



國家實驗研究院

國家高速網路與計算中心

National Center for High-performance Computing

115 年人工智慧高速計算主機建置案

公開徵求資料 (RFI) 說明會



簡報大綱

1. 購案概要
2. 購案目標
3. 細部需求
4. 徵求意見項目

購案概要(1/2)

- 本中心115年度預估投入總金額**新台幣24.05億**用以建置本案。
- 本案依國研院科學技術研究發展採購作業要點，以**公告招標評審標**方式辦理，採**不定底價、固定金額給付**，由**評審委員會**訂定或審定招標文件之評審項目、評審標準以及評定方式進行評審，評定最優廠商決標。
- 本案允許廠商共同投標（不超過三家為限）或單獨投標。
- 本案不允許外國廠商、大陸地區廠商、第三地區含陸資成分廠商及在臺陸資廠商投標。
- 本案建置地點
 國網中心**台南分部「國網雲端算力中心」**機房

購案概要(2/2)

- 本案履約時程
 - ◆ 115年3月上旬，RFP公告與招標。
 - ◆ 115年5月中旬，決標。
 - ◆ 115年10月中旬，完工報驗。

- 付款方式（暫定）
 - ◆ 預付款（30%）
 - ◆ 硬體交付（50%）
 - ◆ 驗收（20%）

購案目標

- AI HPC通用型計算平台：80 PFlops (Rpeak)

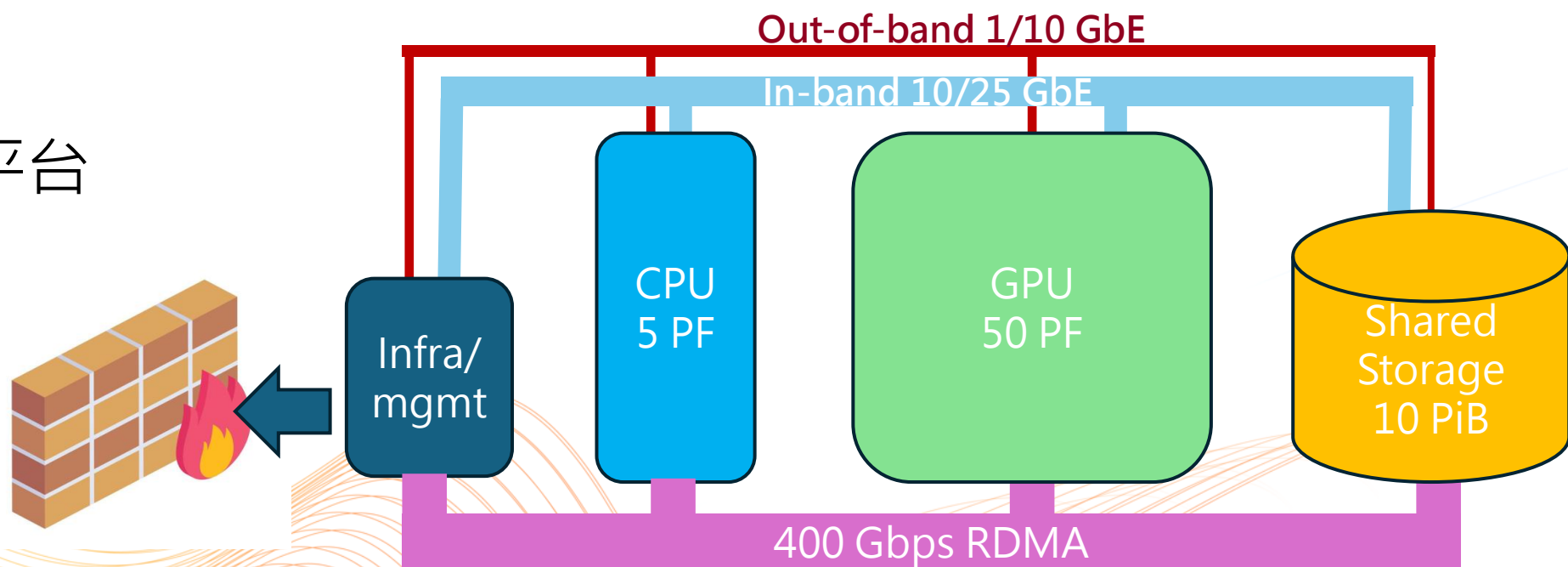
- ◆ CPU計算機群
- ◆ GPGPU計算加速機群
- ◆ 高速網路系統
- ◆ 儲存系統
- ◆ 系統軟體

