



ALE DC Initiative

ALE 資料中心計畫

本文件為 ALE DC Initiative 簡報之繁體中文整理版。技術術語、產品型號及規格保留原文。



本文件由 惠商實業股份有限公司 Nmars International
協助翻譯製作。

資料中心為何成為 ALE 的策略新重點

資產組合多元化

從舊有技術轉型至具有前瞻性的策略成長領域

更強的經濟效益

更大的交易規模、更長的客戶生命週期價值、多年期合約

策略關聯性

數位轉型的核心所在：AI、雲端、主權倡議

資料中心解決方案現況

傳統企業 DC

內部自建設施，由單一組織擁有及運作，用於 IT 工作負載，資本支出高、擴展性有限。

雲端與超大規模 DC

由主要公有雲供應商託管，提供隨需應變、可擴展的運算與儲存，為 SaaS、混合雲及分散式工作負載的骨幹。

共置與受管服務

第三方設施，企業租用空間、電力與網路連線，提供無需自行管理基礎設施的靈活性。

邊緣與混合 DC

靠近終端使用者的小型分散式設施，降低延遲並支援 IoT、即時分析與 5G 應用的在地化處理。

AI 與高效能 DC

專為以 GPU/加速器執行計算密集型 AI 工作負載而建，需要先進冷卻、高功率密度及高速網路能力。

資料中心投資正轉向 AI

AI 驅動的基礎架構

GPU 運算、高速網路及先進冷卻現已成為主要投資驅動力，支援尖端 AI 工作負載。

AI 未來成長

未來五年，大多數增量資料中心支出將與 AI 相關，由生成式 AI、分析、自動化及主權 AI 倡議推動。

資本策略配置

傳統資料中心投資維持穩定，新資本則策略性地導向 AI 能力與基礎架構的加速建置。

重點摘要：「儘管傳統資料中心仍不可或缺，幾乎所有資料中心投資的新增成長均由 AI 驅動。」

第 6 頁

AI 技術概覽

人工智慧 (AI)

執行需要人類智慧的任務：推理、學習、感知、決策、互動。

機器學習 (ML)

從歷史資料學習模式以進行預測或決策，用於分類、回歸及異常偵測。

深度學習 (DL)

利用多層神經網路學習高度複雜的模式：圖像識別、語音識別及視訊分析。

電腦視覺 (Computer Vision)

解讀來自相機和圖像的視覺資料，實現物件偵測、臉部辨識、圖像分類、視訊分析及品質檢測。

自然語言處理 (NLP)

處理與理解人類語言，支援文字理解、翻譯、情感分析及語音轉文字。

生成式 AI (GenAI)

根據學習到的模式生成原創內容，涵蓋文字、圖像、影片、音樂乃至程式碼。

大型語言模型 (LLMs)

在海量資料集上訓練以預測下一個詞，驅動對話式 AI、問答、摘要及程式撰寫輔助。

檢索增強生成 (RAG)

將 LLM 與企業或外部資料來源結合，提供具上下文感知、即時且準確的回應。

多模態 AI (Multimodal AI)

設計用於理解並生成跨文字、圖像、音訊及視訊等多種模態的內容，實現全面的跨媒體推理。

強化學習 (RL)

透過回饋與獎勵學習最佳行動，主要用於機器人技術、遊戲 AI 及自主系統的動態最佳化。

自主與代理 AI (Agentic AI)

指能在最少人工干預下規劃、決策並執行複雜任務的 AI 系統，推動工作流程自動化與多步推理。

第 7 頁

AI 帶來實質商業成果 (一)

醫療保健

- 協助醫生從醫療影像和檢驗結果診斷疾病
- 根據緊急程度自動排定病患優先順序
- 透過分析大型臨床資料集支援醫學研究
- 最佳化醫院運作，如排程與床位使用
- 分析人口健康資料以預測疫情與風險

交通運輸

- 利用視訊分析即時預測並偵測事故
- 根據乘客需求及交通模式最佳化列車時刻表
- 自動化交通管理、旅客流量及信號最佳化
- 以 AI 驅動的旅客資訊及聊天機器人改善客戶體驗
- 使用數位孿生模擬鐵路運作，部署前先測試情境

政府與城市

- 即時最佳化交通號誌並減少壅塞
- 改善公共安全、事件偵測及緊急應變時間
- 最佳化廢棄物收集及城市服務
- 監測空氣品質、噪音及污染
- 公民服務 AI 聊天機器人
- 自動化文件處理與公共行政

教育

- 數位教育平台與個人化學習 (自適應學習、AI 輔導)
- 安全評量與學術誠信 (自動評分、抄襲偵測、線上監考)
- 研究與 AI 技能發展 (AI 研究實驗室、學生沙盒)
- 運營效率與成本最佳化 (招生預測、課表最佳化、能源管理)

第 8 頁

AI 帶來實質商業成果 (二)

金融



- 即時偵測詐欺及可疑交易
- 改善風險評估及信用評分
- 使用預測模型最佳化交易策略
- 透過 AI 聊天機器人提供即時客戶支援
- 在提升安全性的同時改善客戶體驗

零售與電商

- 預測產品需求以最佳化庫存
- 自動化補貨及供應鏈決策
- 生成產品描述及行銷內容
- 提供個人化購物推薦
- 啟用對話式購物助理

製造業

- 預測設備故障以減少非計畫停機
- 在生產過程中自動偵測缺陷
- 在實體部署前模擬製造流程
- 實現更智慧、更自主的機器人系統
- 透過 AI 驅動的模擬加速產品設計

能源

- 預測能源需求以平衡供需
- 即時最佳化智慧電網運作
- 最大化再生能源的生產與使用
- 決定儲能或釋能的最佳時機
- 透過預測分析減少能源浪費

第 9 頁

AI 資料中心中的訓練與推論

AI 模型經歷兩個關鍵階段：訓練 (Training) 與推論 (Inference)，各有不同的網路需求。

訓練 (Training)

涉及將海量資料集輸入 GPU 叢集，在 GPU 與伺服器之間產生極大量的東西向 (East-West) 網路流量。這要求高頻寬、低延遲及無損網路，以實現最佳 GPU 利用率。

推論 (Inference)

相反地，推論是部署已訓練的模型進行即時預測。通常比訓練對延遲更敏感，但頻寬需求較低，專注於對用戶的快速回應。



AI 需要專用基礎架構

傳統 IT 基礎架構無法滿足現代 AI 工作負載的獨特需求。

大規模運算 (Massive Compute)

訓練大型 AI 模型需要數百萬 GPU 小時，運算資源費用高達數千萬美元。

平行處理 (Parallel Processing)

GPU 加速神經網路訓練速度比 CPU 快 100 至 1000 倍。

極大資料規模 (Extreme Data Scale)

如 GPT-5 等模型在數兆 token、數百 TB 文字資料上訓練，需要專用資料管線。

低延遲推論 (Low-Latency Inference)

生產環境 AI 系統需要在毫秒內完成預測，同時處理數百萬個並發請求。

AI 基礎架構的分層視圖

維運層 (Operations)

CI/CD Pipeline：自動化測試與部署

監控與可觀測性：模型效能即時指標、漂移偵測及基礎架構使用率

安全性與合規：IAM 角色、靜態/傳輸中加密、稽核日誌

軟體層 (Software)

框架 (Frameworks)：PyTorch, TensorFlow, JAX

加速函式庫：ROCm, MIOpen, RCCL, CUDA, cuDNN, NCCL, oneDNN

容器與協作：Docker, Kubernetes

MLOps 工具：MLflow, Kubeflow, Weights & Biases

硬體層 (Hardware)

運算 (Compute)：GPU 用於矩陣運算及平行處理

儲存 (Storage)：高速 SSD、分散式物件儲存、針對 AI 最佳化的特徵儲存庫

網路 (Networking)：RDMAv2 無損乙太網路

ALE AI 資料中心方案

運算 (Computing)

用於訓練大型語言模型 (LLM) 及推論的 AI 伺服器。

網路 (Networking)

供 AI 叢集全速運行、無延遲或瓶頸的網路基礎架構。

第 14 頁

ALE AI 資料中心機架元件

主要元件

OS6900 — 管理交換器 (Management Switch)

OS7900 (EC 重新品牌) — AI Fabric Switch

OS6920-D32

OT2290 (未來)

OmniCompute — GPU 伺服器

Transceivers and Cables — 收發器與線纜

SW Orchestrator — 軟體協作器

Switch NOS — 交換器網路作業系統

GPU Framework — GPU 框架

電力

每台 GPU 伺服器 10kW (實際工作消耗為 6~8kW)

網路

AI 運算後端網路 (AI Computing Backend Network)

管理/儲存前端網路 (Management / Storage Frontend Network)

第 16 頁

OS7900-064 & OS7900-032 後端 Fabric 交換器

交換容量：51.2 Tbps 及 25.6 Tbps (高頻寬、低延遲、進階負載均衡)

每埠 8 條 SerDes 通道，各 @100Gbps

Broadcom Tomahawk5 (TH5)，最適合 AI 工作負載

LPO (線性可插拔光纖)：低功耗、降低成本、更低延遲



OSFP 可升級至更高速度 1.6Tbps

第 17 頁

OS7900-D64 & OS7900-D32 後端 Fabric 交換器

交換容量：25.6 Tbps 及 12.8 Tbps，每埠 8 條 SerDes 通道 @50Gbps

Broadcom Tomahawk4 (TH4)，最適合 AI 工作負載

動態負載均衡

適用於頻寬需求較低的 HPC/GPU 部署

第 18 頁

ALE OmniCompute [OC8600] 規格

規格項目	詳細資料
GPU	AMD 8 x MI325X GPUs with Infinity Fabric
CPU	AMD 2 x EPYC™ 9005/Turin Series Processor
CPU NIC	2x BCM957608-P2200GQF00 Dual-Port 200GbE
GPU NIC	8x BCM957608-P1400GDF00 Single-Port 400G
儲存	2x1.92TB U.2 2.5 NVME4 1DWP (non SED)
外型/冷卻	8U 氣冷 (Air Cooling)
效能	FP32: 163.4 TFLOPS FP16: 1300 TFLOPS FP8: 2610 TFLOPS
每 GPU 記憶體	256GB HBM3e
記憶體頻寬	6 TB/s
GPU 峰值功耗	1KW
伺服器實際工作功耗	6-8KW

第 19 頁

ALE 交換器與網路軟體

NOS 主要功能

現代化 NOS 架構

直通式交換 (Cut-through switching)

自適應路由與動態負載均衡



PFC (優先流量控制)、RoCEv2 (RDMA over 融合乙太網路 V2)、ECN (明確壅塞通知)
進階遙測 (Advanced telemetry)
VXLAN-EVPN 多租戶支援

協作器 (Orchestrator) 主要功能

後端、前端及管理網路的機架協作
業界標準 Day-0、Day-1 及 Day-2 維運
意圖式網路佈建 (Intent-based provisioning)
網路事實來源 (NSOT) 以維護網路完整性
整合數位學生
使用進階遙測進行監控與維護

雙 NOS 策略

ALE SONiC：適用於大規模雲端原生企業，完整掌控技術堆疊
AOS-X：提供功能深度與成熟度，適合小型團隊/企業的運營簡易性

第 20 頁

GPU 後端 Fabric : [16 GPU] [實驗藍圖-1]

單機架 AIDC 解決方案，含兩台運算伺服器。伺服器、後端網路及管理共用同一機架。

規模

GPU 數量：~16
GPU 伺服器：2
機架數：1
交換器數：1

網路

機架內 AI Fabric：400G/800G DAC
管理 Fabric：1G/10G

第 21 頁

GPU 後端 Fabric : [16 GPU] [實驗藍圖-2]

雙機架 AIDC 解決方案，可擴展至 8 個機架。獨立的管理/Spine 網路機架。機架內使用 DAC，機架間使用 AOC/光纖線纜。

規模



GPU 數量 : 16
GPU 伺服器 : 2
機架數 : 3
交換器數 : 4

第 22 頁

GPU 後端 Fabric : CLOS 架構 [256 GPU]

