



# AI 轉型新紀元

CubeCOS AI 雲落地，驅動校園應用與治理實踐

Sharon Lin

Bigstack 關鍵客戶經理

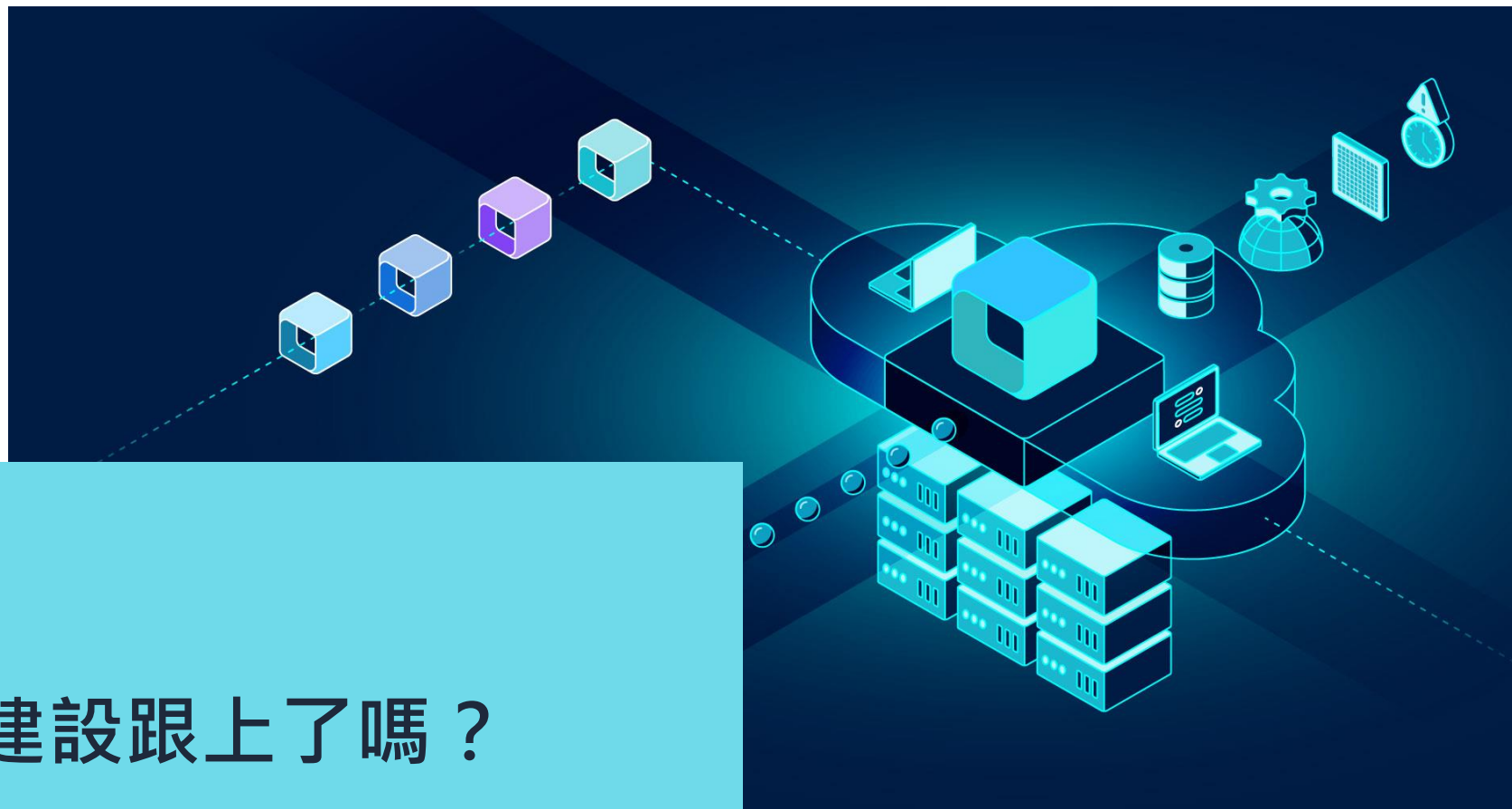


# 今天談什麼

1. 國家押注了，學校的基礎建設跟上了嗎？ 教育部與國科會的 AI 預算分配
2. AI 落地的五塊拼圖 您的學校現在缺哪幾塊
3. Bigstack 補的是哪層 基礎建設層 · Platform · 主權
4. 真實案例：從實驗室到全校：兩所大學的實踐 東海大學 · 陽明交通大學電子所
5. 我們正在做的下一步 GPU 資源管理強化
6. 學校是生態圈的節點，不只是用戶 招生 · 育才 · 產學，三件事同一個答案

