透過數據及AI助力行業轉型與加速創新

- 技術副總經理- 台灣戴爾科技集團
- 梁匯華/ Eric Leung

@ar_wa2000

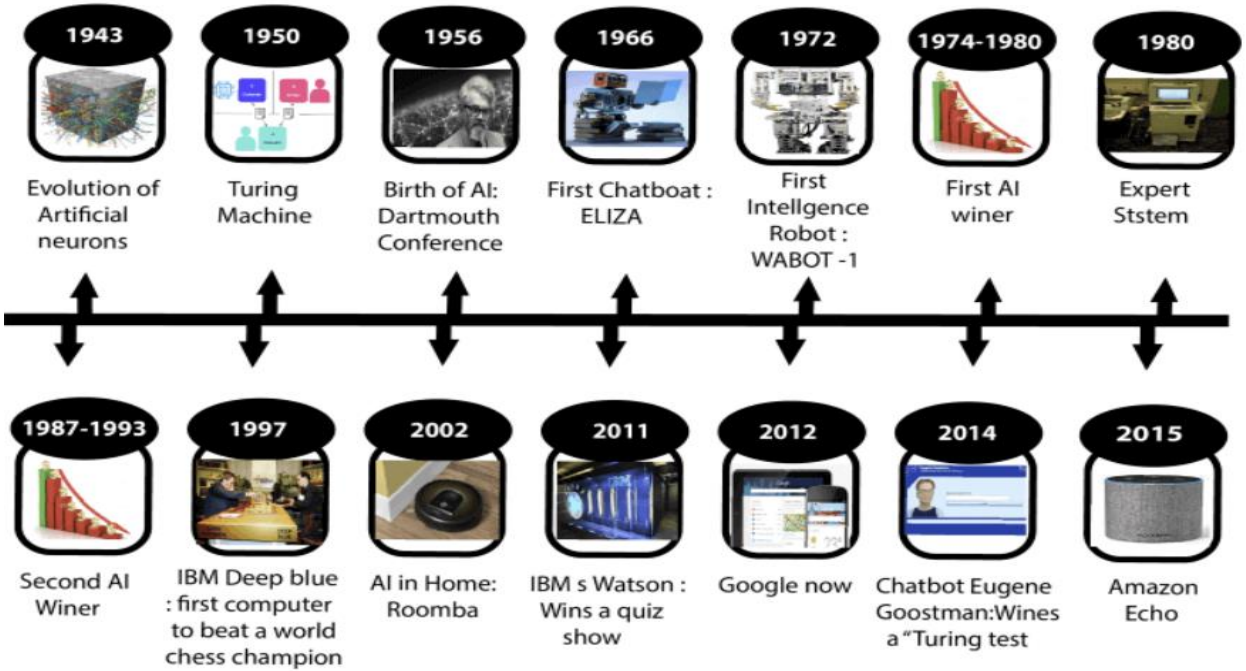
DELL Technologies

第一次的AI 對奕

Bertie the Brain pitted humans against an artificial intelligence in a game of tic-tac-toe

人工智慧史

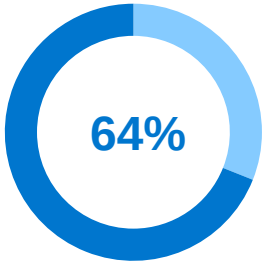
人工智慧歷史上的一些里程碑



- *人工智慧不是一個新詞，也不是研究人員的新技術。這項技術比你想像的要古老得多；甚至還有古希臘和埃及神話中機械人的神話。*

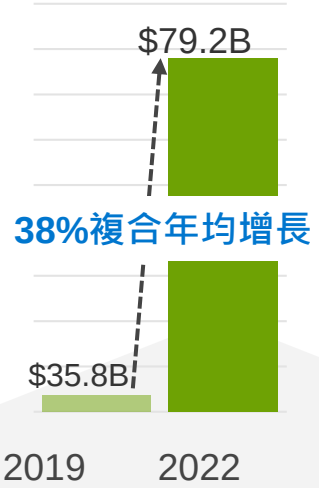
我們正處於AI 發展的臨界點

誰會參與其中?



在未來3年中，CIO將投資於機器學習領域

全球在人工智慧系統上的支出



交易規模

整體方案

- **~$265K** Ready Solutions for AI avg
- **~$200K** for a proof-of-concept configuration

AI 優化儲存設備

- **~$250K** starting point
- **~$600K** average

伺服器

- **~$5K** workstation starting point
- **~$65K** 4-GPU server
- **$+++** growth

[1] ServiceNow: The Global CIO Point of View, 2019
[2] IDC. Press Release: "Worldwide Spending on Artificial Intelligence Systems Will Grow to Nearly $35.8 Billion in 2019."

人工智慧的未來就是現在

一場完美的風暴，推動著客戶的需求

更多數據

有越來越多的數據可用於為人工智慧-每秒生成更多數據。其中大部分是數據為非結構化數據。

大大增強的計算力

多線程 GPU 現在提供電源演演算法，即時處理，便於快速識別趨勢和模式。

AI 創新

現在，我們可以訓練機器使用數據來感知、學習、推理、預測和創新之應用.

智慧保險顧問催生AI 應用

智慧保險顧問指的是奠基於人工智慧及大數據分析技術，依據顧客基本資訊提供個人化風險分析、投保建議的創新型保險服務

- 相對較公正客觀
- 全面掌握顧客需求
- 不受時空限制

數位化也促使新生代客群對保險公司的依賴程度下降。全球市場上，不只有更多的顧客傾向使用自動化工具獲取投保建議，新生代客群當中，更有將近9成在購買保險時選擇聽取朋友或家人的意見，而不是保險公司的建議

全球市場上顧客對於自動化投保建議採取開放態度(資料來源：Accenture 2017市場調查)

案例1: 產品外觀檢測

Dell筆電Latitude A cover 組裝線

[AI 機器視覺識別]

TRIVISION AI 模型，篩查十大類常見缺陷，並告知缺陷類別和位置以及可能成因,並繼續學習新缺陷，其中：

- 灰度演算法，篩查嚴重不良；
- 比良演算法，篩查未經學習的缺陷，比對未經學習過的缺陷，不良防堵以防漏失。

Data scale: **每10秒**完成一個 A cover組裝，產生**250MB**數據，1天產生**2TB**數據。資料需保留**2年**，**約1PB**數據儲存在 **Isilon**上

[具體成效]

- ✓ 該環節的良品率 yield rate 提升**10%+**
- ✓ 漏殺率 missing rate (把不良品誤判為良品) **從 0.5% 到 0.3%**；
- ✓ 過殺率 overkill rate (把良品判為不良) **從 5 % 到 3%;**
- ✓ 節省組裝線人力配置: 一條流水線設 **5** 個人工檢測站，機構件廠還需要外派人員到ODM做入料篩查和隨線篩查。

案例2: 產品組裝件檢測 - Dell 筆電組裝線

Before
手工檢測

在組裝前，62個檢測點，手工的一個個檢測。

After
CCD+AI 檢測

配上光學設備+AI演算法，自動檢測

[結果]

a. inspection item.

