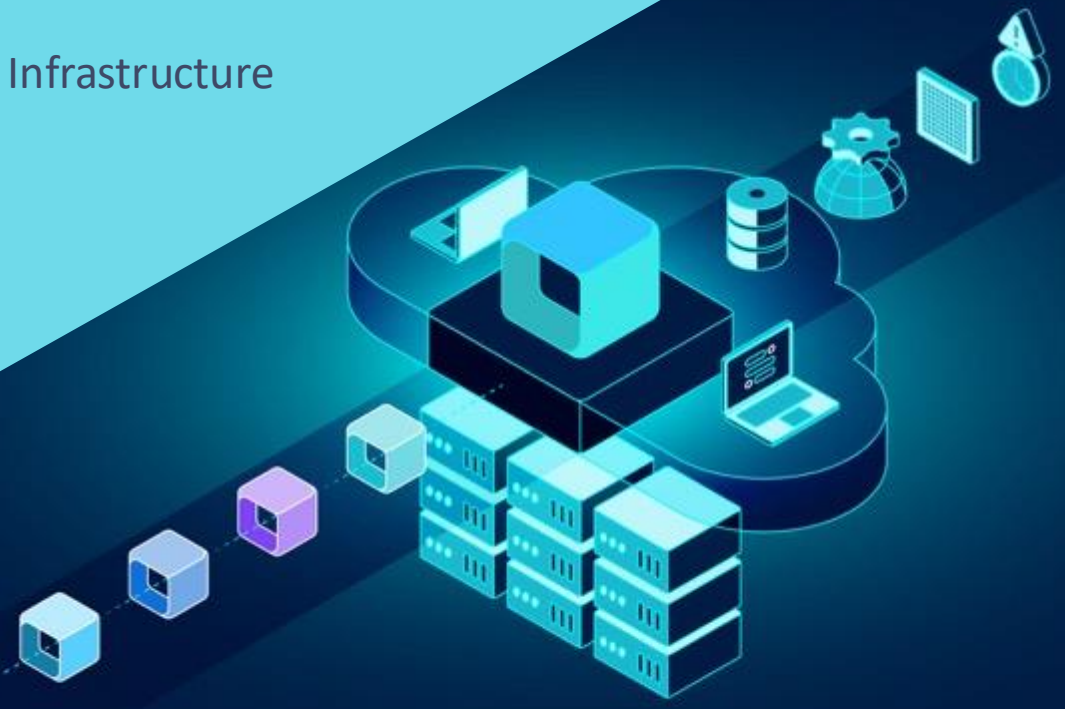
高額運營成本

軟體年度訂閱、硬體更新與維護費用持續增加，造成沉重經費壓力。

Bigstack

CubeCOS / CubeVDI
校園智慧基礎設施的穩固基石

CubeCOS / CubeVDI 作為運行環境底層基礎架構，為上層 Agentive Infrastructure 智慧管控層提供穩定虛擬化資源，實現自動化調度與智慧服務。



cubeCOS

Cube App Framework

Cube Cloud Operating System (COS)

Bigstack HEX bedrock©

Open-Infra Software

直觀好用

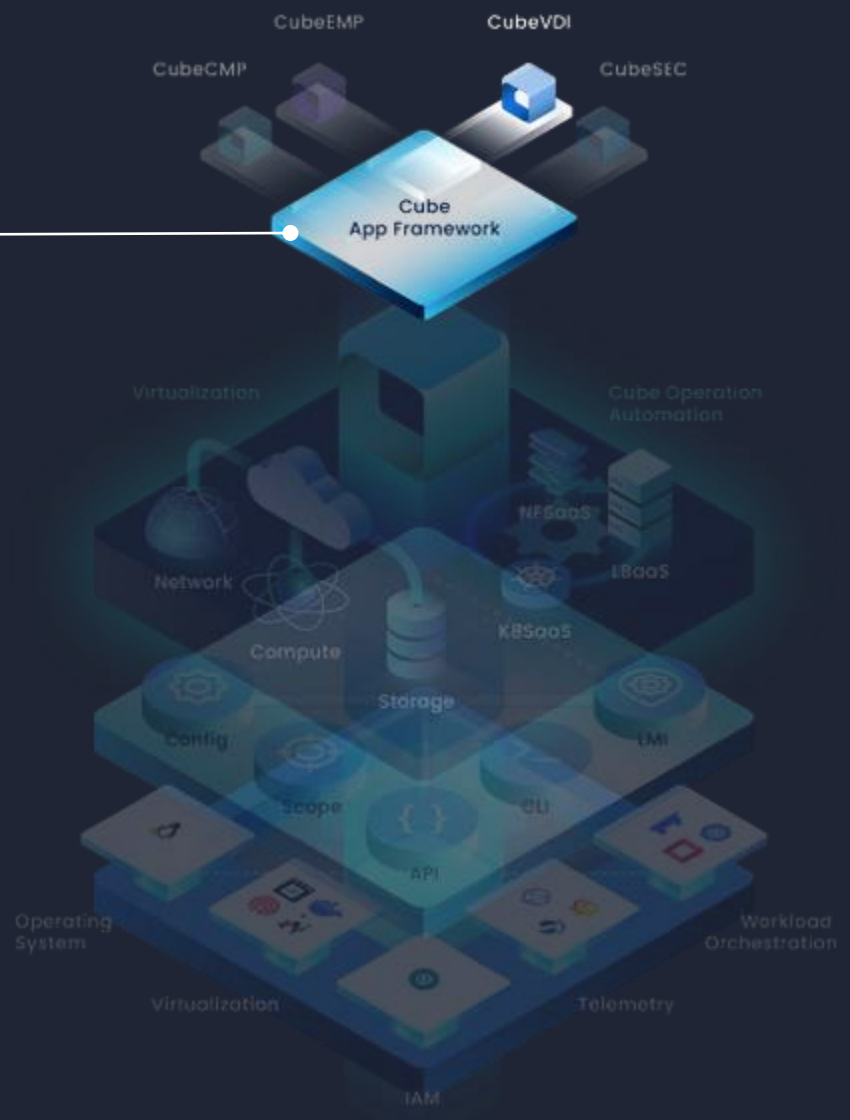
簡單建置

高穩定性

易於維運

高自主性

Cube App Framework



cubeVDI

- 遠端桌面管理
- 資料管理
- 高可用性

三大核心優勢：虛擬、穩定、自主

高穩定的虛擬化平台

提供虛擬運算、虛擬儲存、虛擬網路服務提供100%純軟體定義架構。

彈性部署與資源調度

依需求快速調整計算、存儲、網路資源分配與架構設計，提升系統靈活性。

集中管理與簡化維護

系統更新、資源監控與使用者管理集中化，大幅降低維護成本與管理複雜度。

高穩定虛擬平台的優勢

自動化與即服務

Automation and XaaS

網路

軟體定義網路 (SDN)

存儲

軟體定義存儲 (SDS)

