

「Taipei-1」算力第4次徵案 說明會

產業技術司 專案計畫辦公室

114年7月10日



說明會議程

	時間	主持人
14:00-14:20	Taipei-1算力申請說明 20分鐘	經濟部產業技術司 專案計畫窗口 趙祖佑組長
14:20 - 15:30	Q&A 70分鐘	請以雲端提問單詢問

第四次徵案

雲端提問單 網址

<https://forms.gle/5k8PRpPYqKEo4y7C7>



- 一. 計畫說明**
- 二. 徵案說明**
- 三. 申請資格**
- 四. 申請主題與審查重點**
- 五. 算力資源與使用說明**
- 六. 整體作業時間**
- 七. 權利與義務**
- 八. 資料下載與諮詢**



- 經濟部推動「領航企業研發深耕計畫」(簡稱大A+計畫)，吸引國際級大廠輝達公司來台設立AI研發中心，與總部同步研發人工智慧核心技術，並於**112年底**完成建置超級電腦**Taipei-1**
- 經濟部積極與輝達協商，促成該公司從**113年7月**起**免費分**享**25%**算力給我國產官學研與新創申請使用，為期**至116年2月止**

Taipei-1 介紹

- Taipei-1是一座NVIDIA在台灣投資建置的超級電腦，具備支持人工智慧與數位孿生運算需求的運算資源環境，將投入提供25%算力支持產官學研及新創研發使用所需運算資源的基礎設備
- 人工智慧運算環境提供NVIDIA各式AI開發環境與套件，例如支持大語言模型與生成式AI開發環境NVIDIA NeMo Framework，或例如支持醫療與生命科學的各項AI開發環境NVIDIA Clara
- 數位孿生運算環境提供NVIDIA 運算需求，建構Omniverse Farm環境執行大量Rendering運算需求



資料來源: NV Taipei算力申請(CFP)說明會

■ NVIDIA DGX SuperPOD

- 16台DGX H100，整體共計128個H100 GPU
- 每台DGX H100主機內具備8個NVIDIA H100 SXM GPU透過最新一代的NVLink以及NVSwitch技術進行整合，任2顆H100彼此之間具備900GB/s的雙向溝通頻寬
- DGXH100主機與主機之間透過8條InfiniBand技術搭配NVIDIA Quantum-2 InfiniBand Switch 提供400Gb/s高頻寬與低延遲高速網路溝通頻寬
- 主要運用在AI運算，生成式AI運算與NVIDIA最新的前瞻技術開發

■ NVIDIA OVX SuperPOD

- 8台OVX 2.0 主機，整體共計64個L40 GPU
- 每台主機主機內具備8個NVIDIA L40 PCIe GPU
- 透過NVIDIA高速網路架構進行整合
- 主要運用在Digital Twins數位孿生與NVIDIA最新的前瞻技術開發

資料來源: NV Taipei算力申請(CFP)說明會

二、徵案說明

■ 徵案期間

- 自公告日起至**114年8月1日(五)下午5時整**截止收件，採**線上申請**
(建議申請者提早3-5天送件，專案計畫辦公室將協助檢查是否有缺漏等事宜，以盡早通知補件上傳)



■ 算力使用

- 預計推薦**6隊**(DGX 4隊、OVX 2隊)
- DGX共分四批：每批涵蓋DGX1隊，每隊依序**使用6周算力**(備註)
- OVX共分二批：每批涵蓋OVX1隊，每隊依序**使用12周算力**(備註)

備註：第四次徵案最後一批次因跨2/16-2/20春節一周假期，因此DGX第四批、OVX第二批算力使用展延一周，至3/13(五)為止。



三、申請資格

- 提供我國產官學研及新創進行研發使用；可單獨一家公司/法人單位，或多家公司/法人單位聯合申請
- 聯合申請需由一家公司/法人單位主導，並擔任專案主持人。聯合申請之主導單位必須具法令地位法人(**legal entity**)，且應具備大算力使用經驗，負責規劃與協調團隊成員的算力資源配置與使用，並綜整聯合提案之結案報告，請各團隊成員予以配合
- 基於NVIDIA算力管理規定，申請單位必須完整使用**16台DGX或8台OVX**，且一次須完整使用**6周或12周**(備註)，為充分運用算力資源，如算力需求未滿**16台DGX或8台OVX**之申請單位，請組隊進行聯合申請