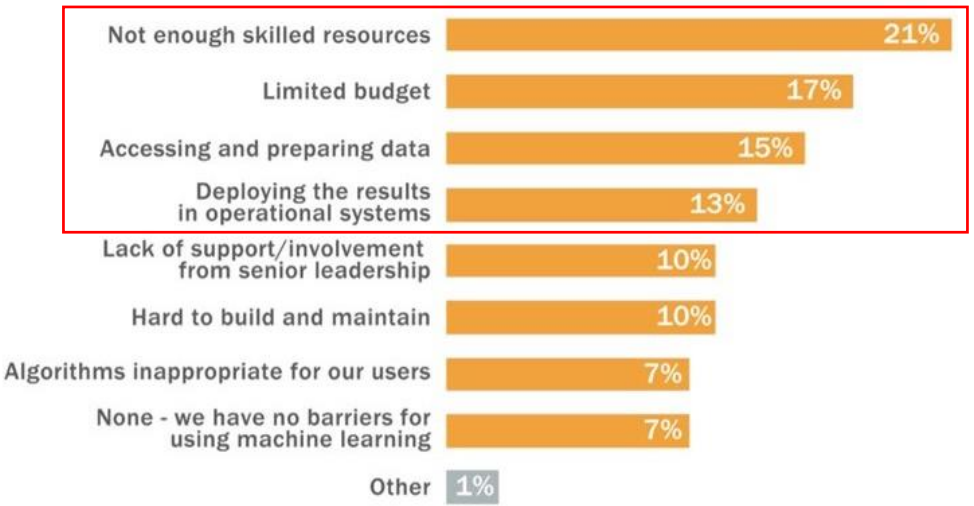
Checking items	數量
Screws	36
Cable	11
Connectors	12
Ruber	3
Summary	62

b. in line inspection
c. CT < 30sec.
d.Coverage:95%
e.overkill:2%
f. missing: 0.5%(by unit)

客戶對人工智慧寄予厚望

Source: 451 Research's AI & Machine Learning Use Cases 2020

然而，它面臨著重大挑戰



Source: 451 Research's Voice of the Enterprise: AI & Machine Learning, Adoption and Use Cases 2018

DELL Technologies

熊貓 vs 魚子醬

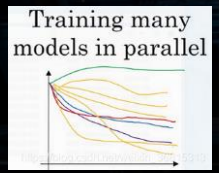
熊貓

- 在DL中，它是具有1：1比率的工作站/伺服器的個別開發人員。
- 數據科學家的主導
- 運營成本昂貴，其中包括機會成本和實際成本
- 數據大小往往被單一實體限制

魚子醬

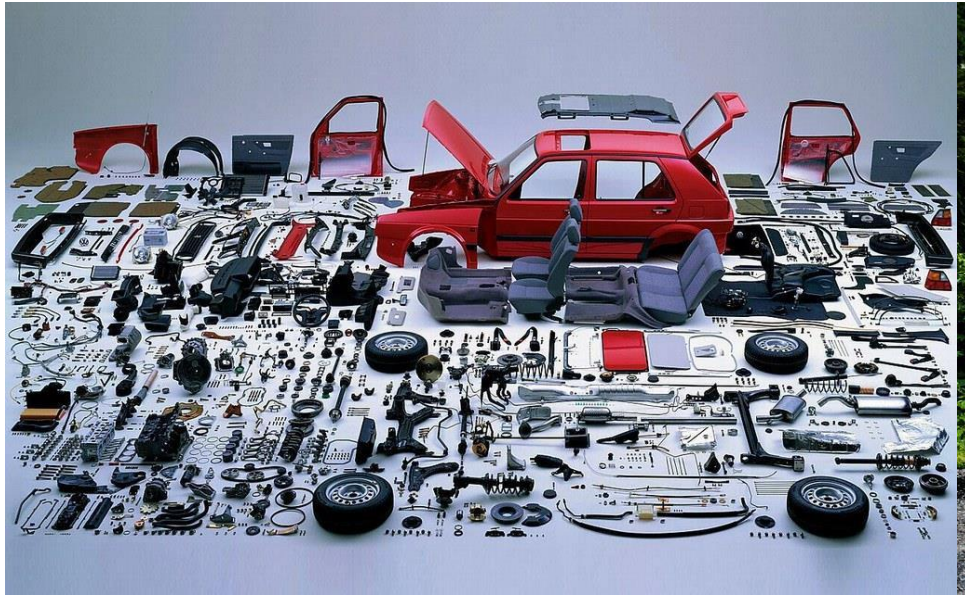
- 魚有上百萬個魚卵
- 如在深度學習中，構建“即服務”的基礎架構團隊
- 沒有個人擁有任何特定的"GPU"
- 由 IT 支援
- 同時顧及安全與共用資源，運營上更受歡迎
- 數據大小是彙集的，而不是孤立的；能處理的數據大小決定在整個集群有多大

All credits to Pandas vs Caviar to Andrew Ng



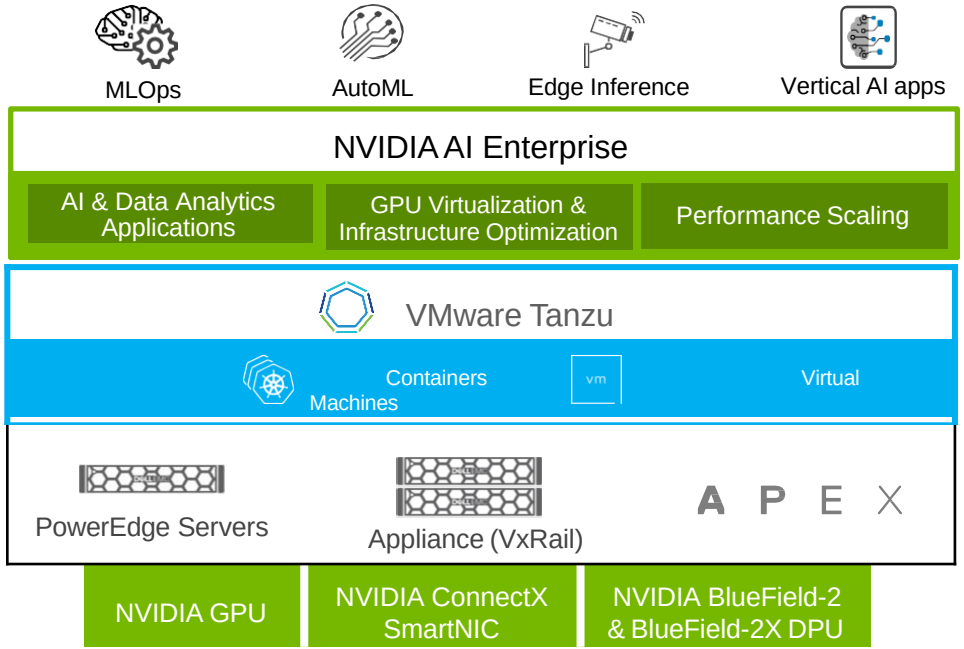
企業 AI 的簡單按鈕

今天的DIY方法



主流人工智慧

AI Enterprise - 簡化 AI 基礎架構的部署和管理



- **Full Stack Enterprise Software Product from NVIDIA**
- **Supports vSphere 7.0u3**
- **Runs on NVIDIA Certified servers from Dell Technologies**
- **Benefits of virtualization with bare-metal performance**

