DOI: 10.6653/MoCICHE.202104_48(2).0009

人工智慧平臺淺析與硬體效能測試

張慰慈／財團法人國家實驗研究院國家地震工程研究中心 助理研究員

人工智慧（Artificial Intelligence，簡稱 AI）已經成為近十年來的顯學，專家學者無不思索這一波科技浪潮將帶來的衝擊和助益。雖然目前人工智慧的軟硬體技術都很容易取得，入場門檻相對其他高速計算領域來得低上許多，但一般使用者對人工智慧平臺架構和其最重要硬體設備 — 圖形處理器通用計算（General-Purpose computing on Graphics Processing Units，簡稱 GPGPU、GP2U 或 GPU）卡 — 的瞭解都可能很是有限。本文將淺析人工智慧平臺架構與軟硬體選擇的思維，再由實際測試兩種 NVIDIA GPU 卡的加速（speed up）效益，期能給予有志於人工智慧的研究人員有所助益。

背景

近十年來人工智慧（Artificial Intelligence，簡稱 AI）的發展如火如荼，已從剛開始的基礎技術發展轉移到個別領域應用，目前各應用領域的至專家們無不思索這一波人工智慧浪潮將帶來的衝擊和助益。拜美商輝達（NVIDIA）所提出的圖形處理器通用計算（General-Purpose computing on Graphics Processing Units，簡稱 GPGPU、GP2U 或 GPU）卡和美商谷歌（Google）在軟體開發上的努力，現在即使是一般使用者也能輕易取得這些已經發展好的軟硬體資源。

由於人工智慧開始有機會真正在各應用領域做出貢獻，有愈來愈多的研究中心或使用者需要建置自有的人工智慧平臺、又或需要以租賃的方式取得人工智慧計算資源。但人工智慧平臺的硬體有其獨特性，並非純以中央處理器（Central Processing Unit，簡稱 CPU）為主要計算資源的傳統高速計算平臺思維所能涵蓋。此外在人工智慧發展的道路上，NVIDIA 所提出的 GPU 卡產品線有 GeForce、Quadro、Tesla 三個系列，除了專注在顯示加速的 Quadro 系列之外，一般通用型 GeForce 系列和伺服系用 Tesla 系列的 GPU 卡都可應用於人工智慧加速之上。但頂級 Tesla 卡的取得價格約為頂級 GeFroce 系列卡的八至十倍上下，其效益差距則眾說紛紜。本文即由人工智慧平臺的架構出發，先討論平臺所需的架構設計、再評析 NVIDIA 的 GeForce 和 Tesla 系列 GPU 卡的效能差別，以供欲建置自有人工智慧平臺或租賃設備的同好一個參考。

平臺架構

人工智慧平臺的架構一般而言是基於主從式架構（client-server model）拓展，也就是經由一系列的「伺服器」（server）和「客戶端」（client）關係結合，某一臺電腦可以為其他的電腦的「伺服器」，也可以是另一臺電腦的「客戶端」。一個稍具雛型的人工智慧平臺架構如圖 1 所示。計算功能和任務相異的各電腦稱為「計算節點」（computing node），首先是需要一個中央行程控制電腦「管理電腦」（manager）來管理這許多的計算節點；管理人員（administrator）是唯一可以登入操控和設定「管理者」的角色，而其他的用戶可能經由個人電腦（personal computer）或筆電、手機、平板等行動裝置（mobile device）間接經由一個權限較低的「前端電腦」（frontend）登入人工智慧平臺；而用戶們也可能需要一個「儀表板電腦」（dashboard）來快速查看人工智慧平臺的使用狀況和附加資訊；如果需要較大的儲存空間，那便需要再新增一座「儲存設

