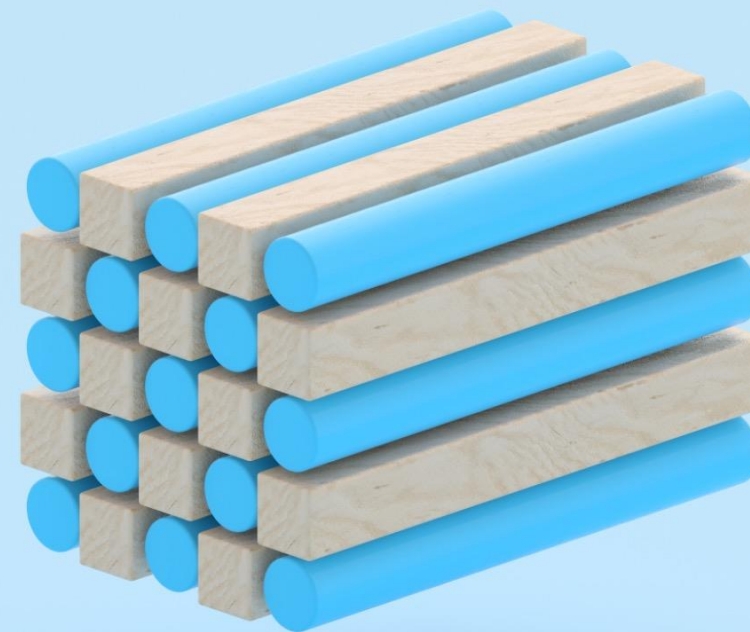


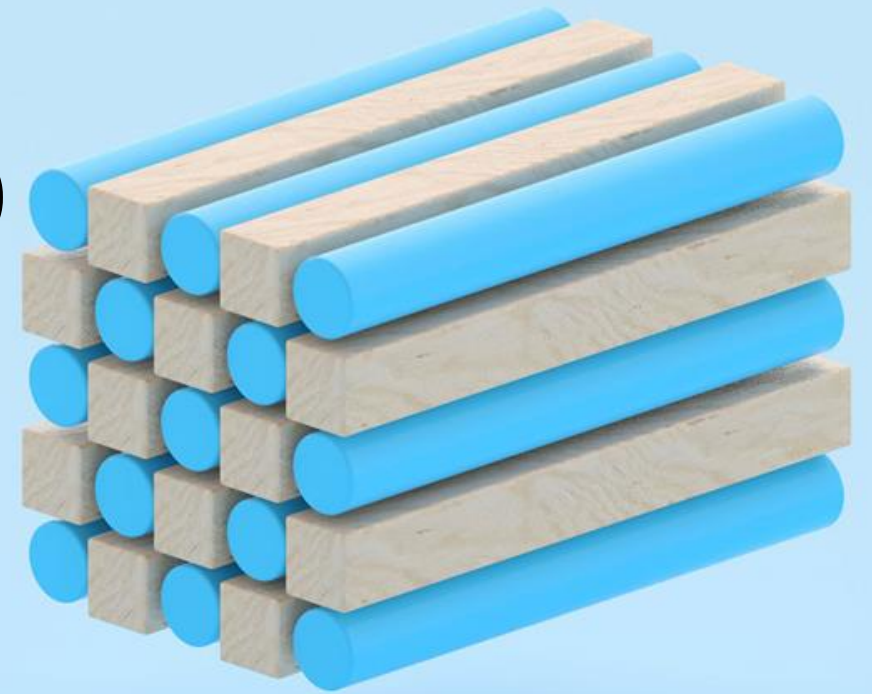
Full Stack AI Data and Experiment Management
Across a Hybrid Cloud