備註：第四次徵案最後一批次因跨2/16-2/20春節一周假期，因此DGX第四批、OVX第二批算力使用展延一周，至3/13(五)為止。

四、申請主題與審查重點

■ 專案主題

- **DGX**專案主題包括但不限於以下範疇：生成式AI相關應用。如大型語言模型開發、物理資訊神經網路、量子計算模擬、藥物發現、醫療保健
- **OVX**(僅限下列主題)：3D動畫與影像渲染、合成資料生成、**Isaac-Sim** 機器人模擬、強化學習

■ 採**批次收件**、**批次審查**，由經濟部產業技術司籌組專家委員會，針對提案申請書進行審議，分就**DGX**與**OVX**進行審查及核定。審查重點如下：

- 專案主題具**前瞻性**與**商業化潛力**
- 提案單位須**具備大算力執行經驗及訓練資料**
- **算力需求合理性**，且能充分運用**DGX 16台**或**OVX 8台**算力
- 預期**產出與效益**

五、算力資源與使用說明 (1/2)

- NVIDIA主要提供兩個運算群: (1) 16台DGX H100 ; (2) 8台OVX L40
- DGX分四批，每批使用**6周算力**(備註)(1周上線、4周運算、1周下線與清理)
- OVX分二批，每批使用**12周算力**(備註)(1周上線、10周運算、1周下線與清理)

DGX		OVX	
第一批 (6周)	1隊	第一批 (12周)	1隊
第二批 (6周)	1隊	第二批 (12+1周)	1隊
第三批 (6周)	1隊	-	-
第四批 (6+1周)	1隊	-	-

備註：

- 每台DGX 配備8個 H100 GPU，主要運用在 AI 運算
- 每台OVX 搭配8個 L40 GPU，主要運用在 Digital Twins 數位孿生
- 每一隊可為單獨一家公司/法人單位申請，或多家公司/法人單位聯合申請
- 第四次徵案算力期間因跨2/16-2/20春節一周假期，DGX及OVX最後一批算力使用各增一周

五、算力資源與使用說明 (2/2)

- 算力分配注意事項：聯合組隊者，由**主導單位**進行規劃與協調算力配置，於**算力啟用前一周**，將**分配方式及帳號提供NVIDIA**進行事前作業，**每個使用單位以2個帳號為原則**
- 資料上傳：以**網路傳輸**方式上傳至平台。然資料量過大，可向**NVIDIA**提出與討論，以**實體硬碟**方式提供
- 技術支援：**NVIDIA**會提供聯繫窗口，安排技術人員協助解決平台使用的問題。服務時間為**每周一至五的上午9時至下午6時**

六、整體作業時程 (1/2)

- 114年6月27日：公告於A+企業創新研發淬鍊計畫網站
<https://service.moea.gov.tw/EE514/tw/aiip/50-6597.html>



- 114年8月1日下午5時整：線上申請截止收件(截止後系統關閉，不予進行新增、更新或補件)
- 9月上中旬：進行審議並擇優推薦，獲推薦單位將公告於A+企業創新研發淬鍊計畫網站

- 各批次之專案執行：依審查結果，**按排序優先者依序選擇使用批次(擇一)**。考量重要假期與系統升級維護等，預估**各批次使用時間如下**，主辦單位保留彈性調整之權利，屆時以本司算力幕僚窗口一周前之通知為準

DGX共分四批

- 第1批次 (114年09月22日至114年10月31日)
- 第2批次 (114年11月03日至114年12月12日)
- 第3批次 (114年12月15日至115年01月23日)
- 第4批次 (115年01月26日至115年03月13日)

OVX共分二批

- 第1批次 (114年09月22日至114年12月12日)
- 第2批次 (114年12月15日至115年03月13日)

