生成式AI風潮下，2024年全球運算系統發展

魏傳虔
產業顧問兼組長
產業情報研究所
財團法人資訊工業策進會
2024.01.17

簡報大綱

- 2024年全球運算系統市場發展
  - PC市場
  - 伺服器市場
- 熱門議題觀測
  - AI PC/NB新興應用
  - AI伺服器市場發展
  - 美中科技禁令對RISC-V之影響
- 結論

全球總體經濟預測持續向下修正

全球經濟成長率,2023-2024

公布機構（公布時間）	2023(e)	2024(f)
World bank (2023/06)		2.4
OECD (2023/11)	2.9	2.7
IMF (2023/10)	3.0	2.9
WTO (2023/10)	2.6	3.2→3.3
美國(OECD, 2023/11)	2.4	1.5
中國大陸(OECD, 2023/11)	5.1	4.7

由於各國中央銀行持續推行**貨幣緊縮政策**和更嚴格的**信貸條件**，預計這些因素將減少企業和住宅投資，並可能會在2023年下半年及2024年進一步**放緩經濟成長**

2023年迄今為止，GDP成長高於預期，但由於**金融狀況收緊**、**貿易成長疲軟**以及企業和消費者**信心下降**，目前成長放緩。近期前景面臨的風險仍偏於下行，包括**地緣政治緊張**局勢加劇

全球復甦依然緩慢，地區差異日益擴大，政策失誤空間不大

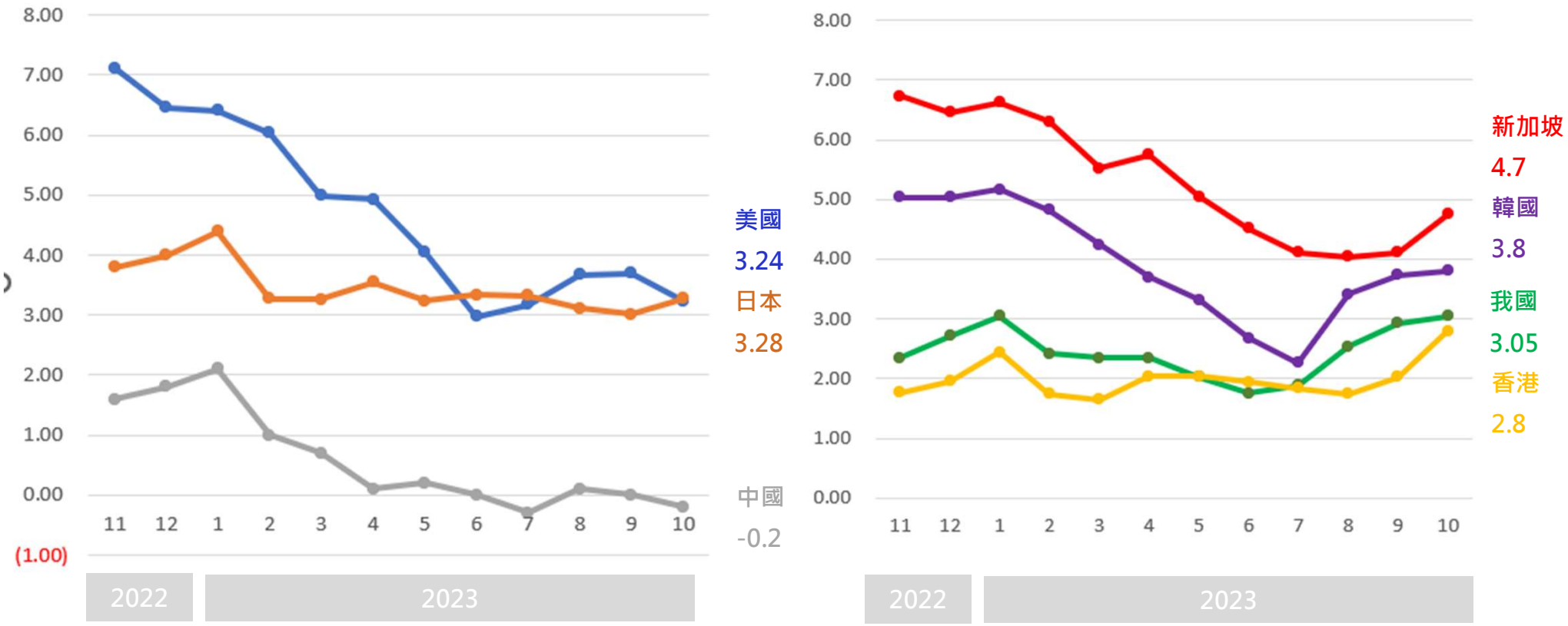
全球經濟一直在應對不斷上升的**通膨和高利率**，特別是在歐盟和美國。儘管能源價格下跌和中國大陸COVID-19大流行使景氣快速反彈，但**房地產市場**和持續不斷的**烏克蘭衝突**也持續為全球經濟帶來壓力

資料來源：各單位，MIC整理，2023年12月

- 整體金融狀況的收緊、貿易成長疲軟和企業及消費者信心下降導致當前經濟成長放緩；而地緣政治緊張局勢加劇，如烏克蘭衝突持續，則增加經濟前景的不確定性
- 全球經濟持續應對高通膨和高利率的挑戰，其中美國和中國大陸對於2024年的景氣前景表現，預估將弱於2023年

多數國家(地區)通膨壓力趨緩

主要國家消費者物價指數CPI年增率(%)



資料來源：經濟部統計處(2023/12/20)，2023年12月

- 國際油價下跌，使能源價格有較明顯回落，核心通膨持續下降，且服務相關細項皆有所下滑。美國CPI年增率3.24%，相較於去年年底減少49.78%；日本3.28%，減少17.95%；中國-0.2%，減少111.11%；韓國3.8%，減少24.37%；新加坡4.7%，減少26.5%，台港CPI年增率分別為3.05%與2.8%，由於基期相對較低，分別相較去年年底增加12.55%與42.77%

全球PC市場發展

PC市場需求影響因素

正向因素

需求

- PC換機週期來臨，特別是在商用市場
- 歐美及新興市場買氣有望恢復增長
- 電競遊戲市場持續擴增
- 教育標案啟動較大規模換機

供給

- AI PC陸續問世
- Windows 10停止支援
- Intel、AMD、Qualcomm等新處理器平台上市

負向因素

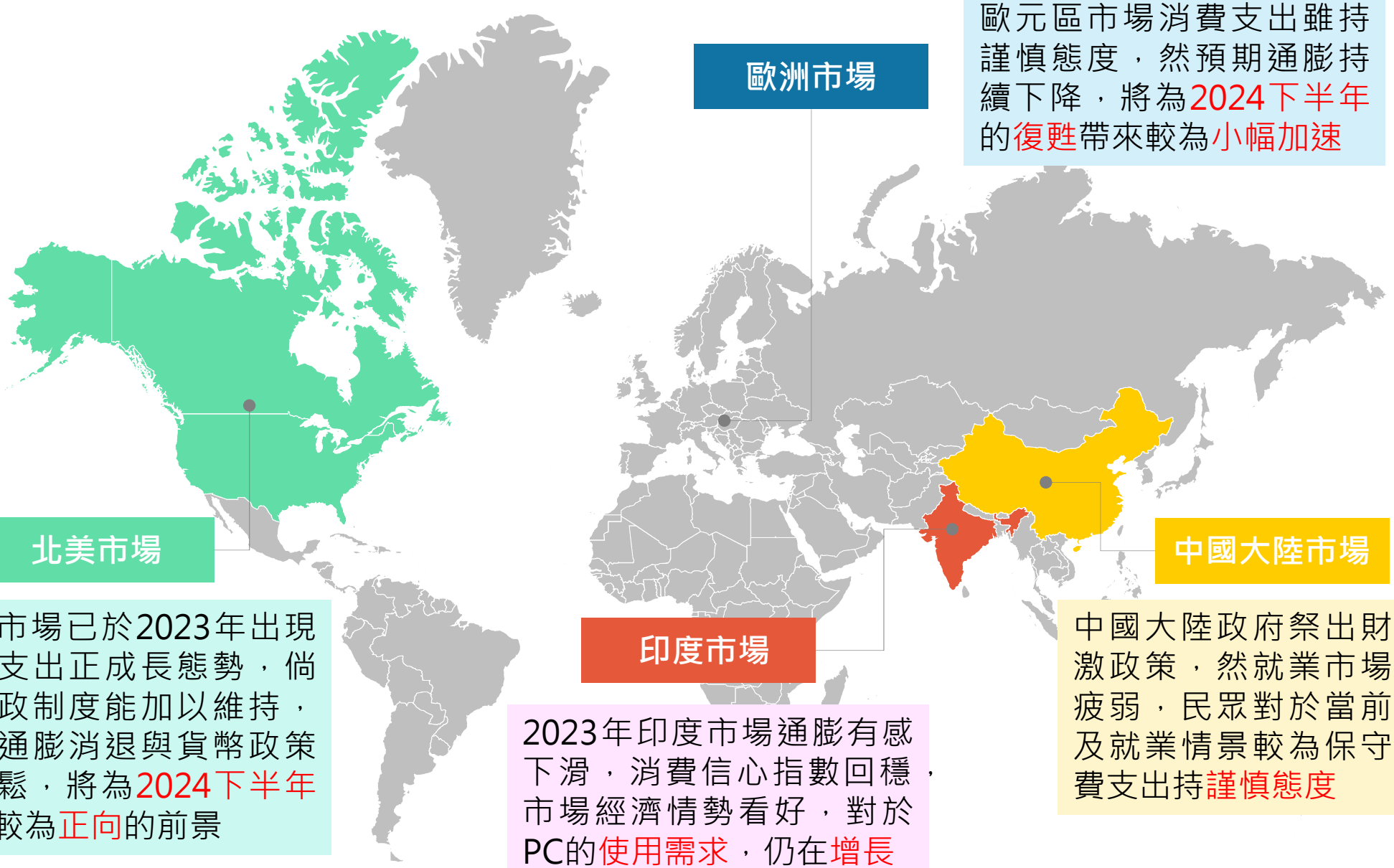
需求

- 全球景氣復甦尚不明朗
- 通路鋪貨受地緣政治與經濟情勢影響，不若過往來的積極

供給

- 新款處理器平台的上市時程與新興AI應用是否拉抬買氣仍待觀望

區域市場復甦進度差異大



歐洲市場

歐元區市場消費支出雖持謹慎態度，然預期通膨持續下降，將為2024下半年的復甦帶來較為小幅加速

北美市場

北美市場已於2023年出現消費支出正成長態勢，倘若財政制度能加以維持，預期通膨消退與貨幣政策的寬鬆，將為2024下半年帶來較為正向的前景

印度市場

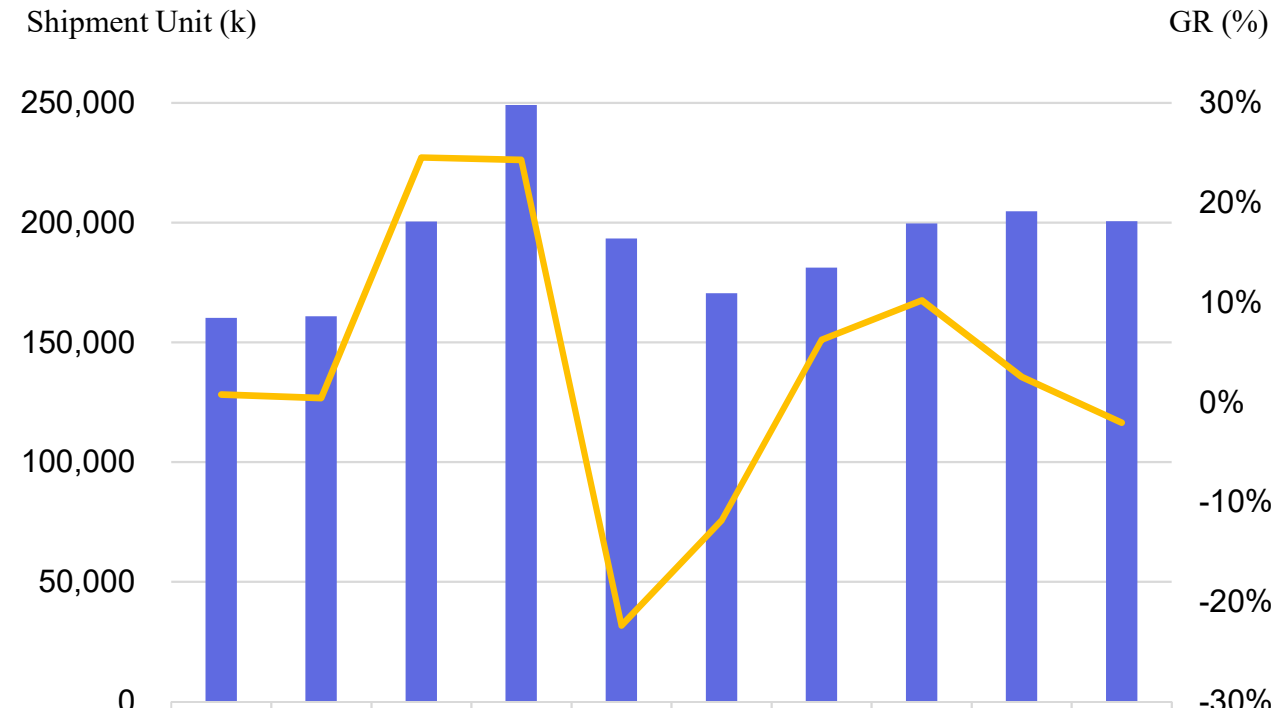
2023年印度市場通膨有感下滑，消費信心指數回穩，市場經濟情勢看好，對於PC的使用需求，仍在增長