運算

虛擬機與容器服務

GPU

虛擬機與容器調度GPU資源

監控

確保全方面的基礎設施，
保持性能和運營需求。

自動化 – Automation and XaaS

透過統一的入口網站輕鬆進行資源配置和管理簡化超融合平台

K8SaaS

- 開發者友善
- Rancher多版本支援

LBaaS

- L4負載平衡
- API支持基礎架構即代碼部署

NFSaaS

- NFS掛載使數據與操作系統分離
- 多租戶架構保持數據訪問隔離

DNSaaS

- 多租戶DNS支援
- RESTful API自動化

VPNaaS

- 跨租戶、遠端訪問、站點對站點
- 租戶和叢集互聯

VDIaaS

- 增強安全性
- 與現有系統的整合

GPU – AI計算平台

優化加速器使用，支援 AI 應用、機器學習與高效能虛擬化

GPU 加速與資源配置

- 支援虛擬 GPU 與實體直通
- 適用於 AI / 深度學習 / 圖形運算

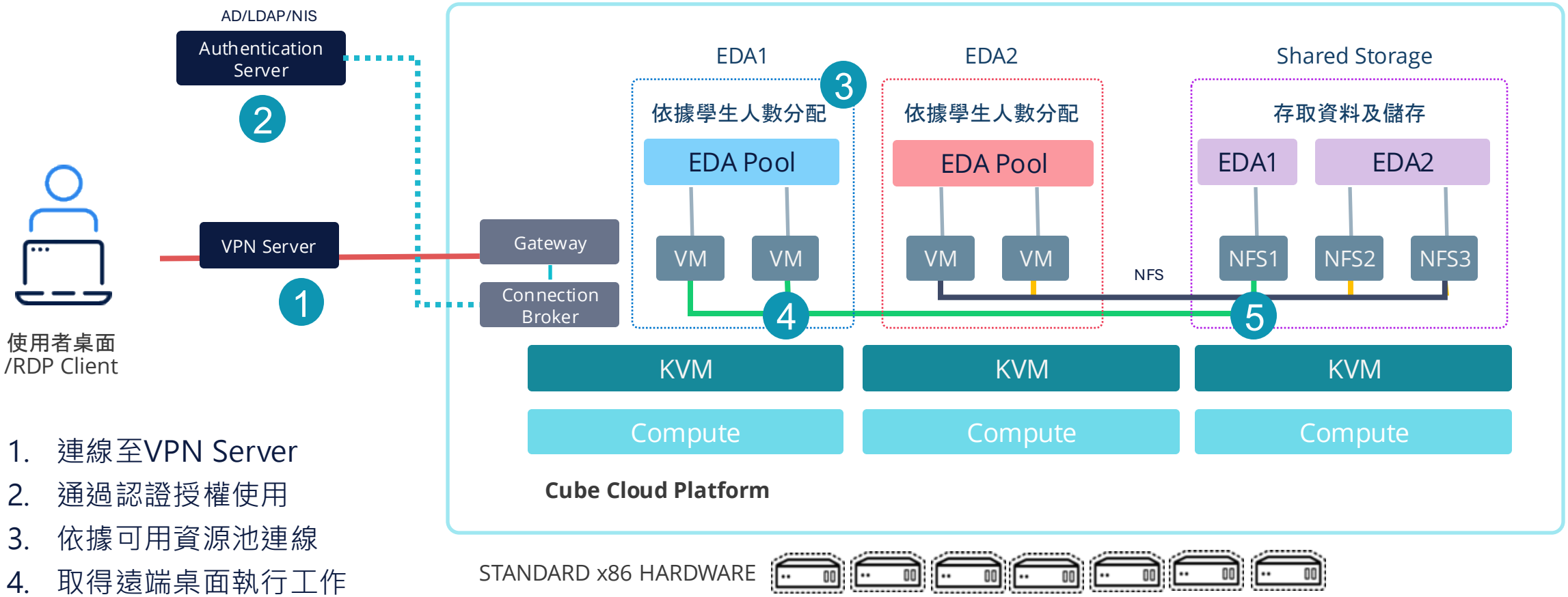
異質運算與硬體加速

- NUMA 與記憶體配置優化
- 異質資源混合部署（CPU / GPU）

統一 AI 平台與運行環境

- 相容 TensorFlow、PyTorch 等主流框架
- 容器化與 Kubernetes 加速執行

cubeVDI 遠端桌面應用 – 教學與研究



1. 連線至VPN Server
2. 通過認證授權使用
3. 依據可用資源池連線
4. 取得遠端桌面執行工作
5. 大型資料存放共享區域

三大核心優勢：虛擬、穩定、自主

高穩定的虛擬化平台

提供虛擬運算、虛擬儲存、虛擬網路服務提供100%純軟體定義架構。

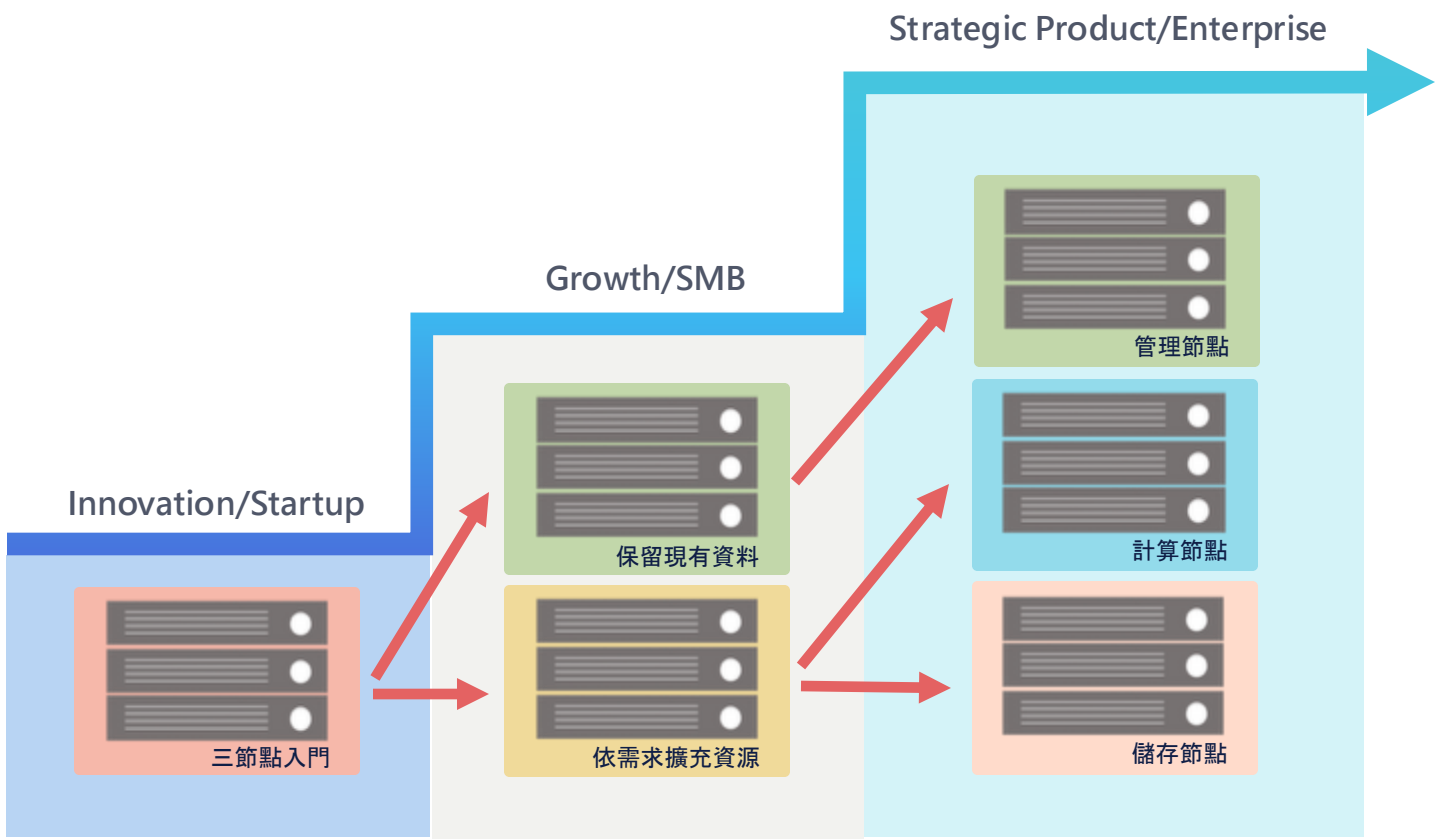
彈性部署與資源調度

依需求快速調整計算、存儲、網路資源分配與架構設計，提升系統靈活性。

集中管理與簡化維護

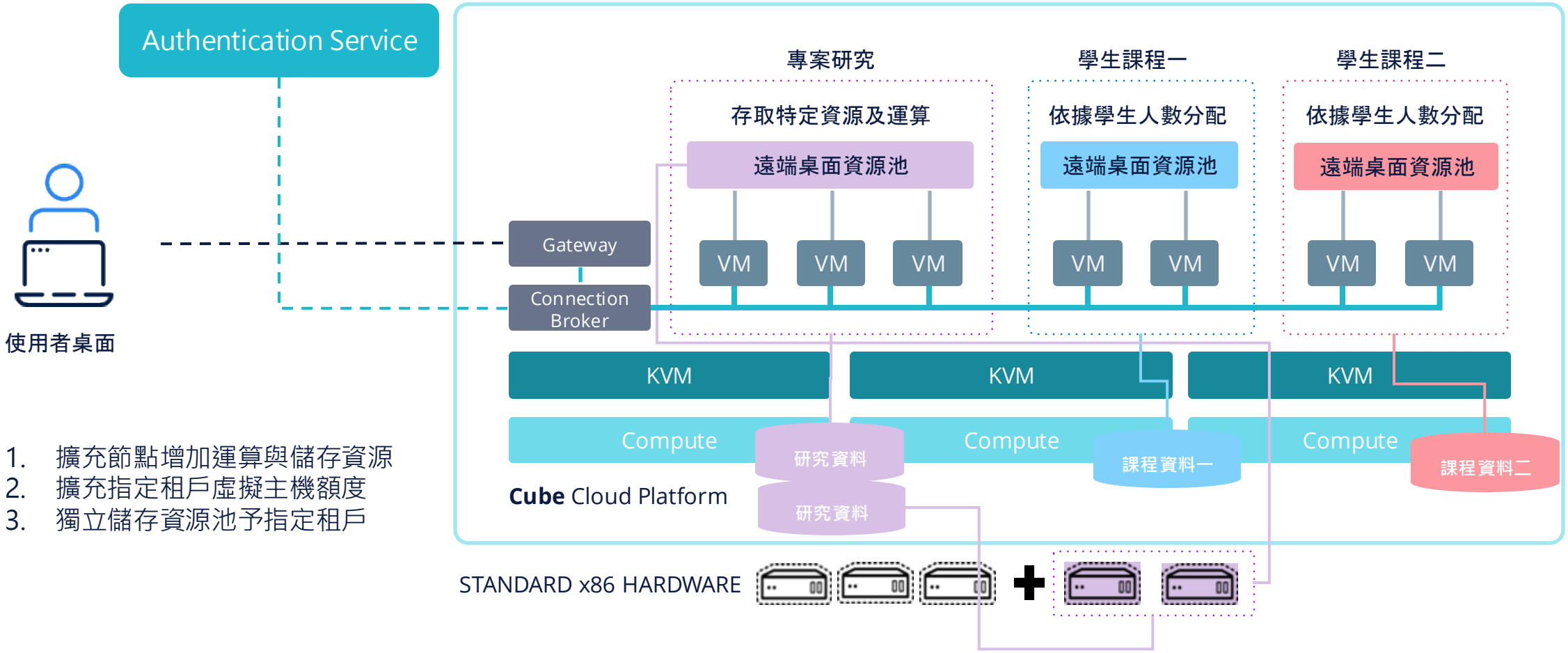
系統更新、資源監控與使用者管理集中化，大幅降低維護成本與管理複雜度。

Pay as You Grow 簡單且彈性的擴充策略



- ✓ 快速安裝
- ✓ 一鍵升級 (叢集滾動. 雙磁區.)
- ✓ 角色分拆提高可靠度
- ✓ 按照需求資源擴充
- ✓ 活化現有IT設備
- ✓ 可搭配任何廠牌 X86 伺服器

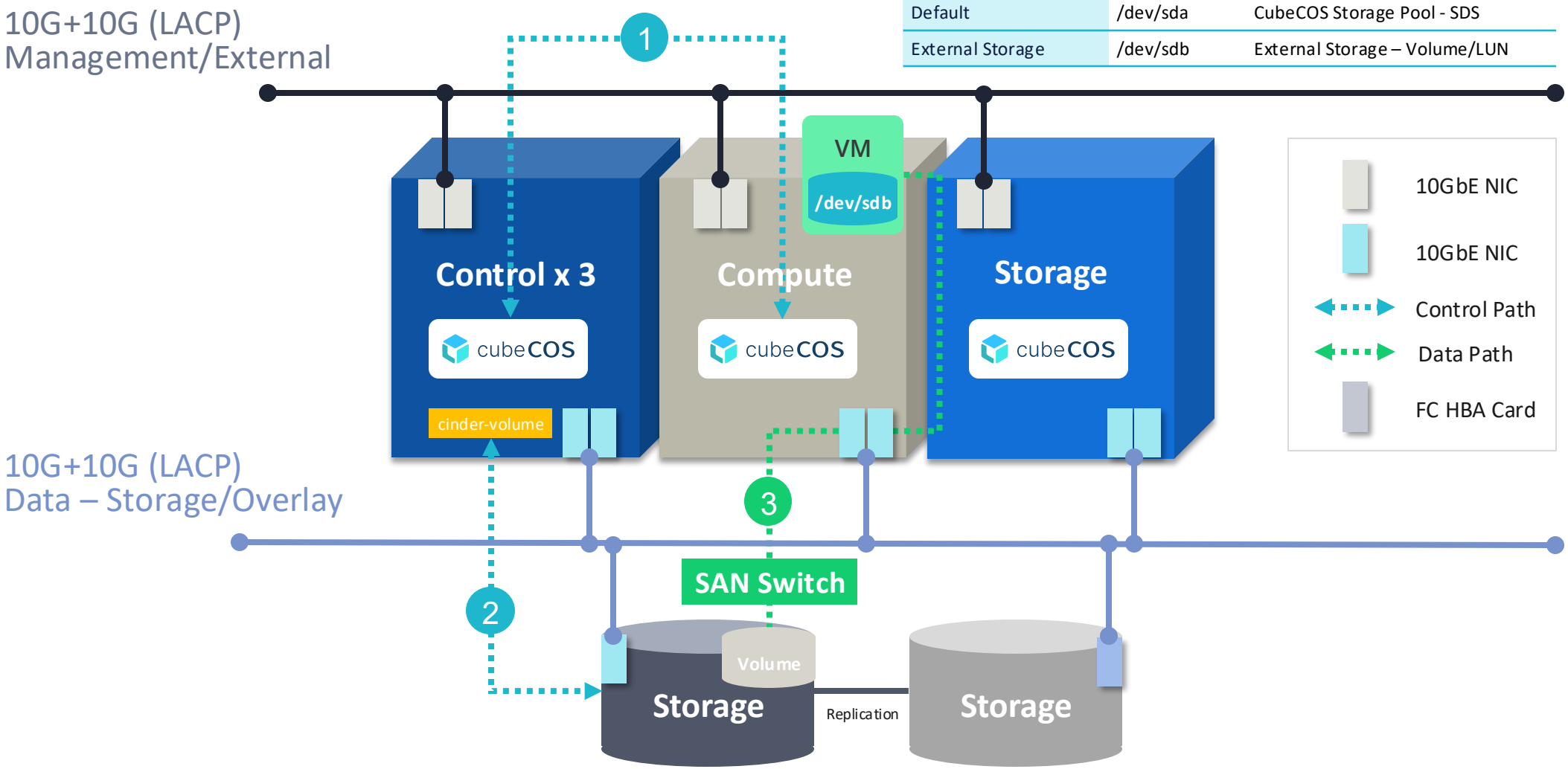
資源管理與獨立擴充性



1. 擴充節點增加運算與儲存資源
2. 擴充指定租戶虛擬主機額度
3. 獨立儲存資源池予指定租戶

External Storage with iSCSI/NFS

VM Storage Type	Device	Data Source
Default	/dev/sda	CubeCOS Storage Pool - SDS
External Storage	/dev/sdb	External Storage – Volume/LUN