GPU 數量 : 256
GPU 伺服器 : 32
機架數 : 9
交換器數 : 10
可擴展至 : 8 個機架

第 23 頁

物料清單與定價

單項產品定價

產品	說明	定價 (美元)
GPU 伺服器	8x MI325x GPU , 含 NIC 卡	\$300,000
800G 交換器 (64埠)	64x 800G OSFP 後端 Spine 交換器	\$132,000
800G 交換器 (32埠)	32x 800G OSFP 後端 Leaf 交換器	\$110,000
800G 光纖	800G SR8 收發器 (50m)	\$879
800G 轉 2x400G DAC	Leaf 至伺服器 , 3m	\$300
管理交換器	6900-T48	\$34,275
NOS	SONiC	\$0
協作器	設備管理軟體	\$7,000

藍圖總價

藍圖類型	ASP 價格 (美元)
Blueprint-1 (1 個機架)	\$753,675
Blueprint-2 (2 個機架)	\$1,141,739



已達成事項與里程碑

第一階段 (PHASE 1)

範疇：具最少品牌標示的 OEM 方案（交換器上的 ALE 標誌、型號名稱、NOS 登入橫幅、管理軟體絲印）

文件：HW/SW/CLI/規格表均修改為 ALE 版本

進度：DR0 已於 1 月 26 日完成，DR1 本週計畫，DR4 預計 2 月 26 日

第二階段 (PHASE 2)

範疇：前端網路轉換為 ALE-NOS (AOS-X)，後端轉換為 ALE SONiC

DR1：2025 年 12 月完成，DR4：預計 2026 年 7 月

第三階段 (PHASE 3)

範疇：後端雙 NOS (AOS-X & ALE SONiC)、ALE 品牌 800/400G 光纖、轉換至 ALE 管理軟體

PRD：2025 年 12 月完成，DR0：預計 3 月

路線圖

時程	階段	主要內容
2026 年 2 月	Phase 1	後端：OEM 800G/400G，OEM EC SONiC 前端：ALE 100G/25G，OEM EC SONiC 管理：ALE 10G 運算：Edgecore AMD Server
2026 年 7 月	Phase 2	前端：+ALE 400G，AOS 10.1R2 後端：ALE SONiC 機架管理：OEM EDM v2.4
2026 年 12 月	Phase 3	前端/後端：AOS 10.1R3，ALE SONiC 光纖：ALE 10/100/400/800G 機架管理：ALE OVCompass

ALE 在 AI-DC 方案上的差異化優勢

端到端 AI 資料中心解決方案：整合硬體、網路設計、協作及支援，超越單純的設備供應。

支援服務作為核心差異化：單一廠商問責制及 AI-Fabric 基礎架構。

意圖式網路協作：簡化設定、減少錯誤，並在整個 Fabric 及解決方案中強制一致性。

Day-2 卓越運營：深度可視性、監控及主動問題偵測。

後端網路實施設計準則

低延遲

選用具 400G/800G 介面的 Tomahawk 交換器

非阻塞式拓撲以提升速度與效率

1:1 超用比，充分利用所有交換器埠位

低功耗

DAC 和 LPO 功耗較低

機架內伺服器至 Leaf 間使用 DAC

機架間使用 LPO

網路設計選擇

兩種模型：常規 CLOS 及鐵軌最佳化 (Rail Optimized)

線纜類型	傳輸距離	功耗
DAC	3m	0.1~0.2W
LPO	~500m	7~8W
ACC	5m	1.2~2.5W
AOC	100m	12~15W

設計選擇準則	推薦設計
簡單部署 (較佳線纜管理及低成本)	常規 CLOS
標準化及熟悉的網路設計	常規 CLOS
低延遲、單跳通訊	鐵軌最佳化

第 27 頁

網路設計摘要

常規 CLOS 網路設計

GPU 數	Leaf 交換器	Spine 交換器	XCVR	DAC	NIC (GPU)
64 (2機架)	1台 (32x800G)	0	0	64台 (800-2x400G)	64
128 (4機架)	4台 (32x800G)	2台 (64x800G)	128台 (800G-800G)	128台 (800-2x400G)	128
256 (8機架)	8台 (32x800G)	2台 (64x800G)	256台	256台	256

鐵軌最佳化網路設計

GPU 數	Leaf 交換器	Spine 交換器	XCVR	DAC	NIC (GPU)
-------	----------	-----------	------	-----	-----------



64 (2機架)	8台 (32x800G)	0	32台 (800G-400G)	64台 (400G-800G)	64
128 (4機架)	8台 (32x800G)	0	64台	128台	128
256 (8機架)	8台 (32x800G)	0	128台	256台	256

第 29 頁

OS7900-032/064 規格

規格項目	詳細資料
埠位介面	800G (OSFP800 或 QSFP-DD800)
ASIC / 容量	Broadcom Tomahawk-5 / 51.2 Tbps
CPU	Intel Icelake-DE 4 核心 2.2GHz with BMC
記憶體 / 儲存	2x 16GB DDR4 / 240GB SSD
工作溫度	0°C 至 40°C
尺寸	440mm x 650mm x 87mm (2RU)
電源 / 風扇	模組化 1+1 AC 電源 / 4 個風扇模組 (7+1 個風扇) , 前進後出氣流
特色	AI/ML/HPC 使用, DC Super Spine, 最多 320 個邏輯埠, SyncE + PTP

第 31 頁

後視圖 - 硬體詳細資料

- 4 個可熱插拔風扇模組, 含 7+1 冗餘風扇
- 每個風扇模組包含 2 個風扇
- 8 個風扇中允許 1 個故障
- 更換風扇模組時限為 2 分鐘, 超過則 MAC 會過熱並關機

第 32 頁

主要功能特色

- SyncE + PTP 支援
- SRv6 支援
- 每埠最高 24W 功率預算, 支援相干光纖 (如兩個電源均接通且總功耗 > 3000W)
- E-fuses 保護收發器及內部元件
- BMC 模組執行 OpenBMC, 支援 Serial-over-LAN
- 待機省電模式



最多 320 個邏輯埠
 認知/自適應路由及動態負載均衡 (AI/ML)
 進階共享緩衝 (Advanced shared buffering)
 可程式化帶內遙測 (Programmable in-band telemetry)
 支援端到端壅塞控制 (AI/ML, 需調校)
 單晶片 5nm 製程, 省電高效
 硬體式連結故障切換, 提升網路韌性及降低 AI/ML 作業完成時間
 VXLAN RIOT 支援

第 33 頁

產品定價

型號	簡短說明	定價 (美元)	標準折扣
OS7900-064	64x 800G QSFP112-DD 後端交換器	\$132,000	70%
OS7900-032	32x 800G QSFP112-DD 後端交換器	\$110,000	70%
OS7900-D64	64x 400G QSFP112-DD 後端交換器	\$115,000	70%
OS7900-D32	32x 400G QSFP112-DD 後端交換器	\$70,000	70%
OC8600-8AM	8x MI325x GPU, 含 NIC 卡	\$300,000	10%
OS-SW-ASON	ALE SONiC NOS	\$14,000	20%
OS-SW-AOSX-AD	AOS-X 進階版	\$15,000	50%
OS-SW-AOSX-PM	AOS-X 旗艦版	\$30,000	50%
OC-SW-NET	OmniOrchestra 設備管理器	\$7,000	50%

第 35 頁

市場進入策略

銷售與通路策略

第一階段：透過現有 ALE DR 及 VAD 的純間接模式
 第二階段：在法規允許且無通路衝突的情況下，對策略客戶進行直接銷售
 合作夥伴招募：具備 DC 專業知識及政府、DCSP、電信商關係

重點國家

第一層：法國、德國、英國、北歐、阿聯酋、沙烏地阿拉伯
 第二層：EUSO/EUNO/MEITA 其他地區及亞太區



第三層：美國/加拿大及中南美洲

目標客戶

企業、私有及區域雲端 AI DC 機會 (第 2-3-4 層)

第 36 頁

客戶輪廓

主權/國家 AI 雲端專案

政府部會

政府管理的雲端倡議

國家 AI 叢集

主權資料中心

第 2-3 層雲端供應商

區域雲端公司

受管服務供應商

新興 AI 雲端平台

共置 + 雲端混合業者

企業 AI 轉型 DC

銀行

能源/公用事業

電信商

航空

製造業

醫療保健

大學/研究機構

交通運輸

AI 優先公司

AI 新創公司

HPC 公司

資料分析公司

GenAI 平台公司

第 37 頁

企業 AI DC (中東地區)

地區	垂直產業	企業範例	應用說明
阿聯酋	能源	ADNOC	儲油層與地震建模 AI 主權
阿聯酋	航空	Emirates Group	運營 AI + 維護優勢
阿聯酋	國家 AI	G42 政府工作負載	主權 AI DC
阿聯酋	醫療	M42 Genome	基因組 AI 必須在主權範圍內
阿聯酋	銀行	FAB	AML + 風險模型須保持私有
沙烏地	能源	Aramco	儲油層建模 IP 為國家策略資產
沙烏地	石化	SABIC	材料創新 AI IP 受保護
沙烏地	政府資料	DGA/NDMO	國家資料主權 AI
沙烏地	智慧城市	NEOM	下一代智慧區域的行業 AI DC
沙烏地	銀行	Saudi National Bank	金融 AI 私有設計

第 38 頁

主權/國家 AI 雲端 DC

國家	代表機構	說明
法國	GENCI / CEA / SecNumCloud (3DS Outscale / OVH Sovereign)	法國管轄下的國家 HPC + AI 模型容量。策略科學 AI + 法規要求的國家資料在地化。
德國	JUPITER Exascale / FZJ Jülich	具國家管控的 Exascale AI — 國家產業 AI + 受保護的研究運算。
英國	UKAI Compute / 國家 AI 研究資源 (NIRR)	英國 AI 研發容量的國家 AI 叢集 — 英國監管的主權模型訓練環境。
阿聯酋	國家 AI 局計畫 + Core42 + Khazna 主權分區	國家 AI 模型訓練 + 國家資料安全優先。下一代 AI 經濟的主權 DC 策略。
沙烏地	PIF 支持的國家 AI DC (Humain 倡議) / SDAIA	國家策略運算 = 主權資產。AI 叢集容量與 Vision 2030 國家轉型計畫對齊。

第 39 頁

第三層 AI 雲端 DC

國家	供應商範例	定位說明
法國	OVHcloud Enterprise / Scaleway	聚焦法國及歐盟的雲端，有限的全球佈局，商業私有雲端託管法規企業工作負載。
德國	IONOS Enterprise Cloud / PlusServer	國家 + 歐盟合規私有雲端，非 AI 專業第二層，非主權 HPC 國家管控。

英國	BT Business / Vodafone Business Cloud	在英商業銷售的本地私有雲端，非全球超大規模，非主權指定工作負載。
阿聯酋	Etisalat/DU Business Cloud	阿聯酋企業工作負載的國家範圍託管，非國家 AI 局，非區域第二層 GPU。
沙烏地	STC Cloud / Zain KSA Enterprise Cloud	企業商業使用的國家電信雲端 DC，非主權授權 DC，非超大規模/第二層 GPU 供應商。

第 44 頁

企業 AI DC (歐洲)

國家	產業	企業範例	說明
法國	銀行	BNP Paribas / Société Générale	主權 + AML + 風險 AI 必須私有
法國	航空	Airbus	安全關鍵模擬 + IP 保護
法國	汽車	Renault Group	製造 AI + 數位學生
法國	製藥	Sanofi	R&D AI 藥物探索資料敏感性
法國	能源	EDF	關鍵國家基礎架構 + 主權 AI
德國	汽車	BMW / Mercedes-Benz / VW	自駕 & 模擬 IP 必須在主權範圍
德國	工業	Siemens	工業 AI + 工廠數位學生
德國	製藥	Bayer	藥物探索 AI 為核心 R&D IP
德國	能源	E.ON / RWE	關鍵基礎架構最佳化 AI
德國	銀行	Deutsche Bank	內部詐欺/風險 AI 必須在內部
英國	製藥	GSK / AstraZeneca	藥物探索 AI IP 受保護
英國	電信	BT	網路 AI 最佳化私有
英國	國防	MoD/DSTL	國家安全 AI
英國	媒體	BBC R&D	生成式媒體 AI IP 優勢