# 國家押注了 學校的基礎建設跟上了嗎？

教育部與國科會的 AI 預算分配





# 政府的錢已經到位 — 接球的準備好了嗎？

160.5 億

教育部 TAICA  
人工智慧學程聯盟

跨校 AI 學程認證、打破校牆限制  
讓非資工學生也能具備 AI 競爭力

12.3 億

國科會  
AI 運算資料中心

115年度第2年經費解決大專校院「  
算力不足、硬體昂貴」痛點

新啟動

AI 超級電腦  
建置與服務計畫

Taiwan AI RAP 雲端開發平台  
供學研團隊快速導入生成式 AI

519 億

晶創台灣 ×  
Taiwan AICOE

國內頂尖大學共築國際級  
AI 學術重鎮大專教授研究計畫

國家的預算方向已經說話了。問題是：你的 IT 基礎建設，準備好接這個球了嗎？



# 招生 · 育才 · 產學 三件事都卡在同一個地方

## 招生

對應計畫：  
教育部 TAICA 學程聯盟

AI 是招生賣點，但學生選校時已開始問  
「你們的實驗室跑什麼？用什麼環境？」  
光靠課程名稱已經不夠。

## 育才

對應計畫：  
國科會 AI 運算資料中心計畫

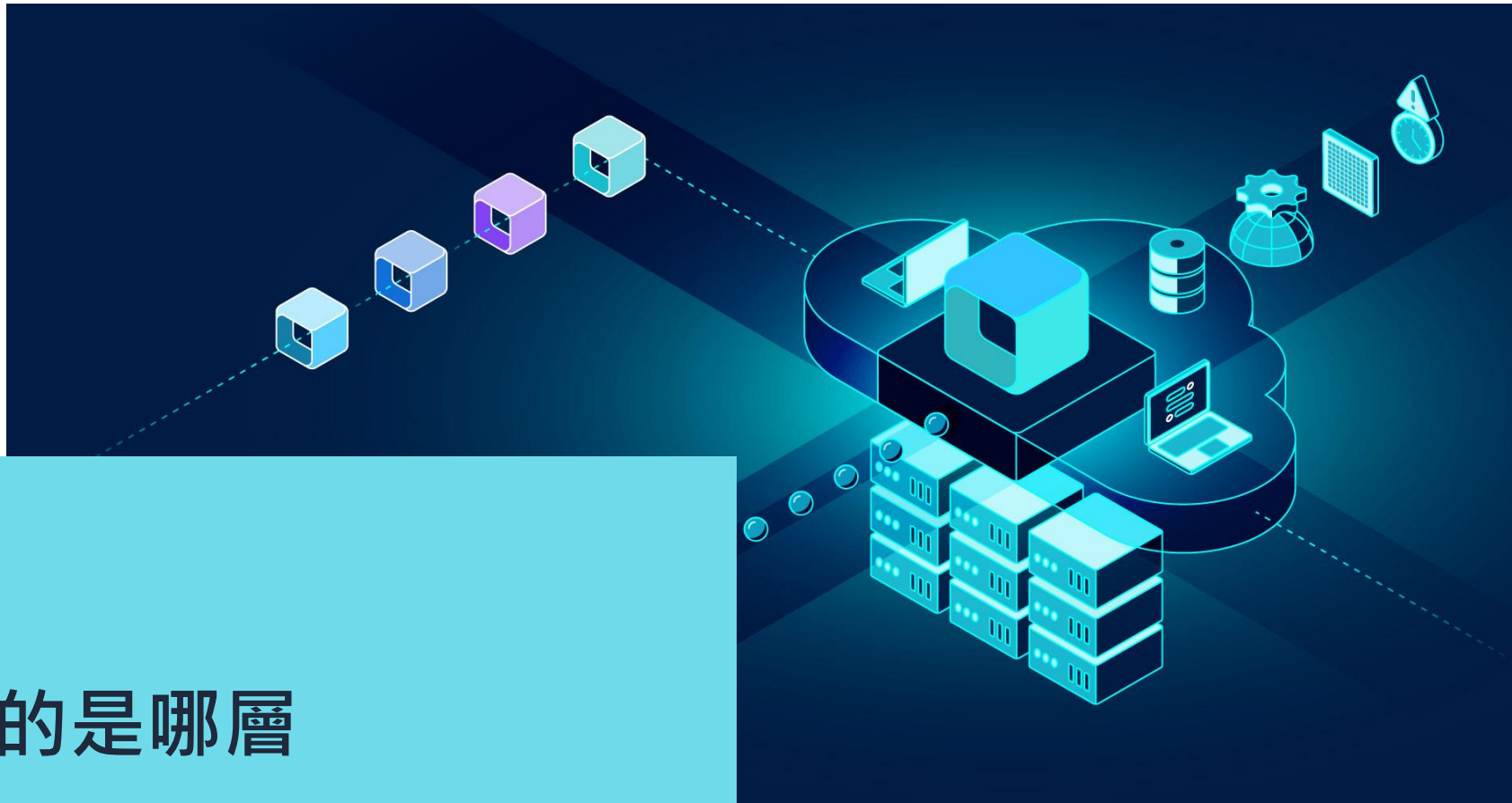
業界要的是碰過真實 Infra 的人。  
學生只用過公有雲 Console，  
進公司第一天面對私有雲就傻了。  
台灣 AI 落地大多在私有環境跑。