- 專案結束：執行結束後**兩周內**提交結案報告，並配合進行成果揭露(輔以影片或展示)；需依據報告格式撰寫。

- 本算力為「**非商業用途**」：本算力以測試、評估和開發為目的，包括模型開發或解決方案的研究，以支持未來產品或服務的研發，相關開發之模型或研究成果，為使用單位所擁有；**禁止**將「Taipei-1」運算資源用於產生營收(例如雲端服務、挖礦等)。
- 使用單位不應在該環境中處理和輸入任何專有或敏感資訊。
- 每個算力使用單位**須與NVIDIA簽署保密協議(NDA)、使用協議(User Agreement)**，相關文件為制式合約無法調整。獲通過之使用單位，**必須**於**on-boarding前**完成**NDA及User Agreement文件**簽署，始得啟用算力，個別使用單位於使用Taipei-1時，仍應遵守簽署之使用協議，不得爭執相關規定未載於公告內。如因未完成簽署而造成算力使用時間短少，由使用單位自行承擔。
- **申請書所提交之各項資料**，請務必據實填寫，如**Token數及記憶體等**，以利進行審查及算力配置。
- Taipei-1為NVIDIA總部營運管理，若接獲總部任何系統維運或軟體升級等作業，將提前2周通知算力使用單位。因各批次排程皆已事先排定，為顧及下批次使用單位權益，應予同意進行排程調整。

七、權利義務 (2/2)

- 本計畫為政府資源，將在不涉及使用單位之研發或營業機密情況下，配合政府要求適當進行成果揭露。
- 使用單位自行進行成果揭露，如提及「係使用Taipei-1 H100 或L40 GPU算力所產生的研發成果」，需提供中英文稿並經NVIDIA總部審閱且獲得同意。
- 依據行政院及所屬機關(構)使用生成式AI參考指引，若有使用生成式AI執行業務或提供服務，應向該業務或服務之有關對象進行揭露，使其知曉。
- 研究團隊於發表論文或研究報告時，請增述(可於 Acknowledgements處)運用到Taipei-1、H100等高階GPU算力。
- 本公告如有未盡事宜，除依法律相關規定外，主辦單位保留修改及補充(包括任何異動、更新、修改)之權利，並以A+企業創新研發淬鍊計畫網站公告為依據。

- 申請須知相關資料，可由本部技術司「**A+企業創新研發淬鍊計畫**」網站(<https://aiip.tdp.org.tw/>)下載取得



- 諮詢窗口：

- 工研院產科國際所 黃小姐
athena.huang@itri.org.tw / 電話：03-5917065
- 工研院產科國際所 鄧小姐
yochingteng@itri.org.tw / 電話：03-5912872

- 算力申請網址：<https://forms.gle/3eLFfvXnbHmRB7yH7>



附件

「Taipei-1」第4次徵案申請-Q&A (1/15)

Q1、徵案時程與使用批次

■ 請問徵案及使用的時程？

- 本案自**113年7月至116年2月止**，免費分享**25%算力**給我國產官學研與新創申請使用
- 第4次徵案之算力使用期間為**114年09月22日至115年03月13日止**，共計有**6隊**使用算力(包括**DGX 4隊**、**OVX 2隊**)
- 考量重要假期與系統升級維護等，預估各批次使用時間如下(主辦單位保有調整之權利)

➤ DGX每批使用時間

- 第1批次 (114年09月22日至114年10月31日)
- 第2批次 (114年11月03日至114年12月12日)
- 第3批次 (114年12月15日至115年01月23日)
- 第4批次 (115年01月26日至115年03月13日)

➤ OVX每批使用時間

- 第1批次 (114年09月22日至114年12月12日)
- 第2批次 (114年12月15日至115年03月13日)

「Taipei-1」第4次徵案申請-Q&A (2/15)

Q2、權利義務

■ 是否需要簽署甚麼合約？可以提前提供嗎？

- NVIDIA要求每個算力使用單位須與NVIDIA簽署保密協議(NDA)、使用協議(User Agreement)，相關文件為制式合約無法調整。獲通過之使用單位，必須於算力啟用前一周完成文件簽署，始得啟用算力，如因未完成簽署而造成算力使用時間短少，由使用單位自行承擔
- 合約文件會於「通過書面審查後」提供
- 合約文件採電子方式簽署，無紙本合約寄送

「Taipei-1」第4次徵案申請-Q&A (3/15)

Q2、權利義務(續上頁)