第 45 頁

第二層 AI 雲端 DC

國家	供應商範例	說明
法國	OVHcloud AI GPU Clusters / NexGen Cloud (EU)	多區域 GPU 雲端擴張模式，AI 運算專業化，比超大規模供應商更快速佈建。
德國	Northern Data (EU HQ)	歐盟規模的 GPU 運算供應商；非國家電信，非超大規模；AI 運算容量銷售商。
英國	Vultr GPU PoPs (倫敦) / DigitalOcean (Paperspace 起源)	GPU 雲端建設商擴張至英國；專業容量供應商，非完整雲端超大規模。
阿聯酋	Vultr GPU 杜拜 / 區域第二層 GPU 新創公司	第二層供應商擴張至 GCC 區域，作為企業快速取得 GPU 容量的銷售商。

沙烏地	區域 PoP 建設的第二層 GPU 業者 / 歐盟第二層進駐 KSA	引進第二層 GPU 叢集至沙烏地 — 專業 GPU 供應，非超大規模，非主權。
-----	------------------------------------	---

附件

Annex — Original Presentation File

【附件說明】

文件名稱 ALE DC Initiative — 原始簡報

文件格式 Microsoft PowerPoint (.pptx)

投影片數量 36 頁

語言 英文 (原始版本)

說明 以下內容為本繁體中文整理文件所依據之原始英文簡報，由 Alcatel·Lucent Enterprise 製作，供參考對照之用。技術規格、定價及路線圖以原始文件為準。

ALE DC Initiative

Why Data Centers Are ALE Strategic New Focus



Portfolio Diversification

Shift from legacy technologies to strategic, future-proof growth areas



Stronger Economics

Larger deals, longer customer lifetime value, multi-year engagements



Strategic Relevance

Core to digital transformation: AI, cloud, sovereignty initiatives

DC Solutions Landscape Today



Traditional & Enterprise DCs

On-premises facilities, owned and operated by a single organization,

For IT workloads,

Involve high CAPEX and limited scalability,



Cloud & Hyperscale DCs

Hosted by major public cloud providers,

Offer on-demand, scalable compute and storage,

Backbone for SaaS, hybrid cloud, and distributed workloads,



Colocation & Managed Services

Third-party facilities where companies lease space, power, and connectivity,

Offer flexibility without the need for direct infrastructure ownership and management.



Edge & Hybrid DCs

Smaller, decentralized facilities located close to end-users.

They reduce latency and support localized processing for IoT, real-time analytics, and 5G applications.



AI & High-Performance DCs

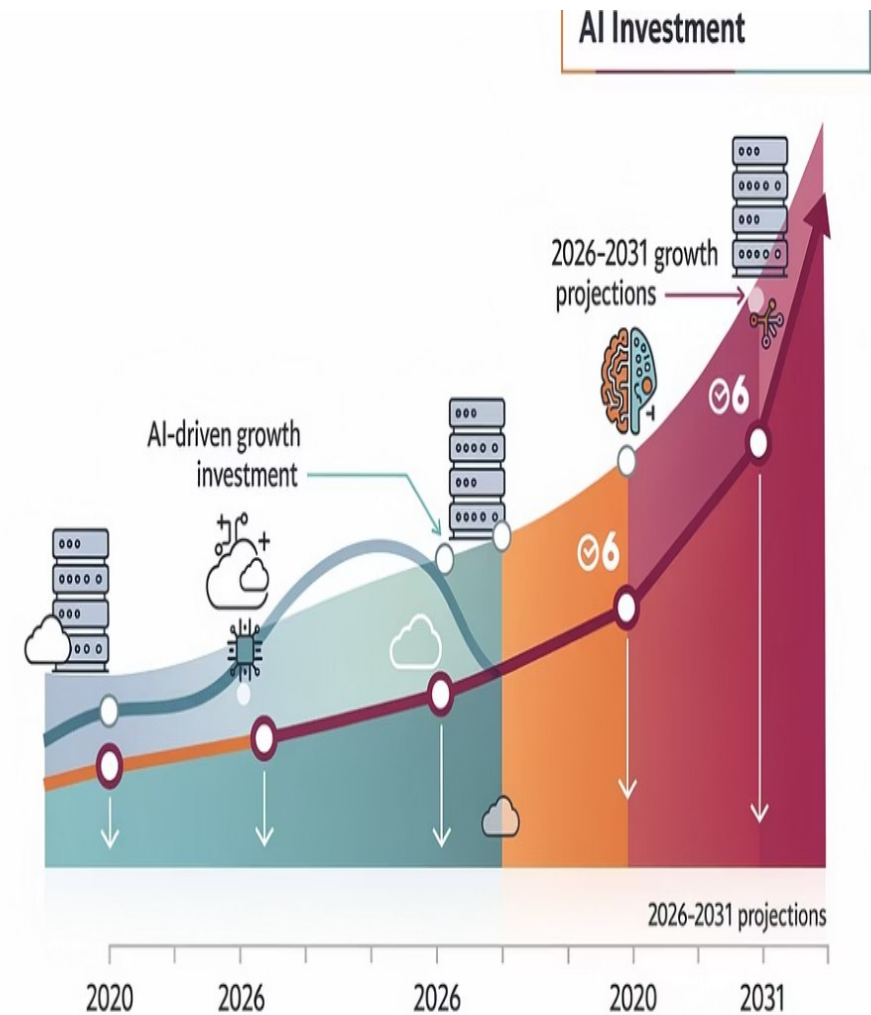
Purpose-built for compute-intensive AI workloads using GPUs/accelerators.

These require advanced cooling, high power density, and high-speed networking capabilities.

Data Center Investments Are Shifting Toward AI

Highlight: "While traditional Data Centres remain essential, nearly all new growth in Data Centre investment is now driven by AI."

- **AI-Driven Infrastructure:** GPU compute, high-speed networking, and advanced cooling are now the primary investment drivers, supporting cutting-edge AI workloads.
- **Future Growth in AI:** Over the next five years, most incremental data center spending will be AI-related, fueled by Generative AI, analytics, automation, and sovereign AI initiatives.
- **Strategic Capital Allocation:** Traditional data centre investments remain stable, while new capital is strategically directed towards accelerating AI capabilities and infrastructure.



Why Everyone Is Using AI Today

Why Everyone Is Using AI Today



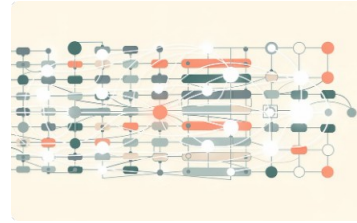
Artificial Intelligence (AI) Machine Learning (ML) Deep Learning (DL)

Perform tasks requiring human intelligence:

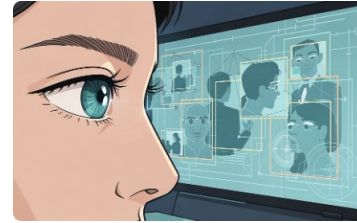
- Reasoning
- Learning
- Perception
- Decision making
- Interaction



Learn patterns from historical data to make predictions or decisions, used for classification, regression, and anomaly detection.



multi-layer neural networks to learn highly complex patterns: In image recognition, speech recognition, and video analysis.



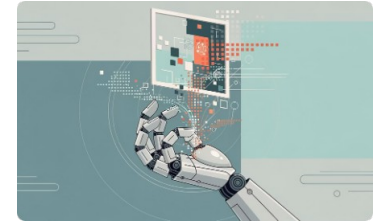
Computer Vision

Interprets visual data from cameras and images to enable object detection, facial recognition, image classification, video analytics, and quality inspection.



Natural Language Processing (NLP)

Processing and understanding human language, facilitating text understanding, translation, sentiment analysis, and speech-to-text conversion.



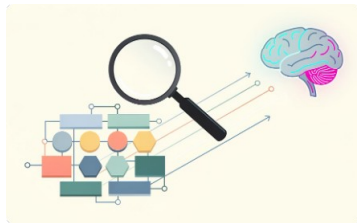
Generative AI (GenAI)

Generates original content based on learned patterns, ranging from text and images to video, music, and even code.



Large Language Models (LLMs)

Models trained on massive datasets to predict the next word, powering conversational AI, question answering, summarization, and coding assistance.



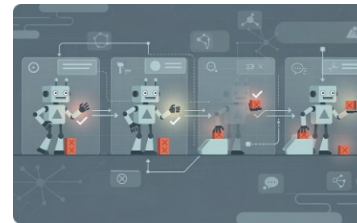
Retrieval-Augmented Generation (RAG)

Combines LLMs with enterprise or external data sources to provide context-aware, up-to-date, and accurate responses for knowledge assistants and enterprise Q&A.



Multimodal AI

Designed to understand and generate content across multiple modalities like text, images, audio, and video, enabling comprehensive cross-media reasoning.



Reinforcement Learning (RL)

Learns optimal actions through feedback and rewards, primarily used in robotics, game-playing AI, and dynamic optimization for autonomous systems.



Autonomous & Agentic AI

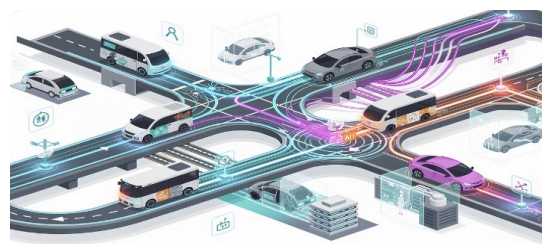
Refers to AI systems that can plan, decide, and execute complex tasks with minimal human intervention, driving workflow automation and multi-step reasoning.

AI Delivering Tangible Business Outcomes



Healthcare

- Assist doctors in diagnosing diseases from medical images and test results
- Automatically prioritize patients based on urgency
- Support medical research by analyzing large clinical datasets
- Optimize hospital operations such as scheduling and bed usage
- Analyze population health data to predict outbreaks and risks



Transportation

- Predict and detect incidents and accidents in real time using video analytics
- Optimize train schedules based on passenger demand and traffic patterns
- Automate traffic management, passenger flow and signaling optimization
- Improve customer experience with AI-powered passenger information and chatbots
- Simulate rail operations using digital twins to test scenarios before deployment



Government & Cities

- Optimize traffic lights and reduce congestion in real time
- Improve public safety, incidents detection and emergency response times
- Optimize waste collection and city services
- Monitor air quality, noise, and pollution
- AI chatbots for citizen services and requests
- Automate document processing and public administration



Education

- Digital Education platforms and personalized learning (adaptive learning platforms, AI tutors)
- Secure assessment and academic integrity (automated grading, plagiarism detection, secure online exams, and AI-based proctoring)
- Research and AI skills development (AI research labs, student AI sandboxes, AI training programs, and cross-disciplinary innovation platforms)
- Operational efficiency and cost optimization (Enrollment forecasting, timetable)

AI Delivering Tangible Business Outcomes



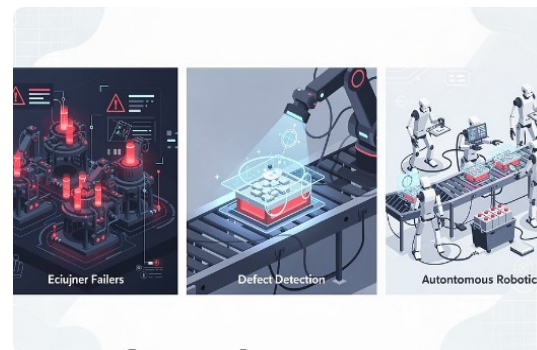
Finance

- Detect fraud and suspicious transactions in real time
- Improve risk assessment and credit scoring
- Optimize trading strategies using predictive models
- Provide instant customer support through AI chatbots
- Increase security while improving customer experience



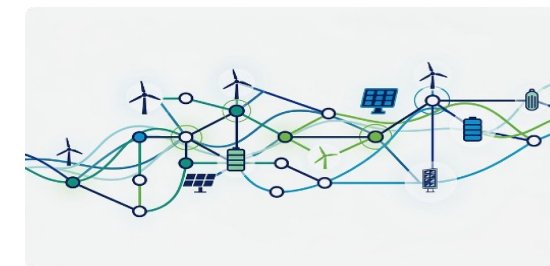
Retail & eCommerce

- Predict product demand to optimize inventory
- Automate replenishment and supply chain decisions
- Generate product descriptions and marketing content
- Provide personalized shopping recommendations
- Enable conversational shopping assistants



Manufacturing

- Predict equipment failures to reduce unplanned downtime
- Automatically detect defects during production
- Simulate manufacturing processes before physical deployment
- Enable smarter and more autonomous robotics systems
- Accelerate product design through AI-driven simulations



Energy

- Predict energy demand to balance supply and consumption
- Optimize smart grid operations in real time
- Maximize renewable energy production and usage
- Decide the best time to store or release energy
- Reduce energy waste through predictive analytics

Training vs Inference in an AI Data Centre

AI models undergo two critical phases: **training** and **inference**, each with distinct networking requirements.