中國大陸市場

中國大陸政府祭出財政刺激政策，然就業市場仍顯疲弱，民眾對於當前經濟及就業情景較為保守，消費支出持謹慎態度

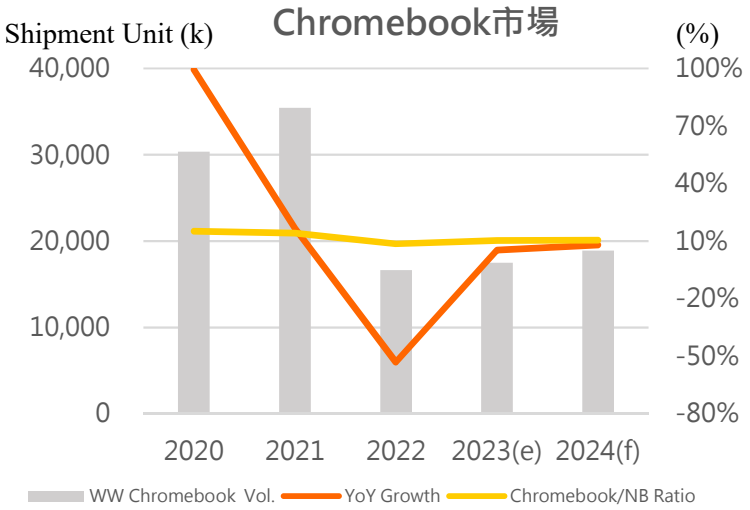
週期換機與AI議題加成，帶動2024全球市場小幅增長

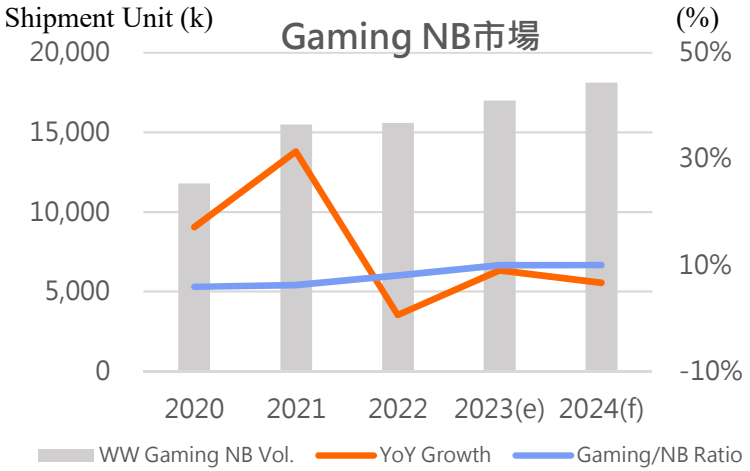
2018-2027年全球筆記型電腦市場預測



	2018	2019	2020	2021	2022	2023(e)	2024(f)	2025(f)	2026(f)	2027(f)
WW Vol.	160,202	160,894	200,354	249,035	193,336	170,470	181,167	199,661	204,738	200571
YoY	0.8%	0.4%	24.5%	24.3%	-22.4%	-11.8%	6.3%	10.2%	2.5%	-2.0%

備註：NB計算範疇包含Chromebook
資料來源：MIC，2023年12月

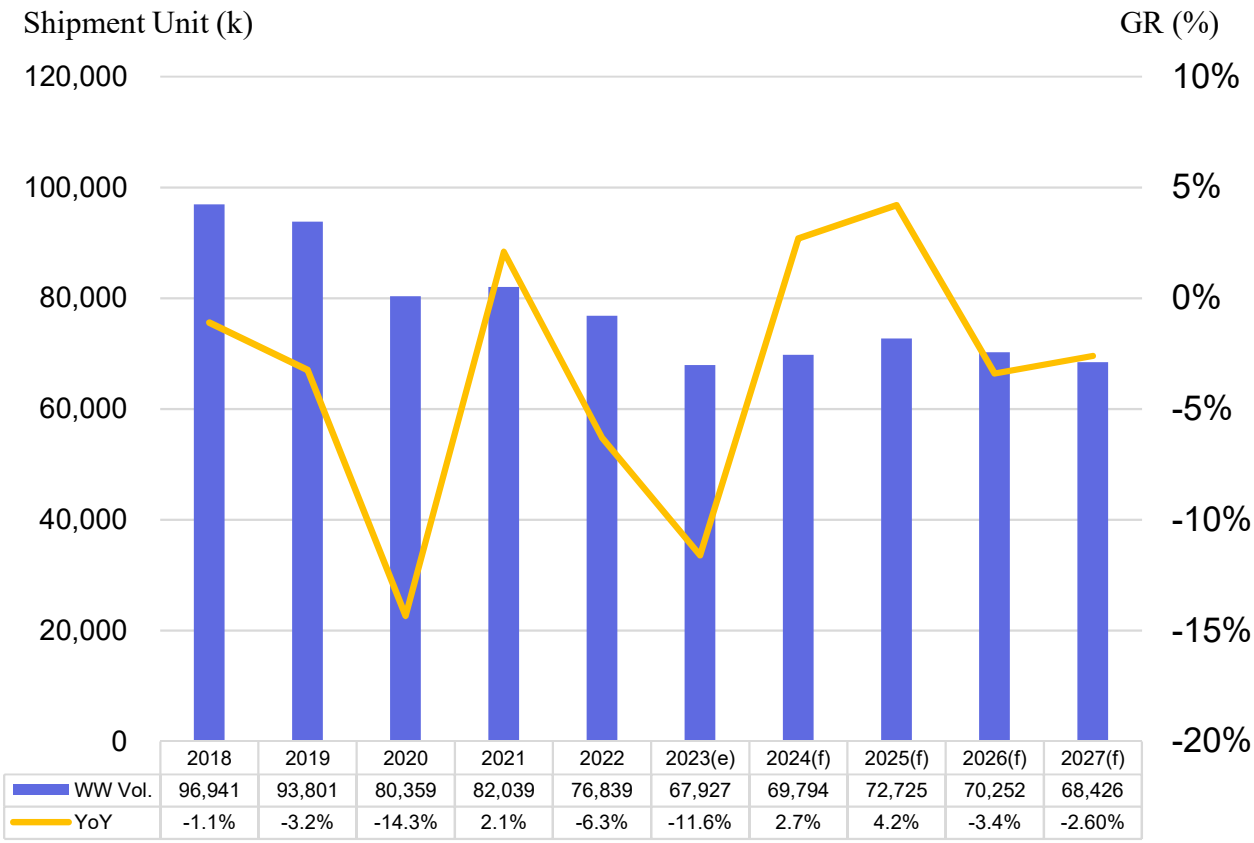




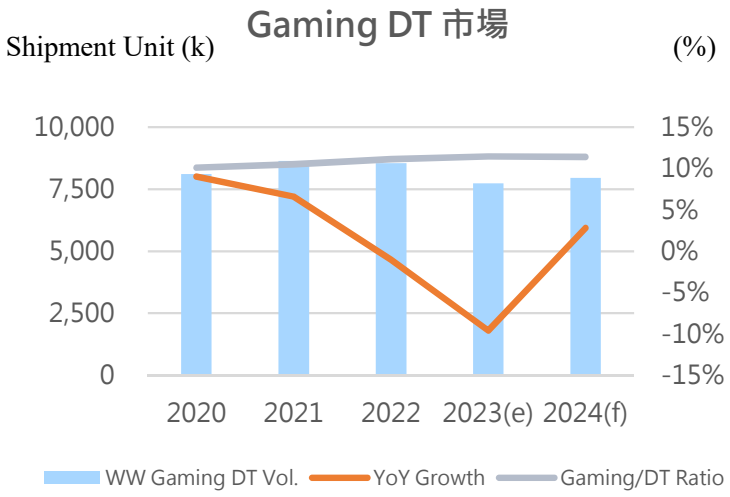
- 展望2024年，部份區域市場顯露復甦曙光，然在經濟情勢尚未明朗前，品牌廠對於市場需求的回穩前景仍持審慎樂觀的態度，市場需求的增長將主要來自於週期換機、Windows 10服務終止，以及AI NB帶來的加乘效果，使2024年全球筆電市場有望小幅增長6.3%，達181,167千台
- 電競筆電需求受惠電競遊戲市場的穩定提升而持續維持增長態勢；Chromebook市場則在2023年起浮現少部分的標案換機，預期2024年也將持續受惠於教育標案的換機力道

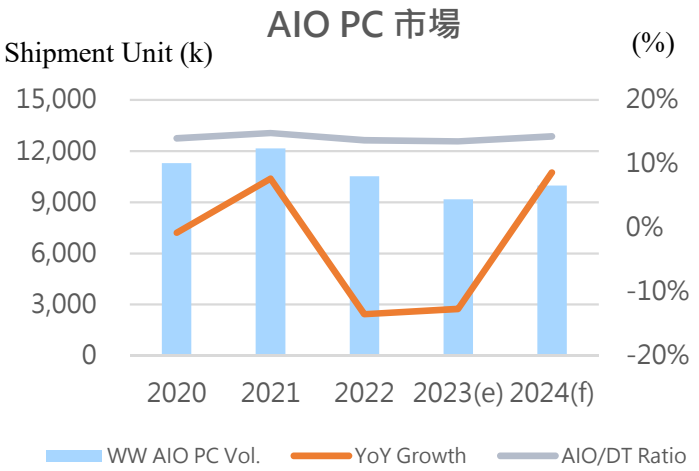
換機潮及CPU新架構，支持2024年市場微幅成長

2018-2027年全球桌上型電腦市場預測



	2018	2019	2020	2021	2022	2023(e)	2024(f)	2025(f)	2026(f)	2027(f)
WW Vol.	96,941	93,801	80,359	82,039	76,839	67,927	69,794	72,725	70,252	68,426
YoY	-1.1%	-3.2%	-14.3%	2.1%	-6.3%	-11.6%	2.7%	4.2%	-3.4%	-2.60%



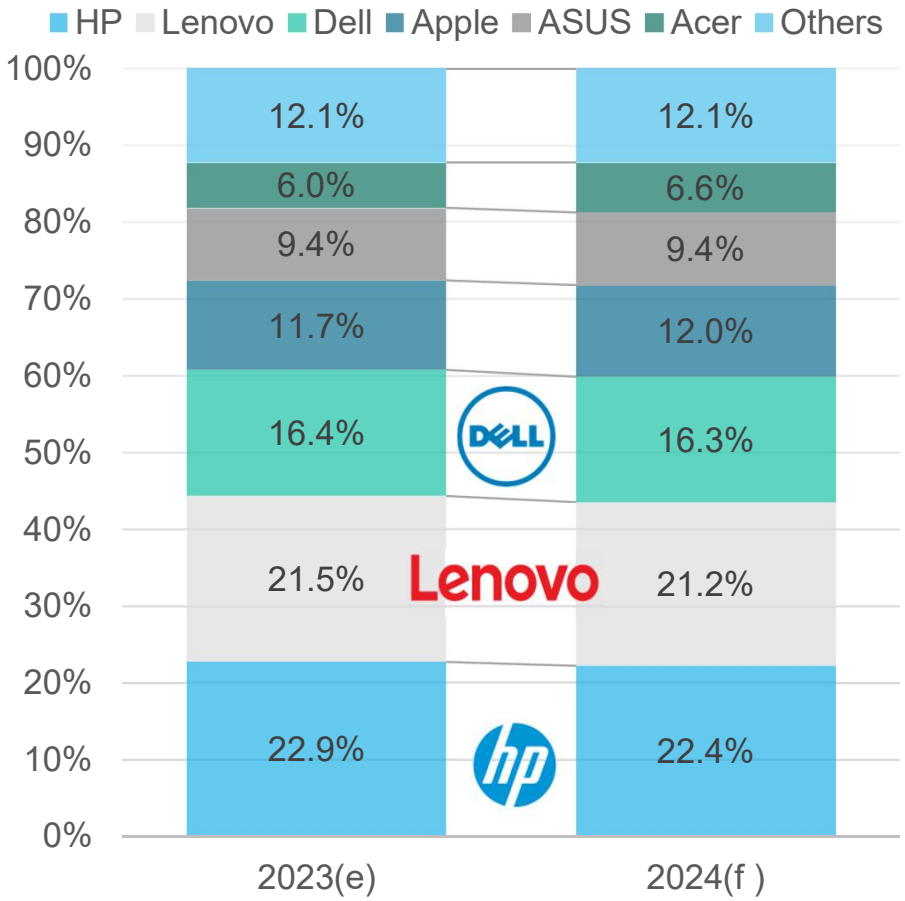


資料來源：MIC，2023年12月

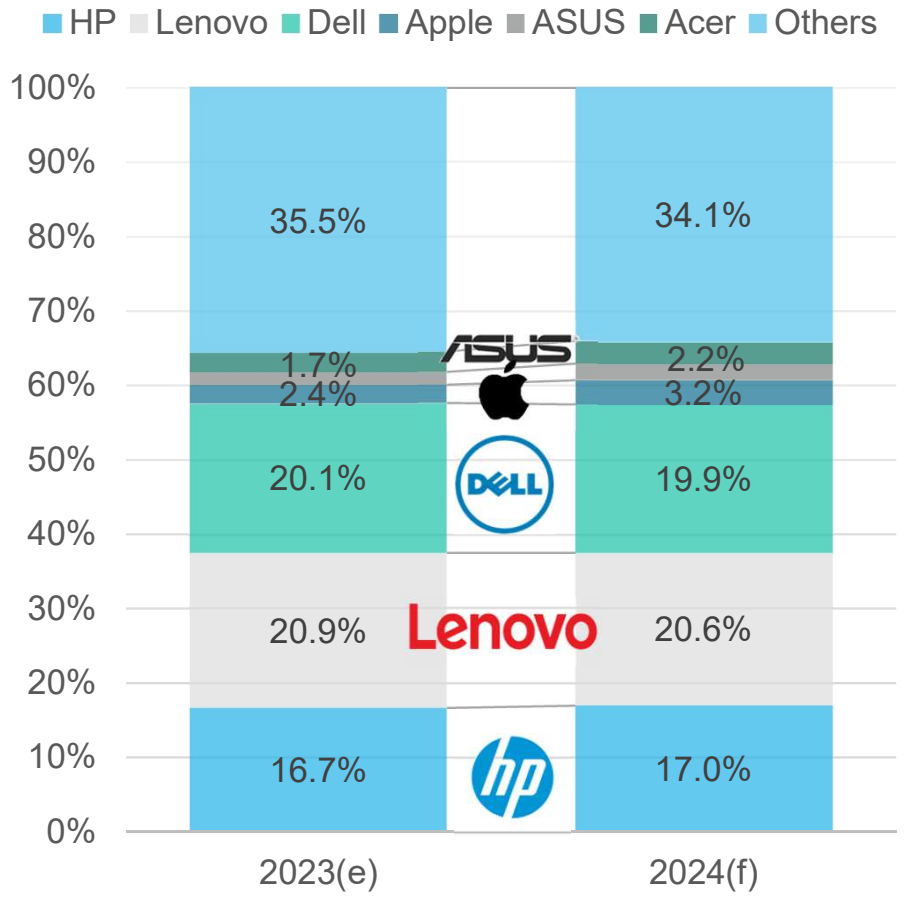
- 2023年全球景氣受通膨影響，電子資訊產品市場消費動能不彰，然而2024年全球經濟景氣狀況仍未見明朗，品牌大廠對於桌機市場出貨持保守態度
- 桌機市場大幅成長之期待會在2025年，隨著2024年底CPU新架構推出及2025年Windows 10終止支援，企業換機需求更為明顯，並預期AI PC應用方向將會更多，有機會吸引商務及一般消費者進行電腦世代交替