## 產學

對應計畫：  
晶創台灣 · Taiwan AICOE

企業願意給資料、給題目，前提是  
資料不能上公有雲。沒有符合資安要求  
的私有環境，很多產學案  
連談都沒辦法談。

三件事的缺口，指向同一個答案：學校需要一套產業等級的私有雲基礎建設

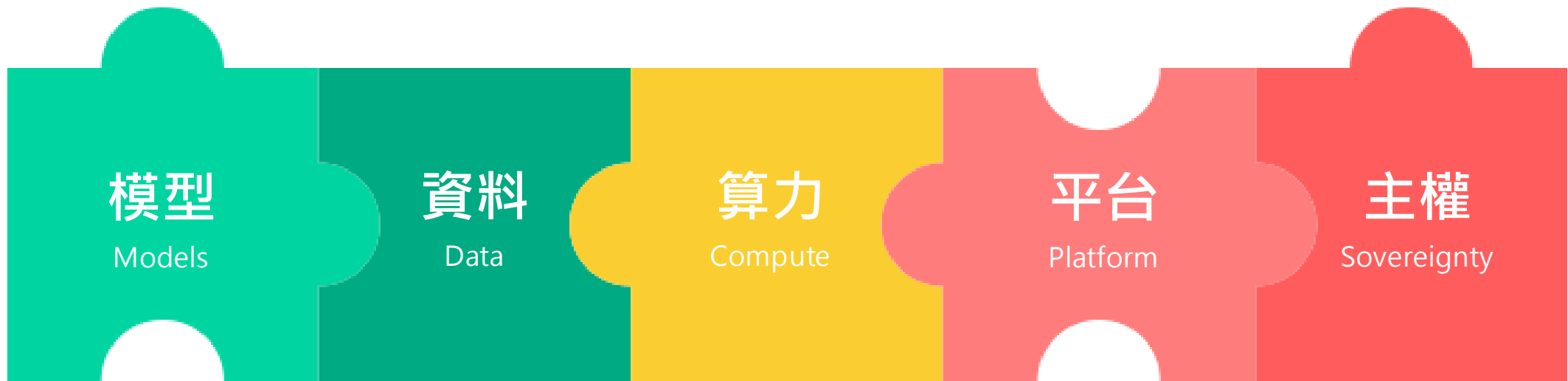


# Bigstack 補的是哪層

Platform + 主權，不是賣伺服器



# AI 落地需要五塊拼圖，你的學校有幾塊？



✅ 有能力

LLaMA · Mistral  
自訓練模型  
研究團隊有能力

✅ 有，但分散

校務資料  
研究資料  
產業合作資料

⚠️ 有設備，缺調度

GPU Cluster 買了  
但利用率 < 30%  
缺統一調度平台

❌ 大多缺

統一管理  
多租戶隔離  
自助服務入口