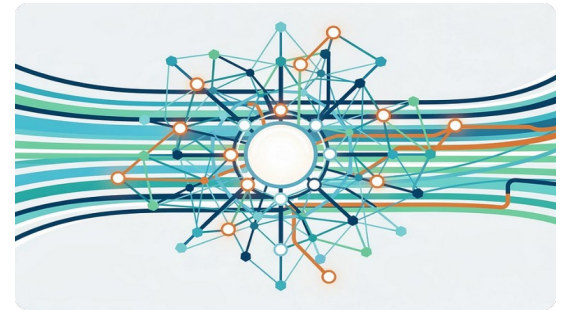
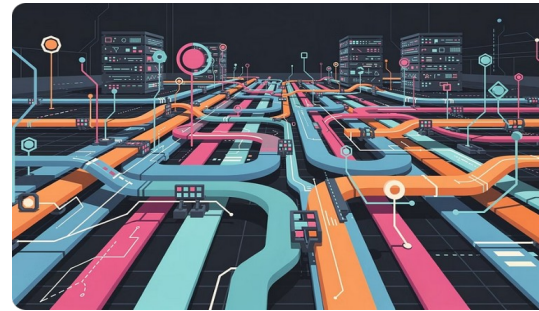
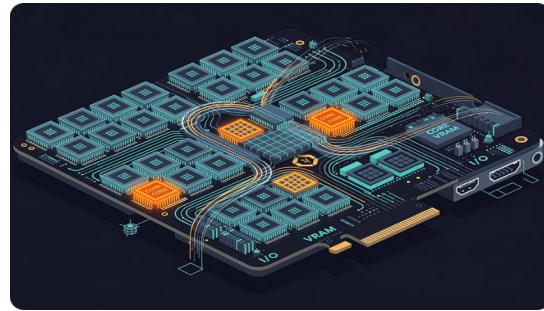
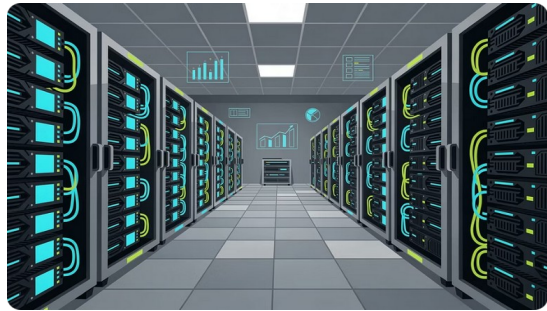
Training : involves feeding vast datasets to GPU clusters, generating extremely heavy **east-west** network traffic between GPUs and servers. This demands high-bandwidth, low-latency, and lossless networks for optimal GPU utilization.

Inference : conversely, deploys trained models for real-time predictions. It is typically more **latency-sensitive** but less bandwidth-intensive than training, focusing on rapid responses to users.



AI Needs Specialized Infrastructure

Traditional IT infrastructure cannot meet the unique demands of modern AI workloads.



Massive Compute

Training large AI models demands **millions of GPU hours** and cost tens of millions of dollars in compute resources.

Parallel Processing

GPUs accelerate neural network training **by 100-1000x** compared to CPUs.

Extreme Data Scale

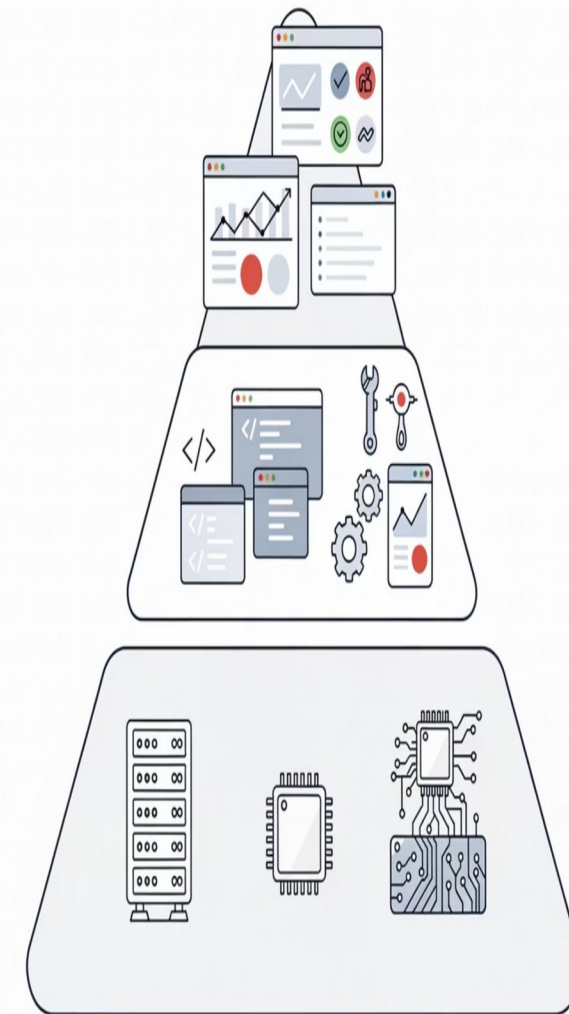
Models like GPT-5 train on **trillions of tokens** from hundreds of terabytes of text, requiring specialized data pipelines.

Low-Latency Inference

Production AI systems need to deliver predictions in **milliseconds** while handling millions of requests simultaneously

The Layered View of AI Infrastructure

- **Operations:** Transform experimental AI into production-grade systems with enterprise reliability
 - ✓ **CI/CD Pipelines:** Automated testing and deployment
 - ✓ **Monitoring & Observability:** Real-time metrics on model performance, drift detection, and infrastructure utilization
 - ✓ **Security & Compliance:** IAM Roles, encryption at rest/in transit, audit logs
- **Software:**
 - ✓ **Frameworks:** PyTorch, TensorFlow, JAX
 - ✓ **Acceleration Libraries:** ROCm, MIOpen, RCCL, CUDA, cuDNN, NCCL, oneDNN
 - ✓ **Containers & Orchestration:** Docker, Kubernetes
 - ✓ **MLOps Tools:** MLflow, Kubeflow, Weights & Biases
- **Hardware:**
 - ✓ **Compute** GPUs for matrix operations and parallel processing
 - ✓ **Storage** High-speed SSDs, distributed object stores, feature stores optimized for AI
 - ✓ **Networking** RDMAv2 lossless Ethernet



ALE AI-DC Offer



ALE AI Data Center offer



Computing

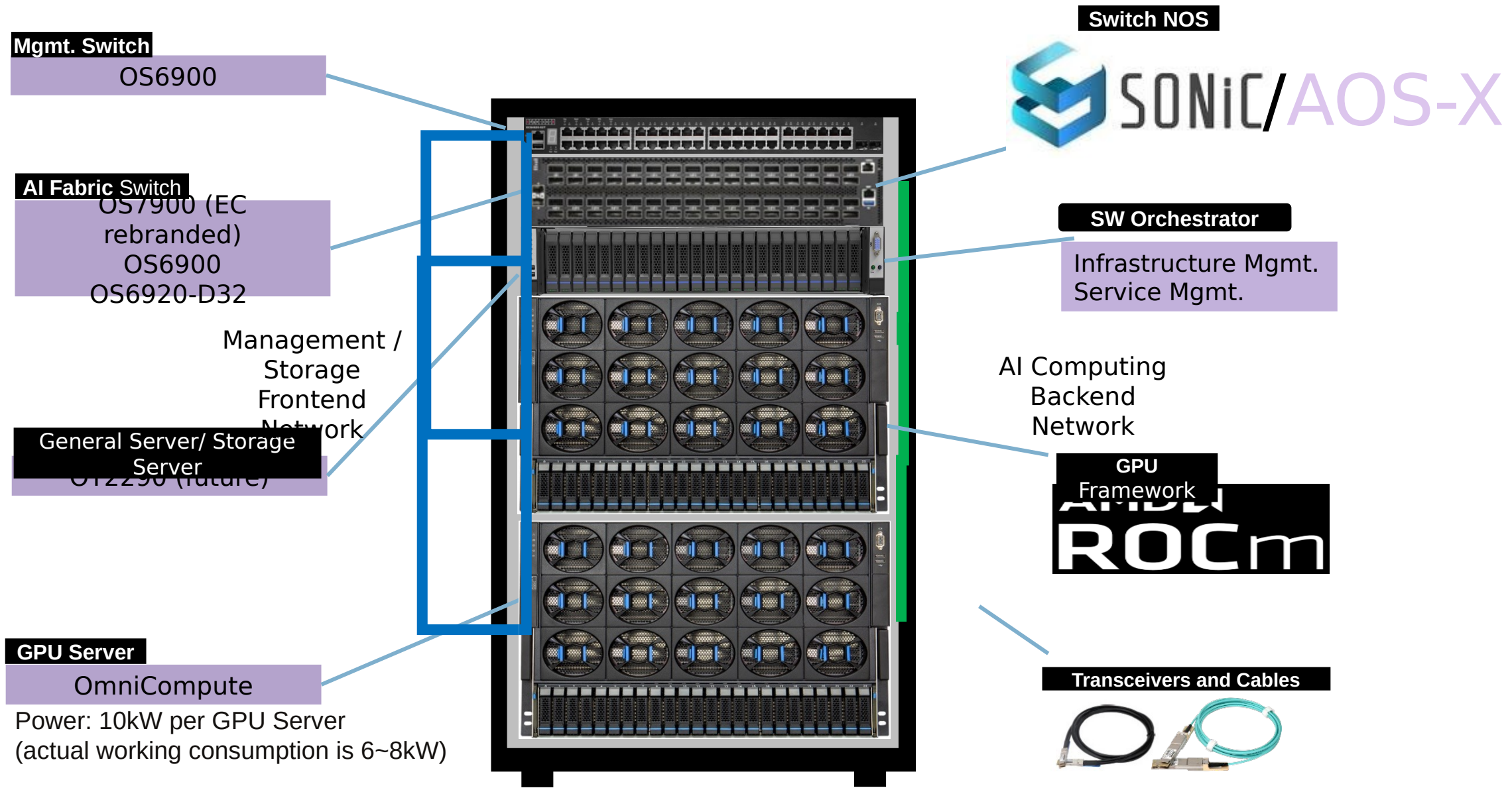
*AI server for training
Large Language Models
and inference*