新品策略影響品牌廠商市占

2023~2024年全球筆電品牌廠商市占率



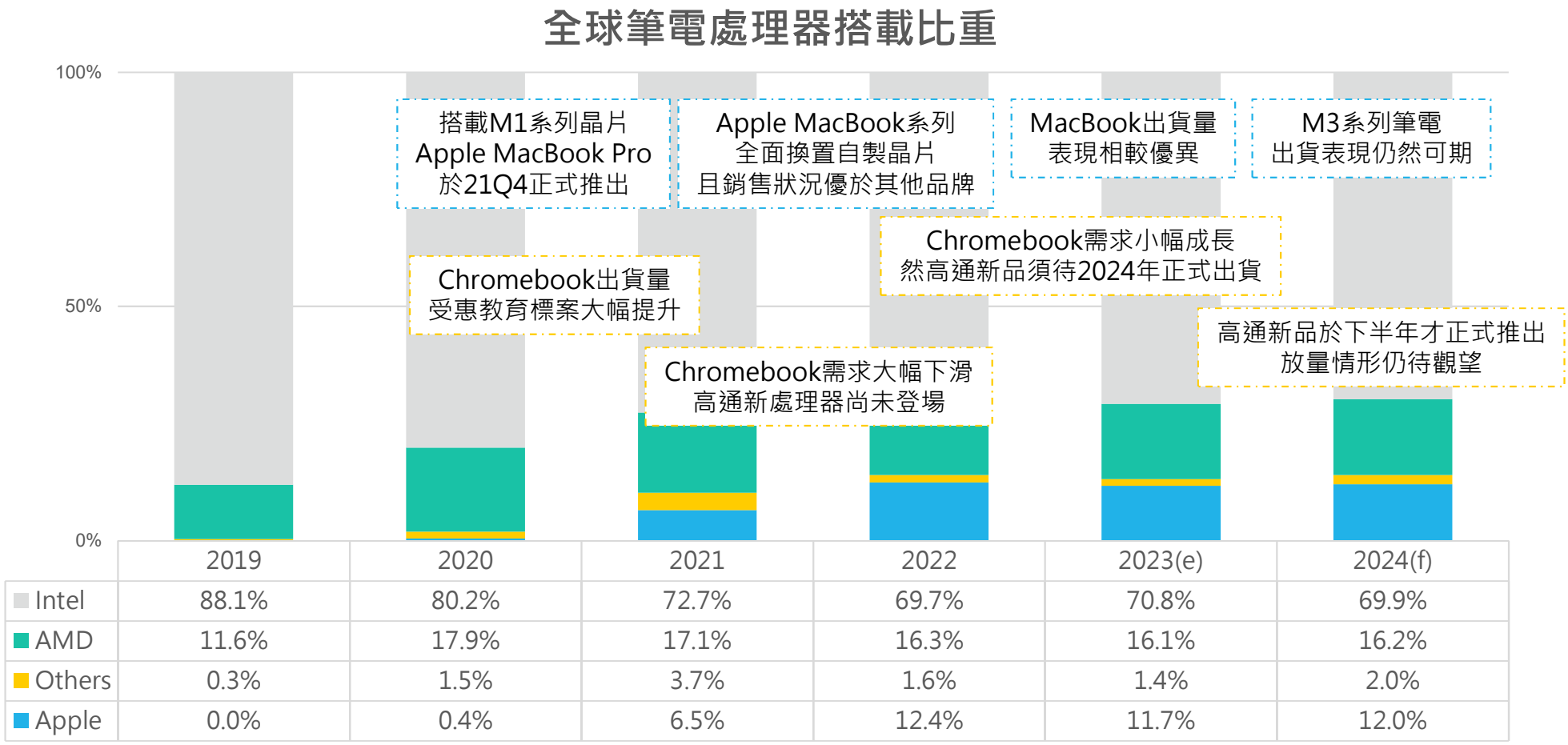
2023~2024年全球桌機品牌廠商市占率



資料來源：MIC，2023年12月

- 面對總體環境的不確定，Dell、Lenovo、HP對於2024年市場展望較為保守；反觀Apple M3處理器於2023年10月31日正式發表後，有望使2023年相對弱勢的市場表現，於2024年重獲新生
- 桌機市場品牌大廠走向類似NB；其他品牌部分，預期在2024年Apple在iMac新品銷售帶動下能拉升低迷出貨量，華碩在2023年第三季取得Intel NUC市場，有望拉近與Acer市場占比差距

Arm架構處理器上市搶搭AI話題，競爭更為激烈

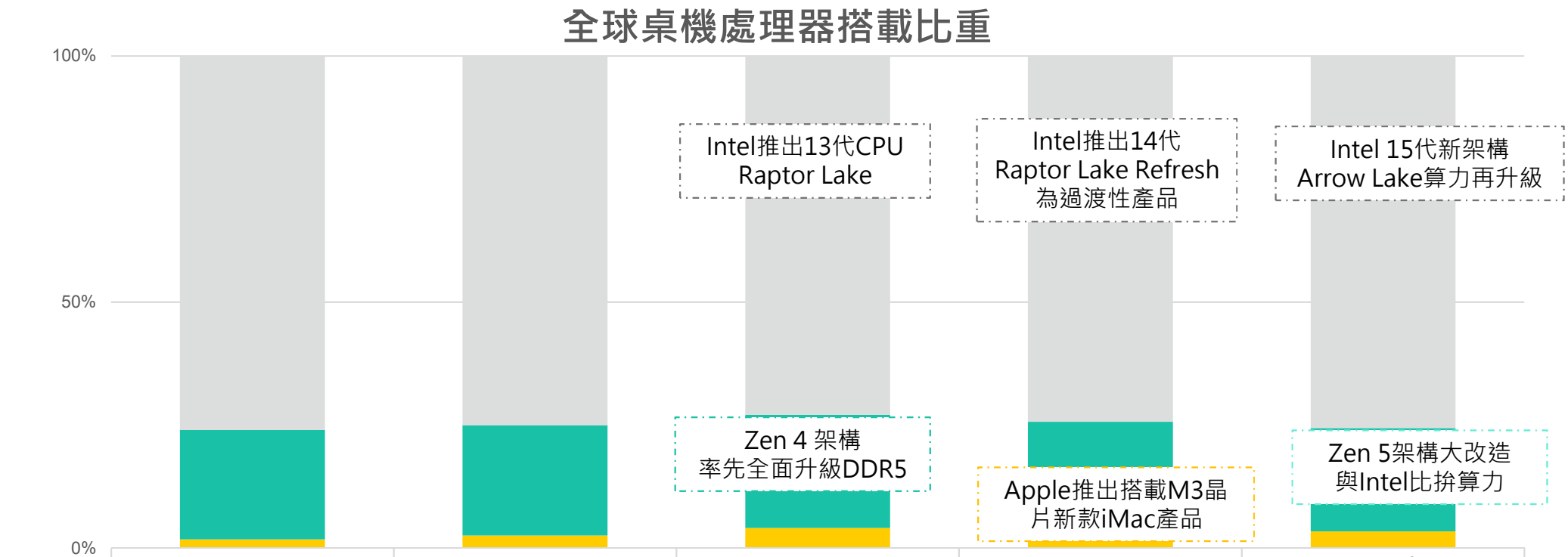


	2019	2020	2021	2022	2023(e)	2024(f)
Intel	88.1%	80.2%	72.7%	69.7%	70.8%	69.9%
AMD	11.6%	17.9%	17.1%	16.3%	16.1%	16.2%
Others	0.3%	1.5%	3.7%	1.6%	1.4%	2.0%
Apple	0.0%	0.4%	6.5%	12.4%	11.7%	12.0%

資料來源：MIC，2023年12月

- 2023年Intel在混合式架構推出後，獲得市場正面回響，因此市佔略有上升；Apple M3系列於第四季上市，加上強調AI效能表現的Qualcomm Snapdragon X Elite正式推出，將有望提升Arm架構處理器的搭載比重，使2024年筆電處理器市場更加競爭

桌機新架構CPU上市，消費級PC算力升級可期



	2020	2021	2022	2023(e)	2024(f)
Intel	76.0%	75.0%	72.9%	74.2%	75.6%
AMD	22.2%	22.4%	23.0%	22.4%	21.0%
Others	1.8%	2.6%	4.1%	3.2%	3.4%

資料來源：MIC，2023年12月

- 在2023年當中，Intel發布14代CPU做為跨入2024年新架構前的過渡性產品；另外Apple在第四季才推出搭載M3晶片的iMac新品，對整年度銷量影響有限，CPU占比較2023年下降
- 展望2024年，預期Intel及AMD將在下半年推出全新架構CPU，在兩大廠CPU架構同步大換新之下，也將帶動周邊零組件換代，DDR5將成為市場主流，固態硬碟滲透率更高

全球伺服器市場發展

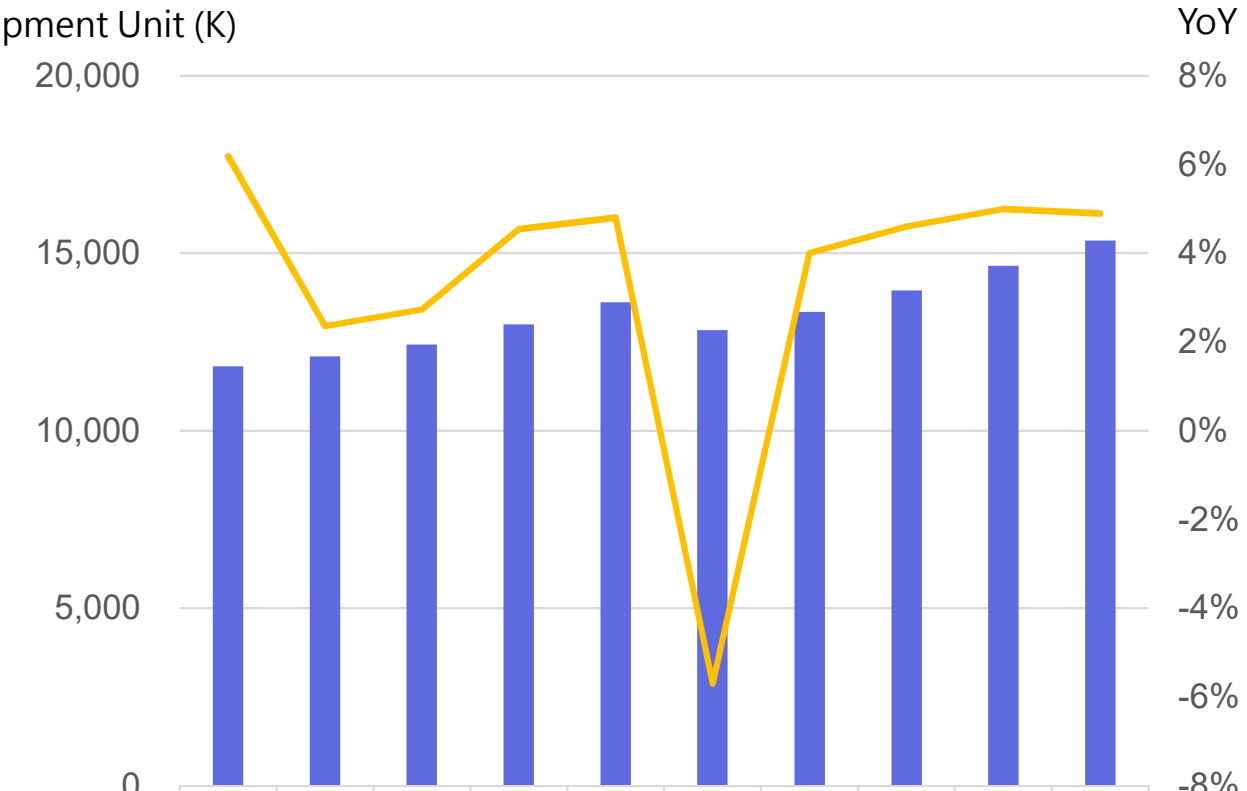
2024年全球伺服器市場藉由AI伺服器重新回溫

影響全球伺服器市場因素

- 搭載Intel、AMD新一代伺服器處理器的伺服器出貨 ↑
- ChatGPT帶動AI伺服器需求 ↑

- 受全球經濟影響，企業減少採購影響伺服器品牌商訂單 ↓
- 雲端服務商延長既有伺服器汰換週期 ↓
- 美國禁令使中系品牌商成長受限 ↓
- 中國大陸雲端服務商受政策限制，伺服器需求下滑 ↓

2018~2027年全球伺服器市場預測



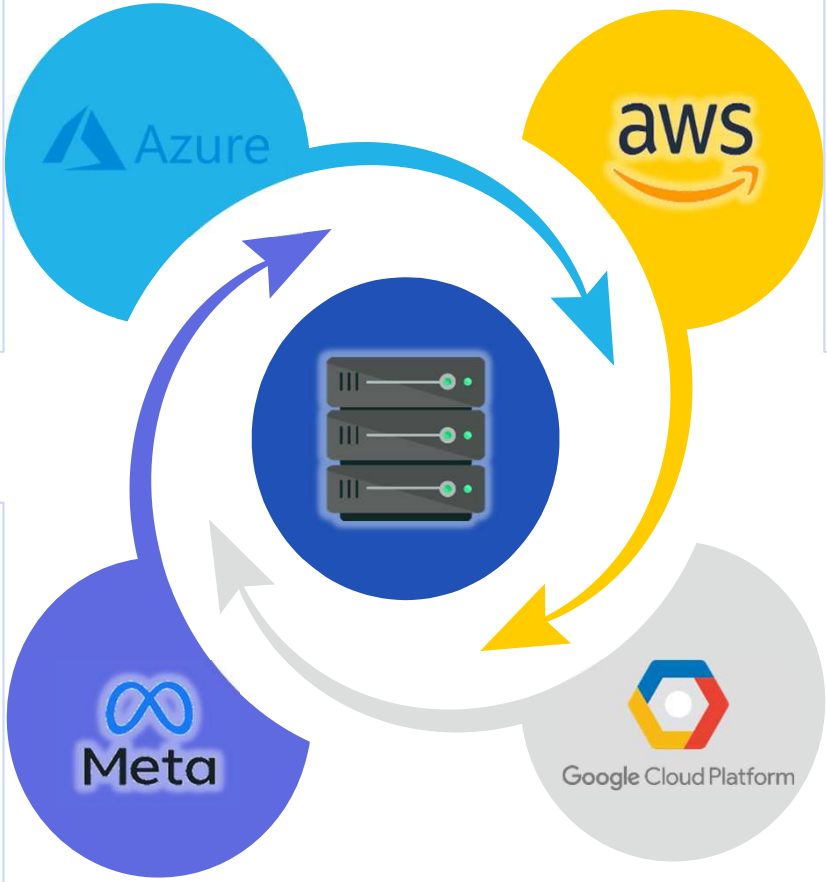
	2018	2019	2020	2021	2022	2023 (e)	2024(f)	2025(f)	2026(f)	2027(f)
WW Vol.	11,814	12,092	12,423	12,987	13,611	12,832	13,343	13,951	14,645	15,356
YoY	6.2%	2.4%	2.7%	4.5%	4.8%	-5.7%	4.0%	4.6%	5.0%	4.9%

資料來源：MIC，2023年12月

- 全球伺服器市場受到企業減少採購、美系雲端服務商延長伺服器汰換週期、調整資料中心建設計畫，中系品牌商與雲端服務商需求不振影響
- 新一代伺服器出貨、ChatGPT帶動等因素，原本預期將可帶動第三季出貨，但出貨實績卻不如預期，2023年全球伺服器市場將相較2022年呈現衰退。因為各方對於GPU的需求攀升，目前要採購搭載H100的伺服器更加困難，或將使第四季AI伺服器的產能遞延至2024年