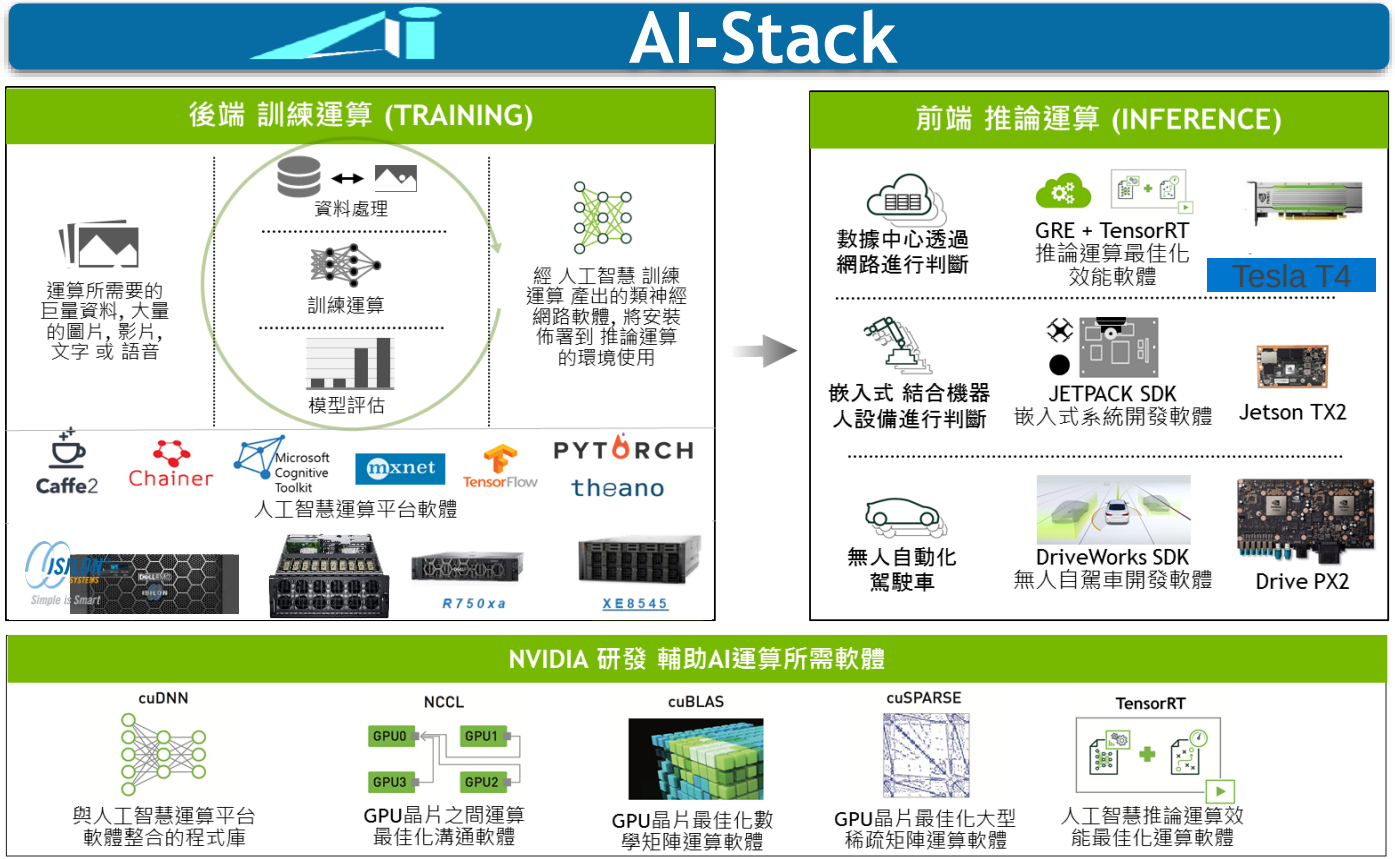
IT ADMIN - BENEFITS

- Use the same tools they use to manage other enterprise apps
- Provision "right size" VMs for data scientists
- Stability of environment through release management and full-stack support from NVIDIA

AI PRACTITIONER/DATA SCIENTIST - BENEFITS

- Industry-leading AI software tools and frameworks
- Focus on data science, rather than IT management
- Guaranteed interoperability across the software stack
- Access to pre-trained models for faster model deployment time