- 微型測試開發平台

- 資通安全防護

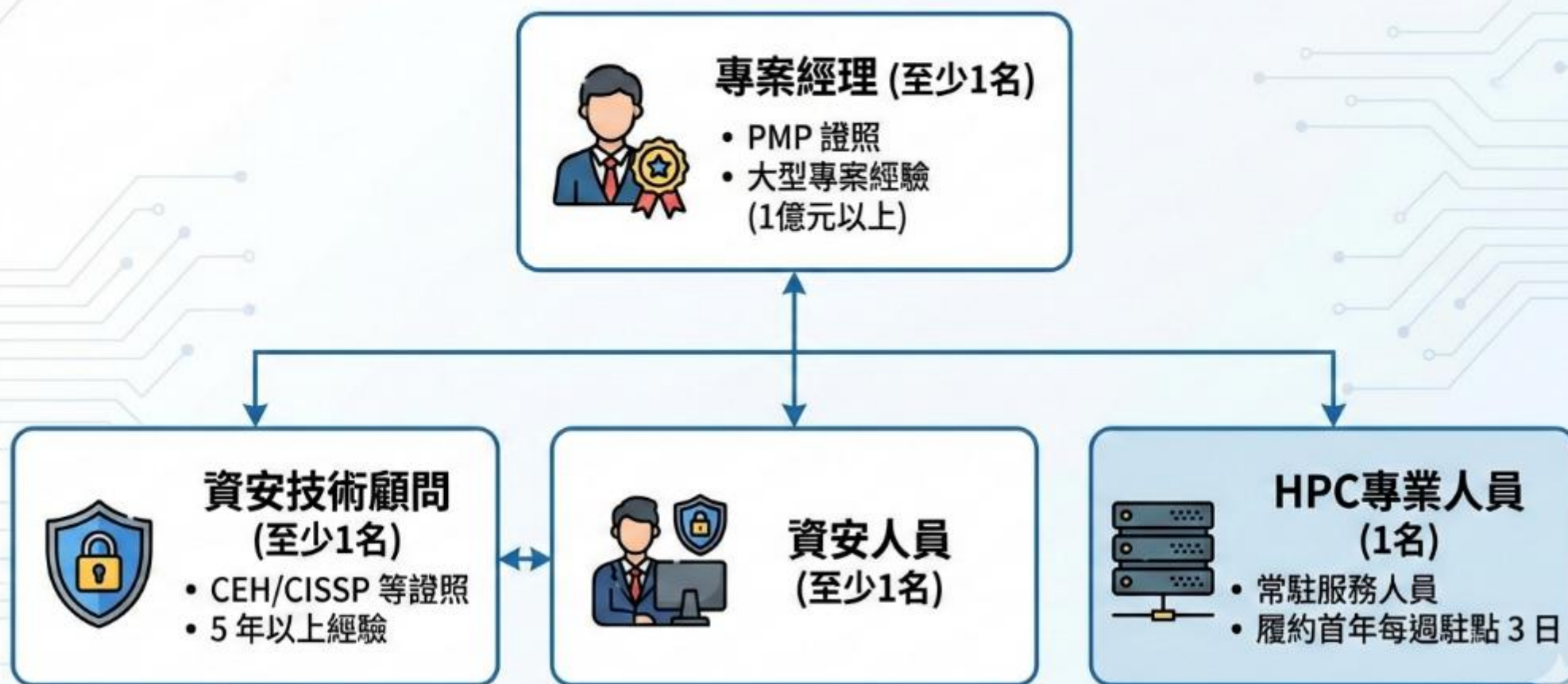
- 電力 $\leq 3.6\text{MW}$

- PUE ≤ 1.25



細部需求 (1/9)

廠商專案組織與人力規劃



細部需求 (2/9)

● AI HPC通用計算平台

◆ 包含4種類型節點

- 服務與管理節點群
- CPU 計算節點群
- GPU 計算加速節點群
- 大記憶體計算節點群

◆ 各式節點通用需求

- 所有節點須提供頻外硬體管理埠 (Out-of-Band Management)
- 電源供應器須具備雙迴路備援與熱抽換功能
- 採用直接水冷 (DLC) 設計

細部需求 (3/9)

◆ 服務與管理節點群

- 各式系統部署、資源管理、工作排程、帳號認證、登入服務
- 高可用性設計架構
- 單節點應具2CPUs、記憶體512GB以上
- 10/25 GbE 2埠連接控制網路
- CPU與GPU計算節點群登入節點各4台以上
- 傳輸節點3台以上
- 登入與傳輸節點具負載平衡分流機制
- UPS供電