雲端服務商增加2024資本支出加速擴建AI算力

為了打造支援**AI的資料中心**，預計2024財年**每個季度增加資本支出**。Azure營收增長速度有望在2024下半年重回逐季增加的態勢，**AI相關服務**佔比亦將逐季提高



2023年資本支出將低於2022年的590億美元、**物流**相關支出預估將呈現年減，2024年**持續投資基礎設施**、以支援AWS客戶需求，包括**大型語言模型（LLM）**和**生成式AI**投資

Meta下調2023年資本支出從300~330億美金下調至270~300億美元，2024年**資本支出**將因**投資資料中心**和**AI**而回升。AI伺服器的專案和設備交付延遲轉移至2024年

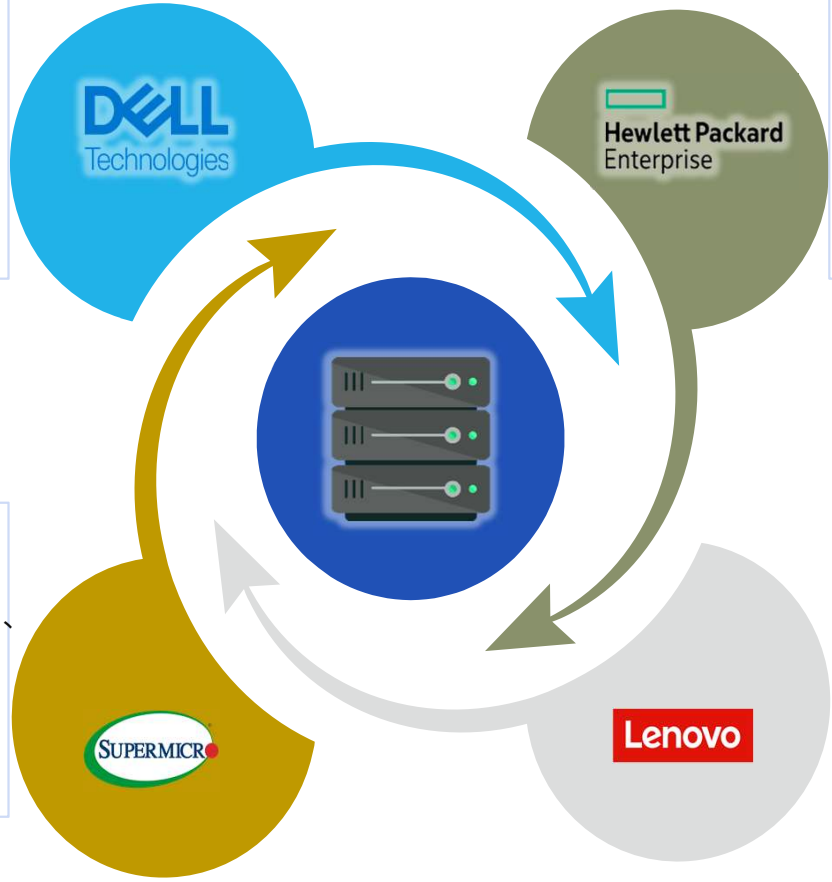
資料中心和伺服器投資步伐在2023Q3加快，並繼續增加。預期2024**全年資本支出將高於2023年**，以滿足其**雲端業務**及**人工智慧**所需的基礎建設

資料來源：各公司，MIC整理，2023年12月

- 為因應生成式AI與大型語言模型的需求，雲端服務商將於2024年擴增資本支出來佈建AI算力，將使2024年AI伺服器市場繼續蓬勃發展

2024年伺服器品牌商透過AI與HPC擺脫市場低迷

AI伺服器訂單需求持續增溫，以訂單來看，第三季**AI伺服器在手訂單**較前季翻倍成長。看好**2024年營收將恢復成長**



預計2024年**HPC**和**AI**的需求將持續改善，預計全年收入增長4%至6%。HPE精簡產品組合和提供**雲原生數據**服務的戰略正在推動增長並提高利潤

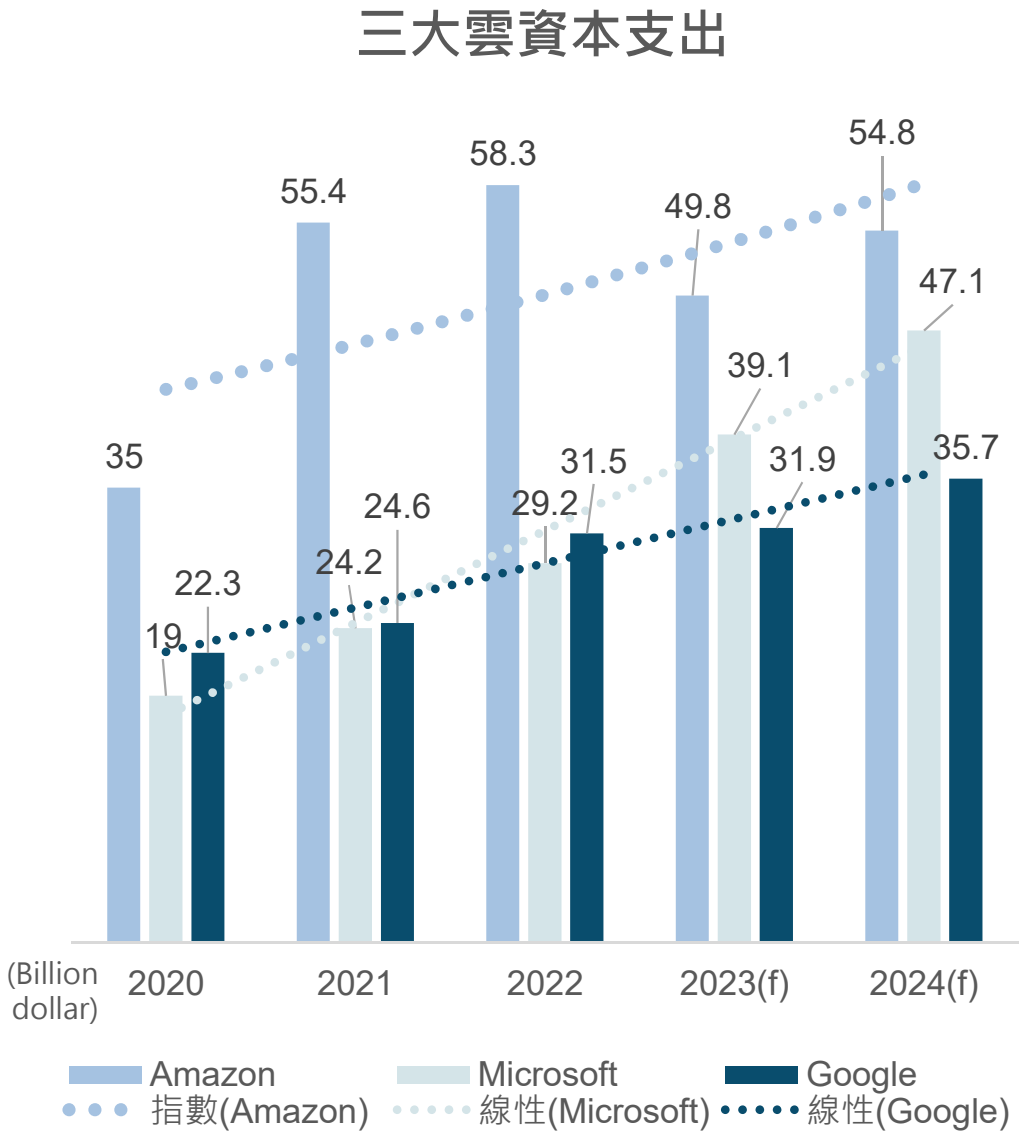
AI伺服器訂單接不完，該公司2023年營收將衝刺200億美元。已看到**AI**、全球企業、**雲端**、**儲存**市場和**5G / 電信**領域，開始接連冒出全新商機

預估2024年**設備即服務市場**保持雙位數增速、**雲端運算解決方案**市場復合增長率18%

資料來源：各公司，MIC整理，2023年12月

- 伺服器品牌商歷經2023年企業級伺服器市場的低迷後，在2024年有望藉由AI伺服器新品使營收回穩，當中雲端及高效能運算為成長的主要因素

三大雲資本支出仍以佈局算法、算力為核心策略



Amazon

2023	調降運輸資本支出；持續投資LLM與Gen AI	
2024		持續布局基礎建設與AI算力

Microsoft

2023	激增對資料中心與AI運算支出	
2024		維持2023年投資策略增加資本投入

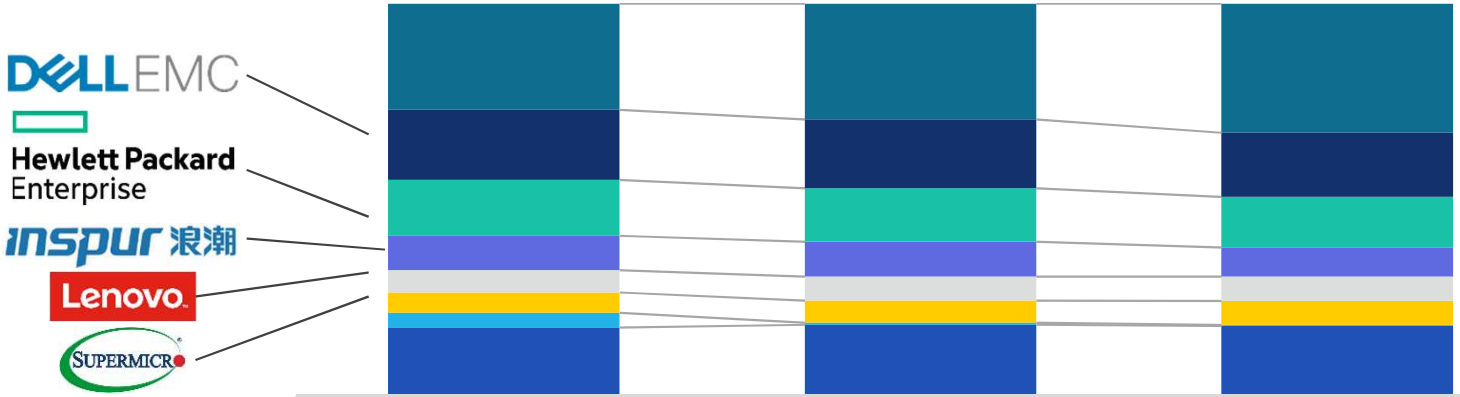
Google

2023	微幅增加投資著重AI領域資本支出	
2024		持續建置資料中心

資料來源：Amazon、Google、Microsoft，MIC整理，2023年12月

ODM Direct出貨於2023年正式超過三成

2021~2023年全球伺服器出貨量百分比 - 品牌別



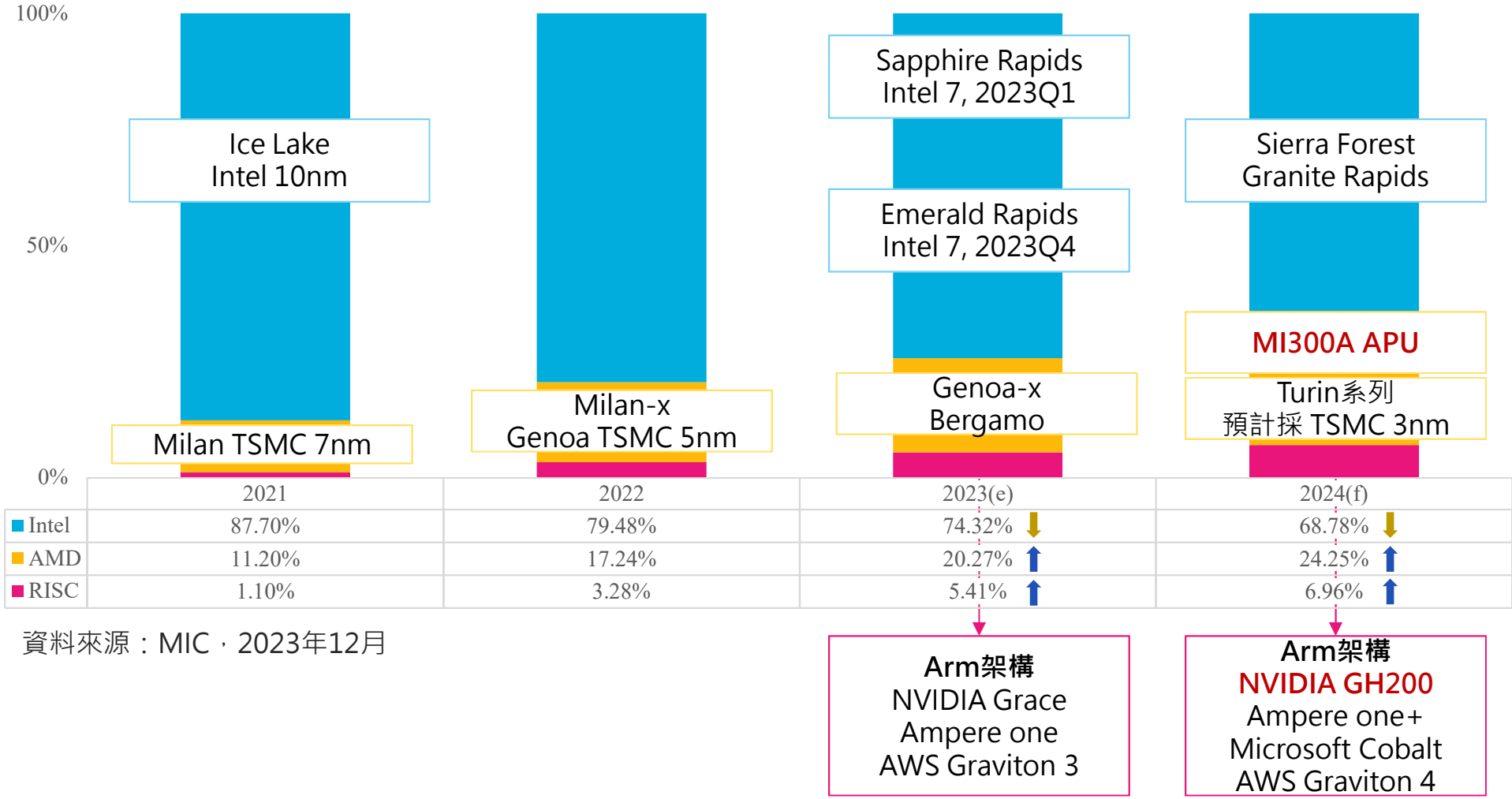
	2021	2022	2023(e)
ODM Direct	27.1%	29.5%	32.9%
Dell EMC	17.9%	17.6%	16.4%
HPE	14.3%	13.6%	12.9%
Inspur	8.8%	8.9%	7.4%
Lenovo	5.7%	6.1%	6.2%
SuperMicro	5.2%	5.7%	6.1%
Huawei	3.7%	0.5%	0.3%
Brand_Others	17.4%	18.1%	17.8%