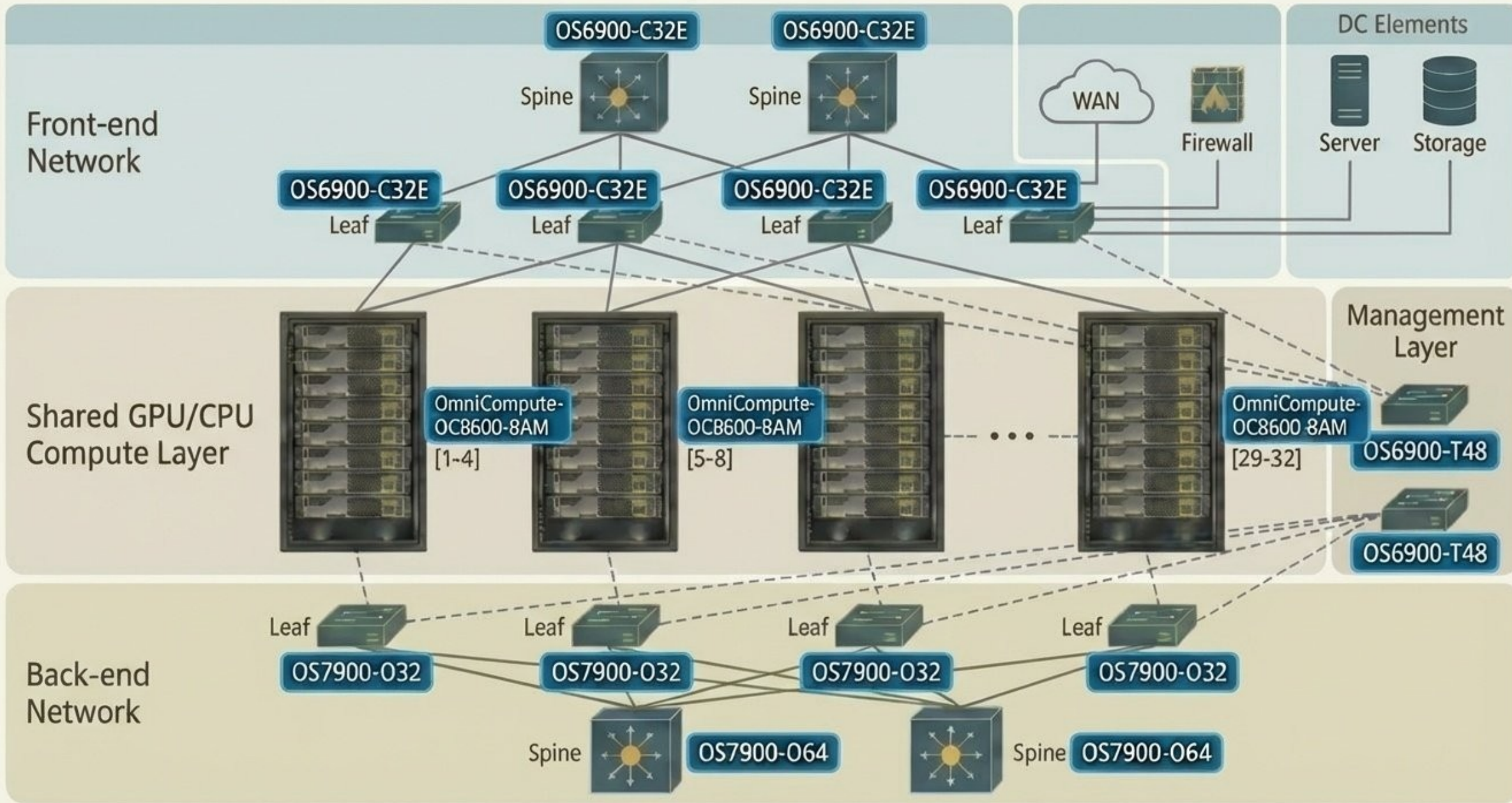
Networking

*Network infrastructure for
AI clusters running at full
speed, without delays or
bottlenecks*

ALE AI DATACENTER RACK COMPONENTS



AI-Data Center Networking [ALE-COMPONENTS]



OS7900-064 & OS7900-032 FOR BACKEND FABRIC



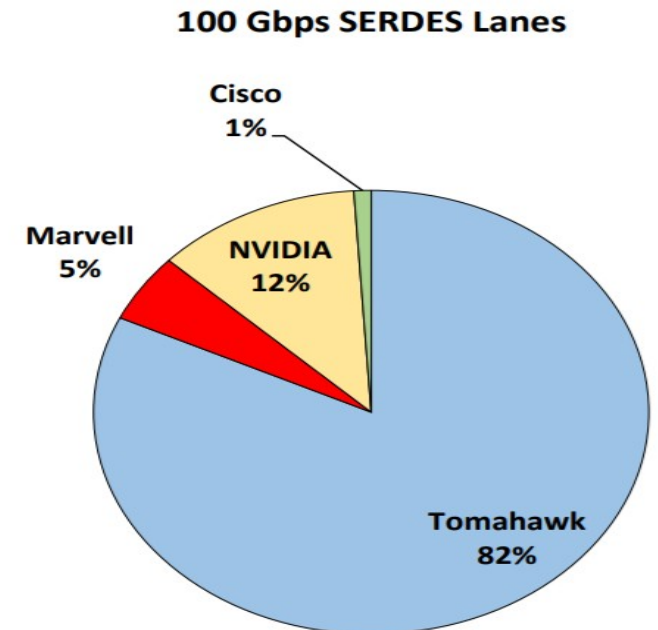
OS7900-064 (OSFP800 Ports)



OS7900-032 (OSFP800 Ports)



- Switching capability of **51.2 Tbps & 25.6 Tbps.**
(**High bandwidth, low latency & Advanced load-balance.**)
- 8 SerDes lanes per port @100Gbps each.
- Broadcom Tomahawk5(**TH5**) ideal for **AI workloads.**
- **LPO**(Linear Pluggable optics) with **low power consumption, reduced cost & lower latency.**
- OSFP **migration** to higher speeds of **1.6Tbps.**



OS7900-D64 & OS7900-D32 FOR BACKEND FABRIC



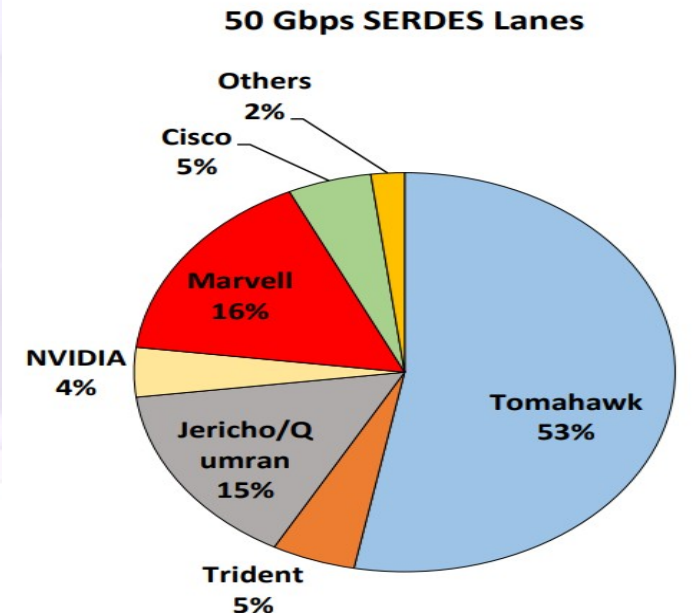
OS7900-64D (QSFPD400 Ports)



OS7900-32D (QSFPD400 Ports)



- Switching capability of **25.6 Tbps & 12.8 Tbps** 8 SerDes lanes per port @50Gbps each..
- Broadcom Tomahawk4(**TH5**) ideal for **AI workloads**
- **Dynamic load-balance.**
- Deployments for HPC/GPUs with lower bandwidth requirement.



ALE OMNICOPUTE [OC8600]



AMD MI325X OAM
CDNA3™
Architecture



Specifications:

- GPU: **AMD 8 x MI325X GPUs** with Infinity Fabric
- CPU: AMD 2 x EPYC™ 9005/Turin Series Processor
- CPU NIC: 2x BCM957608-P2200GQF00 **Dual-Port 200GbE**
- GPU NIC: 8x BCM957608-P1400GDF00 **Single-Port 400G**
- Storage: 2x1.92TB U.2 2.5 NVME4 1DWPD (non SED)
- Form Factor / Cooling: 8U Air Cooling
- FP32 TFLOPS: 163.4 FP16 TFLOPS: 1300 FP8 TFLOPS: 2610
- Compute Memory per GPU: **256GB** HBM3e
- Memory Bandwidth: **6 TB/s**
- Peak Power Consumption per GPU: 1KW
- Actual Working Consumption per Server: 6-8KW
- Compute Fabric Interface: Addition External NIC

MI325X vs H200 (Key Specs)

NVIDIA H200 SXM	AMD MI325X OAM
HBM3 Memory: 141 GB	256 GB
Memory Bandwidth: 4.8 TB/s	6 TB/s
TF32 Peak: 494.7 TFLOPs	658.7 TFLOPs
FP16/BF16 (Tensor/Matrix) Peak: 989.4 TFLOPs	1207.4 TFLOPs
FP8 Peak: 1978.9 TFLOPs	2614.9 TFLOPs

www.amd.com

ALE SOFTWARE FOR SWITCHES & NETWORKS



AOS-X

Dual NOS Strategy:

SONiC : ALE SONiC

- For large scale & cloud-native enterprise for **full control of stack**.

AOS-X :

- Provides **feature depth & maturity**.
- **Operational simplicity** for smaller teams/enterprises.

Key Features for NOS :

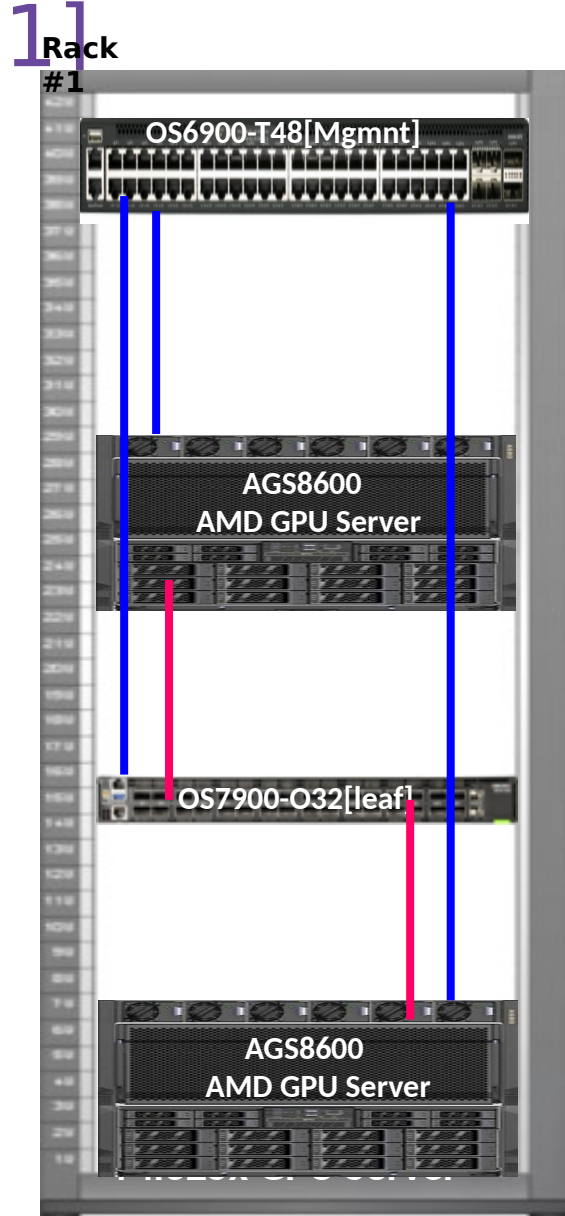
- **Modern NOS** principles.
- **Cut-through switching**.
- **Adaptive routing & dynamic load balancing**.
- **PFC**(Priority flow control),
- **RoCEv2**(RDMA over converged ethernet V2),
- **ECN**(explicit congestion notification).
- **Advanced telemetry**.
- **VXLAN-EVPN** for multi-tenancy.