三大核心優勢：虛擬、穩定、自主

高穩定的虛擬化平台

提供虛擬運算、虛擬儲存、虛擬網路服務提供100%純軟體定義架構。

彈性部署與資源調度

依需求快速調整計算、存儲、網路資源分配與架構設計，提升系統靈活性。

集中管理與簡化維護

系統更新、資源監控與使用者管理集中化，大幅降低維護成本與管理複雜度。

cubeCOS 儀表板與服務總覽

服務總覽

最新事件、登出

資料中心健康檢視

資料中心與管理

版本資訊

首頁/儀表板

節點資訊與監控

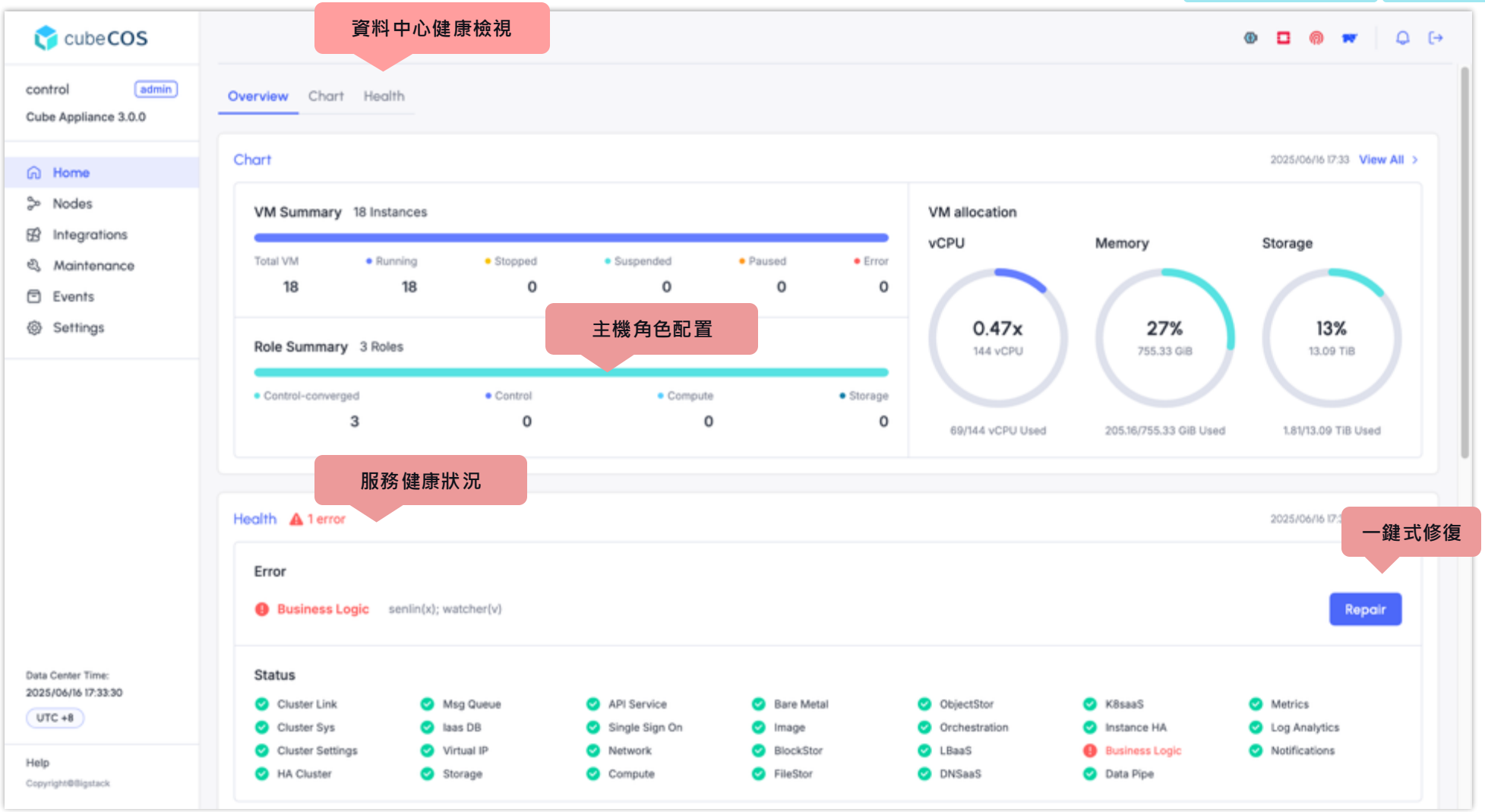
服務與整合

維運與管理

事件分析與告警

系統設定管理設定

技術支援

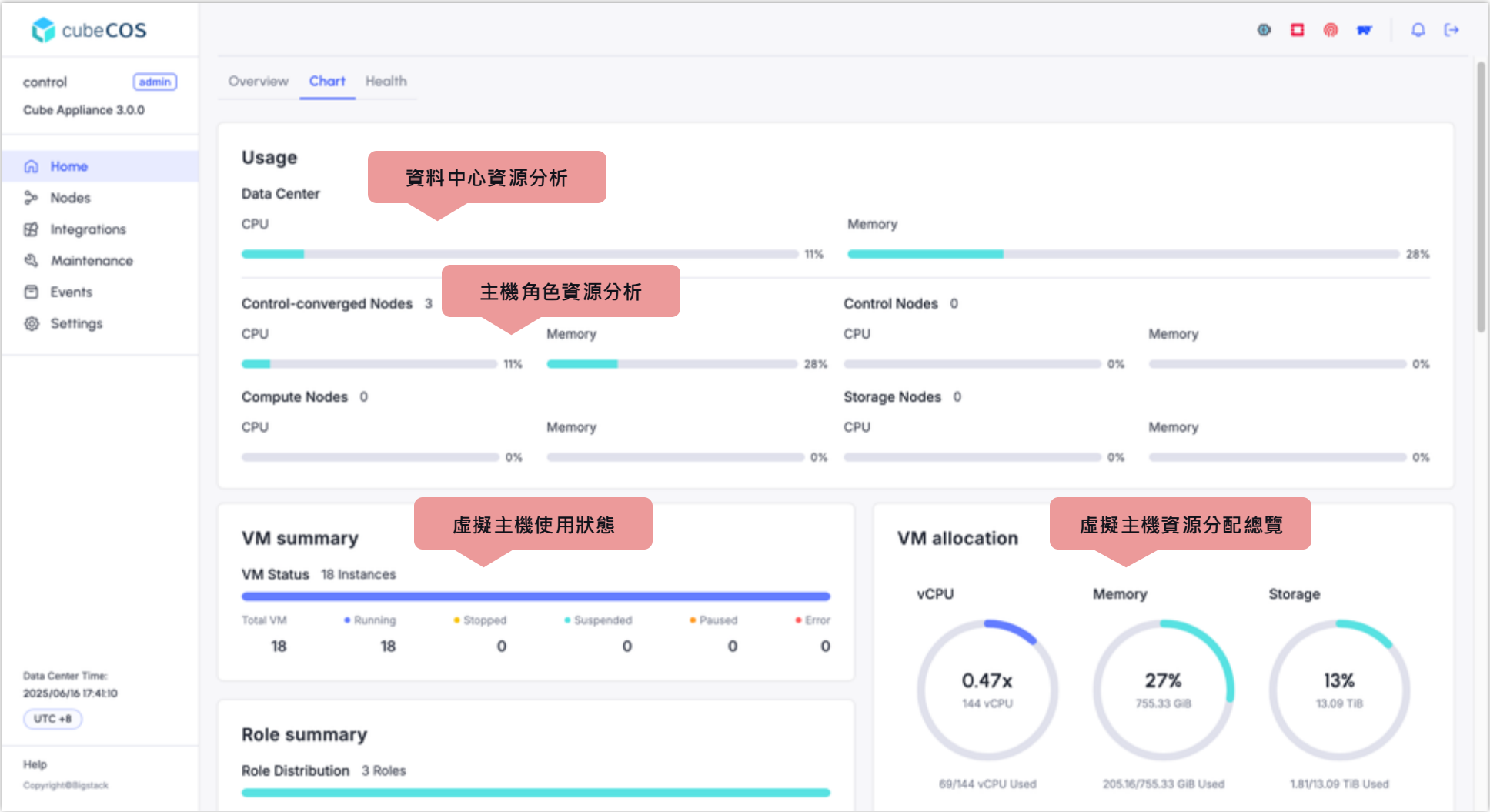


主機角色配置

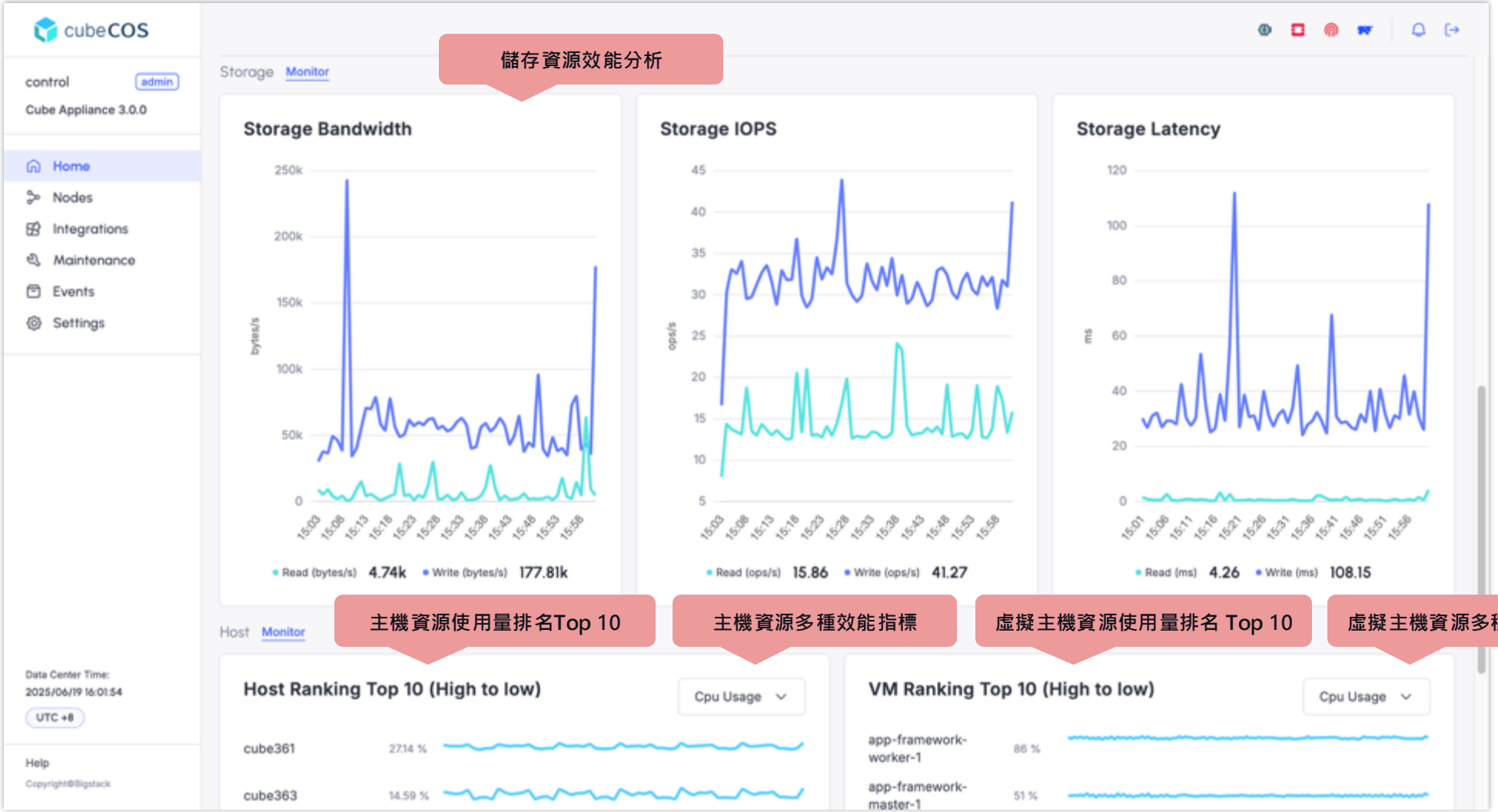
服務健康狀況

一鍵式修復

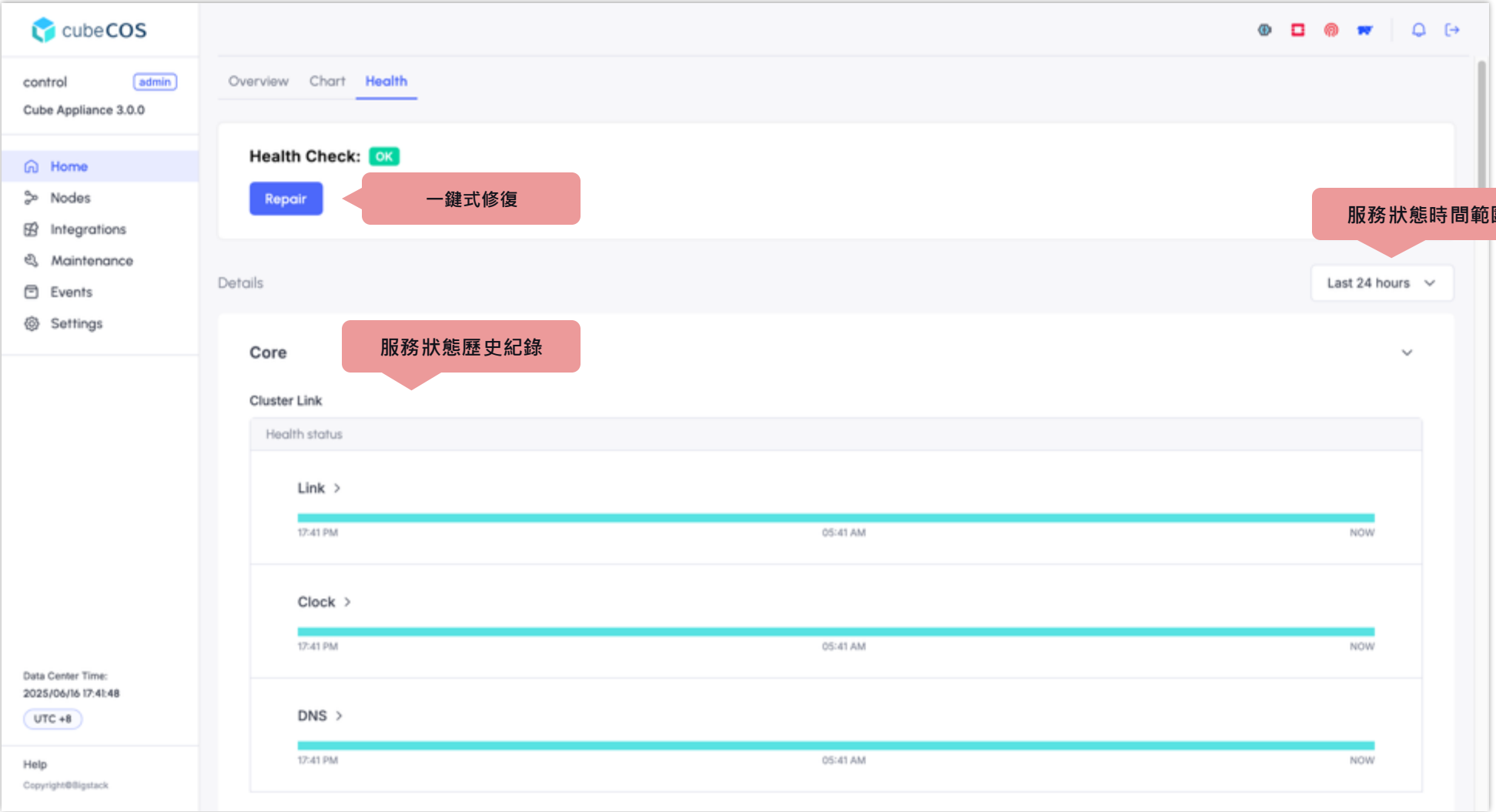
cubeCOS 資料中心與虛擬主機資源



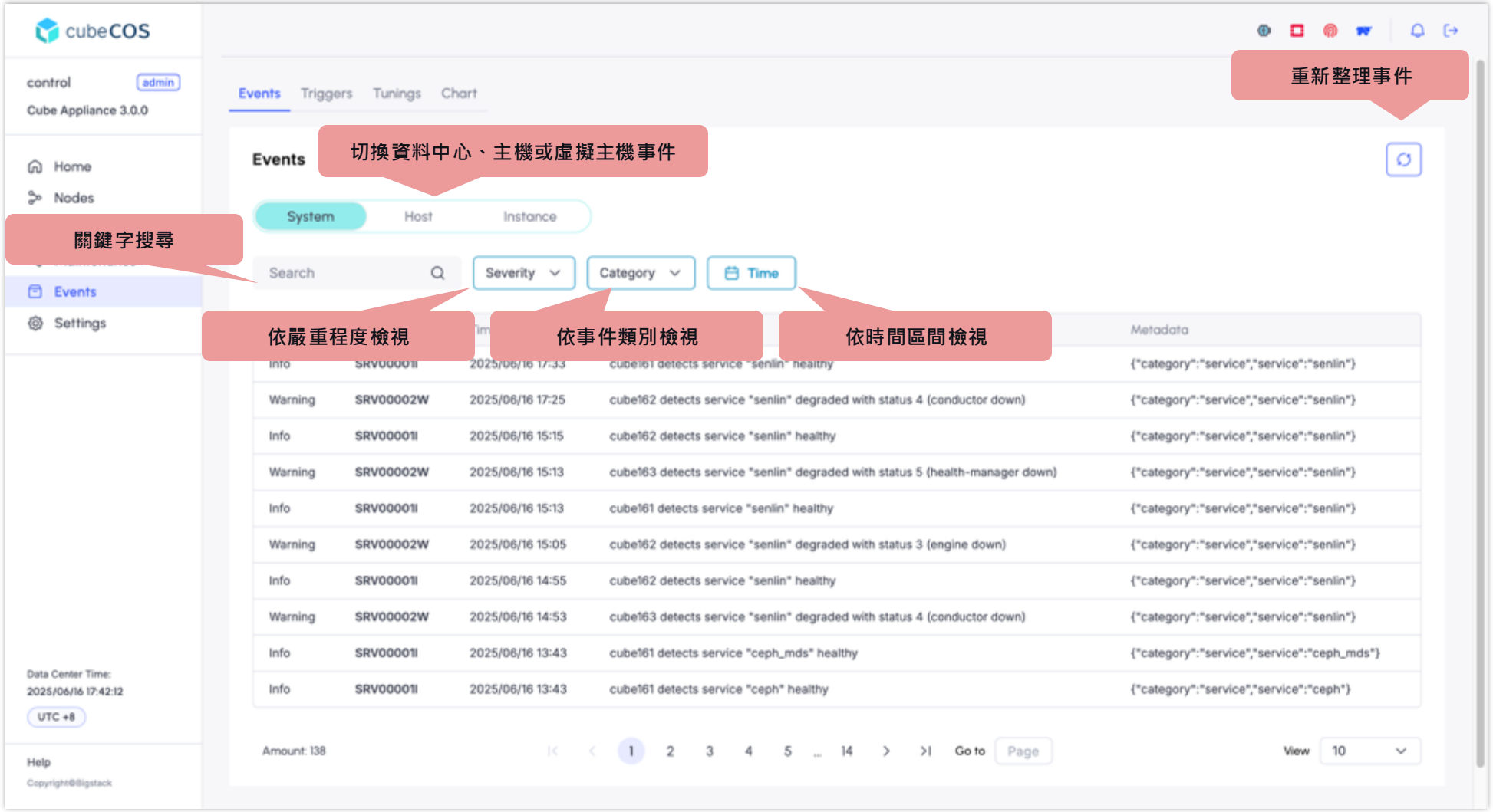
cubeCOS 效能分析與使用量排名



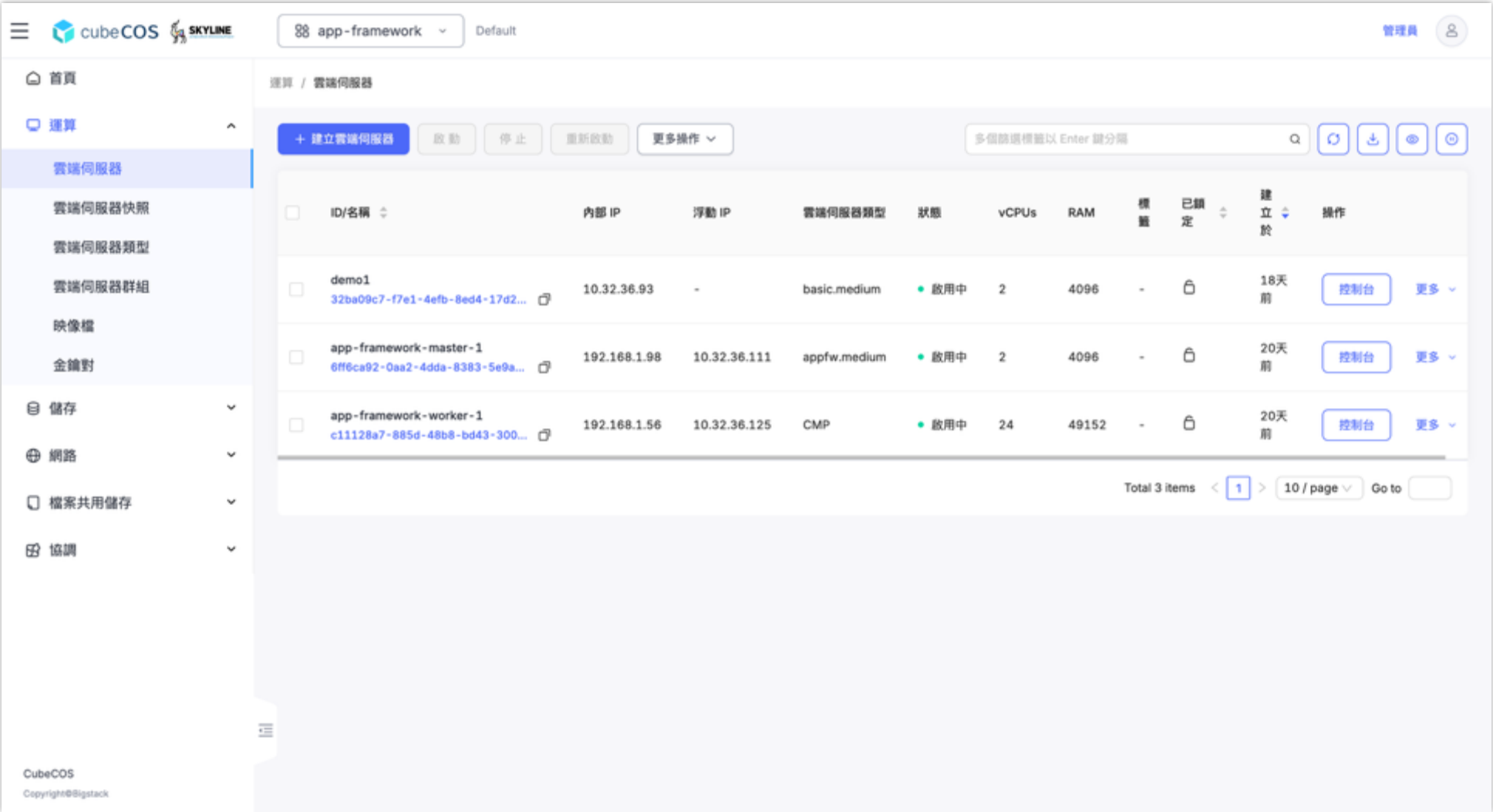
cubeCOS 資料中心服務紀錄與一鍵式修復



cubeCOS 事件紀錄檢視



cubeCOS 中文介面管理



三大核心優勢：虛擬、穩定、自主

高穩定的虛擬化平台

提供虛擬運算、虛擬儲存、虛擬網路服務提供100%純軟體定義架構。

彈性部署與資源調度

依需求快速調整計算、存儲、網路資源分配與架構設計，提升系統靈活性。

集中管理與簡化維護

系統更新、資源監控與使用者管理集中化，大幅降低維護成本與管理複雜度。

自主權與開源優勢

校園可完全掌握虛擬平台自主權，依自身需求安裝或擴展功能，降低對外部廠商的長期依賴。

採用開源架構

增加系統透明度與可控性

避免廠商鎖定

支持長期可持續運營

彈性擴充能力

靈活應對政策與技術變化需求

Get Start to Use

開始使用 CubeCOS

- GitHub 下載連結
- CubeCOS 安裝手冊
- 安裝 CubeCOS 要求之地端規格

使用指引

- Day 01
- Day 02
- Day 03