◆ CPU計算節點

- FP64總計算效能 (Rmax) 5PF
- 最新世代的 x86-64 CPU 2顆，每核心記憶體頻寬 4.5GB/s 以上並配置 6GB 以上記憶體
- NVMe SSD 960GB 系統碟 1顆、3 TB以上 本機scratch空間
- 400 Gbps高速網路埠 1埠、10/25 GbE 2埠連接控制網路

細部需求 (4/9)

◆ GPU計算加速節點

- FP64總計算效能 (Rmax) 50PF
- 單節點搭載2 顆 CPU 與 8 顆以高速網路技術互連 GPU
- 單節點記憶體、磁碟、GPU 對 GPU 高速網路埠以及 GPU對儲存系統高速網路埠，依加速晶片原廠建議規格提供

◆ 大記憶體計算節點

- 4節點以上
- 每核心記憶體配置大小範圍最低 56GB，最高 86GB
- 每節點總記憶體 8TB 以上
- NVMe SSD 1.92TB 系統碟 2顆，採用RAID 1；3 TB以上 本機scratch空間
- 400 Gbps高速網路埠 1埠、10/25 GbE 2埠

細部需求 (5/9)

◆ 網路系統

- 內部高速網路
 - 支援RDMA 400Gbps 網路
 - 計算節點間Non-Blocking Bandwidth
- 控制網路 (In-band)
 - 10/25 GbE
 - HA架構設計
- 管理網路 (Out-of-band)
 - 1/10 GbE
 - 管理專用，需與一般使用者完全實體隔離
- 對外網路
 - 100 GbE
- 系統韌性評估分析
- 均須具備圖形化使用者介面監控與管理工具

細部需求 (6/9)

◆ 儲存系統

- 可用空間10PiB以上
- 讀取 600GB/s、寫入400GB/s以上
- 支援協定NVIDIA GDS、POSIX、NFS
- 可分割為三個獨立檔案系統
- 具高可用性與資料保護機制與設計
- 支援配額與檔案存取控制功能
- 支援與SLURM整合，依個別用戶、節點或計算工作查詢用量
- 提供圖形化管理介面
- 稽核、Email告警通知功能

細部需求 (7/9)

◆ 系統軟體與功能

- 合法完整授權或訂閱服務
- 作業系統應為 64 bits Linux且可支援至2032/12/31
- 登入介面需提供命令列介面 (CLI) 與圖形化介面
- 工作排程使用SLURM
- 程式開發環境、工具與函式庫
- 環境部署功能：節點組態管理、全系統規模部署、全自動化流程、軟體與環境管理
- 系統管理：遠端主控台、平行批次遠端維護管理功能
- 提供圖形化叢集監控介面，監控叢集系統健康狀態、系統效能、Log收集與警示功能。
- 資訊設備管理：提供整合所有節點頻外管理模組之 GUI 管理介面、網路設備及儲存設備須具備完整管理功能

細部需求 (8/9)

◆ 效能與情境測試

- HPC效能測試 (CPU/GPU)
 - HPL : CPU計算機群 5PF、GPU計算機群 50PF
 - HPL-MxP (FP16以上)
 - HPCG
- AI 效能測試 (GPU)
- 大氣科學 (CPU/GPU)
- LAMMPS (CPU/GPU)
- Quantum ESPRESSO (CPU/GPU)
- 量子模擬 (GPU)
- 電漿模擬情境測試 (GPU)
- 網路效能與壓力測試
- 儲存系統效能與情境測試

細部需求 (9/9)

- 微型測試開發平台
 - ◆ 實體隔離且架構與功能需與正式環境一致相容。
- 電力、冷卻與節能
 - ◆ 全系統電力不超過3.6MW、PUE 不超過 1.25
 - ◆ 所有設備皆應使用雙迴路電源
 - ◆ 計算節點須採用直接水冷設計並使用常溫水進行冷卻
 - ◆ 提供電力使用、發散熱量與冷卻系統報告
 - ◆ 與本中心既有環控系統介接
- 保固
 - ◆ 驗收合格日起，全案保固4年

徵求意見項目 (1/3)

- 系統建議架構方案

- ◆ 整體系統設計拓樸圖
- ◆ 系統規模建議
- ◆ 估價單

- 資通安全防護系統建議

- ◆ 資料中心防火牆
- ◆ 應用程式和網路效能監控系統
- ◆ EDR與MDR
- ◆ 大型語言模型（LLM）安全防禦系統
- ◆ 入侵偵測與防禦系統
- ◆ 資安日誌處理與系統關聯性分析系統
- ◆ 網路資安風險管理系統
- ◆ 資安情資服務

徵求意見項目 (2/3)