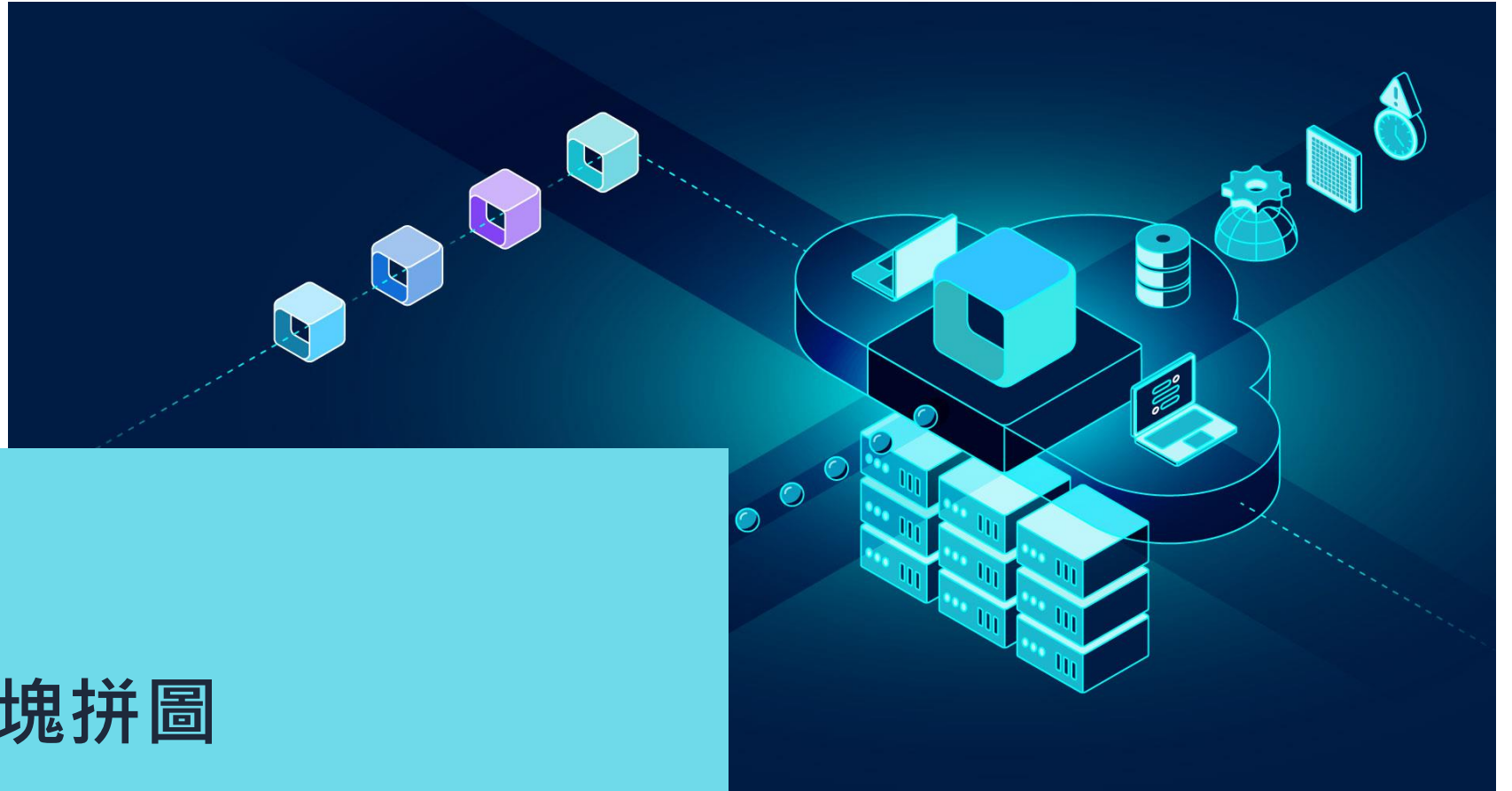
❌ 公有雲做不到

資料不出境  
合規 · 自主維運  
產學合作前提

「平台」與「主權」這兩塊，幾乎每所學校都是空格 —— 這才是今天要談的事

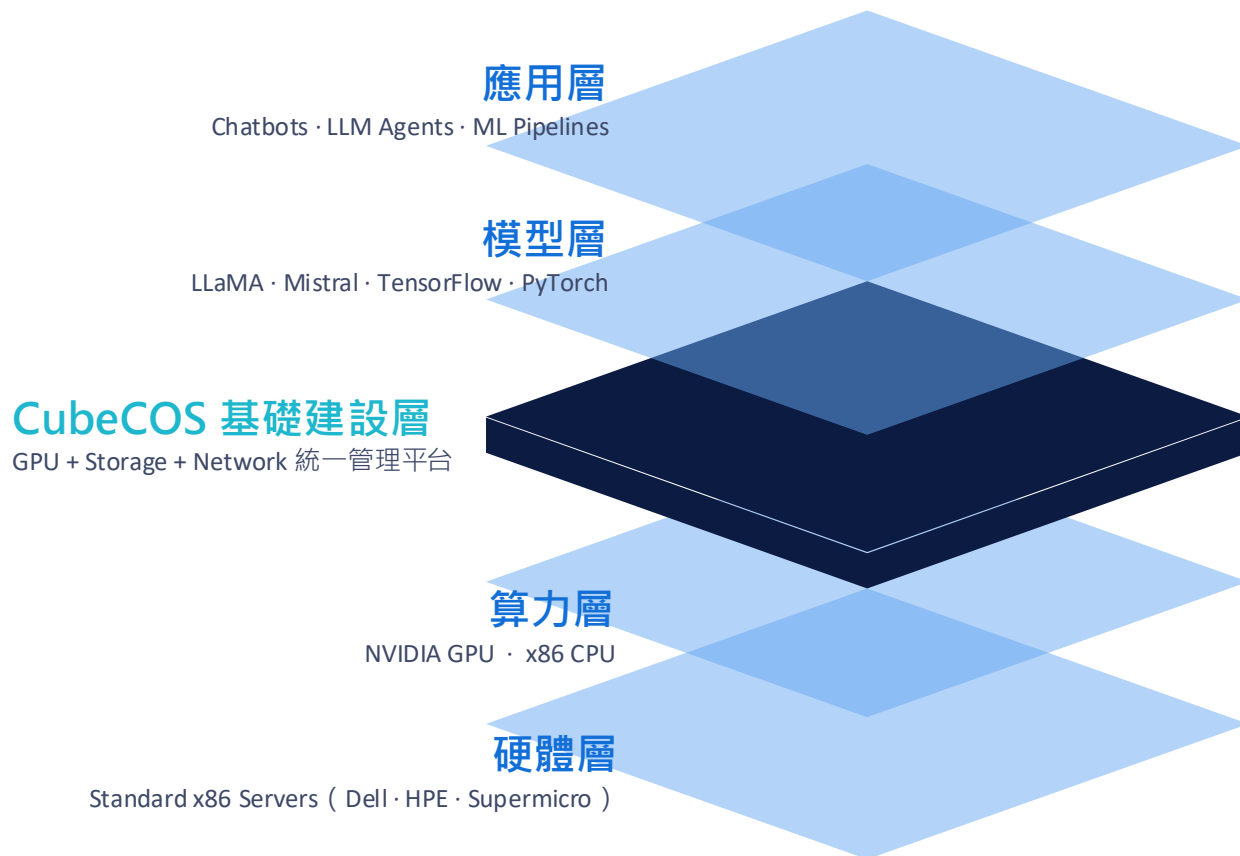
# AI 落地的五塊拼圖

您的學校現在缺哪幾塊





# CubeCOS 在 AI 工作流的位置



## cubeCOS 核心元件

- Rancher (RKE2) · Kubernetes distribution
- Ceph RBD / CephFS · CSI storage
- OVN SDN · tenant isolation
- Harbor OCI Registry · built-in
- NVIDIA vGPU Manager v580
- Pass-through / vGPU / MIG 三模式