打造新一代自動化機器學習平台

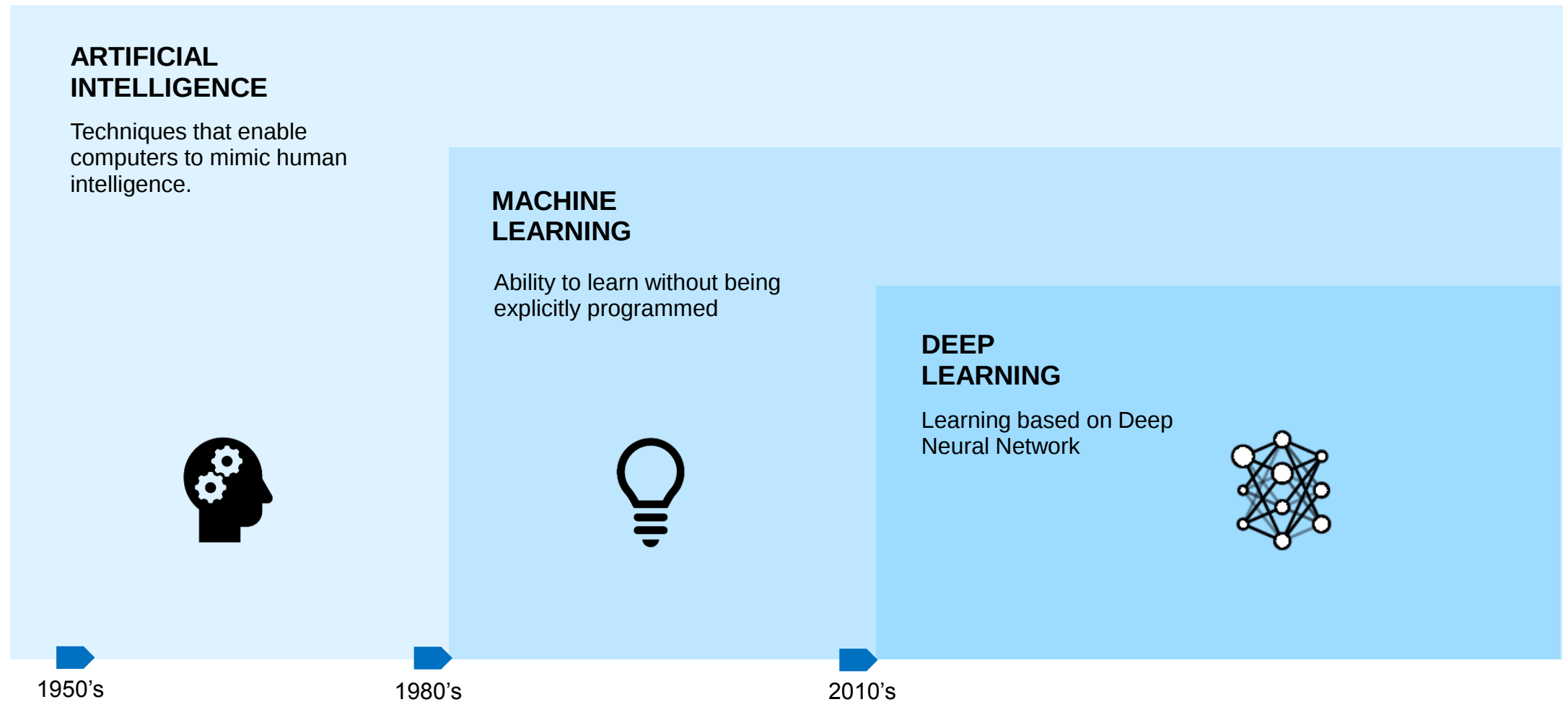
Lawrence Hsu/許宏俊
資深技術顧問



Machine Learning WorkFlows(AI 模型)