- 效能與情境測試建議
 - ◆ HPL、HPCG、HPL-MxP測試與優化方法
 - ◆ AI效能：徵求訓練與推論測試的擴散模型(Diffusion Model) 與大型語言模型(LLM)
 - ◆ 大氣科學：數值天氣預報模型、氣候模擬模式、AI/ML 模型
 - ◆ 高速網路效能：延遲 (latency) 數值、網路介面是否支援虛擬化？
 - ◆ 儲存效能測試：IOR/Mdtest 測試參數、FIO 建議測試項目與參數
 - ◆ 穩定度與壓力測試：明確且可行測試方式、計畫與標準
- 機房需求與能源效率
 - ◆ 提供 80 PFlops 算力下之預估用電量
 - ◆ 冷卻方式比例 (DLC vs 氣冷) 及 PUE 量測方式建議

徵求意見項目 (3/3)

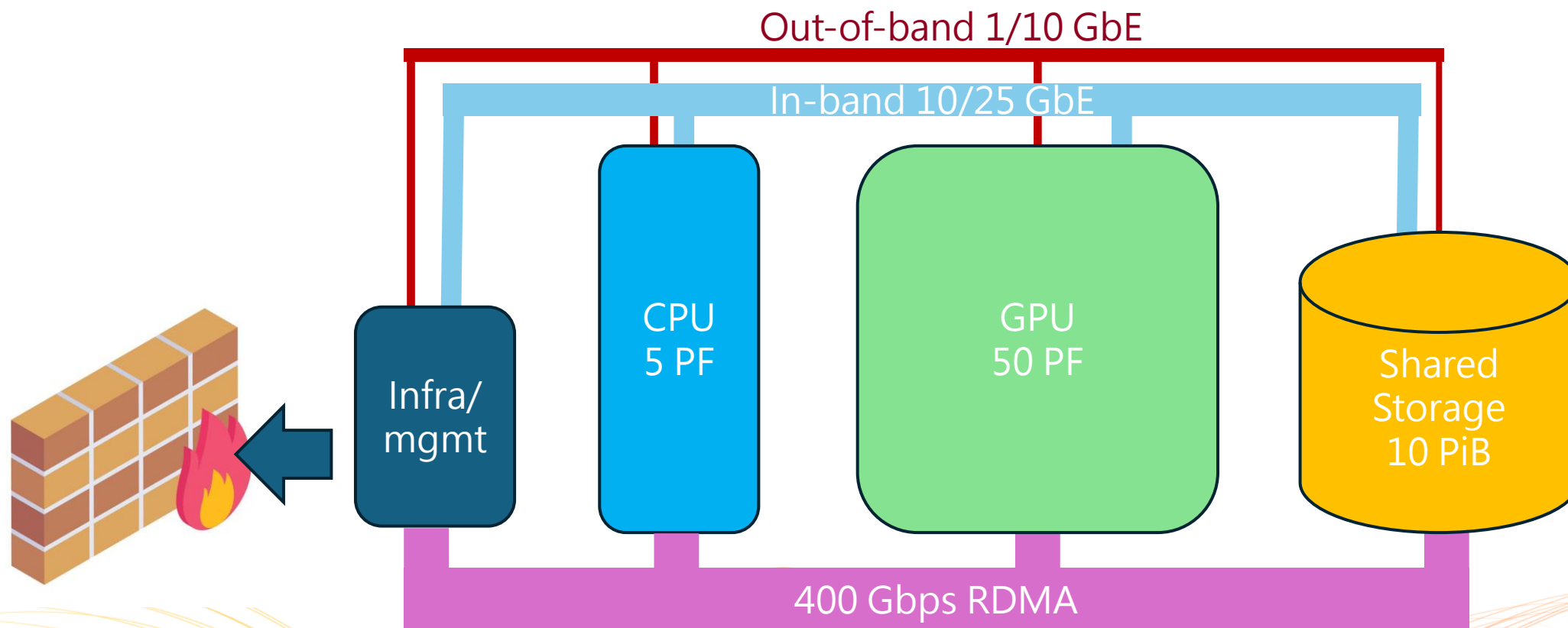
● 評審項目建議建議

- ◆ 請依AI HPC系統應具備之特點，提出適合本案之建議評審項目、內容與配分比例，並詳細說明各評審重點及配分要點與理由。
- ◆ **資安作為與企業社會責任**為必要項目，且資安作為須佔總分 10%
- ◆ 符合相關採購法規之規定，同時確保整體評審之公平性與透明度，歡迎提出更具區辨性與差異化之評審計分方式

簡報結束

敬請指教

- CPU/GPU clusters 使用相同的高速網路



- CPU/GPU clusters 使用不同的高速網路