施」(storage),這座儲存設施除了連接「管理電腦」以存放資料外,也能同時提供不同的資料給「前端電腦」、「儀表板電腦」和其他「計算節點」,以確保資料存取的效率和安全性。這裡的「管理電腦」、「前端電腦」和「儀表板電腦」可以是同一臺電腦扮演該角色,但是為了安全性考量還是應當把「管理電腦」獨立出來,因為我們難以確保一般使用者的個人電腦裝置都有良好的防護,當他們的帳號密碼外洩時,「前端電腦」便成為被攻擊的第一站;將「管理電腦」和「前端電腦」分開的設計就隱含了「壯士斷腕」但的思維 — 一旦「前端電腦」被攻陷,至少可以切斷「管理電腦」和「前端電腦」間的連結來確保後方電腦的安全。

過去曾發生過這樣的案例:一個使用者的個人電腦不幸被入侵,駭客用該使用者的帳號密碼登入「前端電腦」,但一般使用者的權限有限 — 既無法窺探其他使用者的資料、也不被允許跳到「管理電腦」上。雖然該次入侵並沒有危急到「管理電腦」和後方的其他「計算節點」,但是畢竟還是可以存取「計算節點」上的硬體計算資源 — 即原本一般使用者就擁有的權利,因此還是利用該人工智慧平臺去計算獲取虛擬加密貨幣「比特幣」(Bitcoin,BTC 或 XBT)獲利,也是是俗稱的「挖礦」(mining),最終由管理人員監控該人工智慧平臺時察覺異狀才制止,可見駭客對人工智慧平臺的覬覦實在是防不甚防,也突顯出不厭其煩地為人工智慧平臺設能兼具效能和安全性的架構是如此重要。

軟硬體選擇

主機虛擬化

在前述的人工智慧平臺架構設計上,因為真正需要負責重量級 CPU、GPU 資源的都是後方的各計算節點,因此其他的「管理電腦」、「前端電腦」和「儀表板電腦」的硬體資源需求便相對來得低上許多。目前有許多虛擬主機的技術就已經可以滿足這三種電腦角色的需求,例如美商威睿(VMware)的各項產品就能提供建置人工智慧平臺「管理電腦」、「前端電腦」和「儀表板電腦」的良好的解決方案,最簡易的方法是在一臺數萬元等級起跳的電腦上安裝裸機虛擬化管理軟體 VMware vSphere Hypervisor ESXi(或簡稱 ESXi)[1],免費版的 ESXi 具有「沒有官方支援」、「單一虛擬電腦最多只能使用 8 核心 CPU」、「不能由進階管理軟體 vCenter 控制」、「不支援高效能儲存管理介面 vStorage API」等限制,但這些恰恰都不是「管理電腦」、「前端電腦」和「儀表板電腦」所必要的部份,也就是選擇 ESXi 就可以節省管理軟體的開銷,而非計算節點也只需要一臺實體電腦即可(如圖 1 的「VM」字樣所示)。

此外,因為外部連線的電腦只需要看到「前端電腦」和「儀表板電腦」,所以只有這兩種角色需要實體網路位址(Internet Protocol Address,簡稱 IP Address 或 IP),甚至可以利用 ESXi 核發取得單一實體 IP、用虛擬 IP 轉發方式將使用者的連結方式分別導向到虛擬的「前端電腦」和「儀表板電腦」上即可,這樣一來就可以節省租賃網路實體 IP 的數量和成本(如圖 1 的「IP」字樣所示)。而防火牆等安全防護也只需要設立在「管理電腦」、「前端電腦」和「儀表板電腦」的對外連線方面(如圖 1 的火焰圖案所示)。