人工智慧的演進



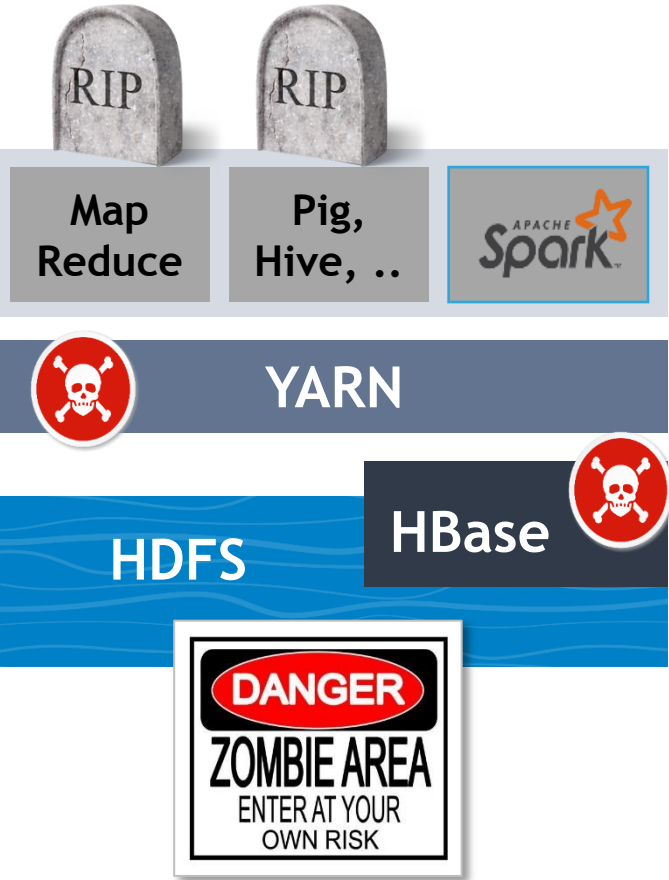
Goodbye Hadoop, hello Auto ML

AutoML成主流AI服務標配

Middleware

Orchestration

Data



Your Business Logic

Any Containerized Microservice