## 學校真正的問題 → CubeCOS 的回答



GPU 買了，用不滿



Pass-through / vGPU / MIG 三種模式，訓練用整張卡，推論共享；利用率從 < 30% 提升至最大化



各系所要環境，IT 應付不來



Self-service Portal：研究生自己開 VM / Container，多租戶隔離，IT 不用逐一開票



產學資料合規 / 主權問題



私有部署，資料不出校園；AI 訓練成果留在自己手上；符合合規要求，產學案子才能談



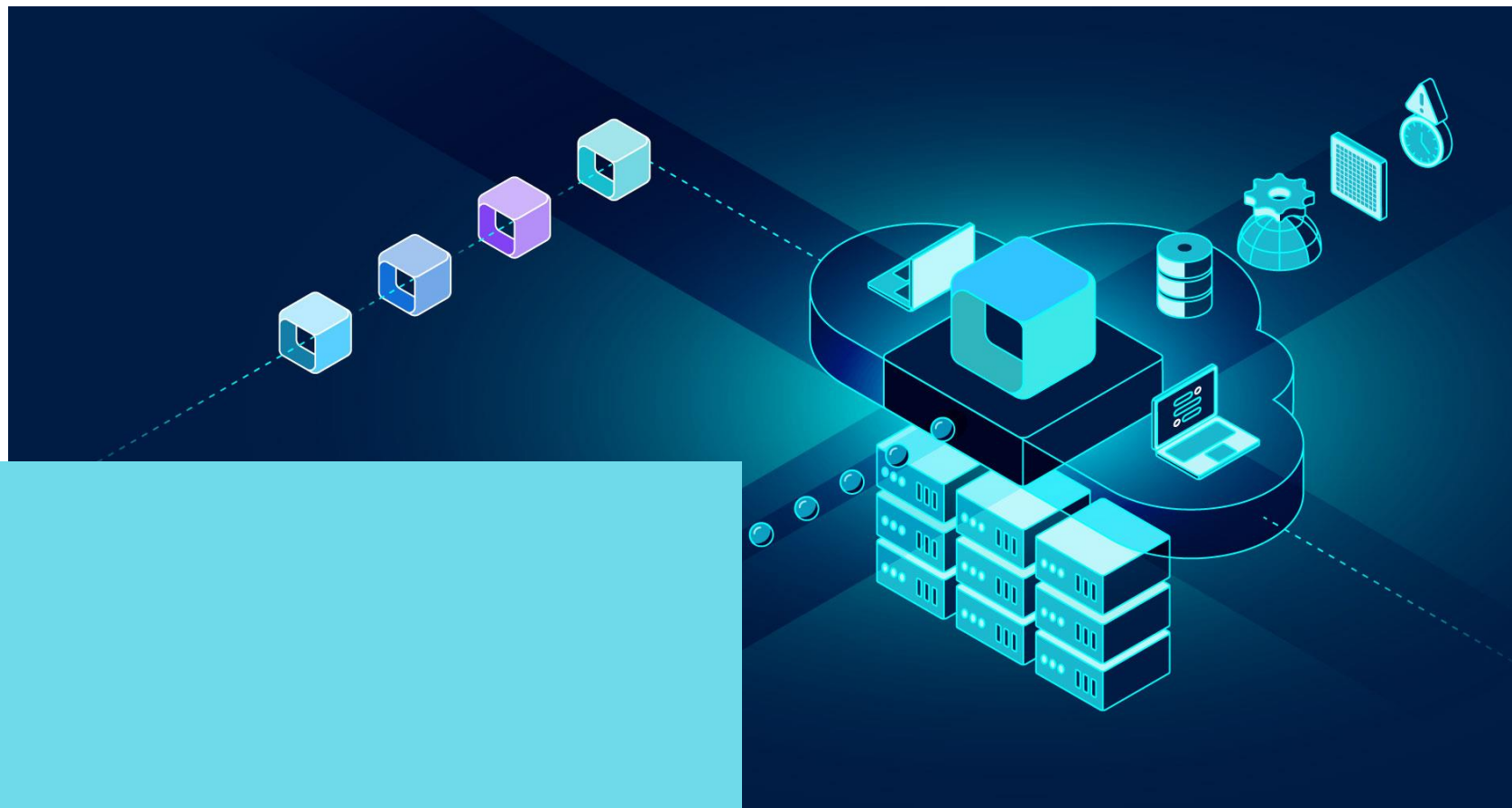
維運人力不足，系統不敢動



One-click 滾動升級 + 告警自動觸發 + IPMI 遠端管理；小 IT 團隊也能穩定維運

# 真實案例

從實驗室到全校：兩所大學的實踐





# 東海大學：從實驗室單機到一站式 AI 研究平台



## 楊朝棟 終身特聘教授

東海大學 圖資長 × 資工系主任

史丹佛全球前 Top 2% 頂尖科學家

- 部署架構：3 節點 AIO 叢集 (Control Converged)
- 7 台 VM，其中 2 台 GPU Passthrough (RTX 2080 Ti)
- YOLOv8 網球賽局辨識系統 (AI 視覺應用)
- 心臟超音波翻譯摘要 LLM 研究服務
- 實驗室官網 via Kubernetes Ingress 對外服務
- Keycloak 身份管理 · Ceph 分散式儲存 · Rancher K8s

### 導入前：傳統單機管理

- ✗ VM、容器、GPU 各自管理，無法統一
- ✗ 單點故障風險高，研究環境無法隔離