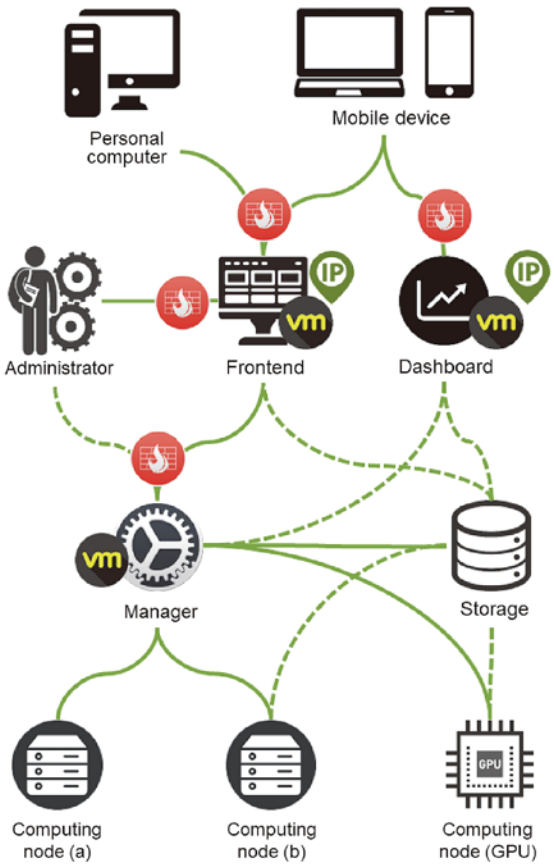

圖 1 人工智慧平臺架構示意圖(虛線是非必要連接方式)

關鍵硬體

至於計算節點的硬體，最重要的是挑選 CPU、記憶體及 GPU 的規格。CPU 以美商英特爾（Intel）與超微半導體（AMD）兩家企業的產品為主，如果平臺使用情境需要著重 CPU 特定的科學計算，必須考量某些會用到的數值函式庫會針對 Intel CPU 特別設計加速；若是以 GPU 加速為主的使用情境則可以考慮「CPU 核心數：GPU 卡數」=「2：1」的搭配，這是因為一般 GPU 加速的人工智慧程式會使用約 1.6 至 2 核心量的 CPU，再增加 CPU 核心的加速效益便不顯著。舉例而言，如果單一計算節點安裝八張 GPU 卡，所需搭配的 CPU 總核心數最小值為 16 核心，但更多的 CPU 核心所能帶來的效益便很有限，關於 CPU 和 GPU 的加速效能測試將會在下一節再行說明。單獨考慮 CPU 時的考量重點為其時脈頻率（clock rate），而不同型號的 CPU 則可以很容易地搜尋到網路的評析報告得知其效能差異。

記憶體方面就較為難以仔細評估，必須由使用情境的程式資料量來計算，最好是先在一般電腦上測試使用案例取得初步記憶體需求量。但需要注意的是：大型程式並不會將所有的資料都同時載入記憶體中，比如說 100 GB 的資料並不會一次載入系統記憶體中，而可能是分批以 1 GB 方式讀入記憶體中使用；對以 GPU 為主的人工智慧計算程式更是如此，常常人工智慧訓練資料集是數百 GB，但是訓練時經常以批次（batch）的方式循序載入，使用者應該計算批次量來估計所需的記憶體量；此外，一張 GPU 卡的記憶體僅有數 GB 到 32 GB 不等，也就是說一次的計算量所會用到的記憶體上限為 32 GB，這時候單一程式對電腦系統的記憶體需求就不會高於 32 GB 多少。整體系統記憶體量的需求要計算多個程式的總量，某些更大型的程式可以再進階考慮串聯多張 GPU 卡的情境使用。

GPU 卡雖然有美商輝達（NVIDIA）和 AMD 兩家企業的產品為主，但是在人工智慧計算領域仍是以 NVIDIA 為主，這是因為 NVIDIA 有提出軟硬體整合技術 CUDA（Compute Unified Device Architecture）[2] 以加速其自家產品的效能，目前世界上所發展的人工智慧軟體也多基於 CUDA 的環境開發，因此在 NVIDIA 的 GPU 卡上的效能便較 AMD 來得突出。此外 GPU 需要考量其核心數，NVIDIA 的 GPU 卡核心又分為 CUDA 核心（即一般核心）和 Tensor 核心，後者對於張量計算表現更為重要。除核心數和記憶體量之外，另外需要考量的有記憶體介面頻寬（frequency of memory interface），NVIDIA 的 Quadro 系列 GPU 卡之所以較不適合人工智慧計算就是因為其記憶體量和介面頻寬較小，即使記憶體量與 GeForce 或 Tesla 系列的卡相若也無法發揮效能；至於所謂的光線追蹤（或稱光跡追蹤，ray tracing）就和人工智慧更沒有關聯。