加入 Bigstack 社群平台

Slack

Discord

Github

YouTube

LinkedIn

與 Bigstack 攜手合作

專屬顧問服務：

規劃架構藍圖，量身設計最佳解決方案。

- Day 0：設計規劃
- Day 1：產品交付、安裝建置
- Day 2：訓練課程、持續營運

Bigstack

官方網站

cubeCOS

使用指引

備份與搬遷 - Trilio

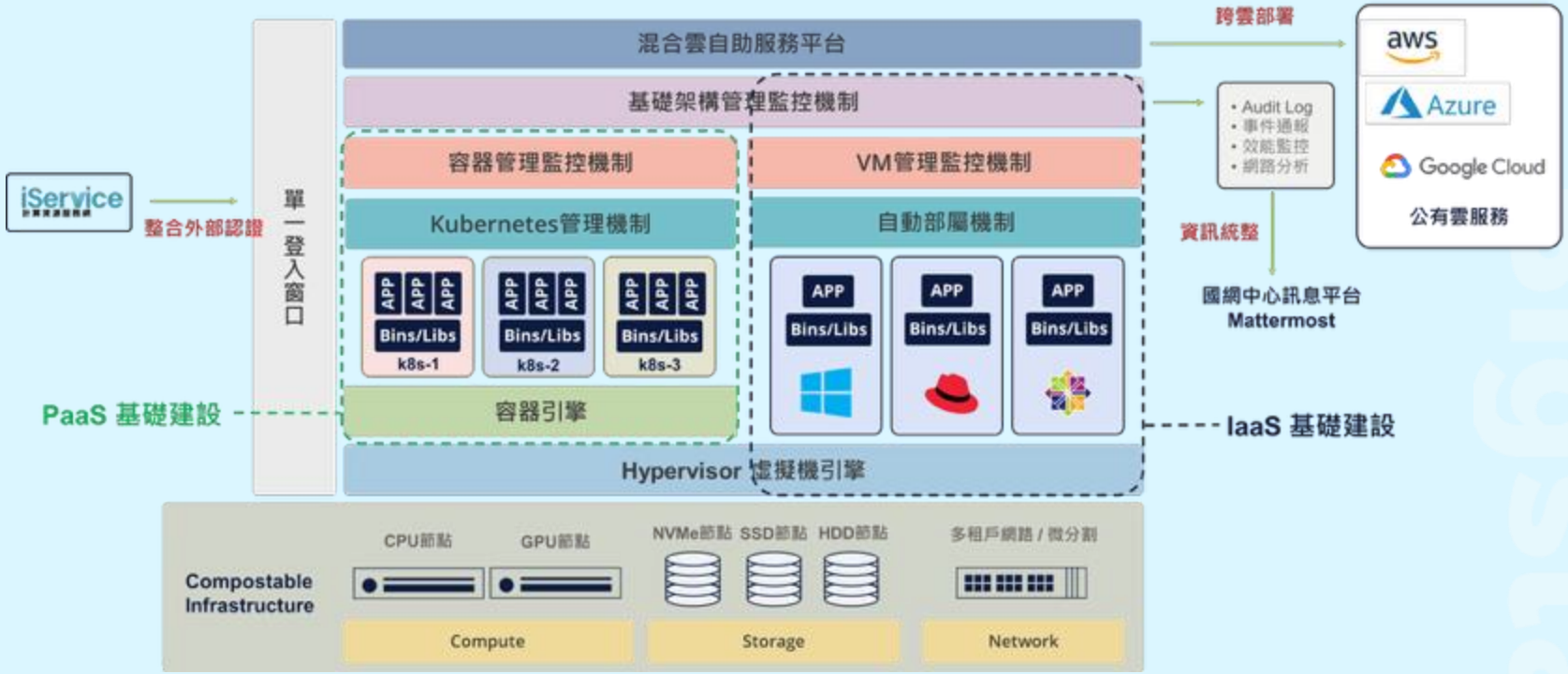
cubeCOS × TRILIO

- 與 CubeCOS 快速整合
- 滿足政府「備份 / 災難復原 / 搬遷」需求
- 大幅縮短開發時間，快速回應市場

國家A級資料中心
整合資源與彈性擴充

使用特色

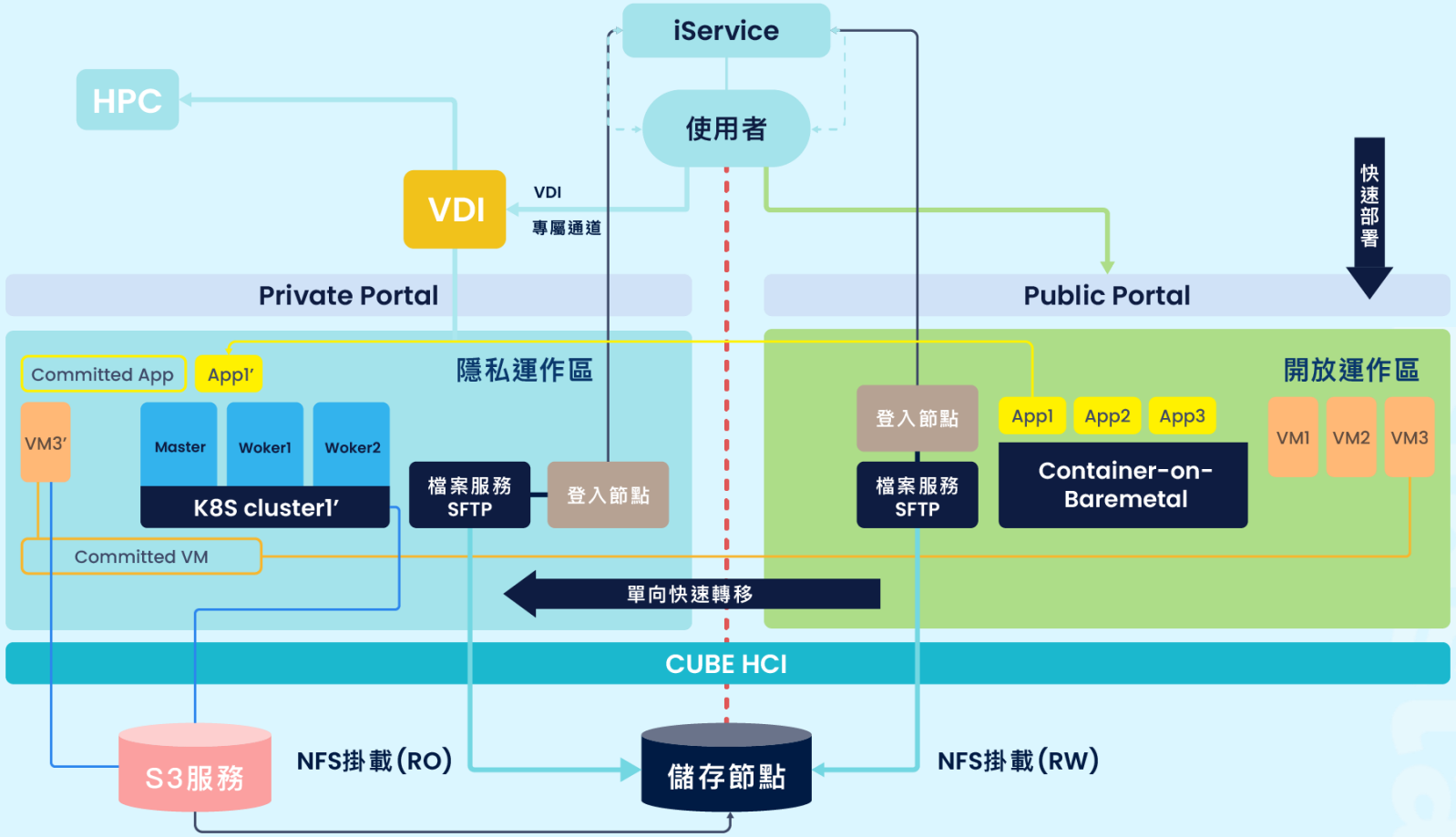
- 打通資料流通：透過統一管理層（如 App-framework）讓資料共享更順暢。
- 整合多租戶環境：提供**租戶**間資料安全**隔離**，同時支援跨系統協同處理。
- 標準化基礎架構：**整合** x86 **硬體**與**虛擬化**，降低成本並提升效率。



國家A級資料中心
實現資料分級管理與高效利用

使用特色

- 將資料區分為**機敏區**與**非機敏區**，滿足內部及外部的合規性要求。
- 提供**統一的資料訪問接口**，科學家可方便地使用機敏區或非機敏區的資料進行操作。
- 符合資料**合規規範**，同時支持**高效的資源調度**，滿足科研與業務需求。



半導體電子設計自動化教學與研究平台

主要競爭對手

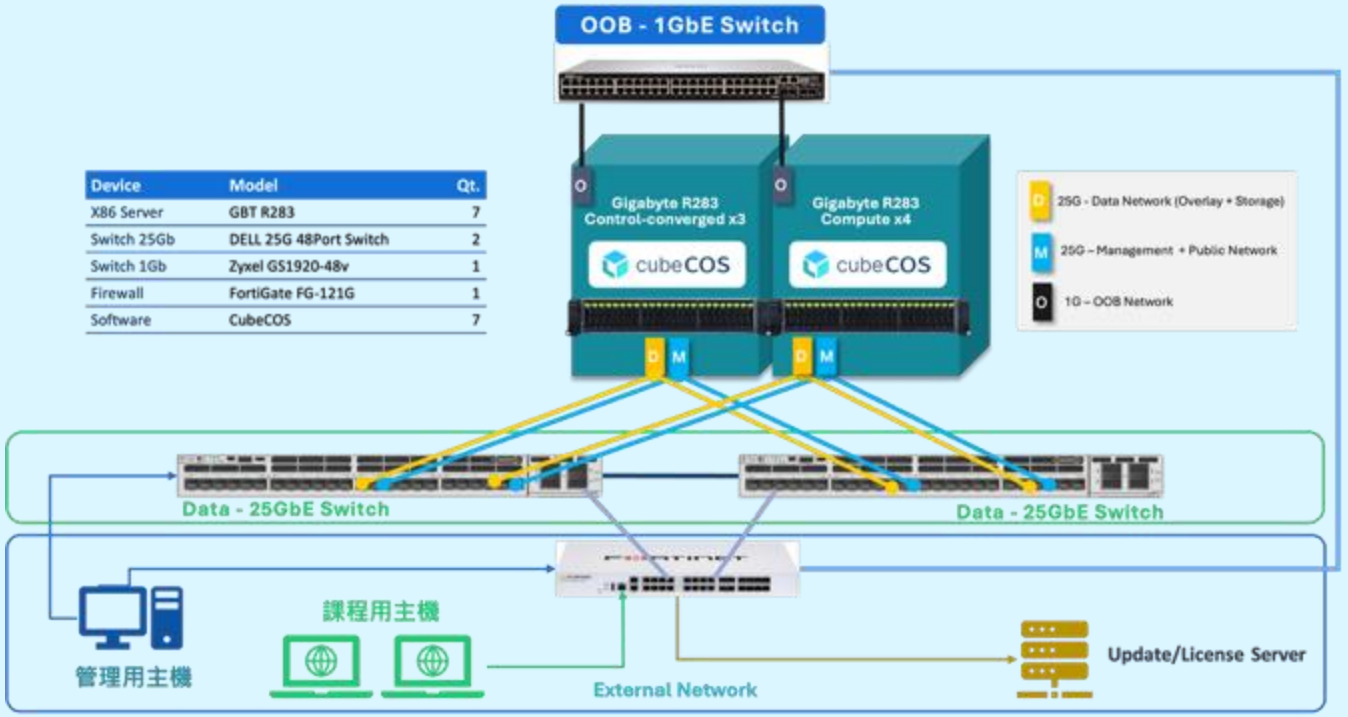
Bigstack

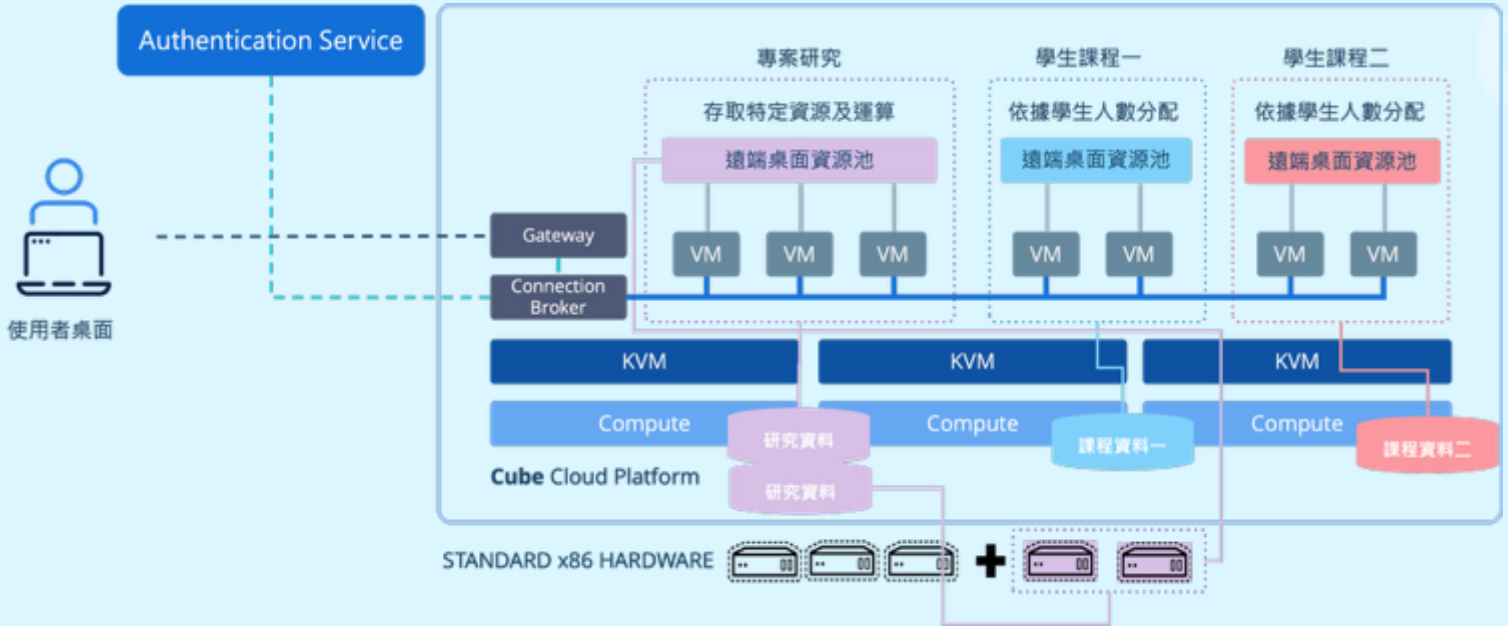
vmware

致勝關鍵

- 確保資源適當分配使用與管理
- 可承擔的建置維運成本
- 結合cubeVDI的加值應用
- 彈性擴充運算或儲存資源

Device	Model	Qt.
X86 Server	GBT R283	7
Switch 25Gb	DELL 25G 48Port Switch	2
Switch 1Gb	Zyxel GS1920-48v	1
Firewall	FortiGate FG-121G	1
Software	CubeCOS	7