■ 專案執行上是否有限制？

- 使用單位(法人)僅得執行自身專案，不得為其他單位(法人)執行專案
- 每個使用單位以2個帳號為原則
- 本算力為「非商業用途」：本算力以測試、評估和開發為目的，包括模型開發或解決方案的研究，以支持未來產品或服務的研發，相關開發之模型或研究成果，為使用者所擁有；禁止將「Taipei-1」運算資源用於產生營收(例如雲端服務、挖礦等)
- 專案須為民用，不得為軍事或情報相關使用
- 依據行政院及所屬機關(構)使用生成式AI參考指引，若有使用生成式AI執行業務或提供服務，應向該業務或服務之有關對象進行揭露，使其知曉
- 研究團隊於發表論文或研究報告時，請增述(可於Acknowledgements處)運用到Taipei-1、H100等高階GPU算力

「Taipei-1」第4次徵案申請-Q&A (4/15)

Q2、權利義務(續上頁)

■ 專案結束後需要繳交甚麼成果？

- 執行結束後兩周內提交結案報告(採線上提報，格式詳徵案公告中附件4：Taipei-1結案報告書)，並配合進行成果揭露(輔以影片或展示)，讓各界了解該專案提升研發能量並增加競爭力之效益

■ 算力使用單位須配合政府要求適當進行成果揭露；於結案報告中成果揭露的程度為何？

- 不需揭露研發之機敏資料(如使用的模型架構細節)，主要說明使用Taipei-1算力的成果效益，如具體產出、使用算力前後之情形、技術擴散、重要效益與亮點案例等

■ 算力使用以能夠完整使用16台DGX或8台OVX為前提，如果沒有完整使用是否有罰則？

- 本徵案算力使用率若未達理想，並無罰則。每一批算力使用率都會紀錄；目的是追蹤本徵案效益、期望發揮算力資源，協助我國各界創新研發

「Taipei-1」第4次徵案申請-Q&A (5/15)

Q3、聯合申請

■ 聯合申請主導單位主要責任是甚麼？

- 聯合申請之主導單位必須具法令地位法人(legal entity) · 且應具備大算力使用經驗；負責規劃與協調團隊成員的算力資源配置與使用，並綜整聯合提案之結案報告，請各團隊成員予以配合
- 聯合申請之主導單位需於算力啟用前一周，將算力分配方式及帳號提供NVIDIA進行事前作業

■ 聯合申請填寫申請書時，因為組隊成員會有不同專案有算力需求，在專案重點說明中該如何呈現？

- 聯合申請之「主導單位」負責協調團隊成員的算力資源配置與使用
- 個別算力使用單位(各公司/法人單位)在專案申請書中各自詳述專案內容

「Taipei-1」第4次徵案申請-Q&A (6/15)

Q3、聯合申請(續上頁)

■ 聯合申請時，有限制公司/法人單位數量嗎？可以產、學界一同組隊嗎？聯合申請的帳號會如何分配？

- 聯合申請時，並無限制公司/法人單位數量
- 組隊聯合申請成員可同時包含：公司/法人單位、大學院校(例如某校某系所某Lab)
- 聯合申請的每一家公司/法人單位都可以有2個帳號可供算力使用

■ 聯合申請時，是每一個公司/法人單位都要簽署合約嗎？

- 不論個別申請、或聯合申請，「每一個公司/法人單位或大學院校」於算力使用前都須完成與NVIDIA之間的合約簽署(包含user agreement以及NDA)

「Taipei-1」第4次徵案申請-Q&A (7/15)

Q4、申請書

■ 申請書可以用email或其他方式繳交嗎？

- **自公告日起至114年8月1日(五)下午5時整截止收件，採線上申請，恕不受理email或書面等其他方式繳交之申請書**

■ 在收件截止後，還可以調整申請書、補件嗎？

- 收件截止日為**114/8/1(五)下午5時整**，屆時系統將關閉不再受理申請
- 申請資料更新或補件上傳**之截止時間亦為114/8/1(五)下午5時整**，逾時恕不受理
- 若於系統上更新申請書，請務必遵循「申請主題需一致」原則。計畫辦公室將以最近日期與時間之檔案為主

■ 申請書是每一個欄位都要填嗎？

- 為使專案審查得以進行，需就申請書相關欄位據實填寫

「Taipei-1」第4次徵案申請-Q&A (8/15)

Q4、申請書(續上頁)