作業系統

人工智慧程式運作的作業系統以自由和開放原始碼的作業系統 LINUX [3] 為大宗，一方面是因為開放原始碼作業系統可以省下作業系統費用，另一方面是系統的效能和安全性遠非美商微軟（Microsoft）的視窗作業系統 Microsoft Windows [4] 所能企及。Microsoft Windows 啟動時的記憶體佔用量為 1 至 2 GB，但人工智慧計算節點使用的 LINUX 一般並不會啟用圖形介面，其作業系統記憶體使用量最多僅有數 MB 而已，而 CPU 的使用量也是如此。

LINUX 的發行版主要考量為 Red Hat 與 Debian 系列。Red Hat 系列為美商紅帽（Red Hat）企業所發行，Red Hat 企業贊助自由軟體社群 Fedora，在匯集 Fedora LINUX [5] 測試的軟體後選擇穩定版本匯集為其企業版 RHEL（Red Hat Enterprise Linux）[6]，發行後會將原始程式碼釋出給全世界供免費使用，著名的衍生發行版為 CentOS LINUX [7]、Scientific LINUX [8] 和 Oracle LINUX [9] 等三者，其中 CentOS 為 Red Hat 企業官方支援，但在 2020 年 12 月 8 日 Red Hat 企業宣佈停止發行 CentOS 穩定版，無疑是斷絕了 CentOS 的發展可能；而美國費米國立加速器實驗室（Fermi National Accelerator Laboratory，縮寫為 Fermilab 或 FNAL）和歐洲核子研究組織（European Organization for Nuclear Research，簡稱 CERN）合作開發的 Scientific LINUX 也已經在 2019 年 4 月宣佈停止開發，目前 Red Hat 系列免費發行版只剩下美商甲骨文（Oracle）企業的 Oracle LINUX 而已，但足以為取代 RHEL 的選擇。Debian LINUX 是由自由軟體 Debian 計畫（Debian project）所發行 [10]，其最著名的衍生版本為英商肯諾（Canonical）與 Ubuntu 社群所發行的 Ubuntu LINUX [11]，也有穩定版本 LTS（長期

支持版本，全名為 Long Term Support）可為人工智慧平臺建置時選擇。

建置人工智慧平臺時應選擇穩定版 LINUX 為宜，同時應優先選擇使用者量大的發行版，在發生問題時才能夠較快速地在相關的網路社群找到解決方案。免費版本 LINUX 在 Oracle LINUX 與 Ubuntu LINUX 二者間的主要考量為是否會使用到僅支援特定系列 LINUX 的軟體，否則兩者的選擇上並無太多差別。

效能測試環境與評估指標

本文使用的 CPU 有 Intel Xeon E5（2.6 GHz）、Xeon Gold（2.1 及 2.6 GHz）等三種；而 GPU 卡則挑選 2020 年最頂級的 GeFroce 卡（RTX 2080 Ti）和 Tesla 卡（Tesla V100-16G 和 Tesla V100-32G）。作業系統為 64 位元版本的 CentOS Linux 7.3，對應之 Linux 核心版本為 3.10.0-1160.6.1.e17；執行 AI 測試的程式為 Python 3.8.7 搭配 Tensorflow 1.12 和 2.1 版。在測試案例與程式碼方面，本研究使用 Andrey Ignatov 所撰寫的 AI 測試套件 ai-benhmark [12]，針對分類（classification）、圖像映射（image-to-image mapping）、圖像分割（image segmentation）、圖像修復（inpainting）、語句分析（sentence sentiment analysis）、文本翻譯（text translation）等六種常見的 AI 案例進行測試，再以這些案例給予綜合評分。