資料來源：MIC，2023年12月

- 綜觀全球伺服器品牌出貨量，2022年雲端服務商訂單增加促使ODM Direct比重持續上升，2023年中系品牌商面臨美國晶片法案、被列入實體清單、中國大陸內需不振等風險，在全球的佔比將會下滑。以伺服器代工業者為主的ODM Direct客戶更多元化，除雲端服務商外有望接獲大型電信商、資料中心託管商的訂單

NVIDIA與AMD接連推出超級晶片 CPU與GPU綑綁銷售

2021-2024(f)年全球伺服器處理器搭載比重-Overall

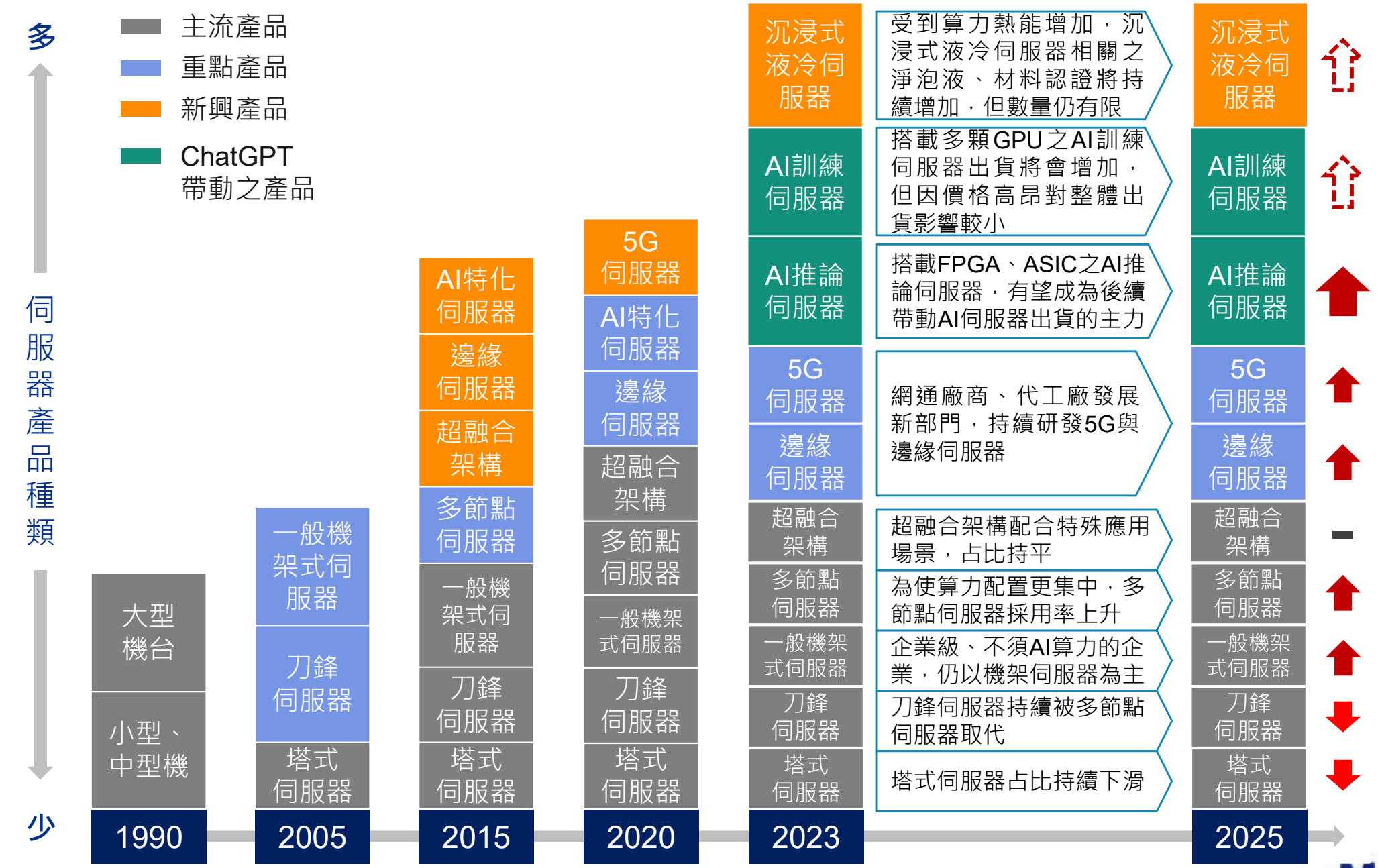


	2021	2022	2023(e)	2024(f)
Intel	87.70%	79.48%	74.32% ↓	68.78% ↓
AMD	11.20%	17.24%	20.27% ↑	24.25% ↑
RISC	1.10%	3.28%	5.41% ↑	6.96% ↑

資料來源：MIC，2023年12月

- 2023年AMD CPU市占率將超過兩成，以Arm架構為首的RISC架構將超過5%，受到生成式AI影響處理器廠商將CPU與GPU共同封裝成超級晶片，將成為2024年的一大出貨主力

ChatGPT催化AI伺服器需求，產品線將更多樣化



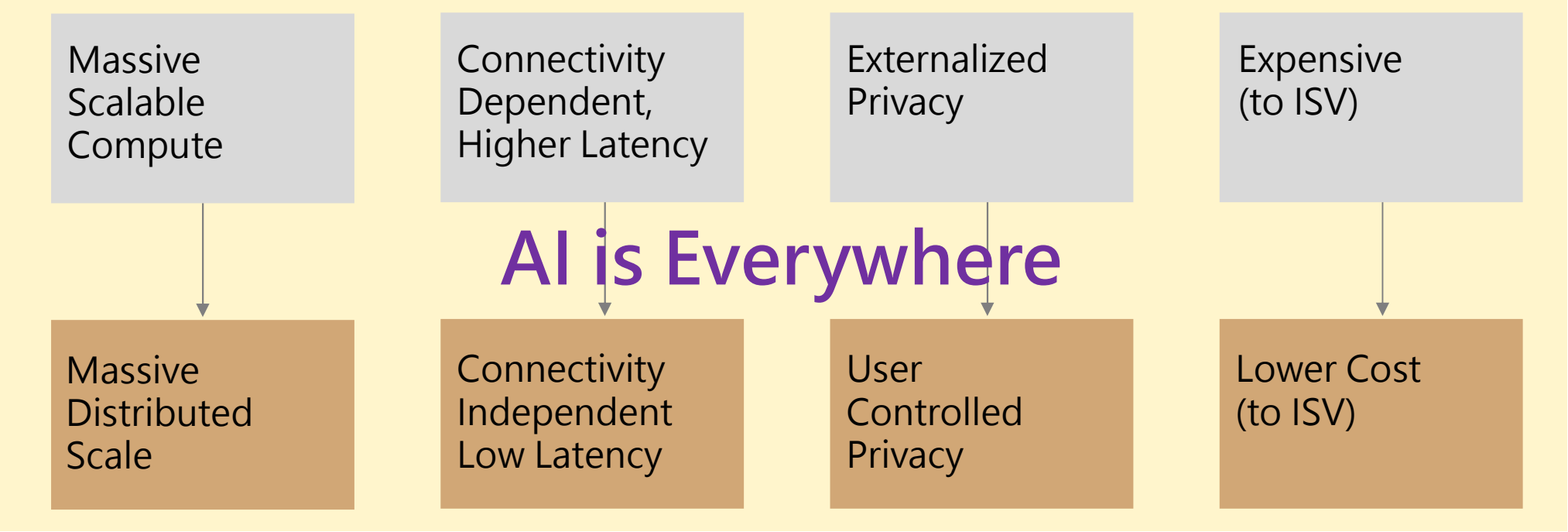
熱門議題觀測(一)

AI PC/NB新興應用

AI運算的範疇

Today：Cloud

- **雲端運算**主要為在**資料中心**或**雲端**進行的大規模運算
- 優勢： (1)運算能力高且可擴充性高，可處理複雜的運算問題和大量數據
  (2)部署難度及維運難度低，較不需考慮設備異構性



Tomorrow：Cloud + Client + Edge

- **邊緣運算**主要發生在**離數據源（如物聯網設備、移動設備等）較近**的地方
- 優勢： (1)低延遲和多連結特性，適用於需要快速反應的場景
  (2)可減少數據傳輸的成本和時間，提高數據的安全性和隱私性

現行AI筆電樣貌：
多數為既有軟體優化與少數生成式AI功能試行

AI筆電

既有功能優化

■AI降噪：消除背景噪音
■AI Camera：偵測眼神、自動模糊背景等

如Acer Swift Edge 16搭載AMD Ryzen 7040系列處理器，透過內建AMD Ryzen AI技術，強化視訊會議功能，包含模糊背景、自動取景和眼神接觸。在音訊端，則有不少品牌廠商均導入一鍵式抗噪的功能，提供視訊會議所需的支援功能

■AI智慧偵測：自動偵測使用程式並調整效能配置，如微星AI Engine

微星AI Engine自動化調整

Light（鍵盤下RGB燈效）、Performance（調整風扇）、Visual（螢幕藍光）、Audio（音效）四種選項

生成式AI功能

■**AI Artist本機版本**：提供本機端自動生成圖片(尚未上市)

微星AI-Artist在本機運算生成圖片

供應鏈寄望藉由打造AI NB，推升新的換機潮

發言者	內容
Microsoft	開始推出Windows 11個人電腦（PC）作業系統的重大更新，將包含名為「Copilot」的生成式人工智慧（AI）聊天機器人，能為用戶操作作業系統功能，並在廣大網路資訊的協助下回答問題，在AI PC競賽中加強軟體攻勢
Qualcomm	透過創新AI合作升級行動運算，為Windows 11 PC帶來行動創新。從過去技術進步趨勢來看，從CPU主導到GPU進展，現在進一步由AI來驅動
宏碁	(陳俊聖)生成式AI出現後，會更普遍應用於生產力的提升或優化上，這就需要用回到筆電，並且將帶來剛性需求
	(施振榮)要看所謂AI筆電能否真正提供新的價值，而不是透過原來的電腦軟體應用。AI筆電應該要賦予新的AI晶片、軟體應用，甚至有新的使用情境和體驗，有可能會是PC發展的新方向
華碩	• 認同AI將帶動新一波筆電換機潮，AI是電腦基礎的再突破，是核心技術，一定要發展最先進技術，影響非常深遠，現在微軟也非常積極把AI導入生產力工具，Google在這方面也非常強 • AI PC在硬體(晶片)與軟體都有額外加值，包括內建NPU，以及記憶體方面，因為要跑大語言模型，會佔據額外記憶體，因此會更大一些，整體MSRP(建議售價)也將提升
HP	• 消費者能獲得的事物將大幅不同，HP正和所有關鍵軟體業者和關鍵晶片供應商合作，重新設計PC的架構，惠普正在打造包含AI效能的PC，讓消費者以歷來最短的時間建立與分析試算表 • 2024年將推出的名為「Z by HP AI Studio」的軟體平台。AI Studio 軟體旨在讓人們更輕鬆地協作和創建人工智慧開發專案。預計還將推出一款人工智慧工作站，稱其「簡化了私有人工智慧模型和應用程式的建置和客製化」。HP表示，將提供首批配備 Nvidia AI 企業軟體平台的工作站

資料來源：各公司，MIC整理，2023年11月

NB供應鏈在AI筆電的發展動態

	廠商	AI筆電發展規劃
作業系統商	微軟	在Windows 11整合更多人工智慧技術，藉此強化隱私安全及更多輔助應用功能，包含將在 Windows 11 中引入 Copilot AI 助手，並將 Bing Chat 插件擴展到 Windows
晶片商	Intel	證實第14代Meteor Lake處理器全面整合NPU加速器的設計，強化AI運算效能
	AMD	已推出首款搭載「XDNA」AI 引擎的Ryzen 7040HS系列筆電處理器
	Qualcomm	宣布與微軟合作，將自動生成式人工智慧技術帶到Windows 11 on Snapdragon裝置，預期能催生不同的使用體驗，(1)透過文字描述自動生成影片，或透過口述方式完成編寫程式，(2)讓裝置執行效率、電池續航表現大幅提升
品牌商	Lenovo	透過Lenovo LA AI 晶片及 AI Engine+ 專用軟體，分擔畫面刷新率的動態調節、增加耐熱度和提升整體效能表現，並應用於Legion系列電競筆電
	HP	正與晶片供應商及重要軟體開發商合作，未來AI PC可直接在本機執行AI運算，藉此降低延遲時間與提高資安防護

推導

- **產品面**：
  - AI筆電的功能優化首要將在(1)提升效能表現、(2)分配雲端與地端運算工作，降低延遲時間與提高資安防護；
  - 未來則可擴充至更多輔助應用的開發，EX：口述及文字生成圖像影片、蒐集個人電腦使用喜好進行體驗調整
- **技術面**：品牌廠或可透過與晶片商合作，或可自行開發AI晶片，開展AI NB的設計
- **應用面(可能應用情境)**：

1.高效能：利用AI晶片，分擔、提高運算效能表現 → 電競筆電提升遊戲順暢度

2.高安全性：在本機執行AI運算，降低延遲並提高安全性 → 商用筆電讓商用客戶擁有個別公司的AI module

AI筆電尋找應用利基，生產效率提升與客製服務有望實現

	智慧化效能管理					音訊、視訊優化		生成式AI
	技嘉	Lenovo				Razer	Acer	微星
機型	AERO系列	Legion Pro 5	Legion Pro 5i	Legion Pro 7	Legion Pro 7i	Razer Blade 14	Swift Edge 16	未定
應用市場	創作者	電競	電競	電競	電競	電競	商用	未定
推出時間	2019年	2023年1月				2023年5月	2023年5月	2023年6月
CPU	第9代 Intel Core 處理器	AMD Ryzen 7000 系列	第 13 代 Intel Core 處理器	AMD Ryzen 7000 系列	第 13 代 Intel Core 處理器	AMD Ryzen 7040HS 系列	AMD Ryzen 7040HS 系列	未定
GPU	NVIDIA RTX 20系列	NVIDIA RTX 40系列	NVIDIA RTX 40系列	NVIDIA RTX 40系列	NVIDIA RTX 40系列	NVIDIA RTX 40系列	AMD Radeon 顯示晶片	未定
AI軟體	Microsoft Azure AI	Lenovo AI Engine+	Lenovo AI Engine+	Lenovo AI Engine+	Lenovo AI Engine+	X	Acer PurifiedVoice AI	MSI AI Engine、MSI AI Artist
AI硬體	X	Lenovo LA AI	Lenovo LA AI	Lenovo LA AI	Lenovo LA AI	AMD Ryzen AI	AMD Ryzen AI	未定
AI功能	分析用戶喜好、自動偵測並調整至最佳化效能情境	與內建軟體搭配，分擔畫面刷新率的動態調節、增加耐熱度和提升整體效能，同時透過機器學習的演算法來最佳化系統表現				即時視訊品質升級(背景模糊、自動取景和自動眼神接觸)	即時視訊品質升級(背景模糊、自動取景和自動眼神接觸)、降噪	最佳化效能模式設定、生成式AI提供快速生成圖片
定價	1,799美元起	1,460 美元起	1,460 美元起	2,000 美元起	2,000 美元起	1,999美元起	1,300美元起	未定