Key Features for Orchestrator :

- Rack orchestration for backend, frontend & management network.
- Provide Industry standard Day-0, Day-1 & Day-2 operations.
- Intent based network provisioning
- Network Source of truth(NSOT) to maintain the integrity of the network.
- Integrated digital twin
- Monitoring & Maintenance using advanced telemetry.

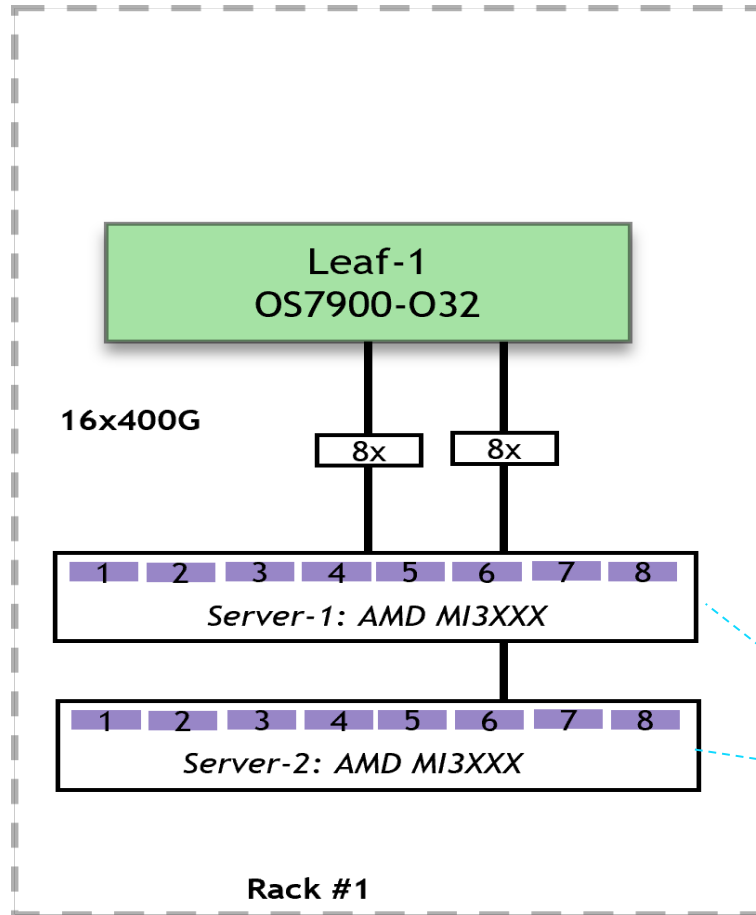
GPU BACKEND FABRIC: [16 GPU] [LAB-BLUEPRINT-1]



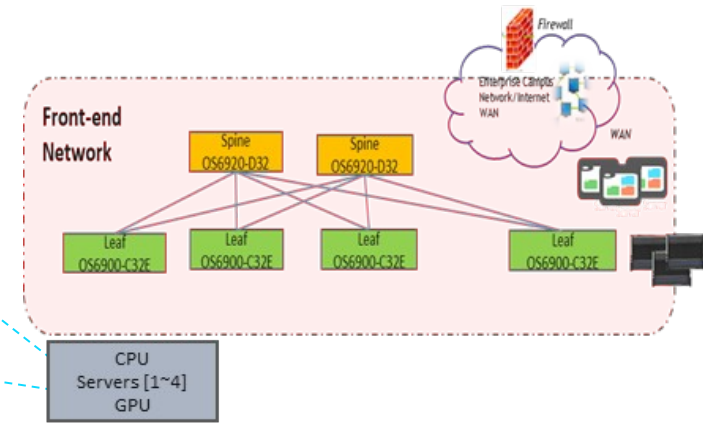
IN RACK AI FABRIC:
400G/800G DAC

Mngmt fabric:
1G/10G

#GPUs: ~16
#GPU Servers: 2
#Rack : 1
#Switches: 1



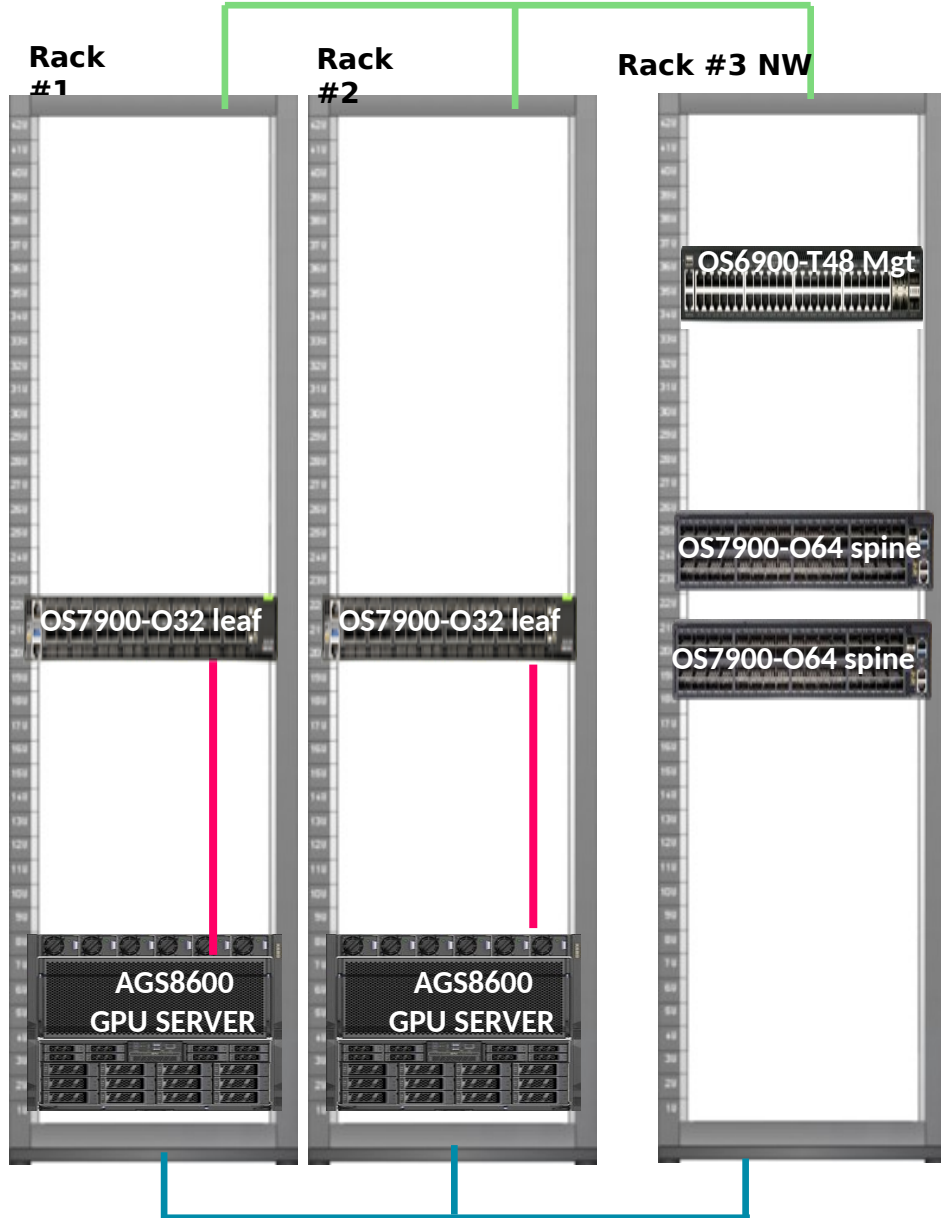
1. Turnkey Single-Rack AIDC solution with two compute servers.
2. Servers, backend network & management share the same rack.



CPU of server connects to front-end network.

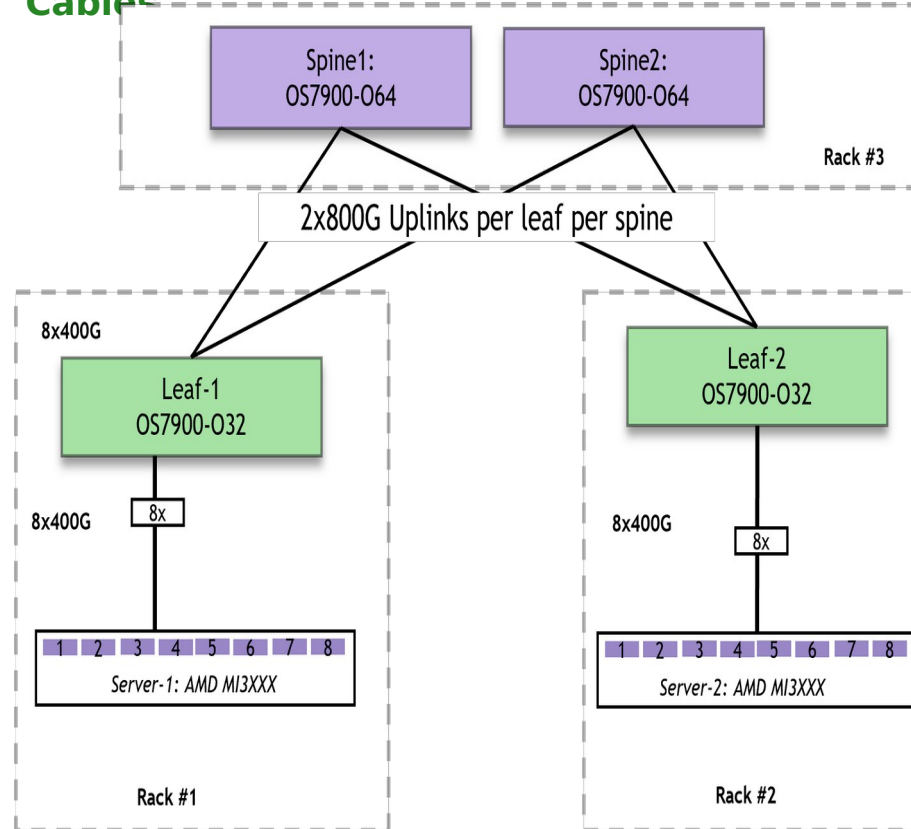
GPU BACKEND FABRIC: [16 GPU][LAB-BLUEPRINT-2]

#GPUs: 16
 #GPU Servers: 2
 #Rack : 3
 #Switches: 4



IN RACK AI FABRIC:
 400G/800G DAC
INTER RACK AI FABRIC:
 400G/800G AOC/Optical Cables