- ✗ 擴充需重建整台主機，停機風險大

### 導入後：CubeCOS 一站式平台

- ✓ GPU、VM、容器、網路統一 Dashboard 管理
- ✓ YOLOv8 / LLM 研究服務穩定上線
- ✓ Kubernetes 對外提供服務，隨需擴充



## 陽明交通大學電子所：53.8% 的學生說，系統讓學習打折扣

415 台伺服器 · 78 位受訪者

26.9% 說「Server 數量太少，常需等」

21.8% 說「License 數不夠，常需等待」



### 效能面：資源閒置 × 瓶頸共存

128 核心伺服器因 XRDP 設計限制，CPU 無法有效利用  
學生集中登入同一台主機，人多即卡頓、當機需進機房手動重啟  
過去至少發生數十次，直接影響課堂進行

### 架構面：單點故障 × 升級困難

所有學生家目錄掛載於單一儲存節點，故障即全班停擺  
維護更新必須全面停機，CentOS 7 → Rocky 8 升級風險極高  
加值服務需額外獨立機器，擴展性受限

### 管理面：帳號雜亂 × 稽核困難

多人同時使用、帳號大量、軟體繁多、版本更新  
工具散落，同學裝不好、TA 也救不完  
帳號 / 權限不一致，資安稽核困難



# 陽明交大：ADFP Cloud 3.0 — 從單機走向虛擬化平台

## ADFP 2.0 — 傳統三層式架構

- 實體機為單位，單點故障風險高
- 環境不可逆，更新 OS 難度高
- 教學與研究混用，無法有效隔離
- 適合小規模單一用途
- 擴充需增加實體機器，改變架構



## ADFP 3.0 — CubeCOS 超融合 + 虛擬化

### L1 基礎層

7+3 台伺服器 ( GIGABYTE + PHISON )