加速效益的評比指標採用傳統的加速值（speed up）：定義增加 N 核心 CPU 或 N 張 GPU 卡後的加速值 s_N 可以表示為：

$$s_N = \frac{T_{ref}}{T_N} < N \tag{1}$$

其中 T_{ref} 為未增加 CPU 或 GPU 時的計算時間（wall-clock time），T_N 為增加 N 核心 CPU 或 N 張 GPU 卡後的計算時間，其理想值為 N。進一步可計算硬體使用效率（efficiency）e_N：

$$0 < e_N = \frac{s_N}{N} < 1 \tag{2}$$

其值介於 0 至 1 之間，愈高表示硬體獲得愈高的使用量，即閒置或冗餘的計算愈少，通常理想的線性代數 CPU 加速效率 e_N 相當接近 1，而科學或工程計算軟體的 e_N 值則多介於 0.7 至 0.9 之間。

測試與討論

CPU 的效益

首先是 CPU 的影響：分別在 Tensorflow 1.12 與 2.1 版測試由一核心 CPU（Intel Xeon E5，時脈 2.6 GHz）增加至 24 核心時的加速效益（表 1），在分類、圖像映射、圖像分割、圖像修復、語句分析方面均確有助益，加速值 s_N 約在 3.6 至 10.2 倍之間，考量增加核心數為 24，硬體使用效率 e_N 落在 0.15 至 0.425 之間，和工程軟體相較都可說是相當低下，可見增加 CPU 對 AI 效率的提昇幫助很有限。再者文本翻譯的加速值 s_N 仍為 1.0，即增加 CPU2 根本無法加速。

Tensorflow 版本的效率差異

再談 Tensorflow 1.12 和 2.1 的差異，分別在一與 24 核心 Intel Xeon E5 CPU（2.6 GHz）環境下測試 2.1 版相較 1.12 版效能加速值 s_N 如表 2 所示。在分類、圖像映射、圖像分割、語句分析上 2.1 版的效能均較佳，文本翻譯則並無差異，而圖像修復上反而較差。但考慮將來 Tensorflow 2.1 版的持續革新，圖像修復演算法效能應該會持續增加，加之以程式開發的便利性與穩定性考量，還是建議 Tensorflow 1.x 的使用者儘快開始使用 2.x 的開發環境。

NVIDIA RTX 2080 Ti GPU 卡的效益

接著測試 NVIDIA 於 2020 年在 GeForce 2000 系列的頂級 GPU 卡 RTX 2080 Ti。在一與 32 核心 Intel Xeon E5 CPU（2.6 GHz）的環境下增加一至兩張 RTX 2080 Ti GPU 卡的加速值 s_N 如表 3 所示，基本上增加 GPU 卡數並不會帶來額外的加速，推測這是因為測試的範例並沒有超過一張 GPU 卡的記憶體量 11 GB。一張 RTX 2080 Ti GPU 卡在分類和圖像分割的加速就可以達到 200 倍以上，而圖像映射更可加速約 460 倍，即使是圖像修復、語句分析和文本翻譯也都能帶來 14 至 32 倍的加速，和純粹增加 CPU（表 1）的個位數加速明顯相差極鉅。

而 32 核心 CPU 環境下，也同樣看到一或兩張 GPU 卡的差異並不大；增加一張 RTX 2080 Ti GPU 卡的加速 s_N 便僅有 2 至 54 倍（如表 3 所示），這是因為 32 核心 CPU 已經有相當程度的加速效益，比較的基數

估計已經較高，因此 GPU 卡的效果便不如單核心 CPU 環境那樣出色；僅有文本翻譯因為 CPU 並不能帶來任何加速（由表 1 文本翻譯的加速值 $s_N = 1.0$ 可知 CPU 並無幫助），因此其加速值約為 13 倍，與單核心 CPU 的 14 倍為同一等級。

NVIDIA Tesla V100 GPU 卡的效益

緊接著測試 NVIDIA 於 2020 年在 Tesla 系列的頂級 GPU 卡 V100，此處測試有 16 GB 和 32 GB 記憶體兩種規格，其結果如表 3 所示。

同樣地，單張或兩張 Tesla V100 GPU 卡的差異並不明顯，推測也是測試的範例並沒有超過一張 GPU 卡的最小記憶體量 16 GB；且 28 乃至 32 核心 CPU 環境下添加 GPU 卡的加速也不如單核心 CPU 來得顯著，道理便和前述的 RTX 2080 Ti GPU 卡相同。

