

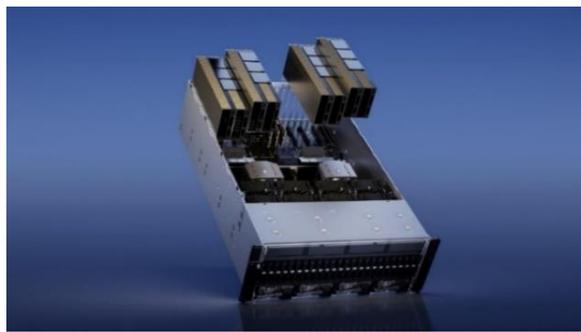
Alpha Helix Asset Management 投資備忘錄 (五十三)

2023.07.31

在 AI 浪潮下的推動下，AI GPU 的發展

AI GPU 的興起

最初 GPU (圖形處理器) 主要用於圖形渲染，以處理和顯示圖像，隨著 AI 和深度學習的發展，GPU 在 AI 領域的地位越來越重要，因為深度學習模型涉及大量的矩陣計算和向量運算，GPU 平行計算的能力發揮關鍵作用，滿足高算力需求。因此，近年 AI GPU 被廣泛應用於機器學習和深度學習等領域，加速模型的訓練和推理過程。



(圖一) NVIDIA H100 Tensor Core GPU

Source : NVIDIA

根據研究顯示，AI GPU 的潛在市場 (TAM) 在 2023 年-2025 年將以複合年增長率+54%成長至 538 億美元 (見表一)。

	2020	2021	2022	2023E	2024E	2025E
Server shipment (mn)	12.2	14.5	15.4	14.6	15.4	16.3
GPGPU Servers %	0.8%	0.8%	0.9%	1.4%	2.2%	3.0%
GPGPU server shipment (mn)	0.10	0.11	0.13	0.20	0.34	0.49
Avg GPU per server	3.4	3.8	6.9	7.6	7.0	6.6
AI Server GPU (mn)	0.3	0.4	0.9	1.5	2.4	3.2
GPU ASP (USD)	7,600	8,360	9,196	14,714	15,891	16,685
Cloud GPU TAM (US\$M)	2,613	3,517	8,402	22,784	37,353	53,857

(表一) GPU 市場規模

Source : HTI

AI GPU 的市場概況

➤ 受限於封裝產能，AI GPU 供不應求

雖然 AI GPU 需求強勁，但是受限於台積電 CoWos 先進封裝產能不足，AI GPU 供

不應求，導致單價上漲且交期延長數月；儘管台積電近期積極擴充先進封裝產能，短期內仍難以改善供不應求的市況。

➤ NVIDIA 具有獨占地位

NVIDIA 在 GPU 的硬體和軟體生態圈（CUDA 開發平台）皆掌握先行者優勢，成熟且完善的 CUDA 平台，降低 GPU 的編譯難度，用戶黏著度高，形成強大的護城河。目前在 AI GPU 市場持有 90% 以上的市場份額，多家雲端服務供應商業者（CSP）皆採用 NVIDIA A100/H100 GPU，如 Microsoft、Meta 等（見表二）。

業者	Microsoft		Google		Meta	AWS	Dell	NVIDIA	
伺服器生產模式	ODM-direct		ODM-direct		ODM-direct	ODM-direct	OEM	OEM	
AI 伺服器類型	OAM型	PCIe型	PCIe型	OAM型	OAM型	OAM型	OAM型	OAM型	MCM型
CPU/基板	2 x Intel Eagle Stream	2 x Intel Eagle Stream	無	2 x Intel Eagle Stream	1 x SXM GH200				
加速器/基板	8 x SXM H100 /A100	8/10/16 x PCIe H100	4 x TPU v4	8 x SXM H100 /A100	8 x SXM H100	8 x SXM H100	8 x SXM H100 /A100	8 x SXM H100 /A100	1 x SXM GH200
L10 ~ L12 供應商	ZT System	美超微	Celestica	Celestica、廣達	緯穎、廣達	ZT System	緯創	緯創	緯創

(表二) CSP 業者大多採用 NVIDIA 高階 GPU

Source : DIGITIMES

➤ NVIDIA 產品價位高，客戶積極尋求其他供應源

NVIDIA 掌握訂價權，AI GPU 產品毛利超過 80%，也讓 CSP 業者有尋找第二供應商與自製 ASIC 的動機，目前包括微軟、Meta、Google 等公司都宣布自行研發 AI 相關的 ASIC 晶片。

而 AMD 成為 AI GPU 的第二供應商，今年推出 MI300 GPU 系列新產品，將在年底開始陸續供貨給微軟、HPE 等廠商，預計明年將取得 AI GPU 個位數的市場份額。

結語

隨著機器學習和深度學習的快速發展，AI GPU 的市場需求和規模快速擴大，目前由 NVIDIA 主導市場，然在產品供不應求、價格居高不下的情形下，客戶也積極尋求其他供應源，像是尋求第二供應商、自製 ASIC 晶片等，未來 NVIDIA 主宰 AI 加速晶片的競爭格局能否延續下去，將取決於競爭對手軟硬體的整合度與客戶對於高價 AI 加速晶片的成本效益考量。