■ 民間財團法人也可以申請嗎？

- 本算力提供我國各界研發使用，我國廠商(公司、法人單位)、大學院校(例如某系所某Lab)、研究機構及新創團隊都可提出申請

■ 同一公司/法人單位可以同時申請DGX跟OVX嗎？

- 只要該公司/法人單位具備DGX跟OVX大算力執行經驗及訓練資料，專案主題具前瞻性與商業化潛力，即可同時申請
- 需注意「OVX申請案僅限下列主題」：3D動畫與影像渲染、合成資料生成、Isaac-Sim 機器人模擬、強化學習

■ 我公司適合單獨一家公司/法人單位，或與多家公司/法人單位聯合申請？

- 審查重點包含「算力需求合理性」及能否「完整使用16台DGX 6周或8台OVX 12周」，若未能完整使用算力，建議自行組隊聯合申請

「Taipei-1」第4次徵案申請-Q&A (9/15)

Q5、資料存取

■ 資料會被存在哪裡？怎麼上傳？

- 資料上傳到NVIDIA的Enterprise Server，並透過OpenVPN讓使用者可以操作。以OVX為例，NVIDIA Omniverse Nucleus 是 Omniverse 平台的資料庫和協作引擎，允許多位使用者同時連接多個應用程式即時協作
- 安全性與管理會透過單點登入（SSO）和SAML身份驗證，並使用SSL/TLS 進行安全的數據傳輸，確保數據安全

「Taipei-1」第4次徵案申請-Q&A (10/15)

Q6、OVX系統使用問題

- OVX運算群使用的是什麼作業系統？可以使用Windows系統的軟體嗎？
 - OVX運算群使用的是Linux作業系統，而非Windows系統。使用者需熟稔Linux作業系統，並確認所使用的軟體可以在Linux環境中執行
 - 申請前，該公司/法人單位必須確保所使用的軟體已取得合法授權，並且可在Linux環境中正常執行
- 申請使用OVX，除了Omniverse渲染等功能外，若需開發其他演算法，是否可在OVX環境下安裝？
 - OVX的環境僅支持渲染需求
 - 與AI相關的開發與驗證必須要是DGX H100的環境才支持，則建議申請DGX環境的使用

「Taipei-1」第4次徵案申請-Q&A(11/15)

Q7、所需 NVIDIA 軟體與開發套件

■ 申請 DGX 運算群時，需要提供哪些 NVIDIA 軟體與開發套件的資訊？

- 申請 DGX 運算群時，需要說明在使用 DGX H100 過程中所需的 NGC (NVIDIA GPU Cloud) 軟體內容
- 參考網址：<https://ngc.nvidia.com/>

■ 申請 OVX 運算群時，需要提供哪些 Omniverse Enterprise Applications 的資訊？

- 申請 OVX 運算群時，需要說明在使用 OVX 過程中所需的 NVIDIA Omniverse Enterprise Applications 軟體內容
- 參考網址：<https://docs.omniverse.nvidia.com/>

「Taipei-1」第4次徵案申請-Q&A (12/15)

Q8、帳號登入與算力平台操作

■ 透過帳號登入，是可以直接連線到主機還是由會透過算力虛擬化平台進行操作？在那個平台上操作的時候有沒有限制？像是container開的數量或是單一個container的規格限制、網路限制？

- 使用者會透過帳號登入 Taipei-1 專屬登入節點(Linux OS 主機)，而非直接連線到 DGX 主機。登入後需透過 Slurm 排程系統搭配Container提交作業來使用 DGX H100 計算節點。所有GPU主機為計算節點，Taipei-1整體環境皆為實體主機環境，沒有使用虛擬化平台，而是以 裸機方式 提供資源。使用者可在作業中使用 conda 環境 或 啟動Container容器 來建立自己的開發環境。NVIDIA已經建立許多搭配Taipei-1最佳化的Container，效能，資訊安全與穩定性都是最佳的開發環境，內容詳閱<https://ngc.nvidia.com/>，會建議申請Taipei-1使用前應該到此網站觀看，選擇最適切的開發環境搭配Taipei-1使用
- 限制說明：
 - 1) 單一container的資源使用會受到作業提交時的Slurm配額限制（如所設定的CPU/GPU數）
 - 2) 網路部分目前支援對外internet連線，允許透過internet下載模型與資料集。不適切啟動大量 container，需考量叢集資源與排程公平性，因為資源分配是以「每個 GPU 為單位」進行，排程會以該單位進行資源配給與限制