但值得注意的是：單核心 CPU 添加 V100 GPU 卡的加速值在分類可達到 186.2 至 197.2 倍，圖像分割可達到 95.7 至 137.2 倍，而圖像映射的加速值更高達 263.6 至 365.5 倍。一如 RTX 2080 Ti GPU 卡的測試結果，圖像修復、語句分析、文本翻譯的加速值 s_N 也都較低 — 約為 10.0 至 32.1 倍。

不同 CPU、GPU 環境下的綜合評分

由於測試不同 GPU 卡環境的 CPU 型號均不相同，且項目有分類、圖像映射、圖像分割、圖像修復、語句分析、文本翻譯等六種 — 其加速效益亦各有優劣，因此使用 ai-benchmark 的綜合評分來通盤比較，其分數愈多表示效能愈好，其結果如表 5 所示。

不借助 GPU 的幫助、單獨使用一核心 CPU 時的綜合評分分別為 123、214、170 分，增加至 8 至 32 核心 CPU 時的分數可達到 1,101、1,598、829，明顯可見 CPU 核心數增加時確實可以帶來數倍的加速。

但增加 GPU 時的加速效益又是另一種等級：RTX 2080 Ti GPU 卡的分數分別為 27,186、26,934、26,013、25,732 分；Tesla V100 GPU 卡的分數分別為 30,141、32,437、32,985、32,797 分（16 GB 記憶體）和 29,258、28,700、30,474、30,288 分（32 GB 記憶體），相較僅有 CPU 的環境再提昇了數十倍之多，可見一張 GPU 卡對 AI 的加速效益完全非多核心 CPU 所能企及。

比較 RTX 2080 Ti GPU 卡、V100 GPU 卡（16 GB 與 32 GB）三種型號的綜合評分，前者的分數介於 25,000 至 27,000 分等級，後者則介於 28,000 至 32,000 分之間。雖然看得出 Tesla V100 GPU 卡的效益確實較 RTX 2080 Ti 卡為高，但其領先幅度也僅約 12% 至 20%，就 Tesla V100 GPU 卡 8 倍 RTX 2080 Ti GPU 卡的價差來說便顯得後者要經濟得許多，足見當案例的記憶體用量在 11 GB 以下時仍應考量使用 RTX 2080 Ti GPU 卡，只有當記憶體量大的時候使用 Tesla V100 卡才能拉大差距。一般使用者常誤以為這裡所說的記憶體大小即為 AI 中的資料集（dataset）檔案大小，實際上 AI 訓練時會分批次（batch）將資料集載入 GPU 記憶體中，每批次的資料集可能只有數 GB，因此真正會達到 11 GB 記憶體用量的時機並不如想像中得多。

表 1　不同版本 Tensorflow，增加 23 核心 Intel Xeon E5 CPU（2.6 GHz）的加速值 s_N

Tensorflow 版本	分類	圖像映射	圖像分割	圖像修復	語句分析	文本翻譯
1.12	6.1	3.9	7.0	9.2	8.6	1.0
2.1	7.9	6.0	7.3	3.6	10.2	1.0

表 2　不同核心數 CPU 環境下，Tensorflow 1.12 升級至 2.1 版的加速值 s_N

CPU 核心數	加速值 s_N					
	分類	圖像映射	圖像分割	圖像修復	語句分析	文本翻譯
1	1.7	1.6	1.5	5.5	1.1	1.0
24	2.2	2.4	1.6	2.1	1.3	1.0

註：統一採用 Intel Xeon E5 CPU（2.6 GHz）

表 3　不同核心數 CPU 環境下增加 NVIDIA RTX 2080 Ti GPU 卡的加速值 s_N

CPU 核心數	GPU 增加數量	加速值 s_N					
		分類	圖像映射	圖像分割	圖像修復	語句分析	文本翻譯
1	+1	271.4	460.7	213.7	14.8	32.2	14.0
1	+2	271.0	458.2	211.2	14.6	31.8	13.8
32	+1	21.1	54.1	17.9	2.1	5.9	13.1
32	+2	20.4	54.1	17.3	2.2	6.0	13.5