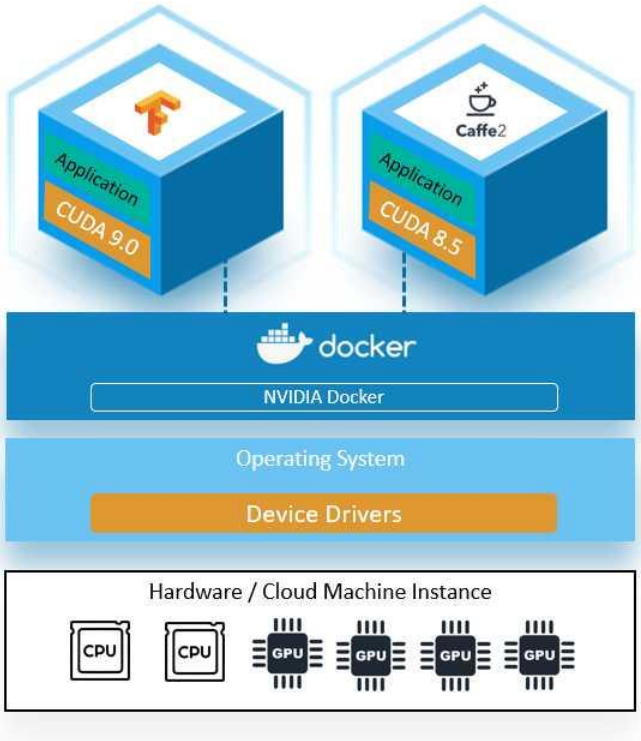
Dell + AI-Stack加速客戶讓AI落地



吳宣儀 博士

數位無限軟體股份有限公司營運長

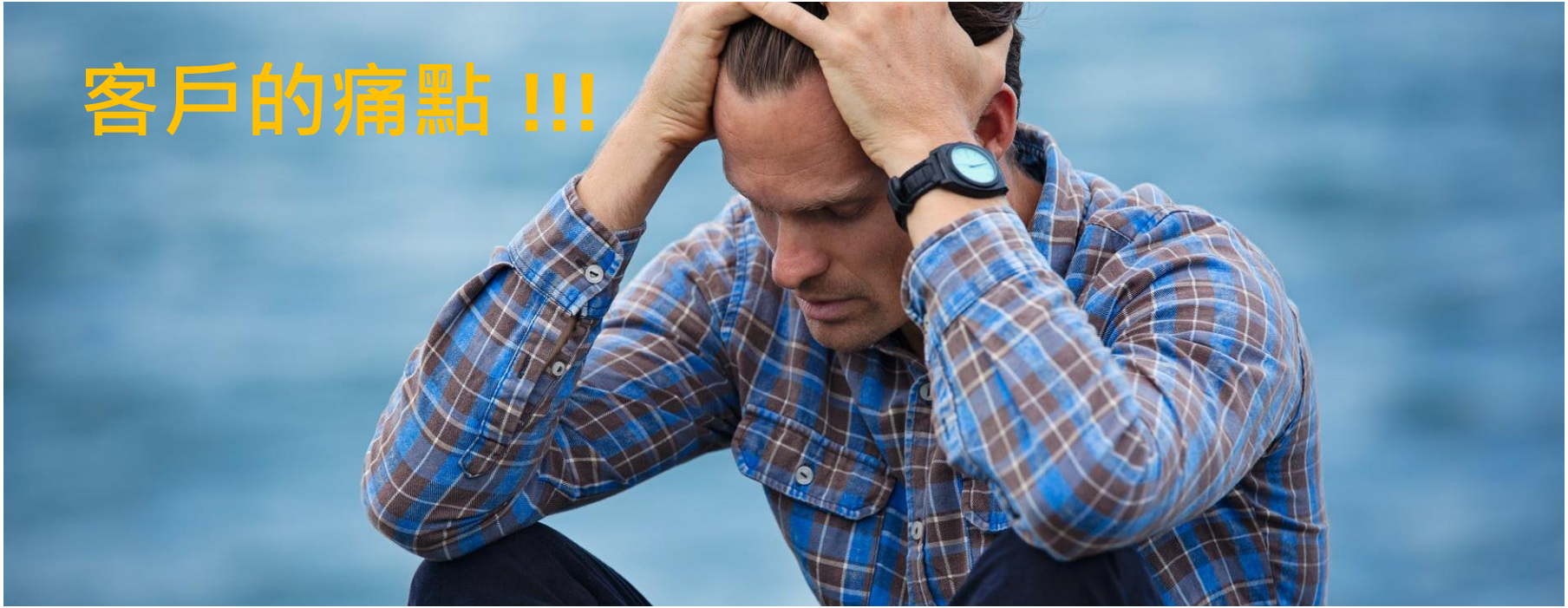
AI基礎設施的趨勢：GPU + 容器



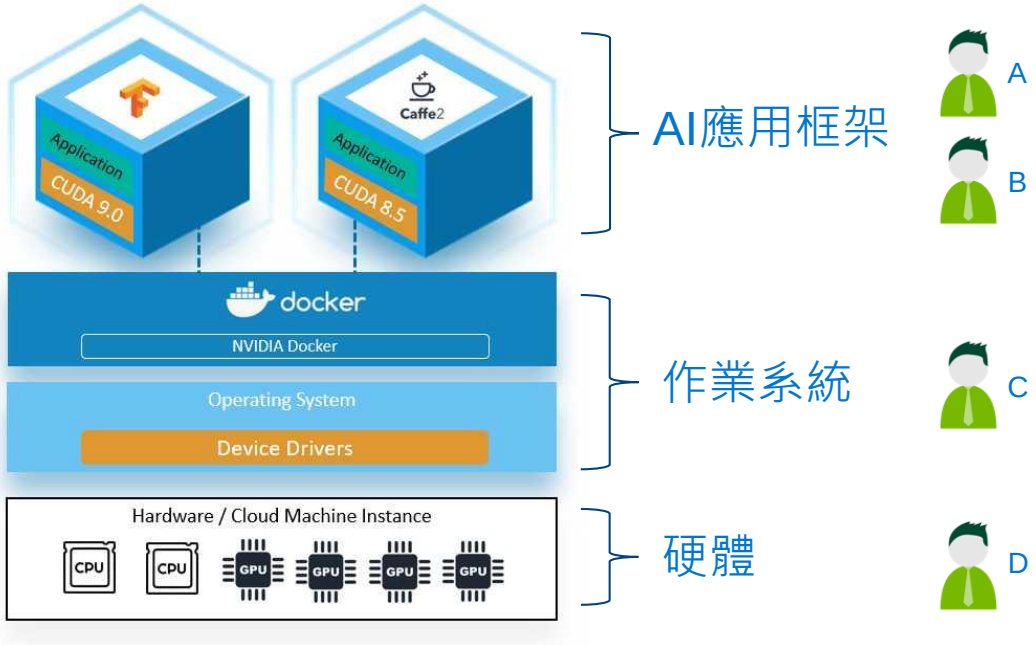
AI應用框架

作業系統

硬體



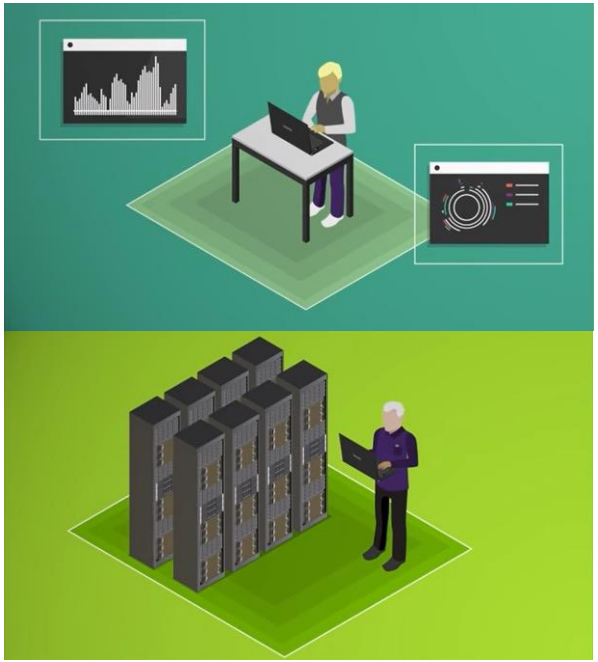
AI 研發人員自用自管

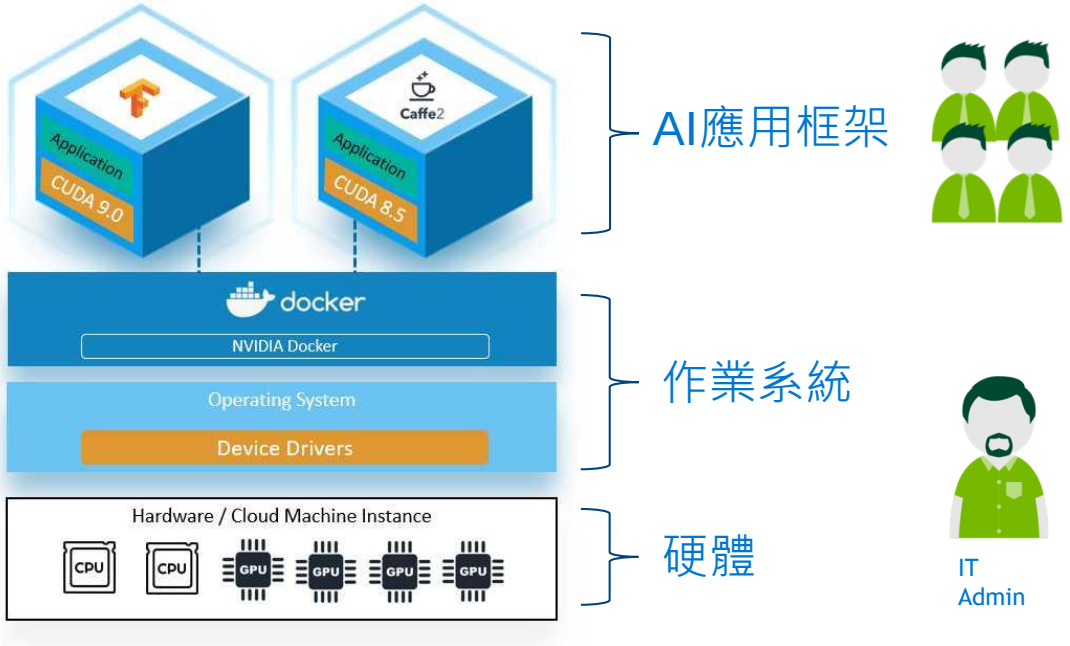


常見問題

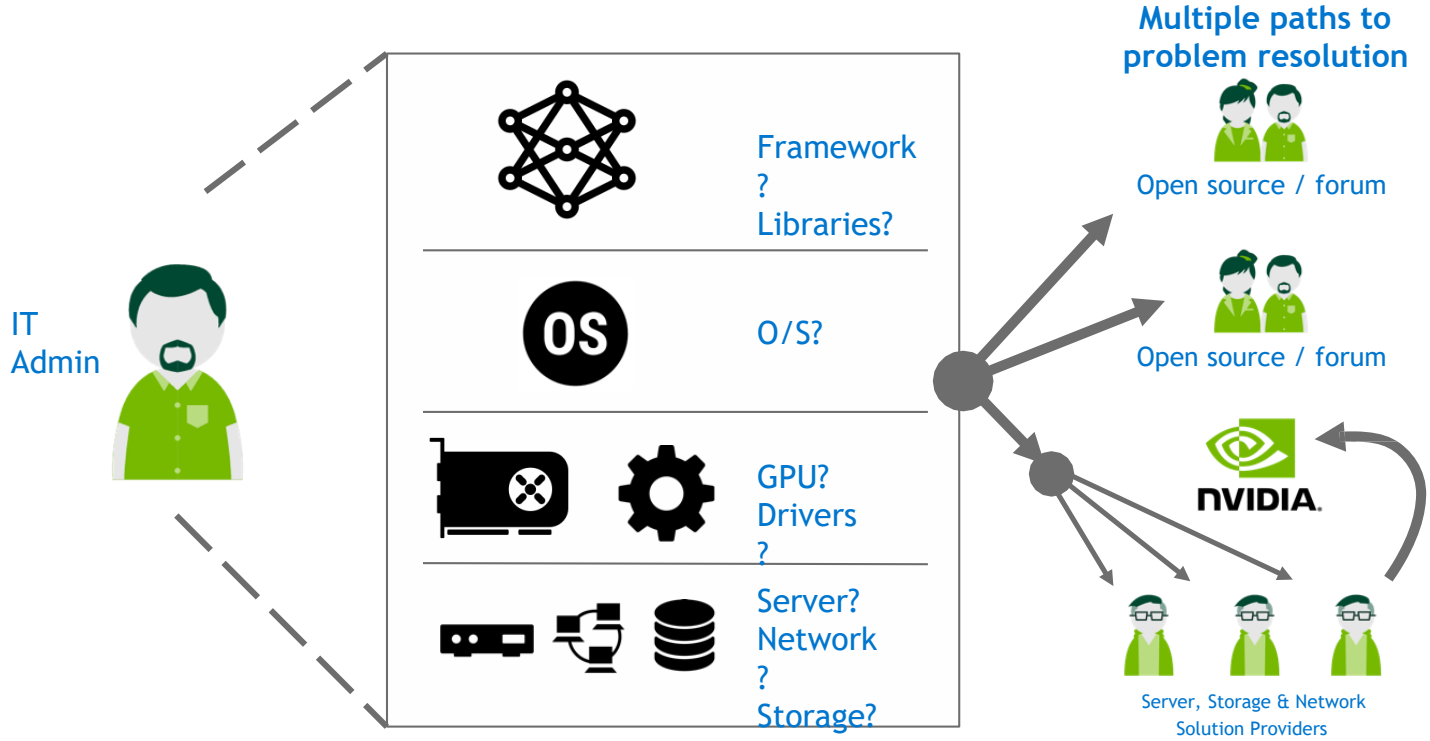
- GPU資源無法隔離，占用到他人資源卻不自知