「Taipei-1」第4次徵案申請-Q&A (13/15)

Q9、DGX主機開發環境與LLM支援

■ DGX主機提供的開發環境為何？例如為了運行LLM，能否支援最新版本的Ollama，要去哪裡找技術資料呢？

- Taipei-1的DGX H100 主機皆是搭配Container進行AI專案開發，如前說明的<https://ngc.nvidia.com/>上的各式Container皆是最適切的開發環境，使用者應該優先使用NGC的Container開發環境
- 搭配NGC Container時使用者可依自身需求，透過conda或容器工具(如enroot)安裝所需的框架與工具，例如Ollama、PyTorch、TensorFlow 等
- 若需使用Ollama 的最新版本，請參考其官方文件進行安裝說明，也可在自行建立的容器環境中安裝特定版本

「Taipei-1」第4次徵案申請-Q&A (14/15)

Q10、平台資安規範與資源擴充機制

■ 平台使用有否資安規範，例如：1.開放對外連線嗎？2.如果DGX上面提供的套件不夠使用，會如何處理？3.如果本身地端儲存資料來源的空間不夠，是否有雲端空間可提供聯外下載訓練資料來源？

- 資安規範：**使用者在 Taipei-1 平台上的一切行為，須符合 User Agreement 之規定，包含不得上傳非法、侵權或機敏資料，亦不得用於營運或商業用途等等。**Taipei-1 平台目前允許對外連線，使用者可自由下載所需的容器映像檔、模型與資料。平台預設不開放主動對外服務（例如直接開啟 SSH 伺服器或 Web 伺服器），但使用者可依需求透過 SSH tunnel 等方式建立受控的遠端連線通道。在進行遠端連線或下載操作時，仍須遵守以下基本資安原則：

- ✓ 不得在平台上開啟未受控的對外服務（如 HTTP、SSH Server 等）
- ✓ 不得下載或上傳非法或未經授權的資料
- ✓ 憑證與 SSH 金鑰須妥善保管，避免洩露
- ✓ 若操作過程涉及機敏資料，請務必加密儲存與傳輸
- ✓ 嚴禁攜帶或上傳任何非法取得之資料，Taipei-1 僅供合法且合乎道德規範之用途使用

「Taipei-1」第4次徵案申請-Q&A (15/15)

Q10、平台資安規範與資源擴充機制(續上頁)

- 軟體套件安裝：建議儘可能使用NGC Container，若平台預設環境中的套件不適用於當前開發環境，使用者可透過 conda 或容器工具（如 enroot）自由安裝所需工具與框架（例如 HuggingFace CLI、AWS CLI 等）。需注意使用者不具備 root 或 sudo 權限，因此建議使用 non-root conda 安裝方式或容器環境來管理開發套件
- 儲存空間：每個梯次的用戶共享 256 TB 高速儲存空間，可作為訓練資料與模型儲存之用，請儘可能下載資料到Taipei-1高速運算用儲存系統，如此才可以有最佳效能進行運算。可以支持從個人雲端空間（如 Google Drive、Amazon S3、Dropbox 等）下載資料，平台亦支援使用 CLI 工具（例如 rclone, aws cli, gdown , aws s3等）將資料下載至平台供訓練使用

第四次徵案 相關文件

【公告】Taipei-1算力申請_第四次徵案申請須知與附件，內容包含：

公告徵案內容及四個附件

附件1：Taipei-1系統說明

附件2：NVIDIA 「Taipei-1 DGX」提案申請書

附件3：NVIDIA 「Taipei-1 OVX」提案申請書

附件4：NVIDIA 「Taipei-1」結案報告

內容及網址	QR code
第四次徵案申請書上傳網址 https://forms.gle/3eLFfvXnbHmRB7yH7	
第四次徵案結案上傳網址 https://forms.gle/aLmggHnxXwzxzCWY9	
第四次徵案說明會雲端提問單 https://forms.gle/5k8PRpPYqKEo4y7C7	

謝 謝！