kubernetes

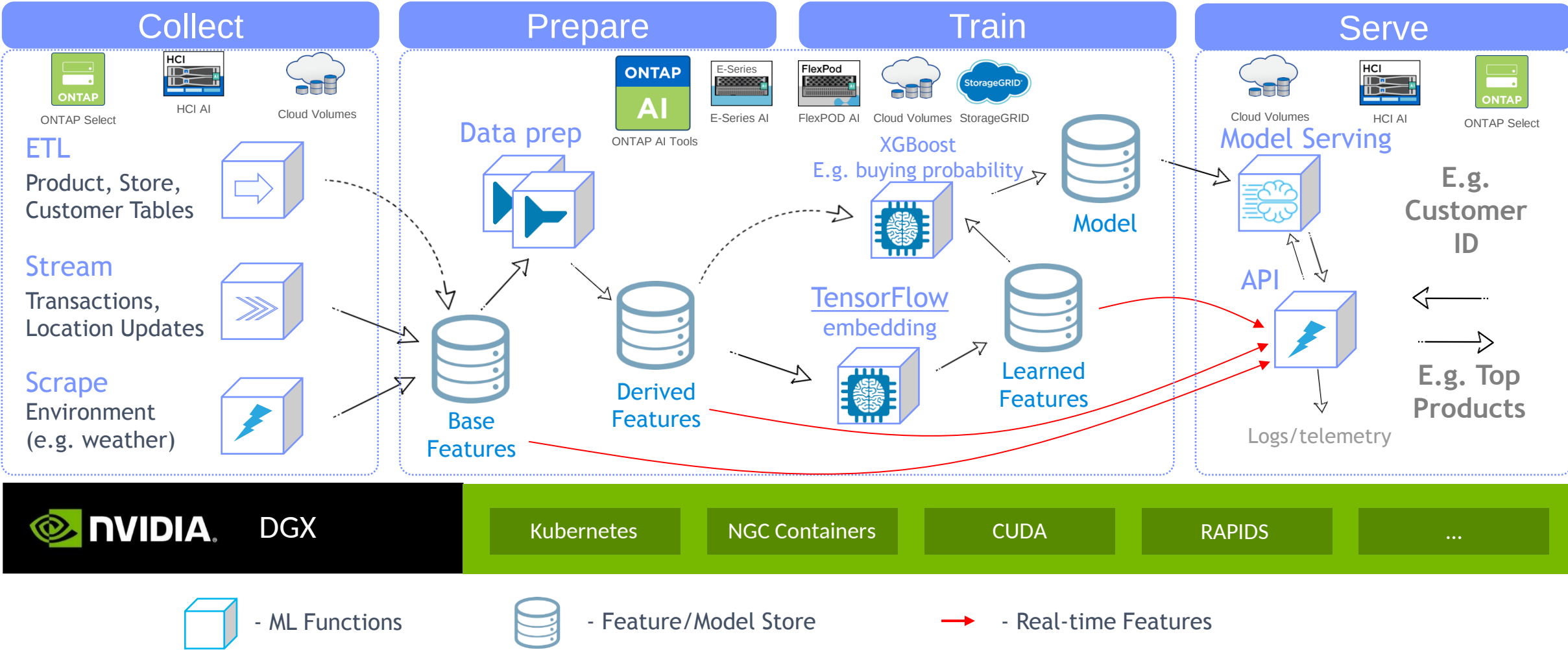
Managed Storage and Databases

Innovate



Consume

AI/ML的流程分析

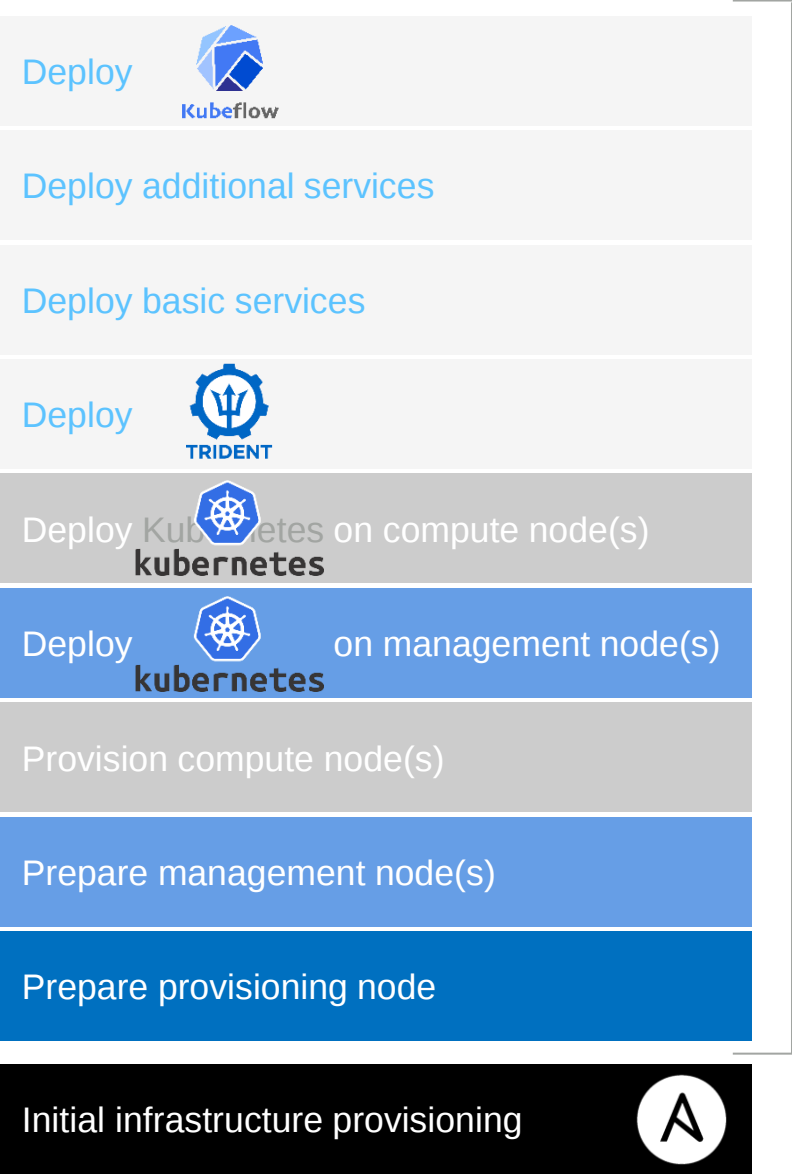
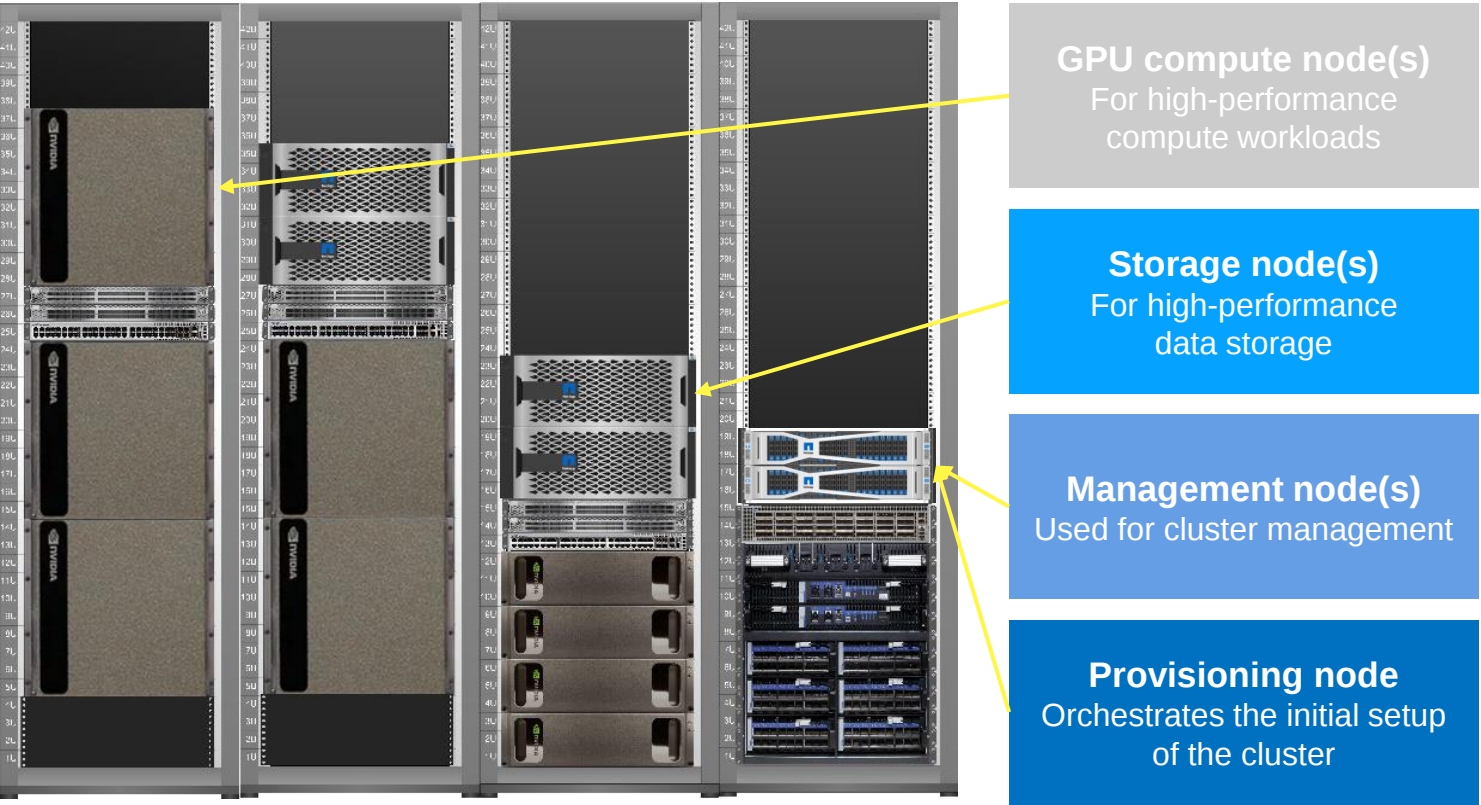