- 需另外學習docker操作指令、自行管理軟硬體環境
- 底層架構權限暴露在外，一人下錯指令，全機無法使用
- 無身分認證，濫用占用也無法追查，只能等待資源釋放
- 各做各的AI應用框架，重複下載浪費資源

IT人員加入管理

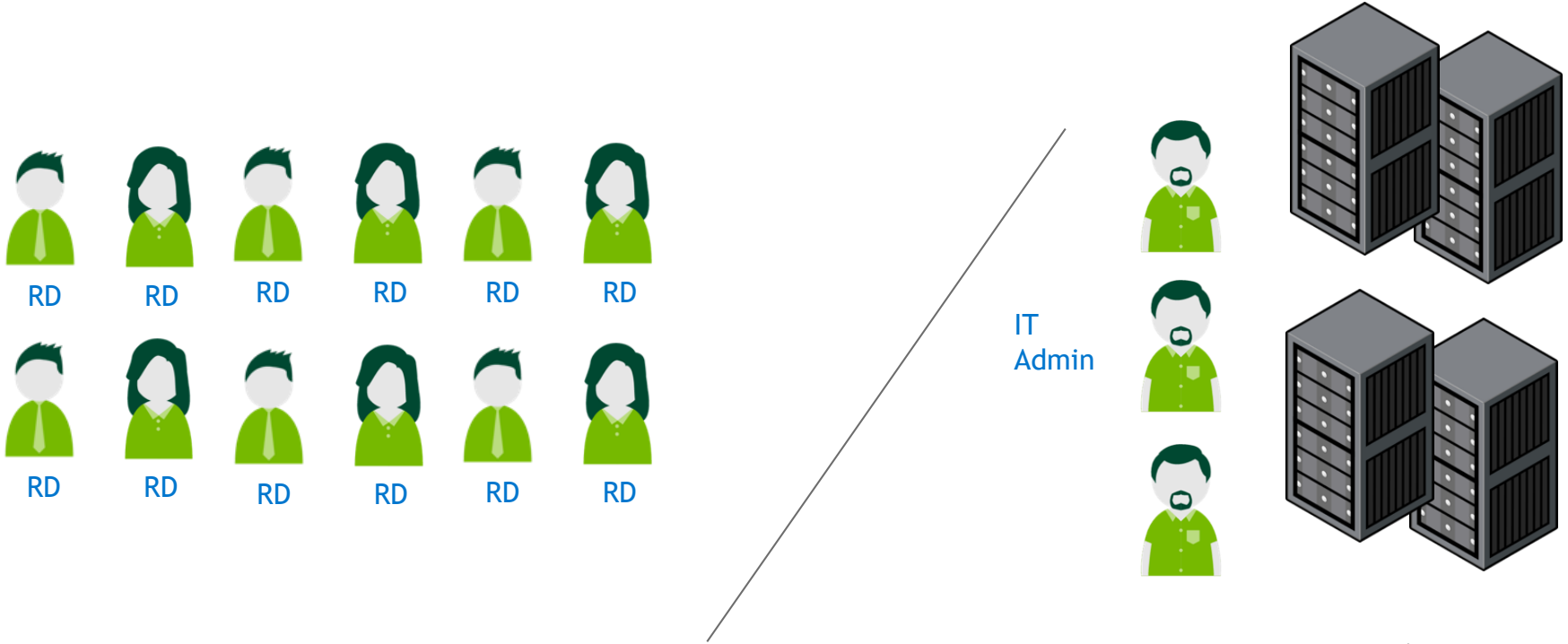




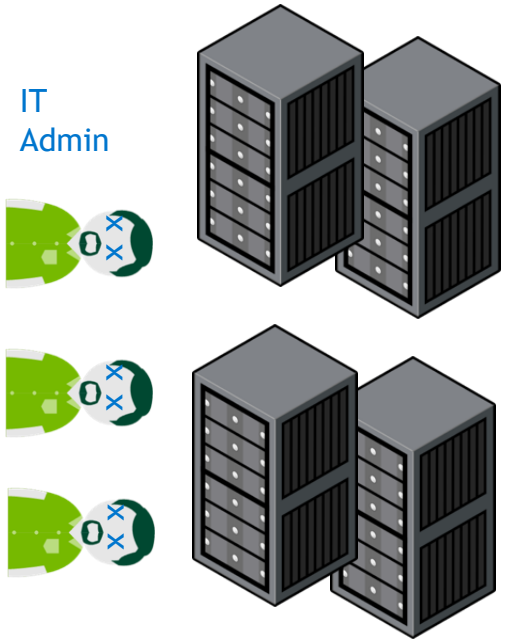
IT 人員需要確認並處理眾多 RD 人員的需求



RD 人員需要 IT 人員提供資源



RD 人員成長需求增加 IT 人員忙不過來



常見問題

- 使用者需求繁複，管理雜亂無章
- 需求又急又趕，打亂工作排程
- 主管關切資源使用狀況，彙整資料不易
- 資源掌握度差，誰再用、用多久、剩餘資源皆不透明
- 需具備docker與kubernetes架構經驗