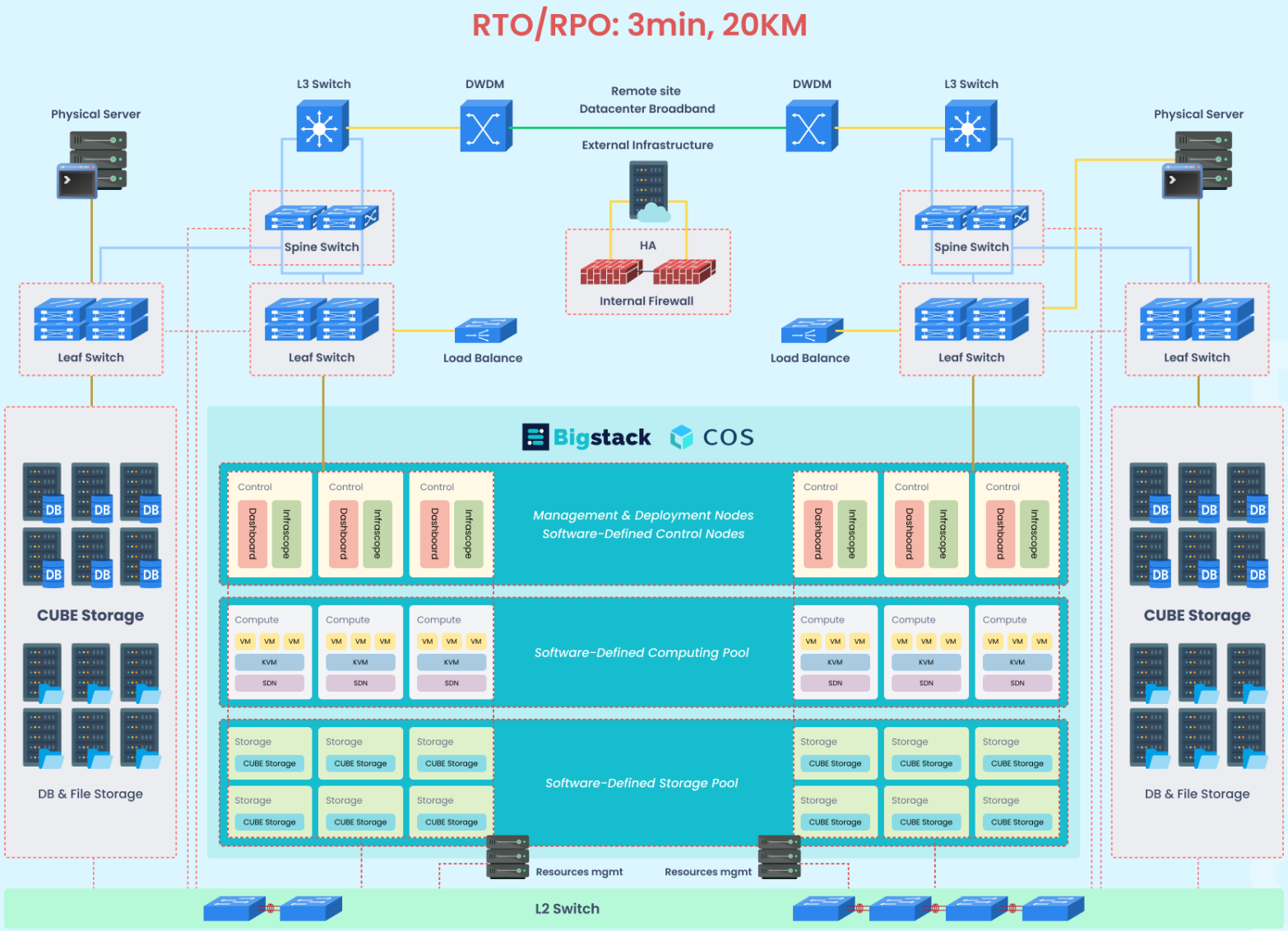




雙活架構與快速 AI 部署：輿情大資料分析雲平台

使用特色

- 雙活資料中心架構：利用產品的高可用性特性，實現**雙活資料中心**設計，提供業務不中斷的資料容災能力。
- 快速 AI 服務部署：提供可快速佈署的 AI 服務，支持靈活的資料分析和**即時決策**。
- 資源池化與靈活擴展：整合計算、存儲和網絡資源，構建**軟體定義**的計算與存儲池。



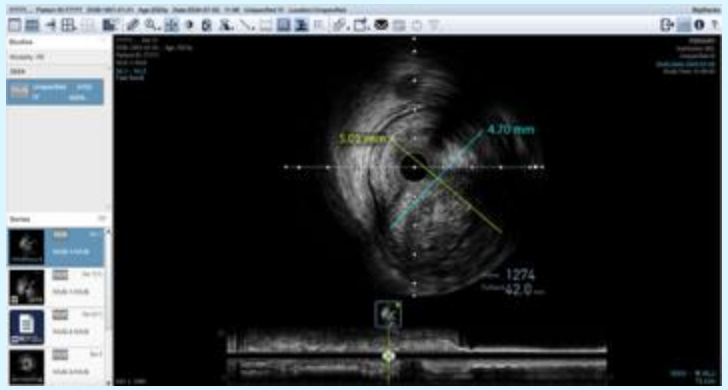
醫學中心PACS醫療影像系統虛擬化超融合雲平台

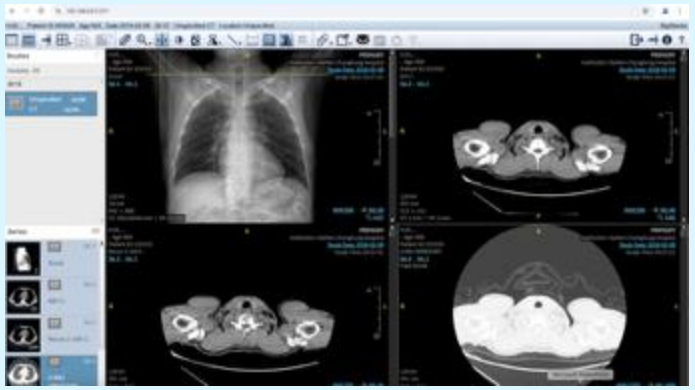
主要競爭對手

Bigstack　vmware

致勝關鍵

- 可承擔的建置維運成本
- 醫療影像系統完整驗證
- 便於技術轉移的維運介面
- 現有設備整合提高設備利用率





Dell 伺服器：
效能與穩定的保證

從基礎架構到 HPC 與 GPU 加速

Dell 伺服器提供完整效能階梯，滿足從一般教學到高效能運算的多元需求

確保穩定性與擴充性

企業級可靠度設計，支持校園持續成長與未來擴展需求

最佳化能源管理

協助校園降低總體擁有成本 (TCO)，實現綠色節能目標

DELL Technologies/Forum

驅動AI新時代：AI運算的標準與基石

2025 年 Dell Technologies Forum

黃敏俊(Moca Huang)

資深技術顧問
戴爾科技集團

30 YEARS PowerEdge EST. 1994

議程

1. 適合企業使用的 Dell AI Factory
2. Dell AI Factory 實際運用
3. Dell PowerEdge XE 產品組合
4. Dell PowerEdge XE7740 和 XE7745
5. Dell PowerEdge XE9780 和 XE9785 簡介
6. Q&A

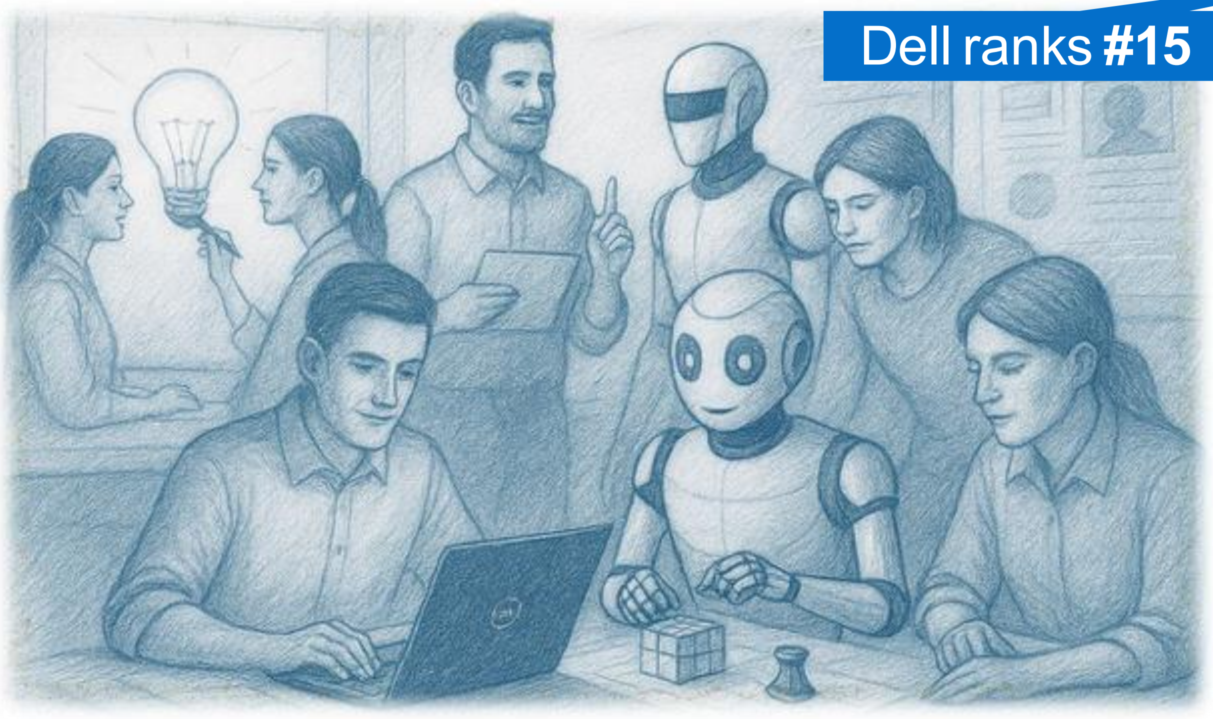
技術創新

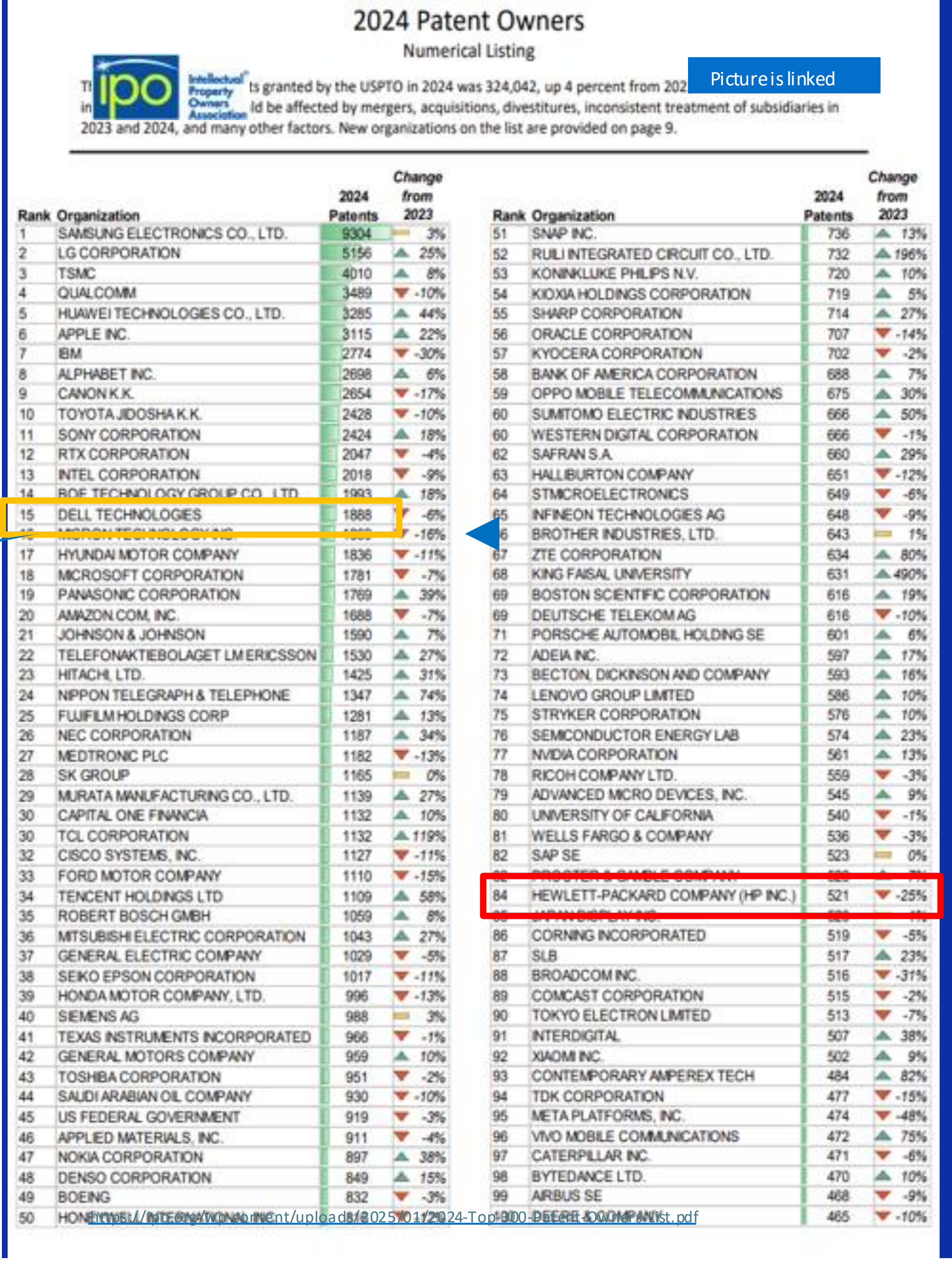
Dell 是伺服器創新領域的領導者

2024年獲得最多美國實用專利的前300個組織, Dell有1,888件專利

168% more than Cisco #32	1127patents
322% more than Lenovo #74	586 patents
435% more than HPE #105	434 patents
✗ Supermicro is not on the list	0 patents

Dell ranks #15





Picture is linked

2024 Patent Owners

Numerical Listing

Th[illegible] ts granted by the USPTO in 2024 was 324,042, up 4 percent from 202[illegible] in [illegible] ld be affected by mergers, acquisitions, divestitures, inconsistent treatment of subsidiaries in 2023 and 2024, and many other factors. New organizations on the list are provided on page 9.

Rank	Organization	2024 Patents	Change from 2023	Rank	Organization	2024 Patents	Change from 2023
1	SAMSUNG ELECTRONICS CO., LTD.	9304	3%	51	SNAP INC.	736	13%
2	LG CORPORATION	5156	25%	52	RULI INTEGRATED CIRCUIT CO., LTD.	732	196%
3	TSMC	4010	8%	53	KONINKLIJKE PHILIPS N.V.	720	10%
4	QUALCOMM	3489	-10%	54	KIOXIA HOLDINGS CORPORATION	719	5%
5	HUAWEI TECHNOLOGIES CO., LTD.	3285	44%	55	SHARP CORPORATION	714	27%
6	APPLE INC.	3115	22%	56	ORACLE CORPORATION	707	-14%
7	IBM	2774	-30%	57	KYOCERA CORPORATION	702	-2%
8	ALPHABET INC.	2698	6%	58	BANK OF AMERICA CORPORATION	688	7%
9	CANON K.K.	2654	-17%	59	OPPO MOBILE TELECOMMUNICATIONS	675	30%
10	TOYOTA JIDOSHA K.K.	2428	-10%	60	SUMITOMO ELECTRIC INDUSTRIES	666	50%
11	SONY CORPORATION	2424	18%	60	WESTERN DIGITAL CORPORATION	666	-1%
12	RTX CORPORATION	2047	-4%	62	SAFRAN S.A.	660	29%
13	INTEL CORPORATION	2018	-9%	63	HALLIBURTON COMPANY	651	-12%
14	BOE TECHNOLOGY GROUP CO., LTD.	1993	18%	64	STMICROELECTRONICS	649	-6%
15	DELL TECHNOLOGIES	1888	-6%	65	INFINEON TECHNOLOGIES AG	648	-9%
[illegible]	[illegible]	[illegible]	-16%	66	BROTHER INDUSTRIES, LTD.	643	1%
17	HYUNDAI MOTOR COMPANY	1836	-11%	67	ZTE CORPORATION	634	80%
18	MICROSOFT CORPORATION	1781	-7%	68	KING FAISAL UNIVERSITY	631	490%
19	PANASONIC CORPORATION	1769	39%	69	BOSTON SCIENTIFIC CORPORATION	616	19%
20	AMAZON.COM, INC.	1688	-7%	69	DEUTSCHE TELEKOM AG	616	-10%
21	JOHNSON & JOHNSON	1590	7%	71	PORSCHE AUTOMOBIL HOLDING SE	601	6%
22	TELEFONAKTIEBOLAGET LM ERICSSON	1530	27%	72	ADEIA INC.	597	17%
23	HITACHI, LTD.	1425	31%	73	BECTON, DICKINSON AND COMPANY	593	16%
24	NIPPON TELEGRAPH & TELEPHONE	1347	74%	74	LENOVO GROUP LIMITED	586	10%
25	FUJIFILM HOLDINGS CORP	1281	13%	75	STRYKER CORPORATION	576	10%
26	NEC CORPORATION	1187	34%	76	SEMICONDUCTOR ENERGY LAB	574	23%
27	MEDTRONIC PLC	1182	-13%	77	NVIDIA CORPORATION	561	13%
28	SK GROUP	1165	0%	78	RICOH COMPANY LTD.	559	-3%
29	MURATA MANUFACTURING CO., LTD.	1139	27%	79	ADVANCED MICRO DEVICES, INC.	545	9%
30	CAPITAL ONE FINANCIA	1132	10%	80	UNIVERSITY OF CALIFORNIA	540	-1%
30	TCL CORPORATION	1132	119%	81	WELLS FARGO & COMPANY	536	-3%
32	CISCO SYSTEMS, INC.	1127	-11%	82	SAP SE	523	0%
33	FORD MOTOR COMPANY	1110	-15%	[illegible]	[illegible]	[illegible]	[illegible]
34	TENCENT HOLDINGS LTD	1109	58%	84	HEWLETT-PACKARD COMPANY (HP INC.)	521	-25%
35	ROBERT BOSCH GMBH	1059	8%	[illegible]	[illegible]	[illegible]	[illegible]
36	MITSUBISHI ELECTRIC CORPORATION	1043	27%	86	CORNING INCORPORATED	519	-5%
37	GENERAL ELECTRIC COMPANY	1029	-5%	87	SLB	517	23%
38	SEIKO EPSON CORPORATION	1017	-11%	88	BROADCOM INC.	516	-31%
39	HONDA MOTOR COMPANY, LTD.	996	-13%	89	COMCAST CORPORATION	515	-2%
40	SIEMENS AG	988	3%	90	TOKYO ELECTRON LIMITED	513	-7%
41	TEXAS INSTRUMENTS INCORPORATED	966	-1%	91	INTERDIGITAL	507	38%
42	GENERAL MOTORS COMPANY	959	10%	92	XIAOMI INC.	502	9%
43	TOSHIBA CORPORATION	951	-2%	93	CONTEMPORARY AMPEREX TECH	484	82%
44	SAUDI ARABIAN OIL COMPANY	930	-10%	94	TDK CORPORATION	477	-15%
45	US FEDERAL GOVERNMENT	919	-3%	95	META PLATFORMS, INC.	474	-48%
46	APPLIED MATERIALS, INC.	911	-4%	96	VIVO MOBILE COMMUNICATIONS	472	75%
47	NOKIA CORPORATION	897	38%	97	CATERPILLAR INC.	471	-6%
48	DENSO CORPORATION	849	15%	98	BYTEDANCE LTD.	470	10%
49	BOEING	832	-3%	99	AIRBUS SE	468	-9%
50	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	465	-10%