Mngmt fabric:
 1G/10G

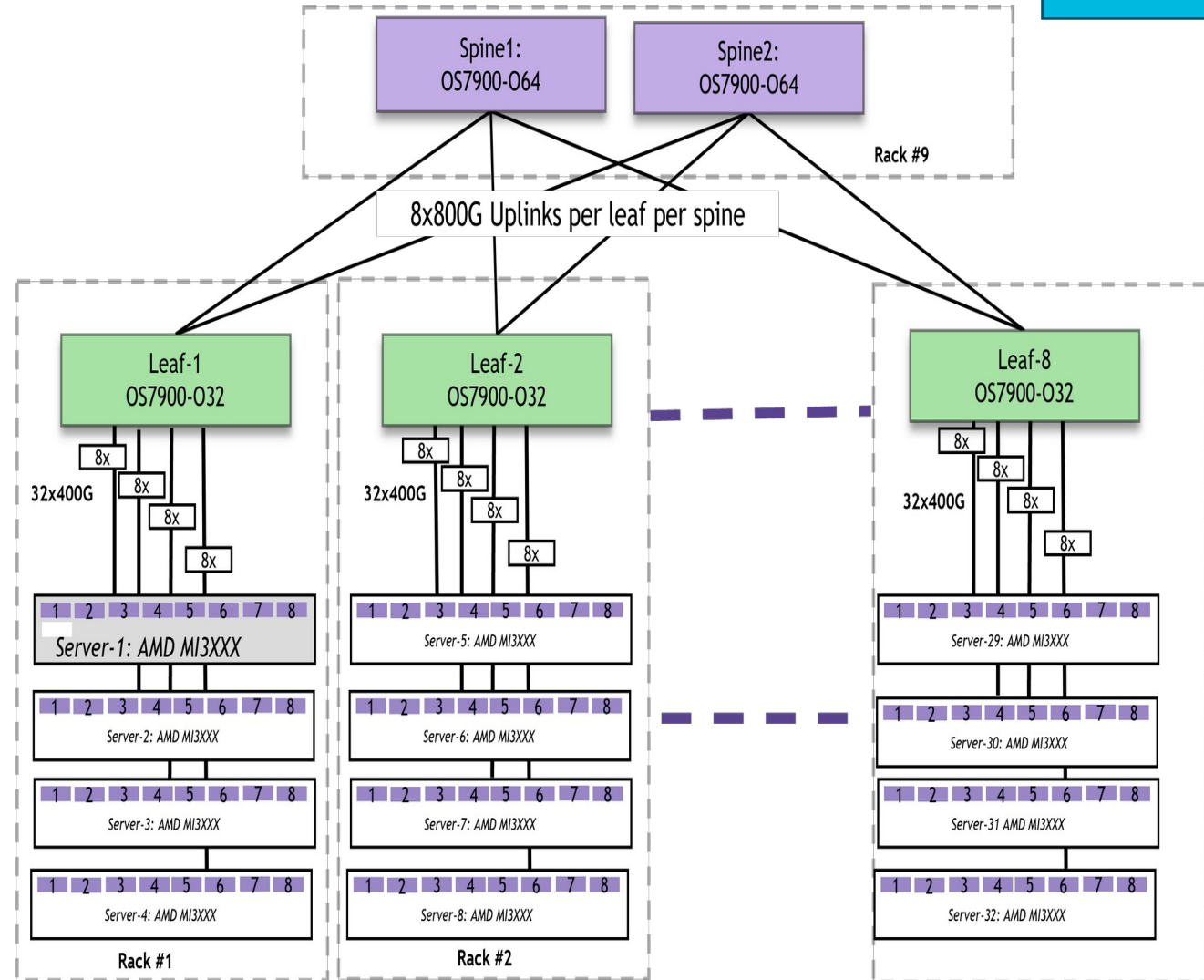
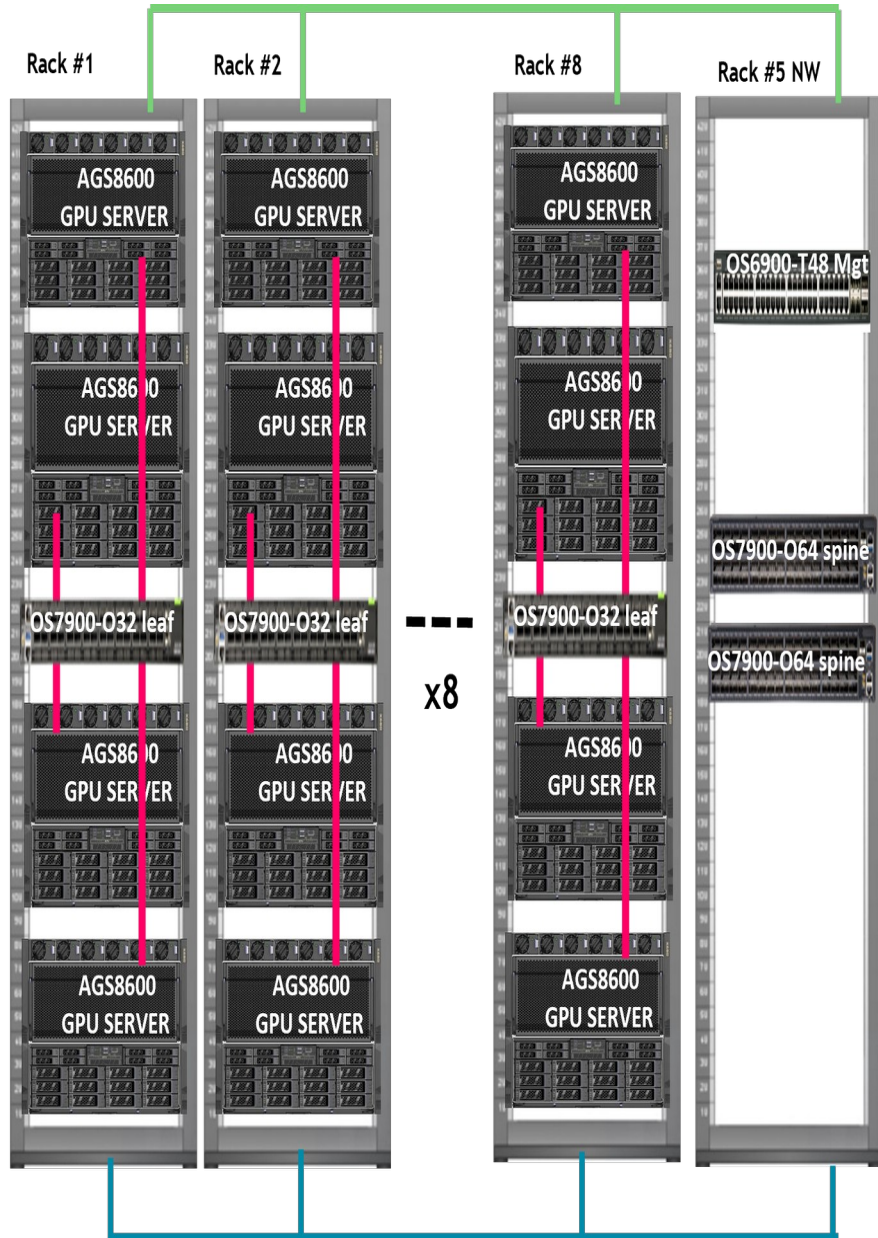


1. Dual-rack AIDC solution scalable to 8 racks.
2. Separate Network Rack for Mgmt/Spine.
3. DAC for intra-rack.
4. AOC/Optical cables for inter rack.

GPU BACKEND FABRIC: CLOS DESIGN[256 GPU

#GPUs: 256
 #GPU Servers: 32
 #Rack : 9
 #Switches: 10

Scales to 8 racks.



BOM AND PRICE-LIST

Product	Description	List Price
GPU Server	8x MI325x CPUs, NIC Card included	\$300,000
800G Switch	64x 800G OSFP Backend switch as spine.	\$132,000
800G Switch	32x 800G OSFP Backend switch as leaf	\$110,000
800G to 800G Optics	800G SR8 Transceiver (50m)	\$879
800G to 2x400G DAC	800G to 2x400G dac for leaf to server, 3m.	\$300
Management switch	6900-T48 to connect to backend nw devices.	\$34,275
NOS	SONiC(start zero but go to license in future).	\$0
Orchestrator	Device Management software	\$7000

Blueprint type	ASP Price
Blueprint-1(1-rack)	\$753,675
Blueprint-2(2-rack)	\$1,141,739

ALE AI-DC SOLUTION MILESTONES / PHASES

ACCOMPLISHMENTS and MILESTONES

PHASE 1

- Scope:
 - OEM offer with minimal Branding on switches.
 - ALE logo on overlay, Model name
 - ALE Login banner in NOS.
 - ALE logo on silk screen in Mgt. software
 - Documentation
 - HW/SW/CLI/Datasheet modified to ALE.
- Progress:
 - Requirement established.
 - DR0: Completed on 26th Jan
 - DR1: Planned this week
 - DR4: Planned on 26th Feb

PHASE 2

- Scope:
 - Transition to ALE-NOS in front-end with AOS-X
 - Transition to ALE SONiC in back-end
- Progress:
 - DR1: Completed on Dec 2025
 - DR4: Planned July 2026

PHASE 3

- Scope:
 - Dual NOS on back-end: AOS-X & ALE SONiC
 - ALE brand optics for 800/400G
 - Transition to ALE mgmt. software

Feb
2026

July
2026

Dec
2026

PHASE-WISE ROADMAP

PHASE 1: Technical Proof of Concept / MVP

- Back-end network : OEM 800G / 400G, OEM EC SONiC
- Front-end network : ALE 100G/25G, OEM EC SONiC
- Management network: ALE 10G,
- Optics : ALE (100G, 400G) OEM (400G, 800G)
- Compute : Edgecore AMD Server (MI325x GPU)

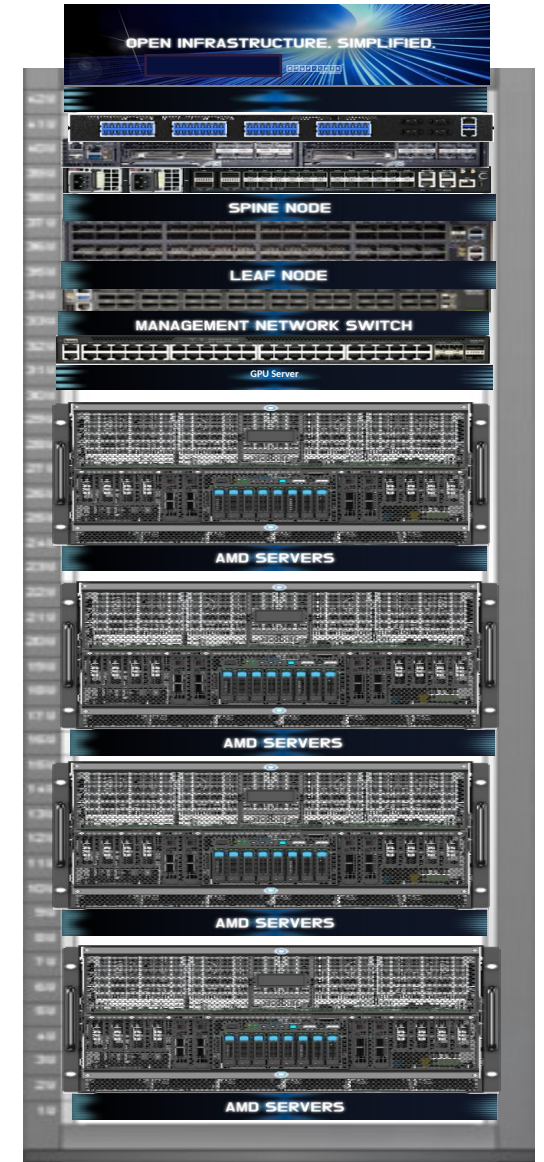
- Rack mgmt. SW : OEM EDM v2.3

PHASE 2: Solution Launch

- Front-end network: +ALE 400G, AOS 10.1R2
- Back-end network: ALE SONiC
- Rack mgmt. SW: OEM EDM v2.4

PHASE 3: SCALE

- Front-end network: AOS 10.1R3
- Back-end network: AOS 10.1R3, ALE SONiC
- Optics : ALE optics 10/100/400/800G
- Rack mgmt. SW: ALE OVCompass



GO TO MARKET



Go-to-Market Strategy

Focus Countries:

- **Tier 1:** *France, Germany, UK, Nordics, UAE, Saudi Arabia*
- **Tier 2:** Rest of EUSO/EUNO/MEITA Regions and APAC
- **Tier 3:** USA/CANADA and CALA

Target Customer:

Enterprise, Private, and Regional Cloud AI DC opportunities (Tiers 2-3-4)

Sales & Channel Strategy:

- **Phase 1:** Indirect-only model through DC specialized ALE existing DR and VAD
- **Phase 2:** Direct sales to strategic customers where permitted by regulation and no channel conflict
- **Partner Recruitment:** Proven DC expertise and active Government, DCSP, and Telco relationships

Customer Profile

Sovereign / National AI Cloud Projects:

- Ministries
- Government-managed - cloud initiatives
- National AI clusters
- Sovereign data centers

Enterprise AI Transformation DCs

- Banks
- Energy/Utilities
- Telcos
- Aviation
- Manufacturing
- Healthcare
- Universities/ Research centers
- Transportation

Tier 2-3 Cloud Providers

- Regional cloud companies
- Managed service providers
- Emerging AI cloud platforms
- Colocation + cloud hybrid players

AI-First Companies

- AI startups
- HPC/
- Data analytics firms
- GenAI platform companies

Enterprise AI DC (ME)

	Vertical	Enterprise Example Type	Why Enterprise AI DC
UAE	Energy	ADNOC	Reservoir & seismic modeling AI sovereign
UAE	Aviation	Emirates Group	Ops AI + maintenance advantage
UAE	National AI	G42 Gov workloads	Sovereign AI DC
UAE	Healthcare	M42 Genome	Genomics AI must remain sovereign
UAE	Banking	FAB	AML + risk models remain private
SAUDI	Energy	Aramco	Reservoir modeling IP strategic national asset
SAUDI	Petrochemicals	SABIC	Material innovation AI IP protected
SAUDI	Gov Data	DGA/NDMO	National data sovereign AI
SAUDI	Smart City	NEOM	Sector AI DC for next gen smart region
SAUDI	Banking	Saudi National Bank	Financial AI private by design

Sovereign / National AI Cloud DC

	Example Sovereign Actor(s)	Why Sovereign / National AI Cloud DC
FR	GENCI / CEA / SecNumCloud certified providers (3DS Outscale / OVH Sovereign)	National HPC + AI model capacity under French jurisdiction. Strategic science AI + national data residency by law.
GER	JUPITER Exascale / Federal AI Research Centers (FZJ Jülich)	Exascale AI with national control — national industry AI + research compute protected.
UK	UKAI Compute (AI Research Resource) / National AI Research Resource (NIRR)	National AI cluster for AI R&D capacity - UK regulated, sovereign model training environment not public cloud multi-tenant.
UAE	National AI Authority programs + Core42 national compute + Khazna sovereign partitions	National AI model training + national data security priority. Sovereign DC strategy for next gen AI economy.
SAUDI	PIF backed national AI DCs (Humain initiative) / SDAIA national data compute	National strategic compute = sovereign asset. AI cluster capacity aligned to national transformation programs Vision 2030.

Tier 3 AI Cloud DC

	Tier-3 Local / National Private Cloud Provider Example Type	Why Tier-3 (Local Private Cloud)
FR	OVHcloud Enterprise Private Cloud instances (non-SecNumCloud) / Scaleway	Country & EU focused cloud. Limited global footprint. Commercial private cloud hosting for regulated enterprise workloads.
GER	IONOS Enterprise Cloud / PlusServer	National + EU compliance private cloud — but not AI specialized Tier-2 and not sovereign HPC state controlled.
UK	UK national telco private hosted cloud (BT Business / Vodafone Business Cloud)	Local private cloud sold commercially to UK enterprises — not global hyperscale, not sovereign designated workload.
UAE	Etisalat/DU Business Cloud private cloud offers	Country-bounded hosting for UAE enterprise workloads. Not national AI authority and not regional Tier-2 GPU capacity builder.
SAUDI	STC Cloud Enterprise hosting / Zain KSA Enterprise Cloud	National telco cloud DC used commercially by enterprises in the country. Not sovereign mandated DC and not hyperscale / Tier-2 GPU provider.

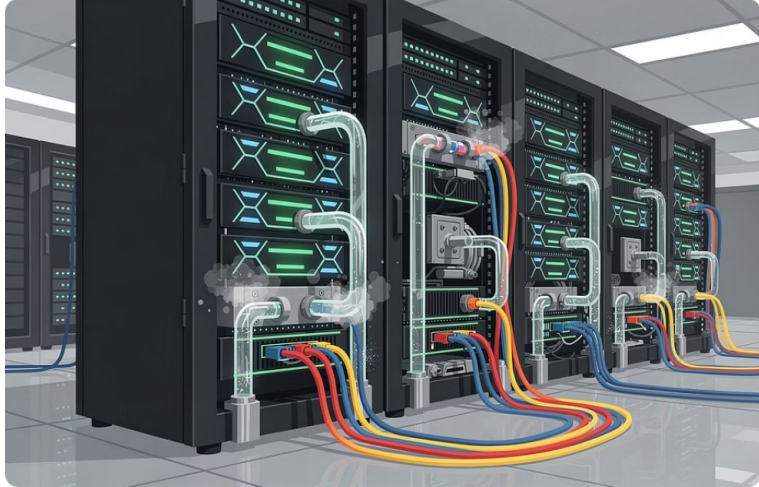


#WhereEverythingConnects

Why Everyone Is Using AI Today

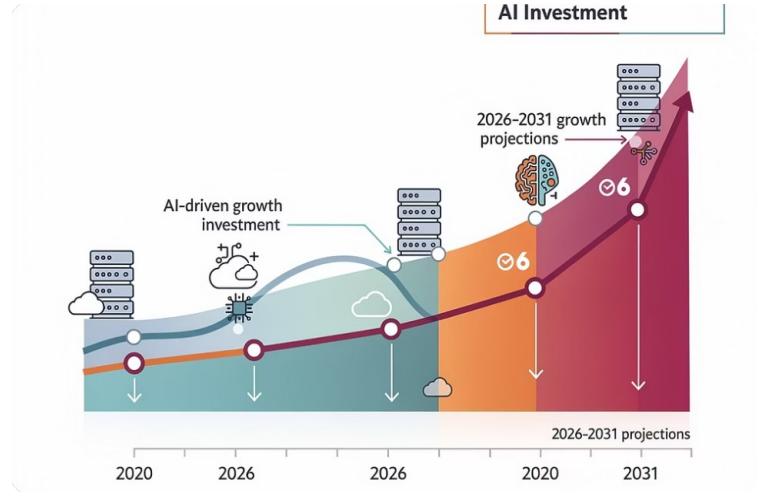
AI Domain	PyTorch	TensorFlow	Industry Reality
Classical ML	✓ (via PyTorch + scikit-learn)	✓ (TF + Keras + TFX)	Both widely used
Deep Learning (DL)	✓	✓	Both core frameworks
Computer Vision	✓ (torchvision, Detectron2, YOLO, OpenCV)	✓ (TF Vision, KerasCV)	PyTorch dominant in research & SOTA
NLP	✓ (Hugging Face, spaCy, Fairseq)	✓ (TF Text, HF TF backend)	PyTorch dominant
GenAI	✓	△ Limited	PyTorch is the de-facto standard
LLMs	✓ (Transformers, vLLM, DeepSpeed, Megatron)	△ Rare	PyTorch overwhelmingly standard
RAG	✓ (LangChain, LlamaIndex, Haystack)	△ Indirect	PyTorch ecosystem dominant
Multimodal AI	✓ (CLIP, LLaVA, Flamingo-like stacks)	△ Limited	PyTorch dominant
Reinforcement Learning (RL)	✓ (RLlib, Stable-Baselines3, CleanRL)	△ Limited	PyTorch dominant
Agentic AI	✓ (LangGraph, AutoGen, CrewAI, OpenAI SDKs + PyTorch models)	△ Rare	PyTorch dominant

Data Center Investments Are Shifting Toward AI



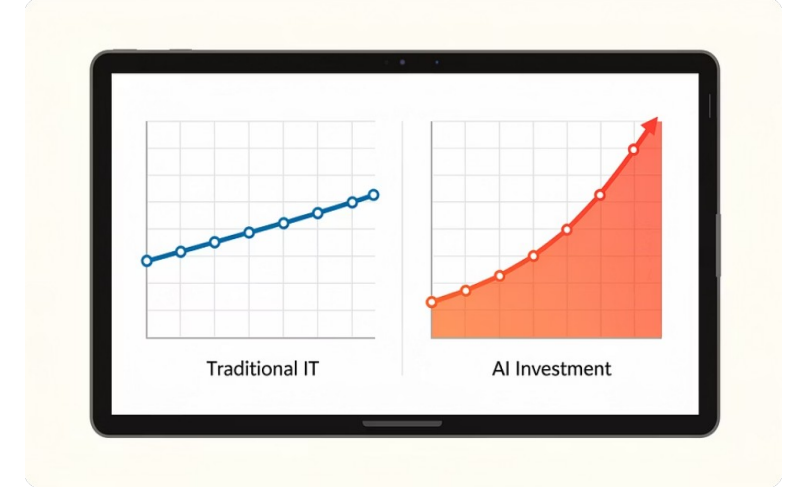
AI-Driven Infrastructure

GPU compute, high-speed networking, and advanced cooling are now the primary investment drivers, supporting cutting-edge AI workloads.



Future Growth in AI

Over the next five years, most incremental data centre spending will be AI-related, fuelled by Generative AI, analytics, automation, and sovereign AI initiatives



Strategic Capital Allocation

Traditional data centre investments remain stable, while new capital is strategically directed towards accelerating AI capabilities and infrastructure

Highlight: "While traditional Data Centres remain essential, nearly all new growth in Data Centre investment is now coming from AI-related spending"

Why Data Centers Are ALE Strategic New Focus

Portfolio diversification and business transformation:

Expanding into Data Centers allows us to diversify away from technologies that are becoming legacy, commoditized, and lower-value, and reposition the company toward strategic, future-proof growth areas.

Stronger economics and long-term value creation:

The Data Center segment offers structurally larger deal sizes, longer customer lifetime value, and multi-year engagements, supporting more predictable and sustainable revenue.

Increased strategic relevance with enterprise customers:

Data Centers sit at the core of digital transformation initiatives (AI, cloud, sovereignty), enabling earlier engagement with C-level decision makers and positioning us as a long-term strategic partner.

Enterprise AI DC (Europe)

	Vertical	Enterprise Example Type	Why Enterprise AI DC
FR	Banking	BNP Paribas / Société Générale	Sovereign + AML + Risk AI must be private
FR	Aerospace	Airbus	Safety critical simulation + IP protection
FR	Automotive	Renault Group	AI for manufacturing + digital twin
FR	Pharma	Sanofi	R&D AI drug discovery data sensitivity
FR	Energy	EDF	Critical national infra + sovereign AI
GER	Automotive	BMW / Mercedes-Benz / VW	Autonomous & simulation IP must remain sovereign
GER	Industrial	Siemens	Industrial AI + factory digital twin
GER	Pharma	Bayer	Drug discovery AI is core R&D IP
GER	Energy	E.ON / RWE	Critical infrastructure optimization AI
GER	Banking	Deutsche Bank	Internal fraud/risk AI must be internal
UK	Pharma	GSK / AstraZeneca	Drug discovery AI IP protected
UK	Telecom	BT	Network AI optimization private
UK	Defense	MoD/DSTL	National security AI
UK	Media	BBC R&D	Generative Media AI IP advantage

Tier 2 AI Cloud DC

	Tier-2 Specialized Provider Example Type	Why Tier-2 (Specialized GPU AI Cloud)
FR	OVHcloud AI GPU Clusters / NexGen Cloud (EU)	multi-regional GPU cloud expansion model, AI compute specialization, faster provisioning vs hyperscalers
GER	Northern Data (EU HQ)	EU scale GPU compute provider; not national telco, not hyperscaler; AI compute capacity sellers
UK	Vultr GPU PoPs (London) / DigitalOcean (Paperspace-origin)	GPU cloud builder expansion into UK; specialized capacity provider not full cloud hyperscaler
UAE	Vultr GPU Dubai / regional Tier-2 GPU startup footprints	Tier-2 providers expanding into GCC region as GPU capacity sellers for enterprises needing faster access
SAUDI	Tier-2 GPU entrants (regional PoP builds) / EU Tier-2 rented into KSA	“bring-in” Tier-2 GPU clusters into Saudi via regional PoPs → specialized GPU supply, not hyperscaler, not sovereign