### L2 虛擬化層

CubeCOS / CubeCMP — 資源池化與高可用性

### L3 系統層

Rocky Linux 8/9 取代 CentOS 7

### L4 服務層

Docker 容器化 ( AMP · Guacamole · DB · LDAP )

### L5 應用層

EDA Tools — Module file 多版本共存



# 陽明交大：為什麼選 CubeCOS，不選 VMware

今天設計和建置這套系統的那個博士生，現在在台積電。

這就是「學生碰到產業等級環境」的最好句子。



## 授權模式更符合教學環境

VMware 以 CPU 核心數為基準、每年續約。

CubeCOS + CubeCMP 採「實體伺服器台數」買斷制。對硬體規模固定、使用年限長的教學環境而言，大幅降低長期營運成本與預算不確定性。



## 開源整合度更友善

建立在成熟開源技術之上，與既有 Linux、NFS、LDAP、XRDP、Web 服務自然整合。

避免導入高度封閉的專有架構，對校內 IT 團隊與後續接手人員更友善。



## 原廠支援配合度高

Bigstack 提供專屬支援窗口，透過固定會議與即時通訊群組協助建置、維運與除錯。

對教育單位而言，「一起討論架構細節」比單純產品交付更具價值。

# 我們正在做的下一步

GPU 資源管理強化





# GPU 資源管理 強化 — 年底推出 Add-on

## 現況：GPU 是學校最貴的採購

- △ GPU Cluster 平均利用率 < 30%
- △ 各研究計畫各自占用，資源無法共享
- △ 排程靠人工協調，缺乏可視化管理

目標：2026 年底正式推出  
現有 CubeCOS 叢集可升級



## 動態 GPU 資源池管理

統一資源池，按需分配；閒置 GPU 自動釋放回池，最大化硬體投資回報



## 跨 Namespace 智能排程

不同系所 / 研究計畫的 GPU 請求，根據優先權與使用率自動排程分配



## GPU 使用率即時監控與自動回收

Dashboard 可視化各 Namespace GPU 使用情況；長時間閒置自動通知並回收



# GPU 調度為什麼對學校特別重要

## 國科會補助算力，學校用得好才受益

國科會 AI 運算資料中心計畫 12.3 億元，目標是補足大專校院的算力不足。  
但補助的是硬體，不補調度效率。  
GPU 買了擺著用不滿，預算就是浪費。

## 多研究計畫同時競搶 GPU，IT 成裁判

教授 A 的 LLM 訓練、教授 B 的 IC 模擬、研究生 C 的 AI 視覺專題，  
三個計畫同時需要 GPU，IT 不可能手工協調。  
沒有調度系統，先搶先贏，資源配置完全不公平。

## TAICA 學程要求學生有真實 AI 環境

教育部 TAICA 學程的設計目標是讓學生具備實戰能力。  
「真實 AI 環境」的最低門檻，  
就是能夠按需取得 GPU 算力、跑得起大模型。

GPU 調度不是 IT 的技術問題  
是學校能不能落實  
AI 辦學承諾的關鍵

# 學校是生態圈的節點

不只是用戶





# 選對基礎建設，學校成為生態圈的節點



## 研究型大學

你的 Private AI Infra  
就是研究的競爭力

- GPU cluster 統一調度，研究資源不浪費
- 資料留在校內，產學合作案子能談
- 跑 LLaMA / Mistral，模型成果自己擁有
- 先進 IC 設計、AI 視覺——都已在跑



## 教學型大學

學生碰到產業等級的  
環境，不是試用帳號

- 學生自己開 VM / Container，實作不靠公有雲
- 畢業出去帶的是業界通用技能
- 建置系統的博士生，進了台積電
- 你在培養台灣下一代 AI 工程師