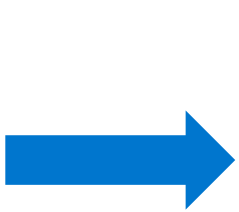
InfinitiesSoft AI-Stack on Dell

最佳化您的 AI 管理平台

AI-Stack on Dell

RD

IT Admin



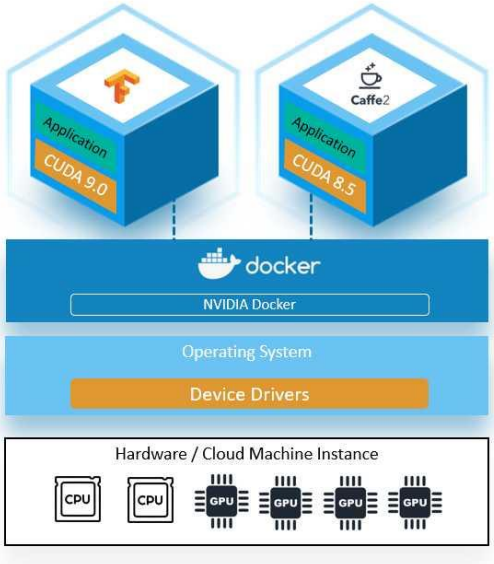
InfinitiesSoft
AI-Stack

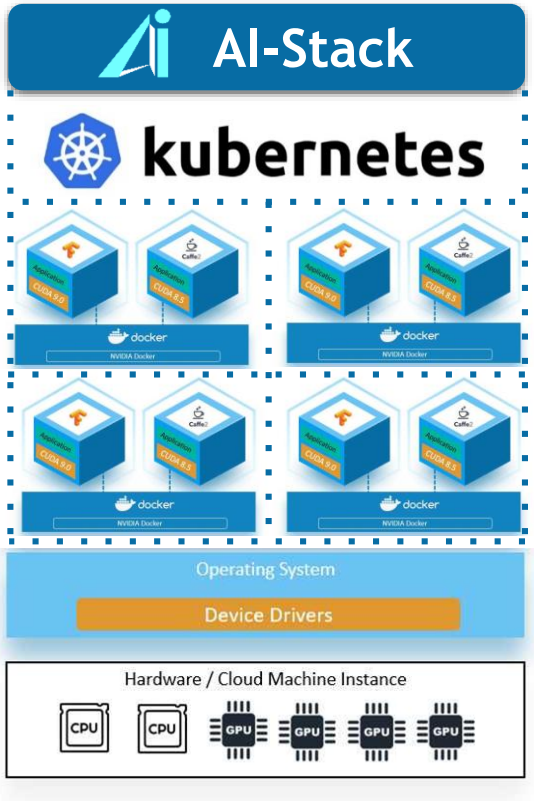
DELL Technologies

“Is there any easy and fast way to use/maintain GPUs and Storage?”

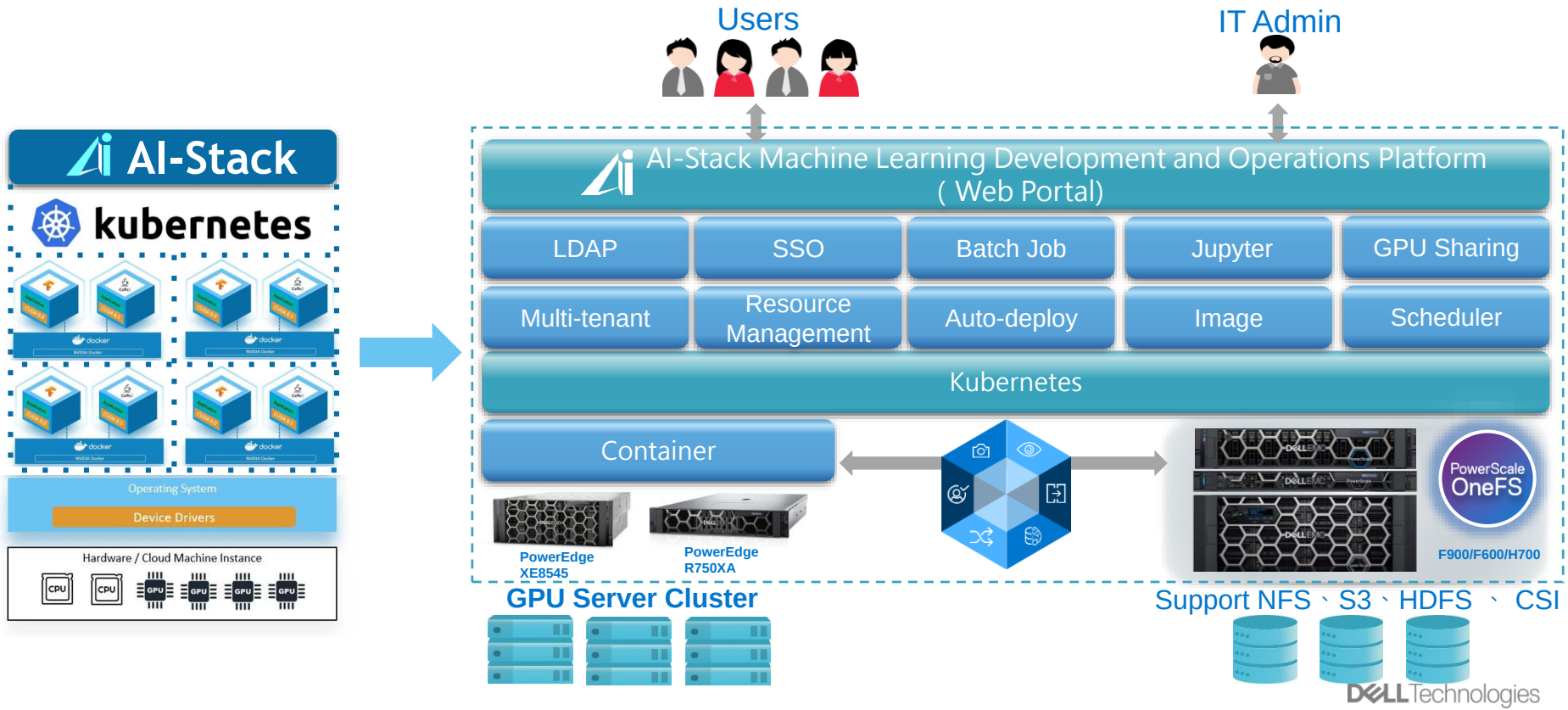
*“ Yes! **AI-Stack on Dell** is your best solution!”*

What is AI-Stack?