資料來源：各公司，MIC整理，2023年11月

AI PC發展浪潮下，國內外業者可推動合作方向

通用性體驗

(一)效能提升：透過處理器本身的效能支援及AI加速器加成，協助筆電進行畫面刷新率的動態調節與分擔

(二)智慧化管理：透過AI軟體導入，讓背景模糊、眼神及臉部偵測、背景噪音的消除等功能更加升級，並可自動偵測使用程式，動態調整效能配置

客製化服務

(一)商務市場建構企業私有訓練架構：品牌廠可與作業系統商合作，針對個別公司打造獨有的AI Module，開發客製AI運算模型

(二)消費市場進行小型模型訓練：在本機端蒐集使用者使用資訊，透過小型訓練模型針對個人的使用習慣加以訓練並提升使用回饋

可推動合作方向

作業系統商

Microsoft 提供具備AI工具、AI模型的作業系統

晶片商

AMD intel Qualcomm 處理器整合AI加速器，滿足運作效能

品牌商

acer ASUS 聯手晶片商，將AI體驗內建於PC產品

代工廠

Istron Quanta Computer PEGATRON Inventec COMPAL 與品牌廠合作，開發設計AI PC

軟體開發商

- 既有PC應用程式加入AI
- 開發新的AI軟體整合方案

資料來源：MIC，2023年11月

MIC

各大廠對於AI PC市場滲透率的看法

廠商	內容
Microsoft	• AI PC換機潮**2024年6、7月**以後可啟動。微軟利用AI強化PC作業軟體，推出Windows 11 PC作業系統重大更新，新版作業系統將包含Copilot人工智慧助手
Qualcomm	• AI PC將於**2024年開始放量**，2025年量能將再放大
Intel	• AI PC換機潮**2024年初就可望啟動**，估計在**2025年前超過1億台**PC上實現AI應用
AMD	• 未來5年內十分看好AI PC商機 • AMD **Ryzen AI** + Windows的筆電，已於**2023年開始出貨**
NVIDIA	• 相關AI PC產品可能於**2025年開始銷售**
Lenovo	• AI PC 有個成熟的過程，按照歷史規律，前期將佔有 **10%** 的市場份額，日後會成為主流，相信**未來每一個電腦都是 AI PC**
ASUS	• 預估AI PC在**2024將會進一步問市**，但是，因為剛開始，其各家的定義還混沌不明，整體的市場滲透率僅會有**個位數百分比**。然而到**2025年將有雙位數**滲透率
HP	• 雖然市場不會立即轉向AI PC，但相信，AI PC 的**普及率將會逐漸上升**，2024 年會出現一些成長，2025 年會出現更多滲透，2026 年甚至會更多 • AI PC將在**三年內**佔到**筆記型電腦總銷量的40%到50%**
廣達	• AI PC是很具有威力的產品，能對市場帶來正面效益；**現階段談出貨何時放量還太早**
英業達	• 預計**2027年前**將能達到在**逾2億臺**PC上實現AI應用的目標，且於**2027年PC的總出貨量中**，逾**60%**為AI PC
和碩	• 2024年雖會看到很多AI PC上市，但**需求真正爆發應該是2025年**

資料來源：各公司，MIC整理，2023年12月

Follow Intel Core Ultra處理器發表 筆電品牌廠已率先發布一波AI PC新品

強調與手機、平板等終端裝置連接性

利用微軟與Intel既有AI體驗

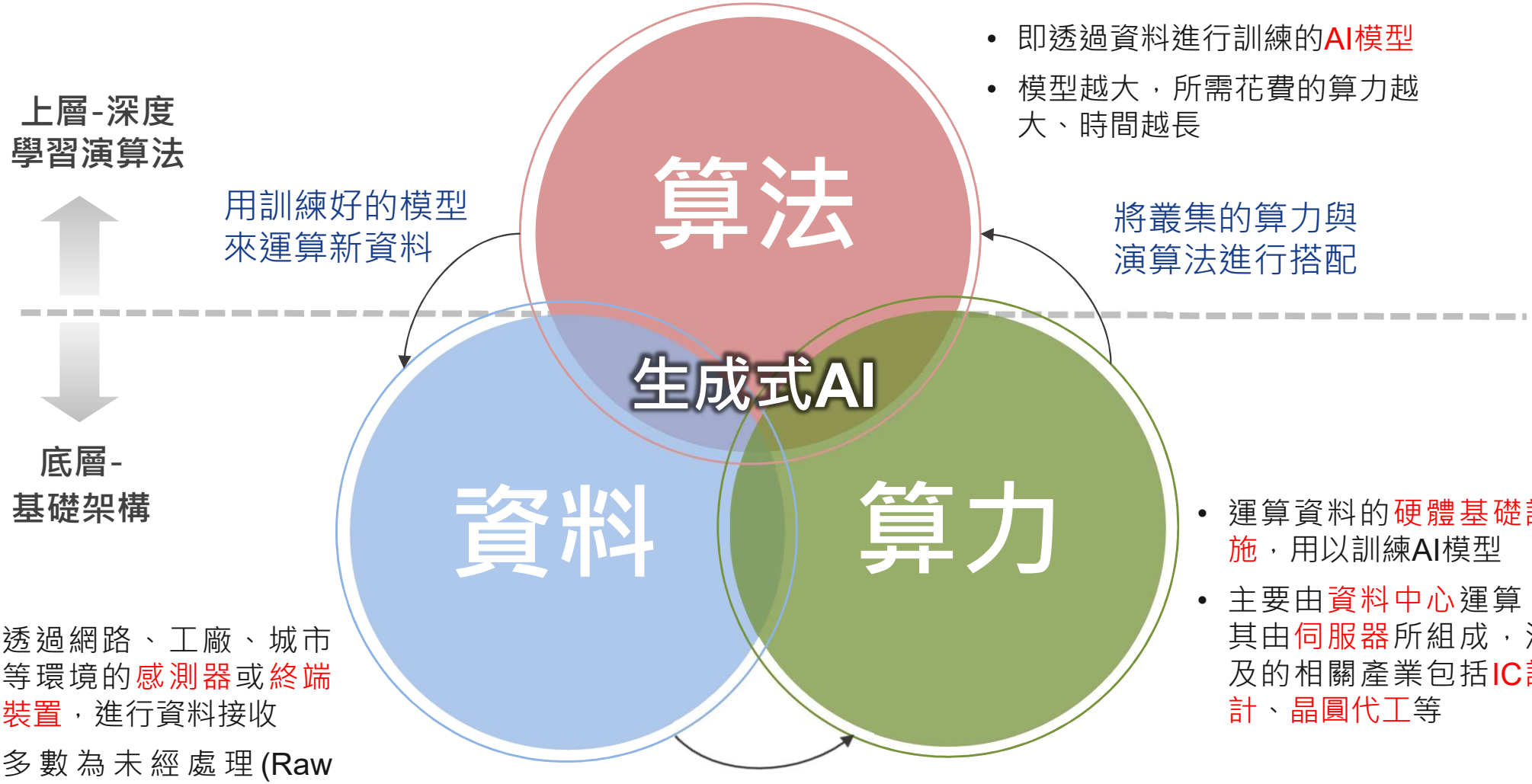
創建自有AI應用

	三星			LG		Acer	ASUS	微星			聯想	
機種	Galaxy Book 4系列			「LG Gram」系列		Swift Go 14	Zenbook 14 OLED	Prestige AI系列			小新 Pro 16 2024	ThinkPad X1 Carbon AI
系列	Galaxy Book4 Ultra	Galaxy Book4 Pro	Galaxy Book4 Pro 360	LG gram Pro	LG gram Pro 二合一			Prestige 16 AI Studio	Prestige 16 AI Evo	Prestige 13 AI Evo		
強調特色	結合了超便攜設計、提升的效能和無限連接性，重塑 PC 體驗並增強用戶能力			透過「LG gram Link」應用程式提供高效能 AI 功能和創新連接		相輔 Intel AI Boost 及 Windows 11 Copilot，加速AI體驗及功能	極「智」效率新生活，利用AI專用NPU達到省電、降低工作負載與提高AI處理能力	引領AI智慧世代 啟動極致運算。在創作、規劃、旅行時皆能提供優異的效能，尤其更能滿足大量增加的生成式AI應用需求。			具備內嵌混合AI算力、創新/增強AI體驗和設備體驗升級三大特點，意味著聯想集團AI PC正式邁入AI Ready階段	

熱門議題觀測(二)

生成式AI應用崛起，帶動AI伺服器出貨風潮

「資料、算力、算法」三大基石構成生成式AI



- 即透過資料進行訓練的AI模型
- 模型越大，所需花費的算力越大、時間越長

- 運算資料的硬體基礎設施，用以訓練AI模型
- 主要由資料中心運算，其由伺服器所組成，涉及的相關產業包括IC設計、晶圓代工等

- 透過網路、工廠、城市等環境的感測器或終端裝置，進行資料接收
- 多數為未經處理(Raw Data)

ChatGPT催化雲端服務商AI算力戰爭(1/2)

以美系雲端服務商而言，當前ChatGPT技術仍佔據領先的位置，然而包含Amazon、Google、Meta均開始加速投入AIGC相關的產品

Meta

2023/02/24，公布AI大型語言模型LLaMA

Microsoft

- 將ChatGPT導入Bing與Edge中
- 2023/03將Open AI技術導入Office服務
- 2023/04將ChatGPT升級為GPT-4

amazon

- 2023/04發表大型語言模型Titan
- 發表平台式服務「Bedrock」

Google

推出Bard後，正在研發各類AI工具，包含「GIFI」AI圖片製作、「Tivoli Tutor」語言學習 AI 聊天機器人等

Baidu 百度

2023/02/07公布「文心一言」，03/27產品正式發表會臨時改成企業閉門溝通會

ByteDance 字节跳动

2023/02/23指出，字節跳動已在AI模型上佈局，主要專注語言和圖像

Tencent 腾讯

2023/03/30正在研發類ChatGPT聊天機器人，並會導入QQ、微信等產品線

阿里云 aliyun.com

於2023/04/12正式推出「通義千問」AI聊天機器人，未來阿里巴巴產品均會導入

資料來源：各公司，2023年7月

中系雲端服務商，以百度的推行速度最快，阿里巴巴於4月正式推出模型，字節跳動及騰訊則尚未有實體產品的出現，然而儘管當前在AI算力上仍得以支撐，未來恐將面臨美國晶片法案的挑戰

ChatGPT催化雲端服務商AI算力戰爭(2/2)

- 2023/07/19 發表新一代大型語言模型**Llama 2**
- 2023/08 發布可支援近100種語言的**文本和語音之間翻譯的AI技術** SeamlessM4T模型
- 2023/09/28，推出首款**生成式AI聊天機器人**

- 2023/07成為Meta在**Llama 2商用化**的特選合作夥伴
- 2023/09宣布與美國雲端病理學AI平臺業者**Paige.ai**將合作建立世界最大癌症影像AI模型

- 2023/09全託管人工智慧平臺**Amazon Bedrock**正式推出，同時也增加Llama 2和Amazon Titan Embeddings新模型
- 語音助手**Alexa將用生成式AI升級**

- 2023/09 將**生成式 AI 整合至 Gmail、和文書處理服務 Docs**之中
- AI開發平臺Vertex AI，將全面轉型為生成式AI開發平臺
- 2023/08/29**生成式AI助手Duet AI**

Meta　Microsoft　amazon　Google

Baidu 百度　ByteDance 字节跳动　Tencent 腾讯　阿里云 aliyun.com

- 計畫推出 **AI 對話軟體**「萬話」
- 2023/08/30旗下 AI大模型產品**文心一言**通過中國《生成式人工智慧服務管理暫行辦法》備案，可正式上線面向公眾提供服務

- 正在內部測試**AI對話類產品**，暫時稱Grace
- **雲雀**大模型通過中國《生成式人工智慧服務管理暫行辦法》備案

2023/09/07推出生成式人工智慧「混元大語言模型」

- 2023/07升級AnalyticDB向量引擎
- 2023/08/25推出可理解**影像**，且能進行**複雜對話**的Qwen-VL-Chat模型
- 2023/09/13「**通義千問**」正式向大眾開放

資料來源：各公司，2023年9月

AI推論、AI訓練伺服器的作用與差異

一般、AI推論、AI訓練伺服器需求情境

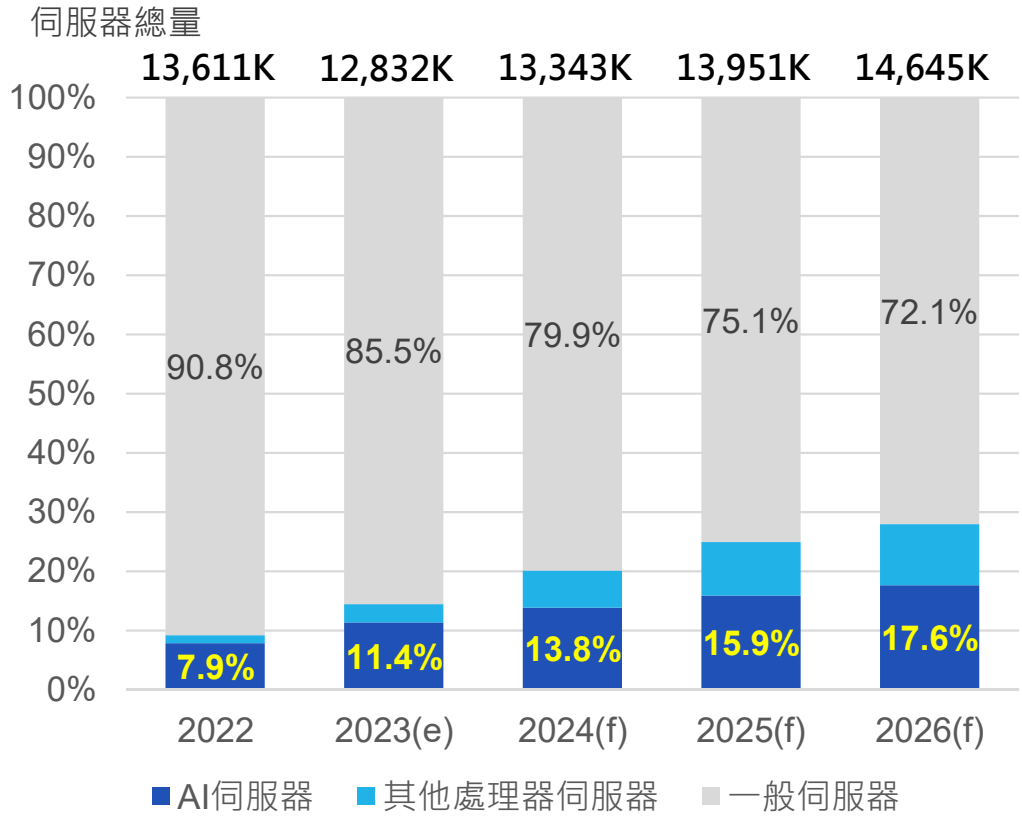
特徵 / 差異	一般伺服器	AI推論伺服器	AI訓練伺服器
主要運算單元	CPU	低階GPU、FPGA、ASIC	高階GPU
運算設備體積	高度以1U、2U為主	高度以1U、2U為主	根據應用，高度可達4U、5U、7U
出貨數量	最多，一般企業均會進行採用	次多，在未來小型AI模型蓬勃發展，有望進一步提升	最少，價格高昂，有大型語言訓練、AI訓練需求的企業才會採購
出貨方式	單台、整櫃出貨	單台、整櫃出貨	除單台、整櫃出貨外，透過超級電腦方式提供(DGX GH200)
伺服器的價格	1,500-3,000（美元）	根據加速卡，價格範圍較大 3,000-20,000（美元）	15萬以上（美元） DGX 8x H100價格高達20萬美元
總耗電量/熱能	最低，平均功耗200w-1,000w不等	平均功耗2,000-4,000w	最高，DGX A100 6.5KW DGX H100 10.2KW
代表性產品	一般市售伺服器	可搭載NVIDIA L4、A2等GPU AMD Xillinx Alveo 加速卡、 Intel Gaudi2和Greco的伺服器	可搭載NVIDIA H100 & A100、 AMD MI200&MI300、Intel Ponte Vecchio GPU的伺服器
作用/應用場景	數據量小，但應用多元且複雜的場景	人臉辨識 & 車牌辨識等，需要迅速進行大量推論的場景	AI大型語言模型訓練，數量龐大且須長時間訓練的數據