註：統一採用 Intel Xeon E5 CPU（2.6 GHz）

表 4　不同核心數 CPU 環境下增加 NVIDIA Tesla V100 GPU 卡的加速值 s_N

環境				加速值 s_N					
CPU		GPU		分類	圖像映射	圖像分割	圖像修復	語句分析	文本翻譯
核心數	時脈（GHz）	數量	記憶體（GB）						
1	2.6	+1	16	186.2	256.9	129.0	10.0	29.7	11.0
1	2.6	+2	16	189.6	263.6	137.2	16.5	43.7	16.6
28	2.6	+1	16	19.7	38.0	17.6	2.9	8.6	10.3
28	2.6	+2	16	19.6	37.9	17.5	2.9	8.8	10.6
1	2.1	+1	32	197.2	365.5	95.7	15.2	28.7	16.2
1	2.1	+2	32	187.5	353.9	101.2	15.2	32.1	15.7
24	2.1	+1	32	31.1	88.2	19.4	3.7	9.6	16.2
24	2.1	+2	32	31.3	86.4	20.3	4.3	9.8	16.7

註：統一採用 Intel Xeon Gold CPU

表 5　不同 CPU、GPU 環境下的 ai-benchmark 綜合評分

CPU（Intel Xeon）			GPU（NVIDIA）			評分
核心數	型號	時脈（GHz）	型號	記憶體（GB）	數量	
1	E5	2.6	—	—	—	123
1	E5	2.6	RTX 2080 Ti	11	1	27,186
1	E5	2.6	RTX 2080 Ti	11	2	26,934
32	E5	2.6	—	—	—	1,101
32	E5	2.6	RTX 2080 Ti	11	1	26,013
32	E5	2.6	RTX 2080 Ti	11	2	25,732
1	Gold	2.6	—	—	—	214
1	Gold	2.6	Tesla V100	16	1	30,141
1	Gold	2.6	Tesla V100	16	2	32,437
28	Gold	2.6	—	—	—	1,598
28	Gold	2.6	Tesla V100	16	1	32,985
28	Gold	2.6	Tesla V100	16	2	32,797
1	Gold	2.1	—	—	—	170
1	Gold	2.1	Tesla V100	32	1	29,258
1	Gold	2.1	Tesla V100	32	2	28,700
24	Gold	2.1	—	—	—	829
24	Gold	2.1	Tesla V100	32	1	30,474
24	Gold	2.1	Tesla V100	32	2	30,288

結論

本文介紹建置人工智慧平臺時的架構與軟硬體評析，同時測試三種 NVIDIA 卡對六種人工智慧案例的加速效益和綜合評比，希望能提供欲建置或租賃人工智慧計算平臺的使用者參考之用。使用者應先瞭解自己的 AI 計算需求才能避免購置過多的高價設備、但又無法得到相應的研究效益。

參考資料

1. VMware. (2001). VMware ESXi, https://www.vmware.com/products/esxi-and-esx.html
2. NVIDIA. (2007). Compute Unified Device Architecture, https://developer.nvidia.com/cuda-zone
3. Linus Benedict Torvalds. (1991). LINUX, https://www.kernel.org/
4. Microsoft. (1985). Microsoft Windows, https://windows.microsoft.com/
5. Fedora Project. (2003). Fedora Linux, https://getfedora.org/
6. Red Hat. (2003). Red Hat Enterprise Linux, https://www.redhat.com/en/technologies/linux-platforms/enterprise-linux
7. Red Hat. (2004). Community Enterprise Operating System, https://www.centos.org/
8. Fermilab and CERN. (2004). Scientific LINUX, https://www.scientificlinux.org/
9. Oracle. (2006). Oracle Linux, http://www.oracle.com/us/technologies/linux/index.html
10. Debian Project. (1993). Debian Linux, https://www.debian.org/
11. Canonical. (2004). Ubuntu Linux, http://www.ubuntu.com/
12. Andrey Ignatov. (2021). ai-benhmark, https://ai-benchmark.com/

版權聲明