https://ipo.org/wp-content/uploads/2025/01/2024-Top-300-Patent-Owners-List.pdf

DELL 17G 伺服器已上市

- 因應新世代 CPU 而生
  - Intel XEON 6 (E core / P Core)
  - AMD EPYC™ 9005
- 採用新的技術設計
  - DDR5 6400
  - CXL Memory
  - PCIe Gen 5
  - EDSFF E3.S
  - Smart NICs
  - M-CRPS Power supplies

PowerEdge Server Portfolio 17G & workload (非CTC勾選範圍)

17G Briefing

CPU	Intel	Intel	AMD	AMD	
GPU	NVIDIA/AMD/Intel	NVIDIA/AMD/Intel	NVIDIA/AMD/Intel	NVIDIA/AMD/Intel	NVIDIA/AMD/Intel
	R470	R670	R6715	R6725	PowerEdge Dense Server
	R570	R770	R7715	R7725	

Intel® Xeon® 6

效率型E-Core (Sierra Forest)

- 單插槽與雙插槽
- 最多配置 144 個核心
- 最高至 330瓦TDP功耗
- DDR5: 最高支援 6400 MT/s
- 先進 I/O: 最高 126 管線 PCIe 5.0/CXL 2.0
- 單執行緒

效能型 P-Core (Granite Rapids)

- 單插槽與雙插槽
- 最多配置 86 個核心
- 最高至 350瓦TDP功耗
- DDR5: 最高支援 6400 MT/s
- 先進 I/O: 最高 136 管線 PCIe 5.0/CXL 2.0
- 多重執行緒
- UPI 2.0

AMD EPYC

Turin

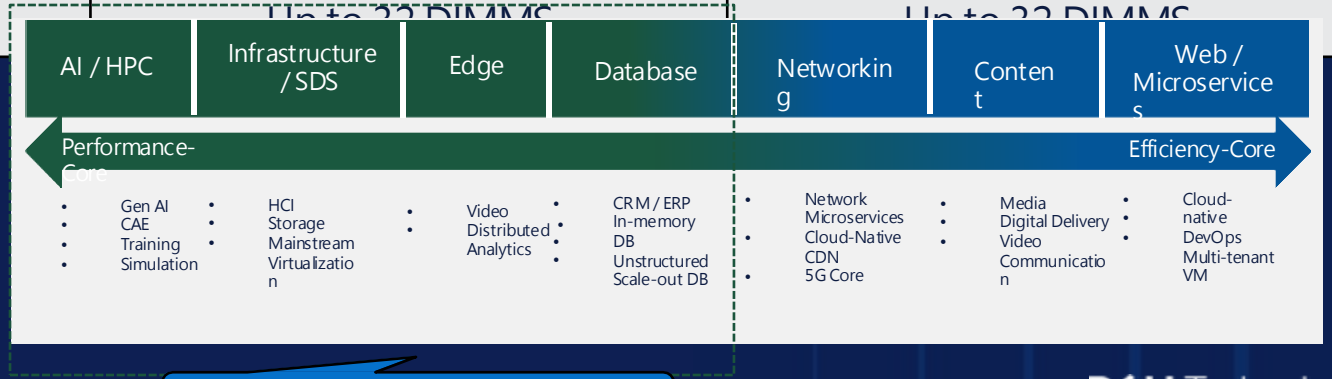
- 單插槽與雙插槽
- 最多配置 192 個核心
- 最高至 500瓦TDP功耗
- DDR5: 最高支援6000 MT/s
- 先進 I/O: 最高 128 管線 PCIe 5.0/CXL 2.0
- 多執行緒

Dell Confidential-shared under NDA to Dell Channel Partners. Not to be shared or distributed

Intel CPU Generational Comparison

Xeon "Ice Lake" (PowerEdge 15G)	Xeon "Sapphire / Emerald Rapids" (PowerEdge 16G)	Intel Granite Rapids –SP (PowerEdge 17G)	Intel Sierra Forrest – SP (PowerEdge 17G)
3rd Generation Intel Xeon Scalable	4th & 5th Generation Intel Xeon Scalable	Xeon 6 – P Cores	Xeon 6 – E Cores
8 to 40 CPU cores	8 to 64 Cores	8 – 86C	64 to 144C
PCIe Gen 4 (64 Lanes)	PCIe Gen 5 (80 Lanes)	PCIe Gen 5 (88 Lanes) R1S/RIO (136 Lanes)	PCIe Gen 5 (88 Lanes)
Max TDP: 270W	Max TDP: 350W	Max TDP: 350	MAX TDP: 330W
Up to DDR4-3200 MT/s; 8-Channel	Up to DDR5-5600 MT/s (1 DPC); 8-Channel	Up to DDR5-6400 MT/s (1 DPC); 8-Channel	Up to DDR5-6400 MT/s (1 DPC); 8-Channel
Up to 32 DIMMs	Up to 32 DIMMs	Up to 32 DIMMs	Up to 32 DIMMs

CTC此次上架主要範圍



一般應用請選P-Core

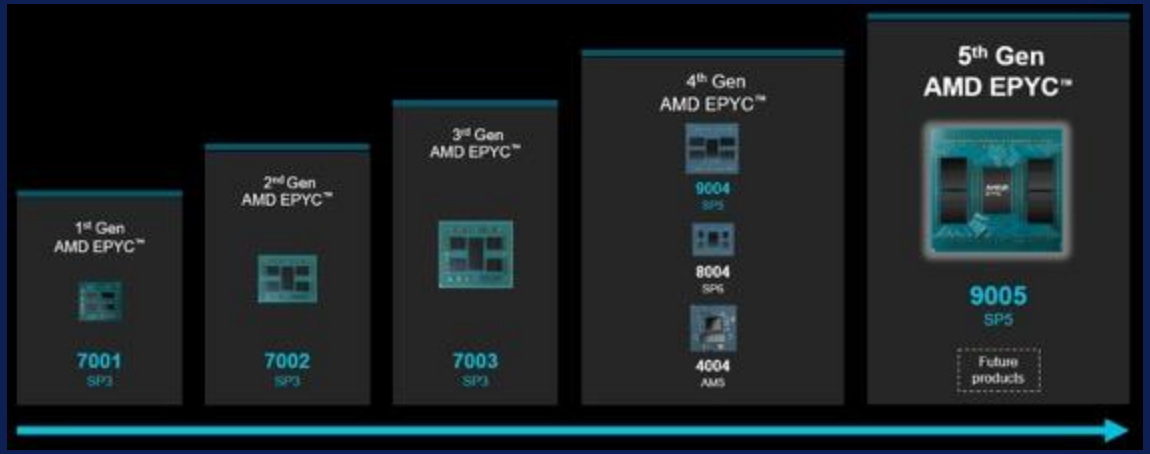
DELL Technologies

AMD CPU Comparison

CTC此次上架主要範圍

SP5 Socket
Compatible with "Genoa"

AMD EPYC "Milan" (PowerEdge 15G)	AMD EPYC "Genoa" (PowerEdge 16G)	AMD EPYC "Turin" (PowerEdge 17G)
3rd Generation AMD EPYC	4th Generation AMD EPYC	5th Generation AMD EPYC
8 to 64 CPU Cores	16 to 96 CPU Zen4 Cores / 128C Zen4c	8 - 192 Zen5-based Cores
PCIe Gen4 128	PCIe Gen5 128 Lanes	PCIe Gen5 128 Lanes
Max TDP: 280W	Max cTDP: 400W	Max cTDP: 500W
Up to DDR4-3200 MT/s; 8-Channel	Up to DDR5-4800 MT/s; 12 Channel	Up to DDR5-6400 MT/s; 12 Channel
Up to 16 DIMMs per CPU	Up to 12 DIMMs per CPU	Up to 12 DIMMs per CPU Up to 24 DIMMS 1P

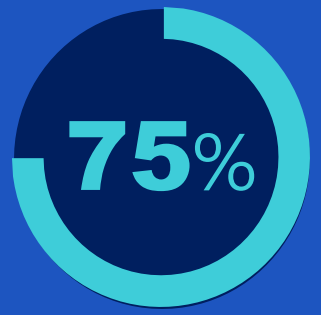


New Platform

專為最新第五代 AMD EPYC 9005 優化而打造的全新平台稱為 "Turin on Turin"

DELL Technologies

企業採用 AI 的趨勢



的組織表示AI/GenAI 已成為其業務策略的關鍵。[1]

AI 進入生產環境

65% 的組織在 2024 年成功將 AI POC 轉移至生產 [2]

位置至關重要

79% 的生產環境AI 使用案例在公有雲之外執行 [2]

資料是差異化關鍵

89% 的組織認為資料是關鍵，且其 GenAI 策略必須優先考慮資料的使用和保護 [1]

組織尋求全方位的AI 廠商

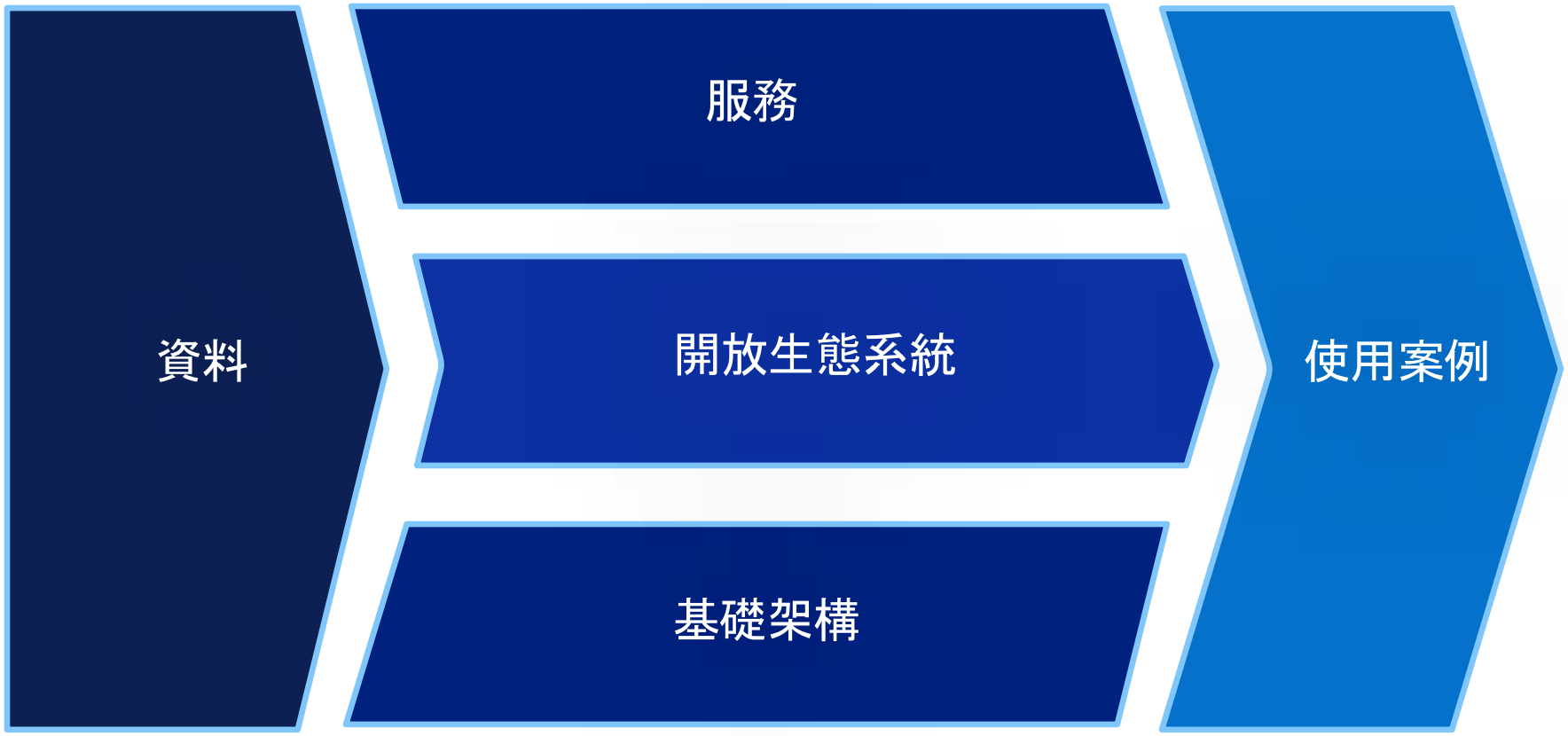
77% 的組織希望其基礎架構供應商能夠在AI 採用過程中提供涵蓋所有層面的功能 [2]

[1] 資料來源：2025 年 2 月 Dell Technologies 對美國、英國、德國、法國和日本所有產業的 750 名業務和 IT 決策者進行的調查。
[2] 資料來源：Dell 首席策略辦公室企業 AI 採用問卷調查，對象為在美國、英國、德國、日本和印度的約 3,800 位 IT 決策者和 AI 從業人員，2024 年 11 月。
ESG：最佳化 AI 總體擁有成本：採用 Dell AI Factory 在內部部署推論，成本效益比公有雲高 2.6 倍