## 所有學校

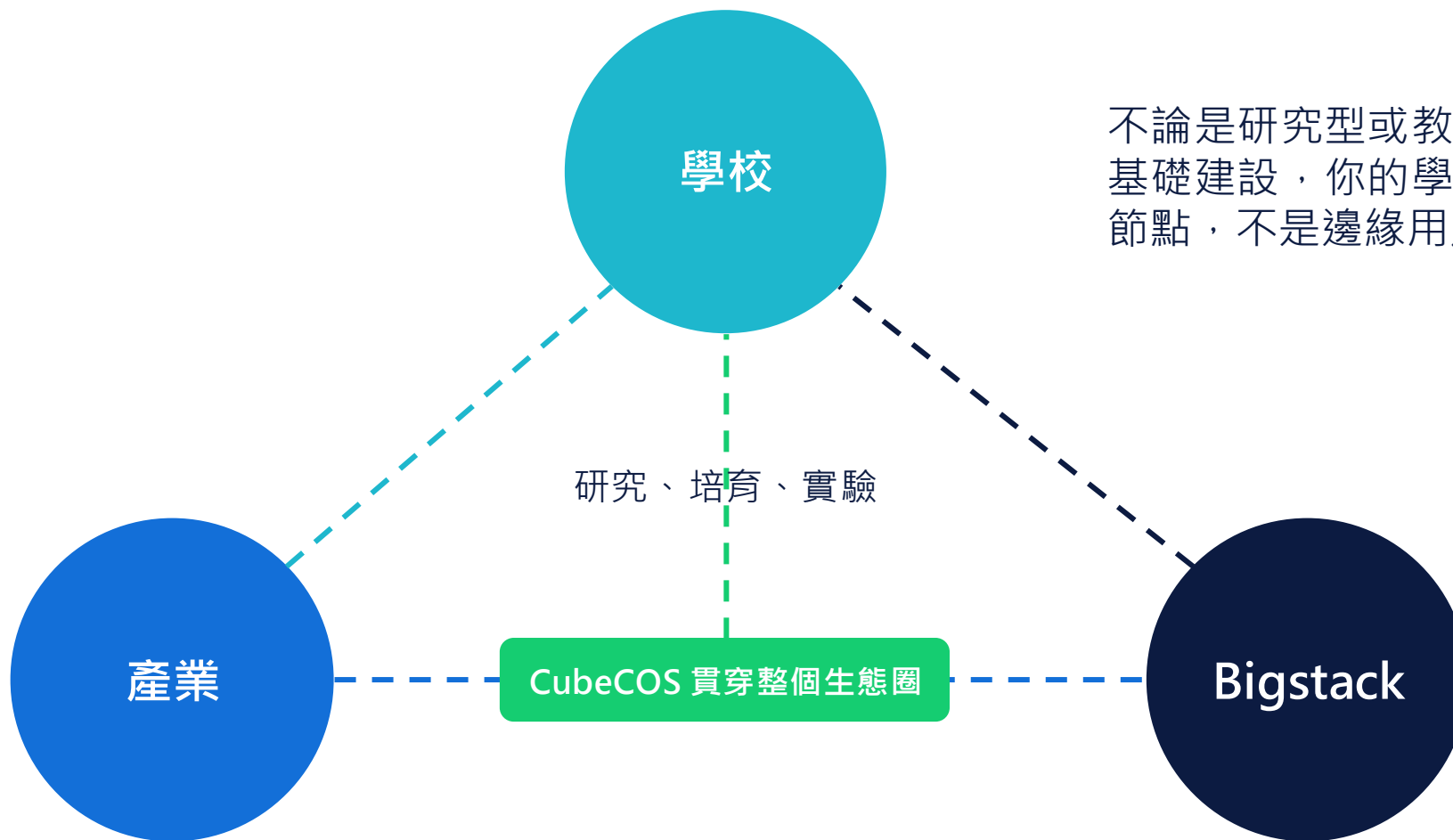
開放標準，  
你的籌碼不被鎖死

- Open source，不依賴單一廠商授權
- 授權買斷制，預算可預測，不怕漲價
- 東海、陽明交大已在跑——你的同行
- 先進場的人，定義標準

「先進場的人，定義標準。」—— Bigshare 2026



# 生態圈的三角：學校 · 產業 · Bigstack



不論是研究型或教學型大學，選對基礎建設，你的學校就是生態圈的節點，不是邊緣用戶。



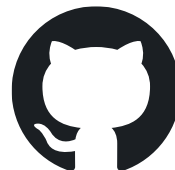
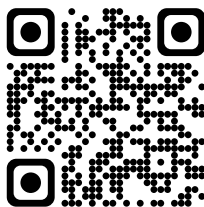
## 加入 Bigstack 社群平台



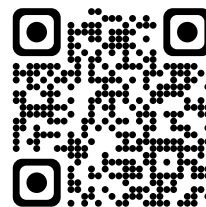
Slack



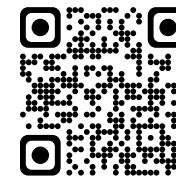
Discord



Github



 YouTube



LinkedIn





Thank you for listening.



## 你的選擇不再單一

基礎建設從來不是主角，  
但它決定了所有主角能走多遠。

招生簡章寫著 AI，辦學方向指著產學，學生值得一個跟得上這個方向的環境。

**有興趣一起盤點你的 AI 基礎建設？我們來談。**

Sharon Bigstack Account Manager  
sharon@bigstack.co