

臺北科技大學靠 AI-Stack 公平分配 GPU 資源 鼓勵老師積極研究、催生創新應用

當前各產業力拼數位轉型，對人工智慧（AI）人才需求越來越大，尤其迫切需要懂得把 AI 技術應用到產業的人才。為此，身負專業人才培育重任的大專院校，紛紛增設相關系所與學程，並致力推動 AI 創新研究，期盼為學生們打造畢業後的未來職涯。

向來在教學上重視實務應用的國立臺北科技大學（以下簡稱「北科大」），不僅急欲強化校內 AI 教學與研究能量，亦順應「教育部優化技職院校實作環境計畫」、成立「北區技專校院 AI School」，負責協助十餘所友校學生，培養參與未來 AI 產業的核心能力，所以必須備妥可觀的 AI 平台運算資源，滿足校內校外使用需求。

北科大計算機與網路中心主任王永鐘表示，歷經長時間籌劃，計網中心展開兩階段建置計畫，首先透過大同世界科技（以下簡稱「大世科」）與 Nvidia 規劃引進五台 NVIDIA DGX-1 GPU 伺服器（內含 Tesla V100 顯示卡），其次進行更重要任務，即是慎選一套優質的 AI 運算資源協作管理平台，否則若無法做到即時且公平的分配，整個計畫註定失敗。後來經過嚴謹的概念驗證（POC）程序，於 2020 年 5~6 月間確定選用數位無限軟體提供的 AI-Stack 平台。



啟用 AI-Stack 至今，為北科大帶來兩大實質成效。首先針對教學輔助部份，已促使電資、管理和機電等至少三個學院，正式開設 AI 課程，達到教育部的要求；同時也滿足 AI School 校外使用者的資源運用需求，確保他們進入北科大內網後，能按照預設路徑存取特定伺服器、特定 GPU 卡及特定儲存區，接著啟用特定軟體，絕無侵犯校園資安隱私之虞。

— 臺北科大計算機與網路中心主任 王永鐘

考量平台客製化需求 嚴格檢驗廠商技術能力

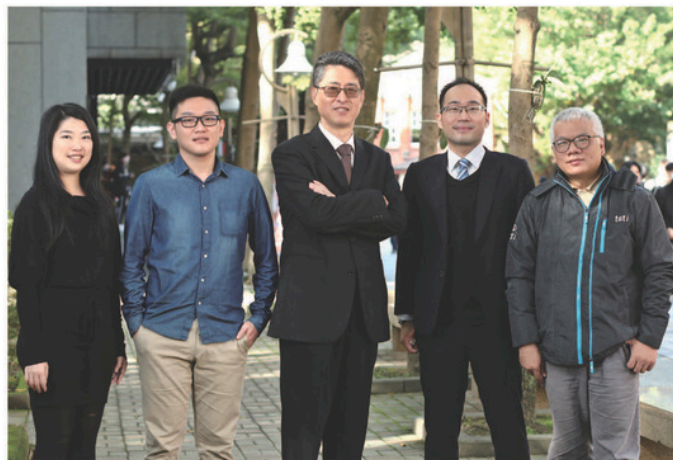
計網中心基於資安考量，一直都有將資訊系統向上集中的想法，但礙於學術單位講求獨立性，遲遲未實現目標，如今正好藉由 AI 運算平台尋求突破。王永鐘主任解釋，此平台對於老師蘊含許多關鍵誘因，一來訴求初期免費申請使用，且使用體驗如同自建環境般順暢；二來幫助老師卸下伺服器維運的繁重負擔，也不再擔心因 GPU 伺服器運作吵雜而影響學習；再者計網中心為平台賦予「主要管理者」（計網中心）、「計畫管理者」（老師）、「使用者」（學生）三階層角色，強調計畫管理者握有資源管理權限，提倡計網中心以專業的機房環境將伺服器等基礎設施統一納管，降低系所維護負擔，並鼓勵老師將計畫及研究量能挹注於此，藉此爭取到這些資源的優先使用權。

上述構思，必須倚靠好的管理平台才能付諸實現。在遴選管理平台軟體時，計網中心以植基於 Kubernetes (K8s) 架構之管理平台為首選，因考量到每次開啟虛擬機所耗資源過大，反觀容器無此問題、且更易於確保運算獨立性，而 K8s 又被喻為最佳的容器調度管理平台，故將此列入第一道篩選要件。

北科大釋出需求後，吸引多家不同廠牌軟體角逐專案，各家提出的管理功能都頗具看頭。但王永鐘主任認為，AI 運算資源協作管理是新穎概念，而各校使用情境不甚一致，很難出現 100% 通用的產品，仍需搭配相當程度的客製化，於是亦將「廠商的技術能力」列為關鍵，以 POC 結果見真章。

而北科大設計的 POC 情境頗為獨特，除了計網中心根據基礎性管理需求，提出諸如帳密整合（單一簽入）、自定義鏡像管理、使用環境的友善性（讓不熟悉 K8s 或 Linux 的使用者也能輕易操作），及如何確保 AI School 校外使用者安全存取運算和儲存資源等深具技術含量的考題外，還邀請全校師生參與測試，從實際使用者的角度反映諸多需求。

面對接踵而來的考題，北科大計算機網路中心技術人員周忻細心彙整計網中心的提問及老師的意見反饋，回頭與原廠數位無限軟體商討因應與調整對策；因數位無限軟體的技術底蘊甚深，屢屢認真回應每一個考題，也正面看待此次專案機會、視北科大為難得的策略夥伴（感念北科大提供場域與 Knowhow，促進 AI-Stack 成長），種種表現深得計網中心的認同，因此當 POC 告一段落



— 臺北科大計算機與網路中心 AI 建置專案團隊

後，便迅速決定與數位無限軟體 AI-Stack 團隊進行合作。

型塑友善使用環境 方便師生聚焦 AI 程式開發

數位無限軟體的用心回應，讓北科大給出的嚴苛考題，逐一出現圓滿解答。比方說設計「自定義鏡像管理」功能，讓使用者即使啟用預設選項外的特殊套件或函式庫，也能將整個環境儲存起來，便於下次直接回復此環境狀態，無需再費時重建。又或者將複雜 Linux 指令轉化為網頁介面上的 1~2 個簡單按鈕，快速進入多數使用者熟悉的 Jupyter Notebook 環境，便於使用者跳過艱澀的平台整合難題，專注執行 AI 程式開發。

此外身為專案夥伴的大世科，從一開始的需求訪談，到後續的規劃、建置及維護的 End-to-end 作業歷程，均展現高度的專案管理能力、溝通協調能力，使 AI-Stack 如期如質在 2020 年 8 月中旬上線。

王永鐘主任透露，考量平台的長遠執行性，北科大校長也很關心專案進展，聽聞能夠達到資源分配的公平性，且確保每一位 AI 教學或研究的需求者，皆有足夠資源可供使用，才放下心中大石。

啟用 AI-Stack 至今，已為北科大帶來兩大實質成效。首先針對教學輔助部份，已促使全校六個學院，全面開設 AI 課程；同時也能滿足 AI School 校外使用者的資源運用需求，確保他們進入北科大內網後，能按照預設路徑存取特定伺服器、特定 GPU 卡及特定儲存區，接著啟用特定軟體，絕無侵犯校園資安隱私之虞，能做到這一步，在國內實屬首創。

其次談到研究部份，平台建置完成後，校方鼓勵老師積極投入研究計畫，利用北科大的資源發展新應用；因而促成北科大推動 AI 醫療及發展智慧製造應用，成果堪稱豐碩。