 版權所有 © Dell Inc. 保留所有權利。

Dell AI Factory 引領您邁向未來

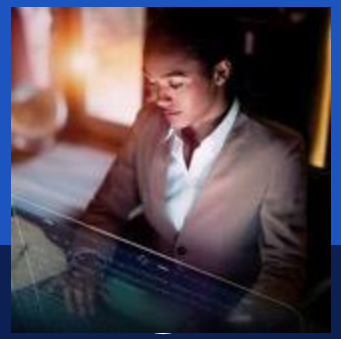
奠定 AI 轉型基礎的基礎架構簡介



永續發展 | 安全性

 版權所有 © Dell Inc. 保留所有權利。

企業 AI 的常見使用案例



內容建立

- 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

數位助理

- 面向客戶的支援
- 訓練和教育
- 排程助理和最佳化
- 追蹤專案截止日期
- 自動處理例行作業

設計與資料建立

- 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

程式碼開發

- 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

電腦視覺

- 改善設施和住戶安全
- 改善永續性成果
- 提升客戶體驗
- 提升營運效率
- 把握新的營收商機

數位分身

- 虛擬產品設計
- 可預測的建模
- 並行產品測試
- 程序變更模擬
- 視覺化大量資料集

 版權所有 © Dell Inc. 保留所有權利。

運用生成式 AI 自動化多步驟流程並加速創新

建立內部AI 工廠

以 GenAI 服務賦予客戶能力。

高達 98% GenAI 模型準確度

使用檢索增強生成 (RAG)。

生產力提升高達 70%

的 GenAI 使用者。

「一致的效能、可靠性和耗電量對我們來說都非常重要。透過搭載 NVIDIA GPU 和網路功能的 Dell PowerEdge XE9680 伺服器能達成所有上述三項目標。」

Scott Koo 先生
雲端服務企業部門執行副總

技術支援提供：

- Dell AI Factory with NVIDIA
  - 搭載 NVIDIA HGX™ GPU 的 Dell PowerEdge XE 系列伺服器
  - NVIDIA InfiniBand 交換器
  - Dell PowerScale

 版權所有 © Dell Inc. 保留所有權利。

開發日文版生成式 AI 與數位廣告服務轉型

加速效能

5. 利用最佳化提高 14-10 倍大型語言模型 (LLM) 效能

高速微調

根據最新資料集的機器學習模式

高效冷卻

配備 6U 外型規格，減少資料中心佔用空間

「我們希望能夠加快更新日文版 LLM。PowerEdge XE9680 配備 NVIDIA HGX™ GPU，效能提升 5.14 倍。」

Daisuke Takahashi
CyberAgent Inc. 集團 IT 部門 CIU 解決方案架構師

技術支援提供：

- 搭載 NVIDIA HGX™ GPU 的 Dell PowerEdge XE9680 伺服器
- Dell ProSupport

 版權所有 © Dell Inc. 保留所有權利。

透過生成式 AI 拯救生命並推動醫療照護革新

改善患者療效

透過在個人、組織與區域層級啟用預測式的主動照護管理

提升臨床應變能力

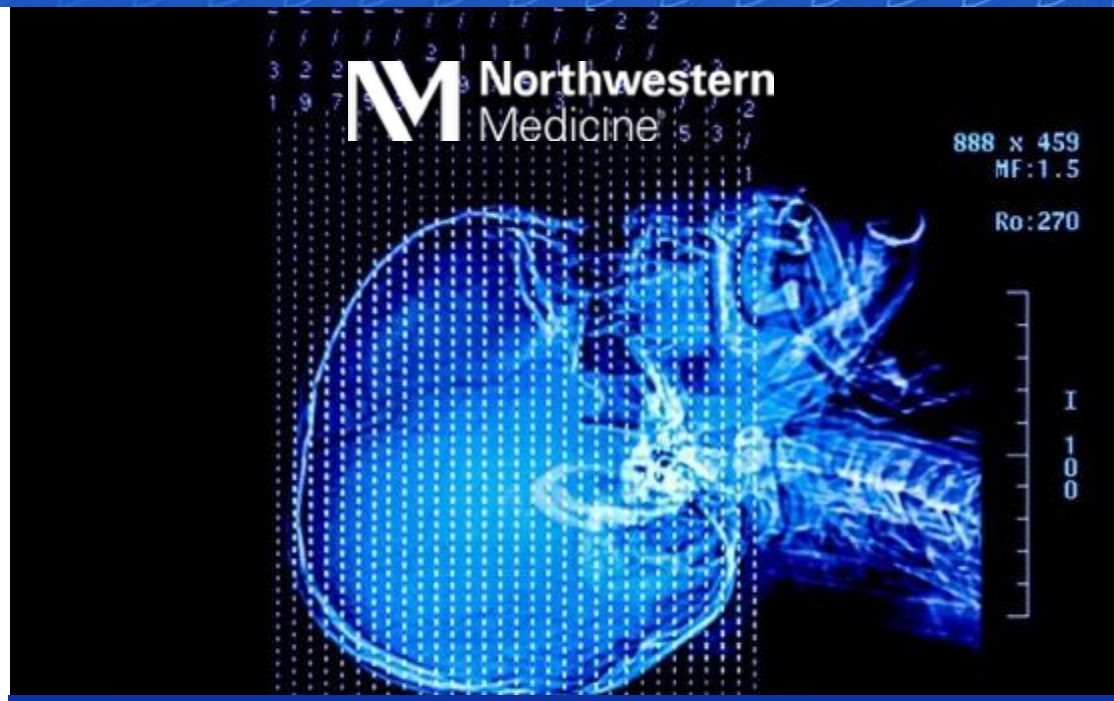
需要立即關注以拯救生命並改善提供醫療服務的效率

率先採用 GenAI

提供涵蓋整個醫療產業的藍圖，提高照護安全性、品質和一致性

「自實施 GenAI 模型以來，Northwestern 的放射師發現放射科工作效率提高了 40%...」

Sam Aboud 博士
Northwestern Medicine

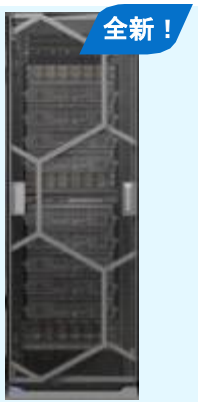


技術支援提供：

- Dell PowerEdge XE 伺服器搭載 NVIDIA GPU
- Dell AI 解決方案
- Dell AI Factory with NVIDIA

 版權所有 © Dell Inc. 保留所有權利。

AI 伺服器 17G 產品組合 – 詳細檢視

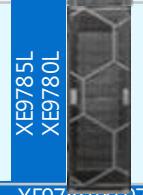


	XE7740 XE7745	XE9780 XE9785	XE9680L XE9685L	XE9780L XE9785L	XE8712	XE9712
GPU	最多 8 個 DW 或 16 個 SW PCIe。H200 NVL、RTX Pro 6000 等	8 個 SXM/OAM GPU：B300、MI350 系列	8 個 SXM GPU：H200、B200	8 個 SXM/OAM GPU：B300、MI350 系列	GB200 NVL4 GPU	GB200 NVL72、GB300 NVL72
機架式	與 19 吋機架相容			與 21 吋 ORv3 機架相容		
冷卻系統	氣冷式	氣冷式	DLC	DLC		
U 高度	4U	10U	4U	每個運算模組 3OU	每個運算模組 1OU	
架構	**單體式伺服器**：運算和電源在單一機箱中			**機架規模解決方案**：運算節點運算模組與非集計電源層架，透過後方匯流排分配電力		
電源供應器	伺服器內建 PSU			非集計機架電源		
部署方案	獨立伺服器，可選用機架整合方案		機架整合式方案	機架整合式方案		

 版權所有 © Dell Inc. 保留所有權利。

PowerEdge GPU 伺服器效能/電源配置

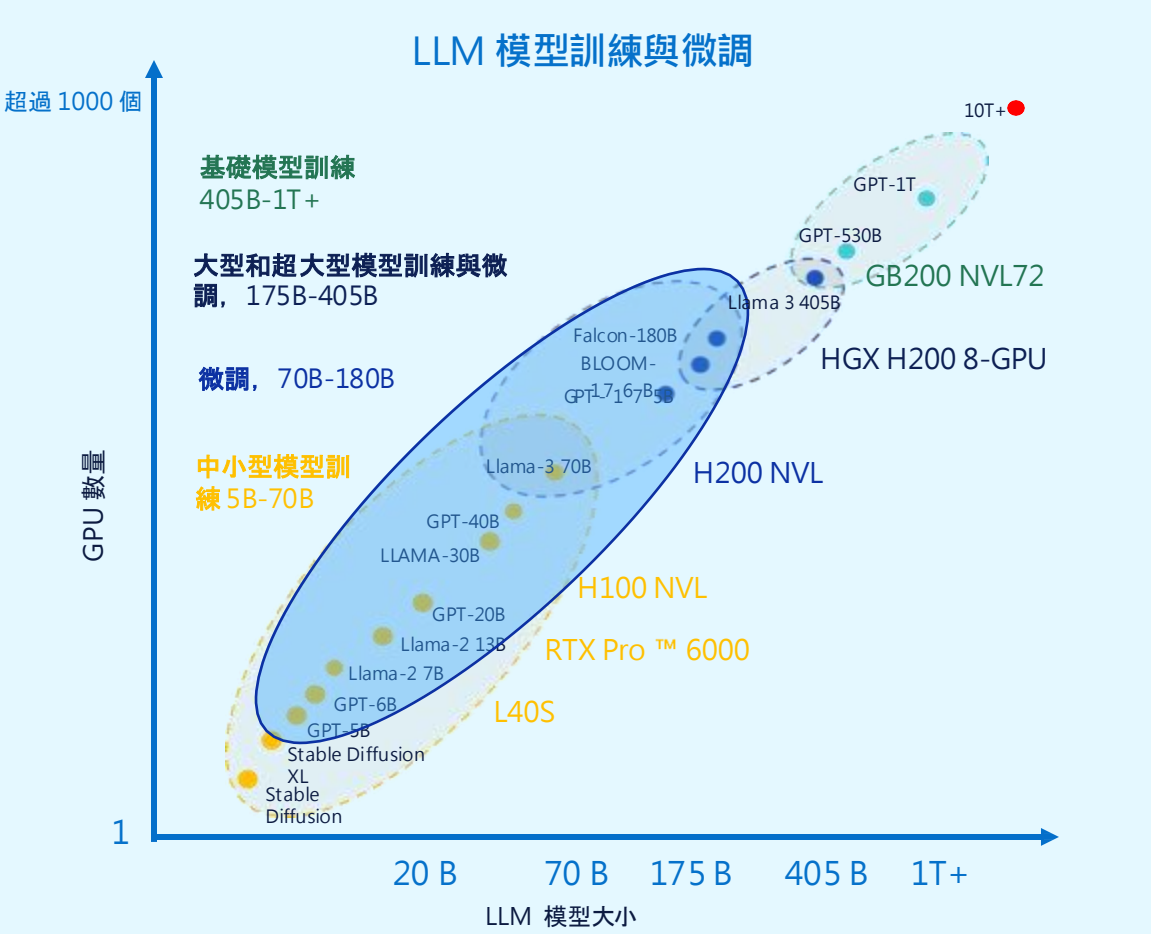
擴展的產品組合新增 DLC 以及 CPU 和 GPU 晶片多樣性選項

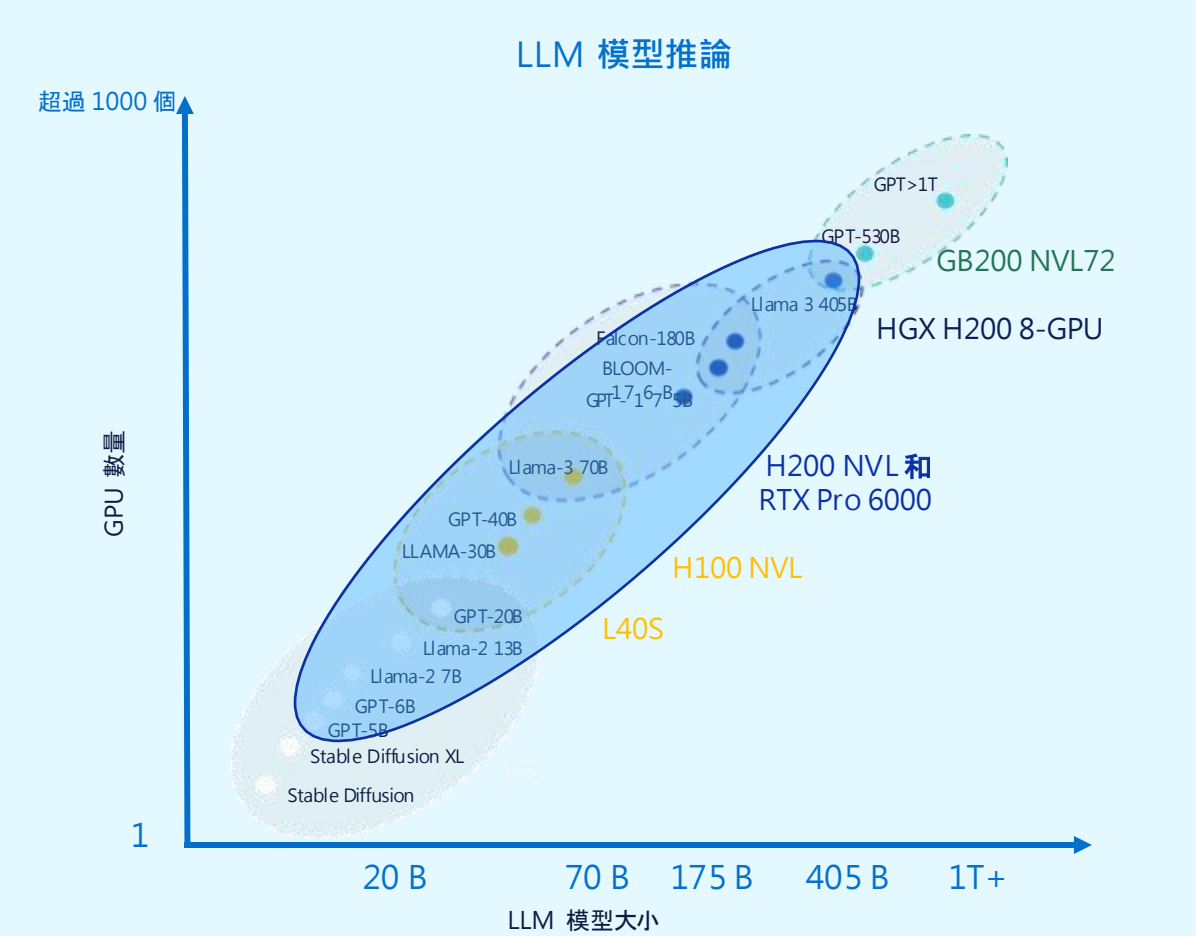
類別	使用案例	細分市場	效能	電源
ARM 型 NVL72 (液冷式)	大規模進行大型語言模型 (LLM) 訓練	CSP 主權 AI	大規模 LLM 訓練的效能最高 (生成式和代理式 AI)	136kW* 176kW*
ARM 型 NVL4 (液冷式)	HPC AI/ML 訓練	企業 HPC 研究 政府機關/教育機構 HFT	大規模 HPC 的效能最高	最高 216kW*
8 路 x86 SXM 或 OAM (液冷式)	生成式 AI/大型語言模型訓練 推薦引擎、 神經網路	CSP 主權 AI、 Enterpris e	LLM 訓練的效能最高，無需為 ARM 架構重構 SW	40 ~ 130kW*  90 ~ 210kW*
8 路 x86 SXM 或 OAM (氣冷式)	生成式 AI 訓練與微調 HPC	Enterprise CSP 主權 AI	LLM 訓練的效能最高，無需為 ARM 架構重構 SW	10 ~ 80kW* 16 ~ 80kW*
最多 16 路 x86 PCIe GPU (氣冷式)	AI/ML 微調與推論 渲染、高效能繪圖 VDI、電腦視覺、 物件偵測	Enterprise HFT	適用於推論的最佳性價比	3 ~ 60kW* 5 ~ 50kW*