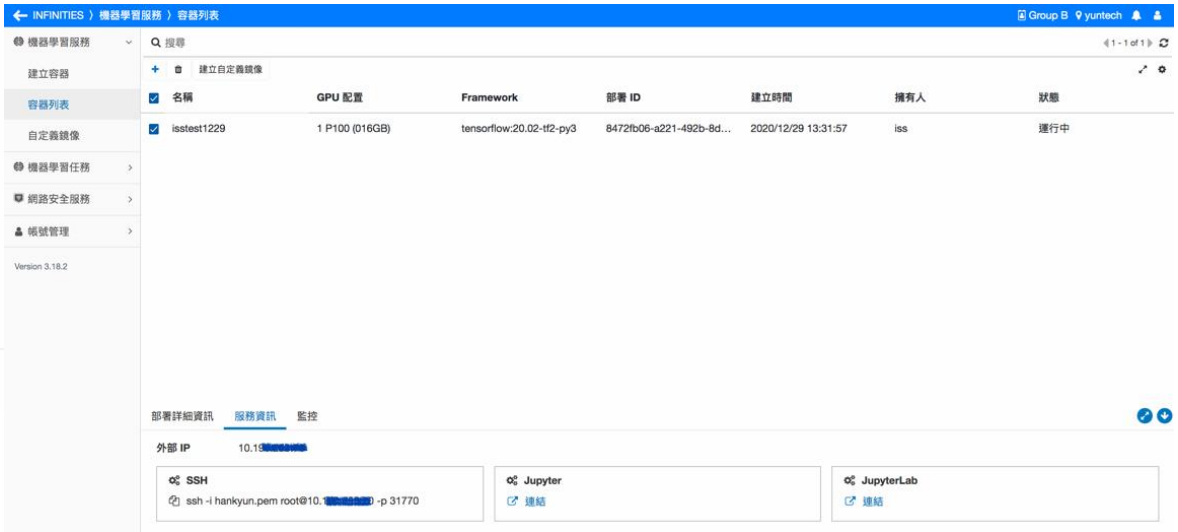
AI-Stack on Dell



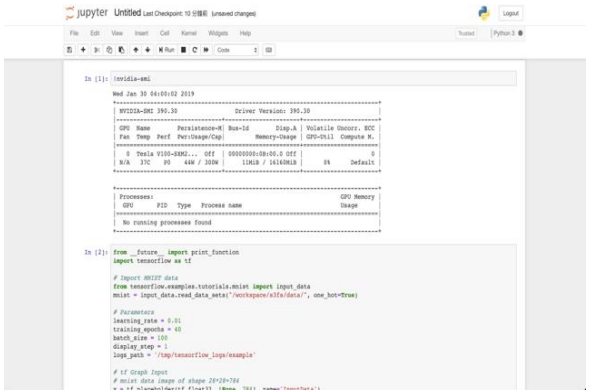
AI-Stack on Dell 滿足用戶需求

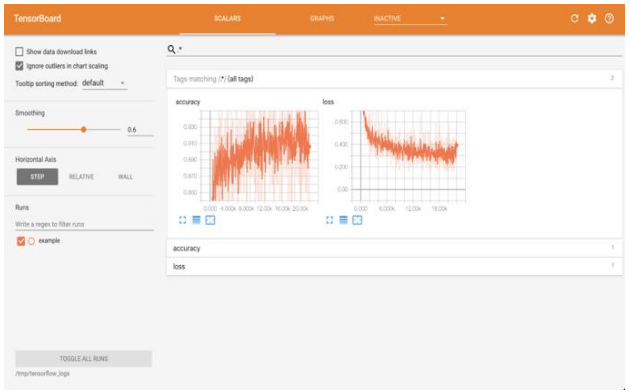
降低使用者門檻	簡易步驟快速建立 (個人/團隊) ML環境
掌握有效資源	彈性IT資源共享、額度限制、工作排程
環境自主權	IT環境隔離、SSH Key或密碼登入環境
易於使用和共用	帳號/儲存整合，批次、預約申請作業
彈性靈活儲存管理	空間管理，放大，複製，快照，備份還原，異地備援