資料來源：MIC，2023年7月

- 一般、AI推論及AI訓練伺服器，各自符合其應用場景。在ChatGPT帶動的AIGC浪潮下，AI訓練伺服器的需求量增加，未來可望帶動AI推論伺服器需求

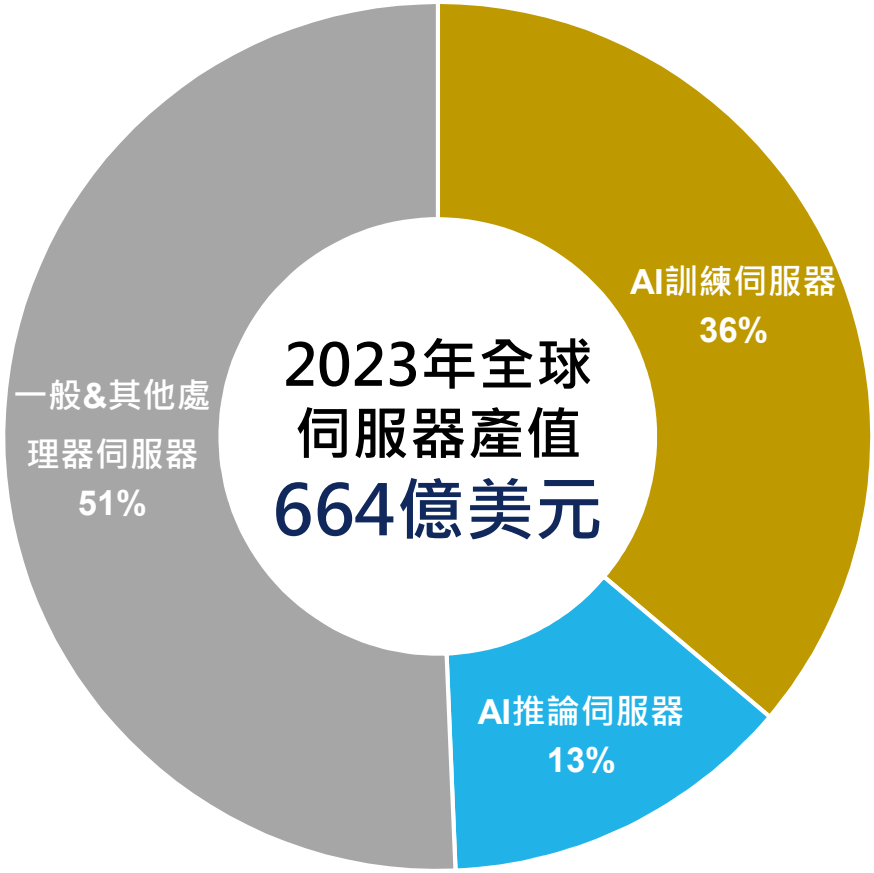
受惠於生成式AI熱潮，AI伺服器出貨與產值提升

2022-2026 全球AI伺服器占比預測



資料來源：MIC，2023年12月

2023全球AI伺服器產值占比預測

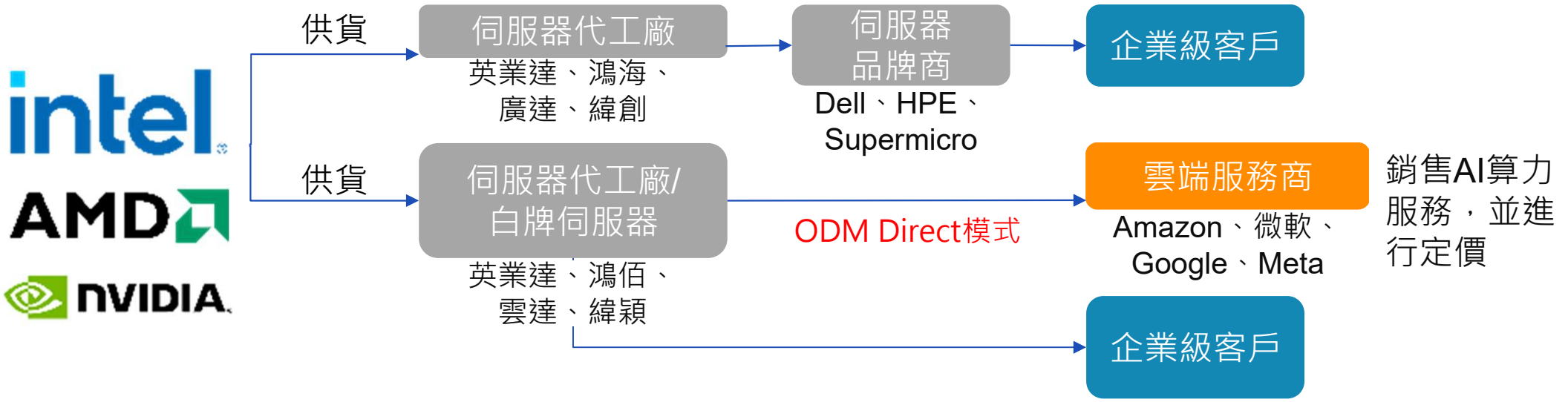


資料來源：MIC，2023年12月

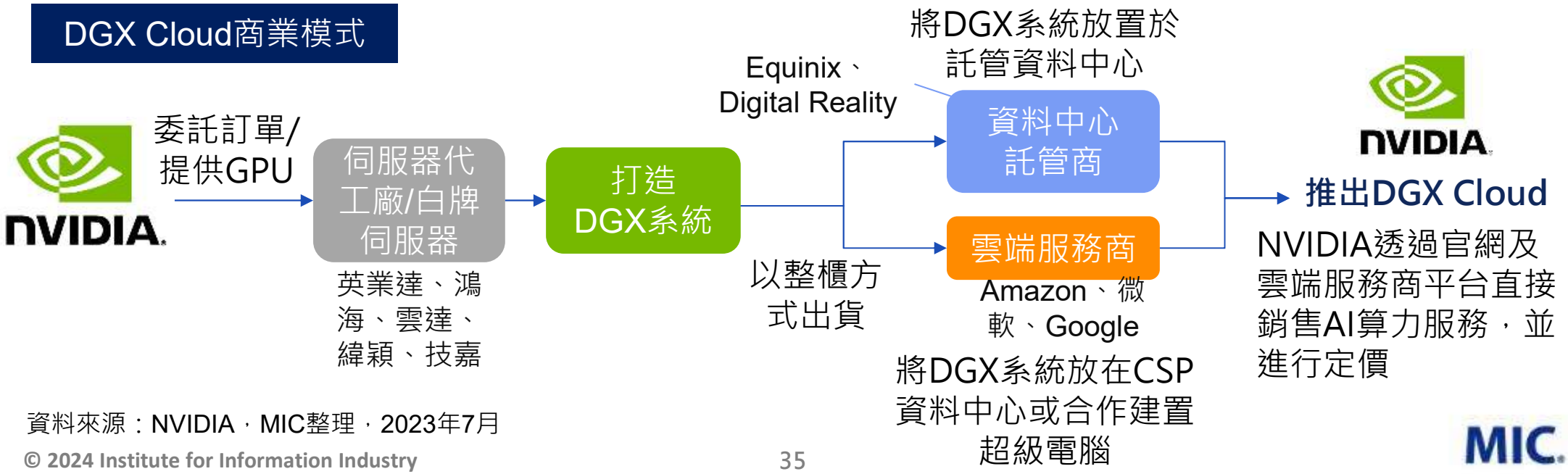
- 因ChatGPT的熱潮帶動生成式AI需求，全球大型企業投資AI訓練，中小型企業則開始導入相關AI應用，全球AI伺服器的佔比有望持續提升，從2022年的7.9%，提升至2026年的17.6%，當中包含價格昂貴並採用高階GPU的AI訓練伺服器，以及採用中低階GPU、FPGA、ASIC的AI訓練伺服器

NVIDIA推出DGX，顛覆傳統商業模式

傳統處理器廠主要商業模式



DGX Cloud商業模式



資料來源：NVIDIA，MIC整理，2023年7月

AI算力需求大幅提升改變伺服器託管型態

資料中心託管商面臨轉型問題

- 考慮高效能運算時，通常會想到在專門設計的設施中部署液冷系統來滿足其功率和散熱要求。有了人工智慧，事情就會變得棘手。雖然系統本身由許多相同的 CPU、GPU 和 NIC 提供支持，但它們部署的環境卻千差萬別
- 人工智慧、機器學習、深度學習和數據分析等工作負載，多年來一直在主機代管設施中存在和部署

資料中心託管商的兩難

DIGITAL REALTY

需平衡基礎設施的部署與客戶需求

電力和冷卻設施過度建設
→耗盡大量未得到利用的電力或冷卻技術

基礎設施落後於客戶需求
→不得不拒絕最挑剔的客戶

伺服器構建靈活性是關鍵，
需要以非常模組化的方式建設資料中心

冷卻系統與電力的調整

一般機櫃功耗7千瓦~10千瓦
Nvidia DGX H100每個機架功耗超過42千瓦

導入水冷背門熱交換器、導入液冷散熱裝置

DGX A100只需一個 A/B 電源，DGX H100需要三個獨立電源
→配電裝置(PDU)模組化

colovore

矽谷新創託管商鎖定GPU運算，
已募資800萬美金，採用液體冷卻

GPU運算/雲提供商抓準即時使用需求

廠商	A100(40GB)/h	A100(80GB)/h	H100/h	可用地區
Google cloud	3.67美金	5.67美金	缺貨無定價	全球
AWS	4.10美金	5.12美金	12.3美金(缺貨)	全球
Azure	3.4美金(缺貨)	缺貨無定價	缺貨無定價	全球
Lambda Cloud	1.1美金	1.5美金	1.99美金	美國
CoreWeave	2.06美金	2.21美金	4.25美金	美國
Cirrascale	2.87美金	3.25美金	4.28美金	全球
Paperspace	3.09美金	3.18美金	缺貨無定價	美國/歐洲
Latitude.sh			4美金(即時) 2.83美金(月/小時)	全球
Vast.ai	1.1美金	1.35美金	缺貨無定價	全球
Leader GPU	1.33美金	2.03美金	5.58美金	芬蘭

資料來源： Cloud ML Ltd、各廠商網站，MIC整理，2023年8月

- 大型雲端服務商應用實例即時價格明顯高於GPU運算/雲提供商，且H100都有缺貨無法提供的情形，可能原因包含自身使用、客戶過多等
- GPU運算/雲提供商在確保即時性的同時，價格相對較低，對於不需長期使用的客戶而言可以進行彈性調配

CoreWeave吸引NVIDIA、MSFT投資合作

成立時間	2017年	總部	美國加州聖塔克拉拉
員工人數	160人	背景	原為以太坊挖礦公司，2019年轉型GPU運算公司
知名投資企業	NVIDIA		

獲得的關鍵投資與合作

2023年4月，NVIDIA參與Coreweave 2.21億美元的B輪融資

2023年6月，MSFT同意未來幾年，付費使用CoreWeave雲端運算基礎建設，金額可能達到數十億美元

被視為Open AI的對手，2023年全球第二大獨角獸（2輪15.25億美金）Inflection AI，與CoreWeave及NVIDIA共同打造全球最大的AI叢集，該叢集最終將由2.2萬個NVIDIA H100 GPU所組成

熱門議題觀測(三)

美中科技禁令對RISC-V之影響

RISC-V興起背景與發展歷程

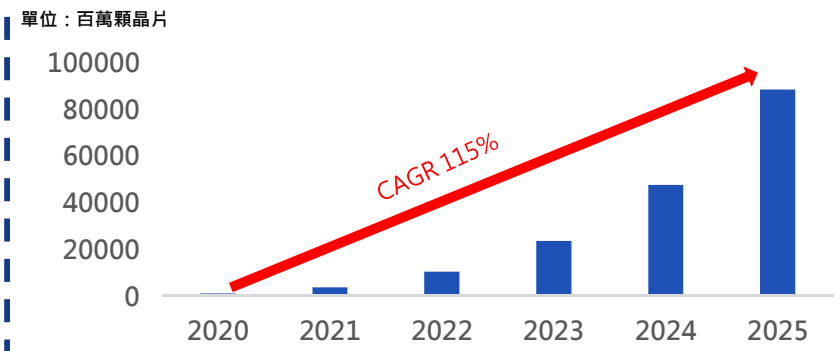
RISC-V 興起背景

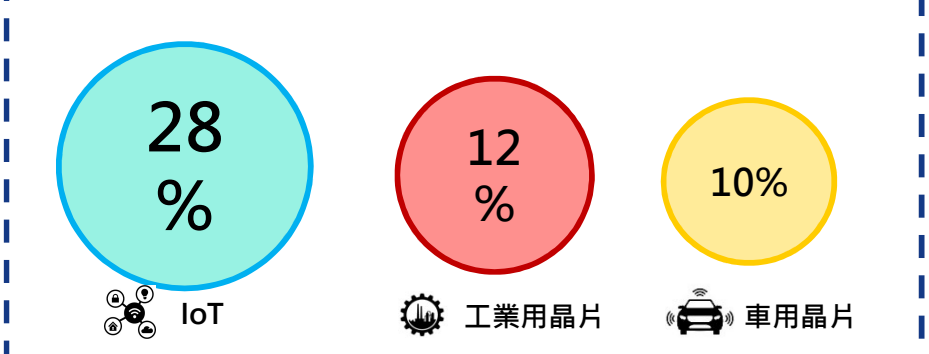
- 因低成本（免授權金）、低功耗比、小面積（較少歷史包袱指令），成為嵌入式與物聯網應用熱門架構
- 逐漸從IoT發展至資料中心、消費電子、高效運算等領域，至2022年7月全球RISC-V晶片總出貨量已達100億顆