*功率數據是目標設計功率估算值，僅供相對位置比較。
尖峰期間系統節流，峰值功率會更高。

您的工作負載適合哪一種解決方案？

根據最適合的模型大小提供產品方向指引





資料來源：NVIDIA

15 版權所有 © Dell Inc. 保留所有權利。

Dell PowerEdge XE7740 和 XE7745

AI 模型微調與推論

XE7740 和 XE7745 伺服器簡介

XE7740 和 XE7745 支援多種 PCIe GPU 和彈性的耗電量組態，
適合採用 x86 生態系統的氣冷式資料中心

彈性、高密度的 PCIe PowerEdge 伺服器

2 插槽功能
- 4U，可安裝在 19 吋的 EIA 機架
- Intel Xeon™ 6 或 AMD EPYC ™ 5

支援最多 8 個磁碟機
- 僅限 NVMe 直接連接

支援高速和記憶體容量
- 32 個 DDR5 DIMM (Intel XE7740)
- 24 個 DDR5 DIMM (AMD XE7745)

GPU 最佳化
- 最多 8 個 DW GPU (600 W)
- 最多 16 個 SW (75 W)
- PCIe GPU 彈性

彈性的 I/O
- 額外 8 個 PCIe 網路插槽
- 1 個 OCP 3.0 網路插槽

- 可安裝 PCIe GPU 的 4U 伺服器
- 最多 8 個高功率 DW、16 個 SW GPU
- PCIe NIC 乙太網路/IB 網路
- 專為氣冷式資料中心和低至中 PPR (5-40kW) 設計
- 一或兩組四路 NVLINK

 版權所有 © Dell Inc. 保留所有權利。

更容易落地的選擇

AI 模型微調與推論

Dell AI 算力伺服器產品組合

NVIDIA | intel | AMD

AI 專用伺服器

8-Way SXM AIR

PowerEdge
XE9680 / XE9780

*PowerEdge XE9785

8-Way SXM DLC

PowerEdge
XE9680L / *XE9780L

*PowerEdge XE9785L

PCI-E NVL 2 / 4 / 8 (DW) Way

PowerEdge XE7745
PowerEdge XE7740

PowerEdge R760XA

Rack Scale (GB300 NVL72) Coming Soon

核心伺服器

1~3 Way PCI-E

PowerEdge R 系列

邊緣伺服器

1~2 (SW) PCI-E

PowerEdge XR 系列

DELL Technologies

GPU Accelerators

Broad multi-vendor portfolio catering to applications ranging from the Cloud to Core to Edge

Solutions for targeted workloads in Gen AI, ML/DL, HPC, VDI, Data Analytics balanced by versatile and entry-level offerings to boost utilization, help the AI journey

Leading edge technology ingredients in core & memory architecture, fabrication technology, air and liquid cooling, interconnect bandwidths to deliver breakthrough performance

Growing ecosystem of frameworks, GPU-accelerated libraries that are optimized & ready-to-deploy and the necessary development tools

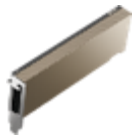
Benchmarking results demonstrate performance leadership

Dell's GPU portfolio on PowerEdge servers

Brand	GPU Model	GPU Memory	Max Power Consumption	Form-factor	2-way / 4-way Bridge	Recommended Workloads
PCIe Adapter form-factor						
NVIDIA	A2	16 GB GDDR6	60W	SW, HHHL or FHHL	n/a	AI Inferencing, Edge, VDI
NVIDIA	L4	24 GB GDDR6	72W	SW, HHHL or FHHL	n/a	AI Inferencing, Edge, VDI
Intel	Flex 140	12GB GDDR6	75W	SW, HHHL or FHHL	n/a	AI Inferencing, Edge, VDI
NVIDIA	A16	64 GB GDDR6	250W	DW, FHFL	n/a	VDI
NVIDIA	L40	48 GB GDDR6	300W	DW, FHFL	N / N	Performance graphics, Multi-workload
NVIDIA	L40S	48GB GDDR6	350W	DW, FHFL	N / N	AI Inferencing, Multi-workload
NVIDIA	RTX Pro 6000*	96GB GDDR7	Up to 600W	DW, FHFL	N / N	AI Inferencing, Multi-workload
Nvidia	H100 NVL	94GB HBM3	400W	DW, FHFL	Y / N	AI Training, HPC, AI Inferencing
Nvidia	H200 NVL	141GB HBM3e	450W - 600W	DW, FHFL	Y / Y	AI Training, HPC, AI Inferencing
AMD	MI210	64 GB HBM2e	300W	DW, FHFL	Y / N	HPC, AI Training
Intel	Gaudi3	128GB HBM2e	600W	DW, FHFL	Y / Y	AI Inferencing, AI Training
SXM / OAM form-factor						
NVIDIA	HGX H100	80GB HBM3 or 94GB HBM2e per GPU	700W	SXM w/ NVLink	n/a	AI Training, Inferencing, HPC
NVIDIA	HGX H200	141GB HBM3e per GPU	700W	SXM w/ NVLink	n/a	AI Training, Inferencing, HPC
NVIDIA	HGX B200 *	192*8=1536GB HBM3e	1000W	SXM w/ NVLink	n/a	AI Training, Inferencing
NVIDIA	HGX B300*	288*8=2304GB HBM3e	1200W	SXM w/ NVLink	n/a	AI Training, Inferencing
AMD	MI300X	192GB HBM3 per GPU	750W	OAM w/ XGMI	n/a	AI Training, Inferencing
Intel	Gaudi3 *	128GB HBM2e per GPU	850W	OAM w/ RoCE	n/a	AI Inferencing, Training
Rack-scale						
NVIDIA	GB200 NVL72 *	Up to 186 GB per GPU	1200W	NVL72	n/a	Large AI Training, Large AI Inferencing
NVIDIA	GB300 NVL72 *	Up to 288GB per GPU	1400W	NVL72	n/a	Large AI Training, Large AI Inferencing
NVIDIA	B200 NVL4 *	Up to 192GB HBM3e per GPU	1200W	NVL4	n/a	HPC, AI Inferencing

- Under development/specs subject to change
- Discuss creating a preface page that will talk about different types of form factors

PCIe Adapter



PCIe with 2-way Bridge

4-way SXM / OAM Baseboard

8-way SXM / OAM Baseboard

Rack-scale

GPU 伺服器選擇

項目	L40S 重視成本與彈性平衡 CTC項次 49 (AMD)	H100 NVL 高效能需求場景 CTC項次 50 (Intel) , 51 (AMD)
任務類型	混合型（AI 訓練/推論 、圖形/多媒體、VDI）	高階AI訓練/推論、HPC分布式運算
典型場景	1. 生成式AI模型開發與內容生產 2. 媒體產業即時高畫質影片編碼/解碼/後製 3. 工程/設計專案的即時可視化與模擬 4. 打造智能助理與視覺AI	1. 雲端AI平台訓練多百億參數大語言模型 2. 生命科學領域的基因組大資料運算 3. 金融機構進行巨量風險分析與即時預測 4. 多GPU協同的AI叢集訓練與推論根本架構
記憶體類型/容量	48 GB GDDR6	94 GB HBM3
記憶體頻寬	864 GB/s	3.9 TB/s
標稱功耗	350W	350-400W
GPU Host 介面	PCIe Gen4 x16 (64GB/s)	PCIe Gen5 x16 (128GB/s)
GPU-to-GPU互聯	無 NVLink（L40S 無支援）	NVLink 4th Gen（最高 900GB/s，多卡互聯）

PowerEdge T560

項次 36

高階直立式伺服器

with 2 CPUs Intel Xeon SP5. Generation

- Workloads
  - 檔案服務
  - 小型應用資料庫 with local storage
  - 分公司據點伺服器
  - 圖形多媒體影像處理與邊緣的 AI 推理運算
- Key Features
  - 最高支援 24x 2,5“或12x 3,5“ 驅動器配置
  - 多達 2 個 DW GPU
  - 可熱插拔的BOSS-N1作為可選的 Boot-device
  - 使用 Lifecycle 控制器和 Redfish API 對 iDRAC9 提供廣泛的管理和腳本支援

HIGHLIGHTS

- 2 個 Intel Xeon 第五代 CPU
- 16 DDR5 DIMMs, 高達 1 TB 的 RAM
- GPU支援 (Nvidia A30 L40 L4 A2)
- BOSS-N1熱插拔
- iDRAC9 系統管理

高度信賴

#1 IDC

x86 Server $
連續 31 個季度[1]

X86 Server units
連續 34 個季度[2]

Mainstream Server $
連續 31 個季度[3]

Mainstream Server units
連續 34 個季度[4]

智慧運行

Up to
$50k
透過電力與管理最佳化
實現節省[5]

80%
PowerEdge 伺服器獲得
EPEAT Climate+ 認證[6]

Up to
110min
每 100 台伺服器所需管
理時間更少[7]

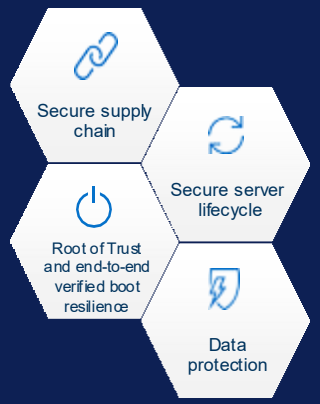
業界領先的智慧管理

- iDRAC10 integrated controller & OpenManage™ Enterprise

資安韌性

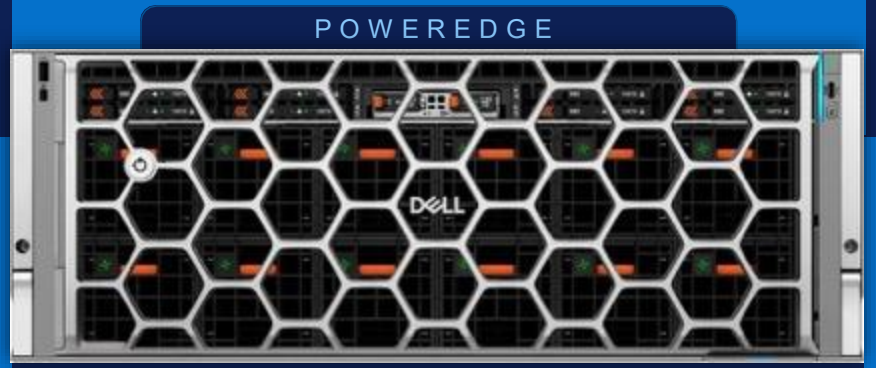
從工廠到現場
安全元件驗證保障

內建安全防護



3.5x
比同級產品擁有更強大
的安全能力[8]

具備**零信任**採用能力



AI 基礎架構領導者
榮登Forrester® AI Infrastructure Wave 報告榜首[9]

為高耗能AI 工作負載做好準備
電源效率最高提升達 **73%**[10]

突破性效能
單一 PowerEdge 伺服器即可完成 7 台舊型伺服器的
工作，且僅耗用三分之二的電力[11]

專為效率而生
PowerEdge 伺服器在過去 8 年間將
能源強度（EI）降低了83% [12]

高效
電源效率最高提升達 73%[13]

成長最快的AI 平台
Dell PowerEdge XE9680[14]

GPU 密度提升 3 倍
搭載液冷技術的Dell PowerEdge XE9680L[15]

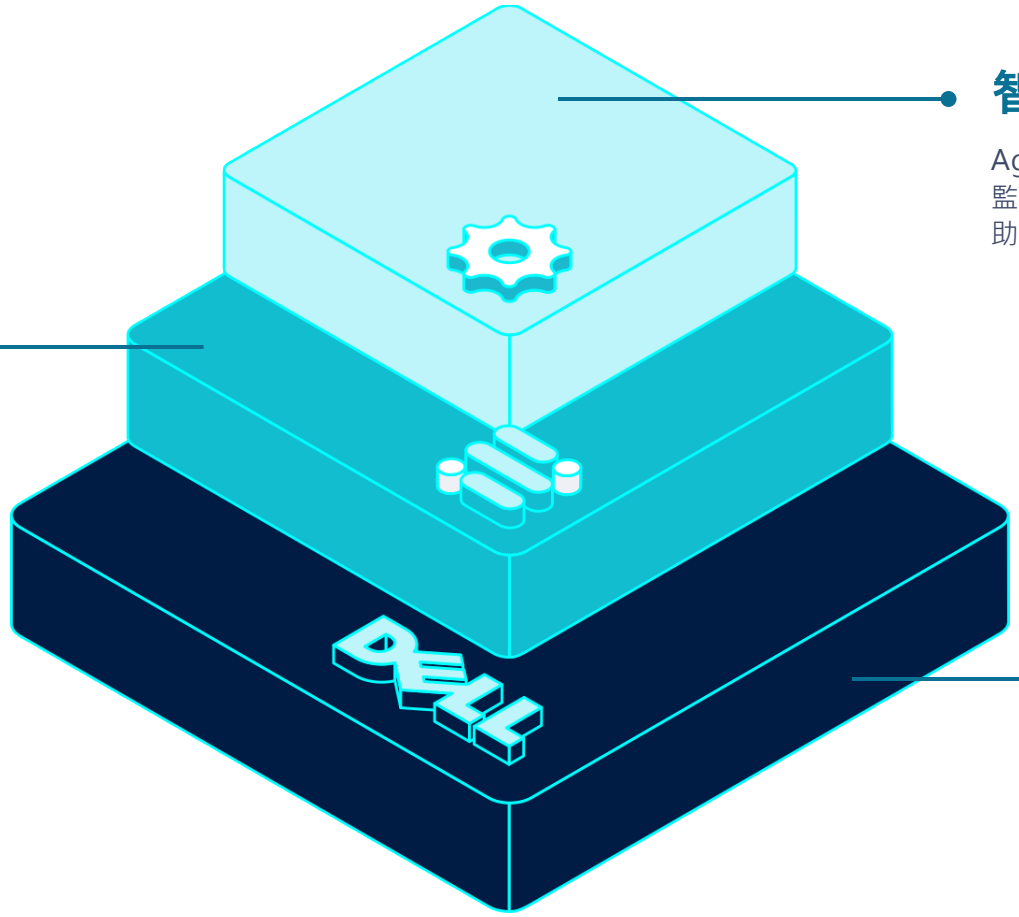
最高提升 6 倍
在 MLPerf 4.0® 基準測試中，相較前幾代產品，搭配
Dell AI Factory with Nvidia[16]

量身打造

永續發展

人工智慧

虛擬化 → 智慧化 → 自主化



智慧管控層

Agentive Infrastructure 在此基礎上實現智慧監控、自動化資源調度，驅動虛擬助理與學習助理等服務。

虛擬化層

CubeCOS / CubeVDI 提供穩定虛擬化資源

基礎硬體

Dell 伺服器提供穩定運算與儲存

從基礎層到智慧層：
邁向智慧校園的未來願景

自我優化的智慧決策系統

Dell 伺服器提供完整效能階梯，滿足從一般教學到高效能運算的多元需求

主動式學習與虛擬助理服務

企業級可靠度設計，支持校園持續成長與未來擴展需求

真正自主化的新世代 IT 基礎建設

協助校園降低總體擁有成本 (TCO)，實現綠色節能目標

政府共同供應契約、校園虛擬化優惠方案

組別	組別名稱	項次	品項名稱
7	虛擬軟體	1	CubeCMP 超融合雲運算平台自動化管理軟體 Add-On - Per Service Unit (含原廠一年技術諮詢與支援)
7	虛擬軟體	2	CubeCOS 超融合雲運算 Network Security – 1 Service Unit
7	虛擬軟體	3	CubeCOS 超融合雲運算平台軟體 - Per Server Node (含原廠 一年Advanced 線上支援) 最新版授權
7	虛擬軟體	4	CubeCOS 超融合雲運算平台軟體 - Per Server Node (含原廠一年 Premium 線上支援) 最新版授權
7	虛擬軟體	5	CubeCOS 超融合雲運算平台軟體 - Per Server Node (含原廠三年 Advanced 線上支援) 最新版授權
7	虛擬軟體	6	CubeCOS 超融合雲運算平台軟體 - Per Server Node (含原廠三年 Premium 線上支援) 最新版授權
7	虛擬軟體	7	高可用Cube 超融合雲運算平台 (3 節點)
7	虛擬軟體	8	高可用Cube 超融合雲運算平台一年更新與支援 (3 節點)
7	虛擬軟體	9	高可用Cube 超融合雲運算平台擴充套件 (1 節點)
7	虛擬軟體	10	高可用Cube 超融合雲運算平台節點一年更新與支援 (1 節點)

Thanks for Listening

Bigstack

官方網站

cubeCOS

使用指引