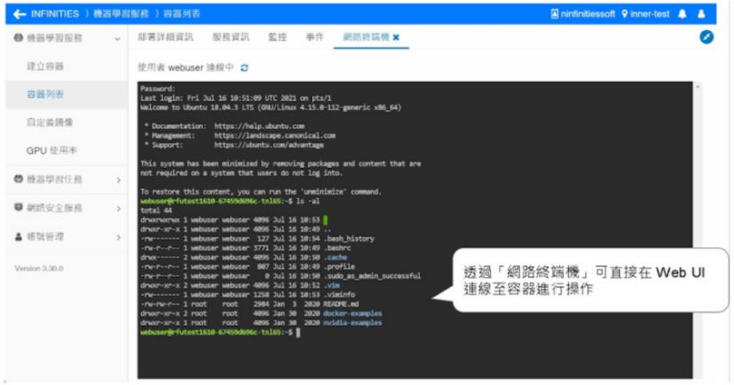
快速建立機器學習框架與運算資源

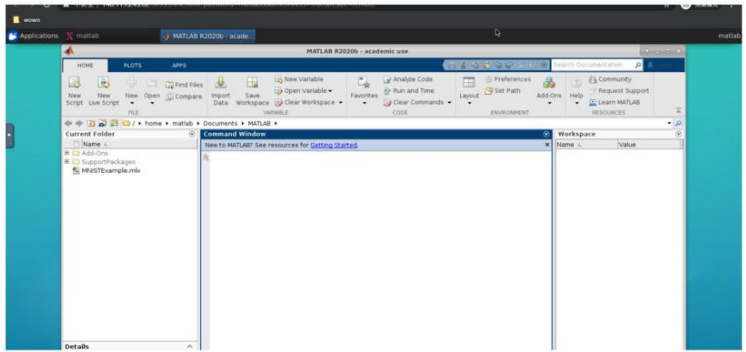


無縫銜接AI/ML研究員常用的開發工具





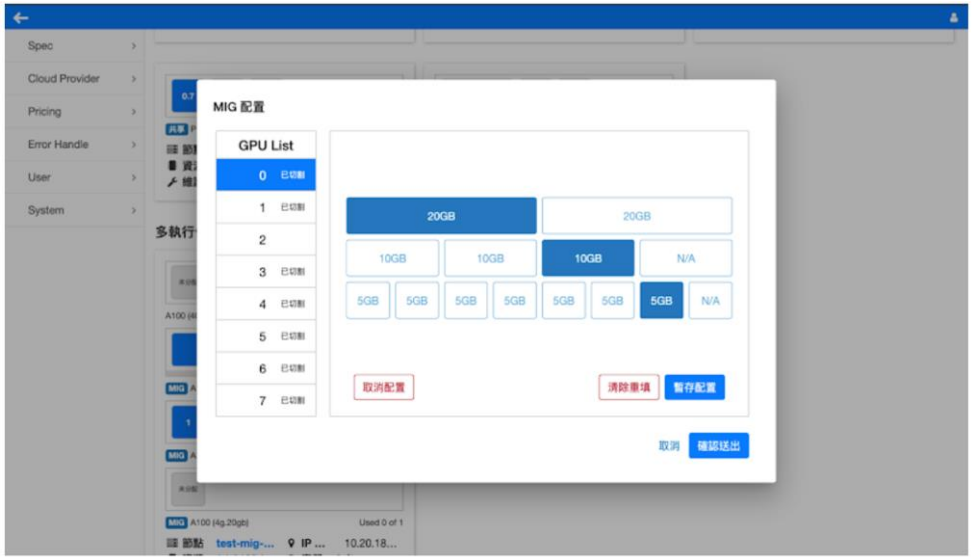




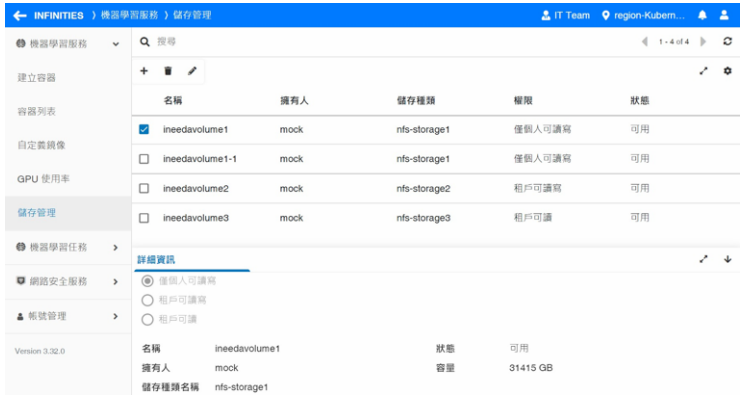
無縫銜接AI/ML研究員常用的開發工具

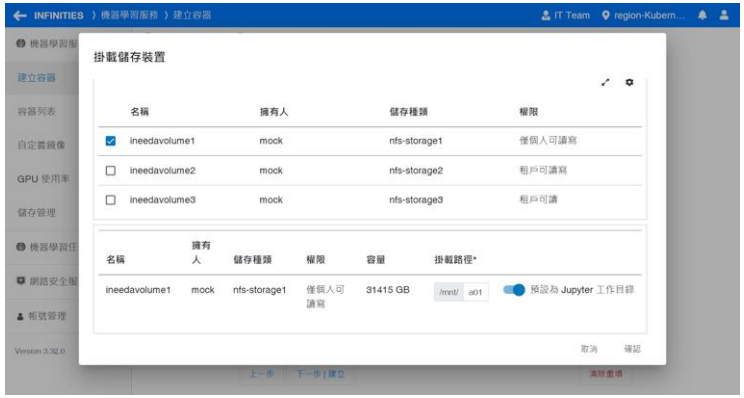
GPU分配利用狀況一目了然

市面上唯一支持MIG技術的AI平台

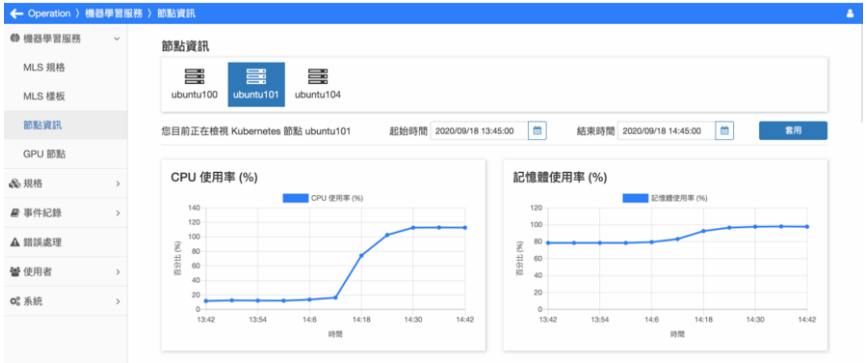


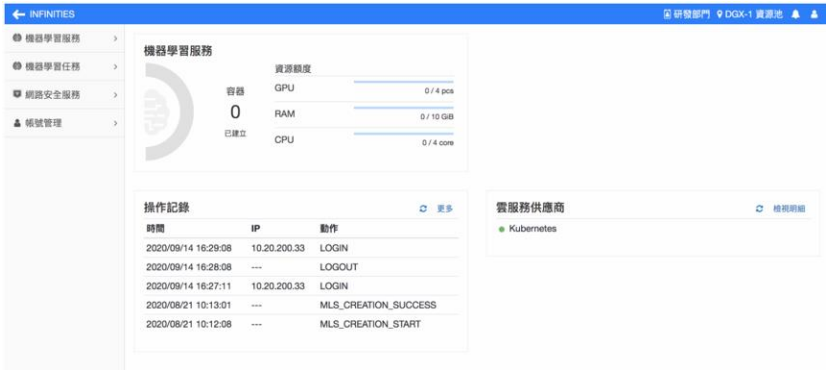
輕鬆掛載儲存設備

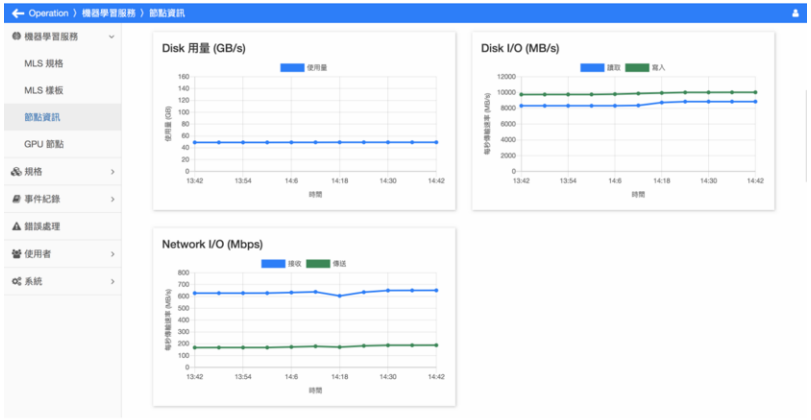


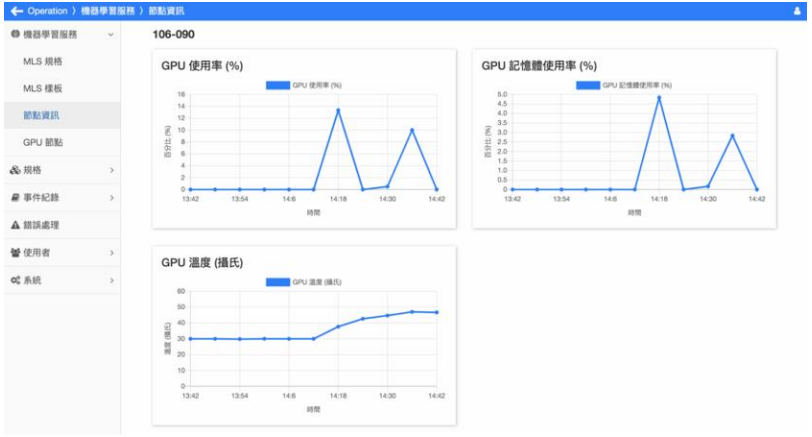


各項資源監控









Dell | AI-Stack 滿足強大易用的需求

GUI 操作/管理

排程管理

DashBoard 資源監控

多租戶/權限管理

Log紀錄

批次推論

報表輸出

單一管理/快速佈署

採用K8S容器管理

Dell + AI-Stack 優勢

- 資源高效分配，透過用戶帳號認證快速將GPU算力交到使用者手上
- 強化用戶資源高度自助使用，降低管理者日常操作負擔
- 精細掌控資源使用成本與報表功能，管理昂貴資源不費力
- 內建資源使用審核功能，讓管理者全方位掌握資源流向
- 掛載外部儲存與GPU伺服器擴容彈性
- 計費計價、成本分析機制
- GPU共享，更有效的提升GPU利用率
- 業界最佳擴展及靈活Dell PowerScale 儲存設備、快速的空間部署及資料管理

關於 InfinitiesSoft

- 亞洲最早開發的混合雲管理平台
- 台灣第一個GPU雲的開發者
- AI-Stack為台灣市佔率最高的AI PaaS

DELL Technologies

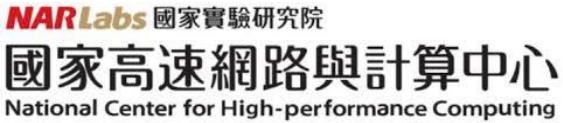
NVIDIA PREFERRED SOLUTION ADVISOR

intel DATA CENTER BUILDER

NVIDIA INCEPTION PROGRAM

重點指標客戶

科研機機



政府機關

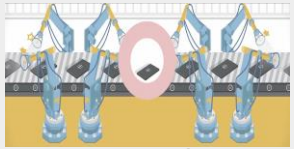
半導體

基建事業

Dell Technologies + AI Stack 解決方案全景圖

AI場景化應用落地解決方案

應用層



AI+製造
產品檢測, 良率提升, 預防維護

AI+零售
商品識別, 智能稱重, 智能貨櫃

AI+教育
人工智慧教學實訓平臺

AI+金融
智慧語音, RPA

AI+醫療
影像輔助診斷

面向AI應用就緒的平臺解決方案

平臺層

資源調度, 雲化,視覺化

Ready solution for AI(Nvidia)

面向GPU的AI就緒解決方案

VMware Bitfusion

GPU虛擬化彈性計算架構

Orion

GPU動態加速雲

Ready solution for AI(Intel)

Intel AI協議棧的就緒解決方案

slurm

kubernetes

深度學習框架

TensorFlow　PyTorch　mxnet　K　Spark　python　jupyter

計算資源管理與優化庫

AI-Stack

GPU　vGPU　FPGA　CPU分散式訓練　Graphcore IPU

端到端AI基礎架構解決方案

基礎架構層

AI計算伺服器
GPU/FPGA/IPU

GPU工作站

AI資料儲存

Isilon

ME4084

資料交換網路

DPS資料保護

VxRail(HCI)

HPC就緒解決方案

DELL Technologies

總結

以數據及AI為核心， 驅動企業創新

- 以數據+AI整合平台、一站購足、附以靈活的付費模式
- 善用雲原生自動化部署策略～加速企業創新 Time-to-Market
- 邊際、核心、雲端場景都能部署～企業能因地制宜，開展具靈活性及彈性
- 員工行動力提升～企業抗疫生存能力提升


DELL Technologies