RISC-V 發展歷程

- 2010 由加州大學推出
- 2015 RISC-V 聯盟成立
- 2016 NVIDIA 採用RISC-V 開發Falcon控制器
- 2019 高通S865 處理器 採用 RISC-V 為 SoC 微控制器
- 2022.2 Intel 加入RISC-V聯盟 投入10億美元創新基金
- 10B+ 2022.7 出貨量達100億顆
- 2022.9 NASA 正式採用RISC-V為航空運算CPU
- 2022.11 歐盟歐洲晶片法 將RISC-V列為發展重點之一
- 2023.1 Google 正式宣布 Android 支援 RISC-V
- 80B+ 2025 RISC-V預計出貨 800億顆

RISC-V 出貨成長 應用領域





資料來源： RISC-V International、各業者網站，MIC整理

RISC-V美中陣營布局分析

美國：朝伺服器與AI應用探索

大廠產品陸續導入 Google Qualcomm intel WD

- Google於2023年Q1正式支援RISC-V Android標準
- Qualcomm 在S865部分導入RISC-V架構，並已出貨超過6億顆晶片
- Intel NIOS處理器以RISC-V架構開發，WD 則以RISC-V開發儲存控制晶片，估計每年出貨10億顆

新創開發高階應用 esperanto.ai VENTANAMICRO

- espranto研發出全球首款RISC-V晶片的AI處理器，含上千RISC-V核心，以低能耗、高效率用於超大規模資料中心布署
- Ventana 以小晶片設計推出RISC-V資料中心處理器，包含16 核心，單核心運行頻率高達3.6GHz

關鍵發展

美國資訊大廠與RISC-V新創投入發展，帶動技術標準革新，並在資料中心、HPC、AI晶片等領域，具發展優勢

VS

- 美國半導體設備、人才、EDA對中禁運限制
- 部分Arm高階晶片IP 對中國大陸斷供

中國大陸：藉RISC-V達晶片自主目標

祭出補助並主導開源社群

- 針對物聯網、工控中低階應用、及支援主流作業系統的中高階RISC-V研發，提供1,000萬至2,000萬人民幣研發補助，及15%至20%銷售補助。
- 大廠聯合投入RISC-V International聯盟，在技術審定委員會取得27席中13席(48%席次)

大廠衝鋒，以大帶小 Alibaba 平头哥

- 阿里於2022推出全球首款RISC-V筆電，另與Google合作，提交全球首個RISC-V Android 標準，並已在行動裝置驗證成功
- 阿里提供RISC-V開發平台（IDE）與資源，以大帶小拉動中國自有RISC-V IP晶片產業鏈發展

關鍵發展

中國大陸業者在PC、智慧手機發展上搶得先機，對外主導RISC-V開源社群，對內由大廠資源帶動國內自有晶片產業鏈發展

資料來源： RISC-V International、各業者網站，MIC整理

路透：RISC-V恐將成為美中科技戰升級的新戰線

- 外電路透報導，RISC-V可能成為美中科技戰升級的新戰線。拜登政府正面臨議員壓力，部分國會議員以國家安全為由，敦促拜登政府在RISC-V採取行動。業界人士解析，倘若美中科技戰戰火延燒至RISC-V，恐將掣肘包括手機晶片、人工智慧（AI）領域等快速發展的機會
  - 根據聯邦眾議院中國問題特別委員會主席Mike Gallagher所提供的聲明指出，**商務部**必須「**規定任何美國人或美國企業在跟中華人民共和國實體就RISC-V技術接洽前，需取得出口許可證**」
  - 眾議院外交事務委員會主席Michael McCaul指出，**中國正在濫用RISC-V技術以規避美國在晶片設計所需的智財權方面的主導地位**。他希望相關部門採取行動，如果沒有的話，他將推動立法

RISC-V終端產品應用展望

終端產品	技術發展及產業現況	終端產品應用展望
IoT 嵌入裝置	IoT與嵌入晶片為低成本的RISC-V目前最大宗應用領域，且因IoT應用情境（如工廠生產線）歧異度高，無生態整合問題，市場需求逐漸增長	▲ RISC-V於IoT產品預計穩定提升市佔率
PC桌機 NB筆電	RISC-V筆電(linux系統)雖問世，但目前RISC-V對x86系統及軟體的整合度、支援度尚不足，除中系業者外尚無大廠投入開發。	▼ RISC-V於消費型PC、NB發展需保守觀望
伺服器 資料中心	新創已成功開發出RISC-V 伺服器，RISC-V的高效率、小面積、節能特性，可能使資料中心成為RISC-V快速發展的終端產品之一	▲▲ RISC-V於資料中心具高度成長潛力
HPC AI晶片	RISC-V具模組異構彈性，同時能在低功耗下，完成以往需要高瓦數的即時AI推論需求，在ESG及淨零大趨勢下，RISC-V具應用優勢	▲▲ RISC-V於AI晶片具高度成長潛力
智慧手機	雖行動晶片與軟體大廠（Qualcomm、Google）皆加強投向 RISC-V，但從Arm到RISC-V設計轉移門檻、作業系統與App優化仍有長路要走	— RISC-V於智慧手機具潛力，但高度依賴大廠（如Qualcomm、Apple）未來投入動向

資料來源： MIC整理

RISC-V未來商機發展

- **RISC-V預期為台灣IC產業鏈、IoT及新創業者帶來契機**
  - 基於RISC-V晶片及系統上層的軟體商、與其下層的硬體製造、解決方案提供商，皆能雨露均霑。除了國內既有IC產業鏈上的業者，能從RISC-V的IP設計（如晶心、力旺）、IC設計與IC設計服務（如創意電子、凌陽、奇景、群聯）、IC代工（台積電、聯電、力積電、旺宏）需求中受惠
  - 在RISC-V迅速發展的IoT、嵌入式領域，國內亦有針對不同垂直領域之業者（如工業AIoT、倉儲AMR等），可在既有對Arm體系的支援外，利用RISC-V的低成本、高彈性設計優勢，提供特定產業的解決方案

- **RISC-V架構仍在起步，然體系逐漸完善，具發展潛力之因素**
  - 包含Intel、Qualcomm等企業正在大力支持RISC-V發展，中國大陸則是因美國禁令的關係，亦會將更多資源投注至RISC-V架構
  - 雲端服務商及企業用戶，因為算力異構化、複雜化等原因，在垂直領域對伺服器對客製化需求增加
  - RISC-V作為開源架構，可以讓使用者打造更彈性化的處理器，符合用戶的需求
  - 因為RISC-V具有設計彈性，可以和Intel、AMD或Arm的CPU進行搭配使用，作為強化算力的手段
  - 至於GPU、FPGA、ASIC等AI加速器，也同樣可以基於RISC-V架構進行設計。在各方的支援下，RISC-V的生態系正在逐步完善，因此看好其長期在伺服器與資料中心的發展

RISC-V成為伺服器與處理器的新商機

intel PATHFINDER FOR RISC-V

2023年1月，Intel因結構調整停止探路者計畫，然對RISC-V開發平台、合作計畫沒有影響

Veyron V1 處理器

採用RISC-V架構的伺服器處理器，可以達到16核，使用TSMC 5nm製程

玄鐵C系列

主打複雜運算，共有4款處理器，針對AI及高效能應用

晶心科AX60系列

可使用於資料中心以及5G基礎設施

Catapult 處理器

專為異質運算設計，適用於資料中心與高效能運算

ET-SoC-1 AI 推論處理器

擁有1,088核的RISC-V處理器，

資料來源：RISC-V Summit、各公司，MIC整理

- RISC-V作為開源架構，擁有彈性設計、無IP授權費、不受美國晶片法案禁令影響的優勢，預期在美系廠商發展之外，中系處理器業者應會積極發展RISC-V伺服器處理器

結論

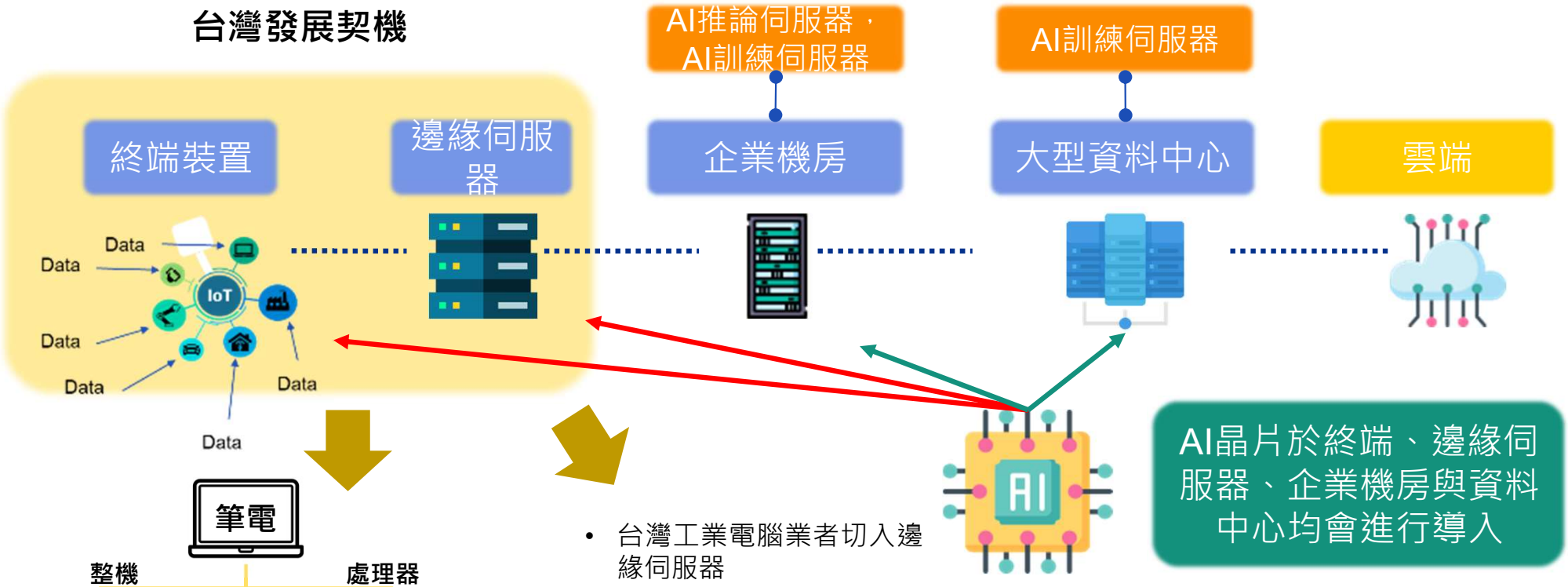
結論(1/2)

- **AI PC議題火熱然應用趨勢未明，2024年PC市場回溫有限**
  - 儘管全球經濟情勢仍未明朗，2024年全球PC市場表現將在多項正向因素支持下，略好於2023年。正面因素如Windows 10將停止支援、商用PC更新週期來臨等，帶來正面影響效應，而AI PC的問世更為PC市場需求帶來一劑強心針，不過仍需軟體支援尋找應用利基，以利滲透率的加速
- **區域市場復甦不同調，新興市場成長看好**
  - 儘管週期換機將有望帶動PC市場需求增長，然全球景氣回溫復甦進度影響PC採購力道。歐美市場通膨趨緩，消費動能在2023下半年已略見好轉，預期2024年PC市場將能有正向表現；新興市場如印度，消費力道受惠數位需求增長而維持強勁；中國大陸則仍陷經濟疲弱，有待復甦
- **處理器三強競逐AI PC主導權，處理器算力競爭白熱化**
  - Intel、AMD及Qualcomm相繼發布具AI運算功能行動處理器，三強將在2024年以AI算力決一勝負；桌機部分，Intel、AMD將發布新架構處理器，在2024年下半年競逐消費級PC產品算力寶座

結論(2/2)

- **2024年AI訓練與AI推論伺服器同步發展，帶動全球伺服器市場**
  - 繼雲端服務商之後，伺服器品牌商亦開始推出AI伺服器產品。然因NVIDIA H100 GPU的產能，供不應求，使得AI伺服器訂單延遲至2024年。展望2024年，各企業將持續搶購高階AI訓練伺服器，同時中小生成式AI模型落地，將有望促成AI推論伺服器的發展，帶動全球伺服器市場出貨
- **2023年企業級伺服器訂單低迷，雲端服務商出貨超過三成**
  - 2023年受到全球景氣低迷的影響，各國企業縮減資本支出使企業級伺服器訂單大幅縮減，影響到伺服器品牌商的出貨，並使雲端服務商訂單的比重持續增加。展望2024年，伺服器品牌商開始擴充產品線，布局AI伺服器市場，預期相較2023年整體出貨會有明顯回溫
- **2024年伺服器處理器市場更加複雜，AMD市占持續上升，更多廠商進入Arm架構生態圈**
  - 2024年伺服器處理器在CPU方面，受惠於產品性價比優勢，AMD市佔將會持續上升。另外Arm架構生態圈吸引許多希望自研處理器的廠商，包含導入多年的AWS、Ampere，研發Grace CPU的NVIDIA，以及Microsoft發表的新處理器Cobalt，預期未來競爭將更加激烈

未來產業發展重點及建議：供應鏈觀點



- **從AI帶動的硬體基礎設施觀察，台廠透過協助客戶生產AI訓練與推論伺服器，已居關鍵地位**
- **長期而言，因低延遲與資安考量，AI亦將往地端(local)發展，台灣IPC業者已切入的邊緣伺服器、以及台廠向來擅長的終端裝置，均有機會參與商機**

資料來源：MIC，2023年12月

MIC® 產業提昇的關鍵力量

Thank You

魏傳虔 產業顧問兼組長 chriswei@iii.org.tw

產業情報研究所

智慧財產權暨引用聲明

- 本活動所提供之講義內容或其他文件資料，均受著作權法之保護，非經資策會或其他相關權利人之事前書面同意，任何人不得以任何形式為重製、轉載、傳輸或其他任何商業用途之行為

- 本講義內容所引用之各公司名稱、商標與產品示意照片之所有權皆屬各公司所有

- 本講義全部或部分內容為資策會產業情報研究所整理及分析所得，由於產業變動快速，資策會並不保證本活動所使用之研究方法及研究成果於未來或其他狀況下仍具備正確性與完整性，請台端於引用時，務必注意發布日期、立論之假設及當時情境