面臨的挑戰

- **資料是新一代的原始碼**
 - 需要能夠保存版本化的AI模型基準
- **數據工作區**
 - 需要有能力對海量數據集進行實驗，而不必擔心弄亂重要資料來源
- **可追溯性和合規性**
 - 需要具備將每個生產模型追溯到用於訓練的確切數據集的能力
- **不同的團隊在不同的位置工作**
 - 需要具備匯總和統一AI計算孤島的能力
- **許多異構數據來源**
 - 需要具備將不同數據源集成到統一AI數據管道中的能力

AutoML 組成元件範例

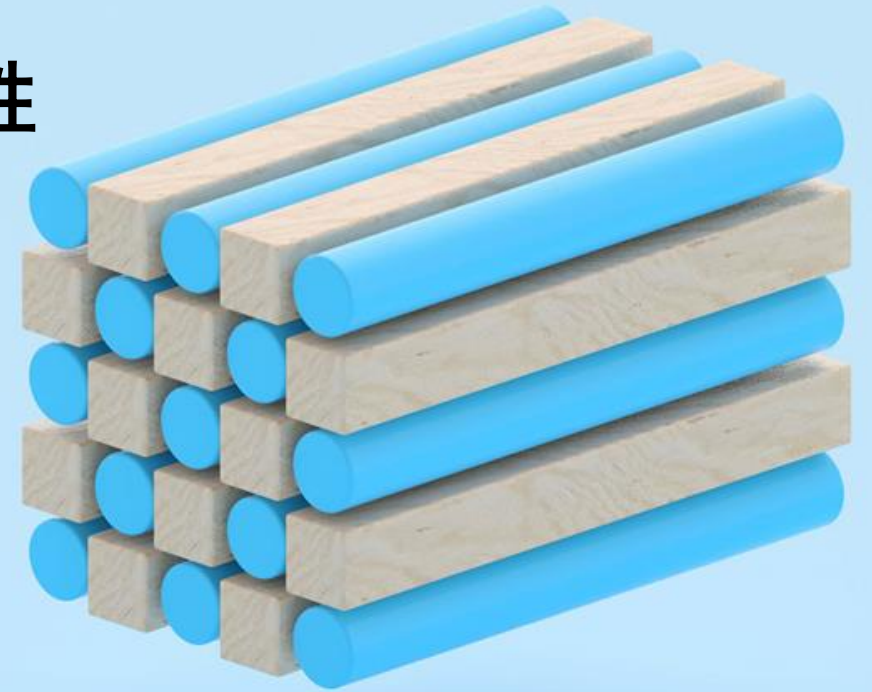
Example for an ONTAP AI environment



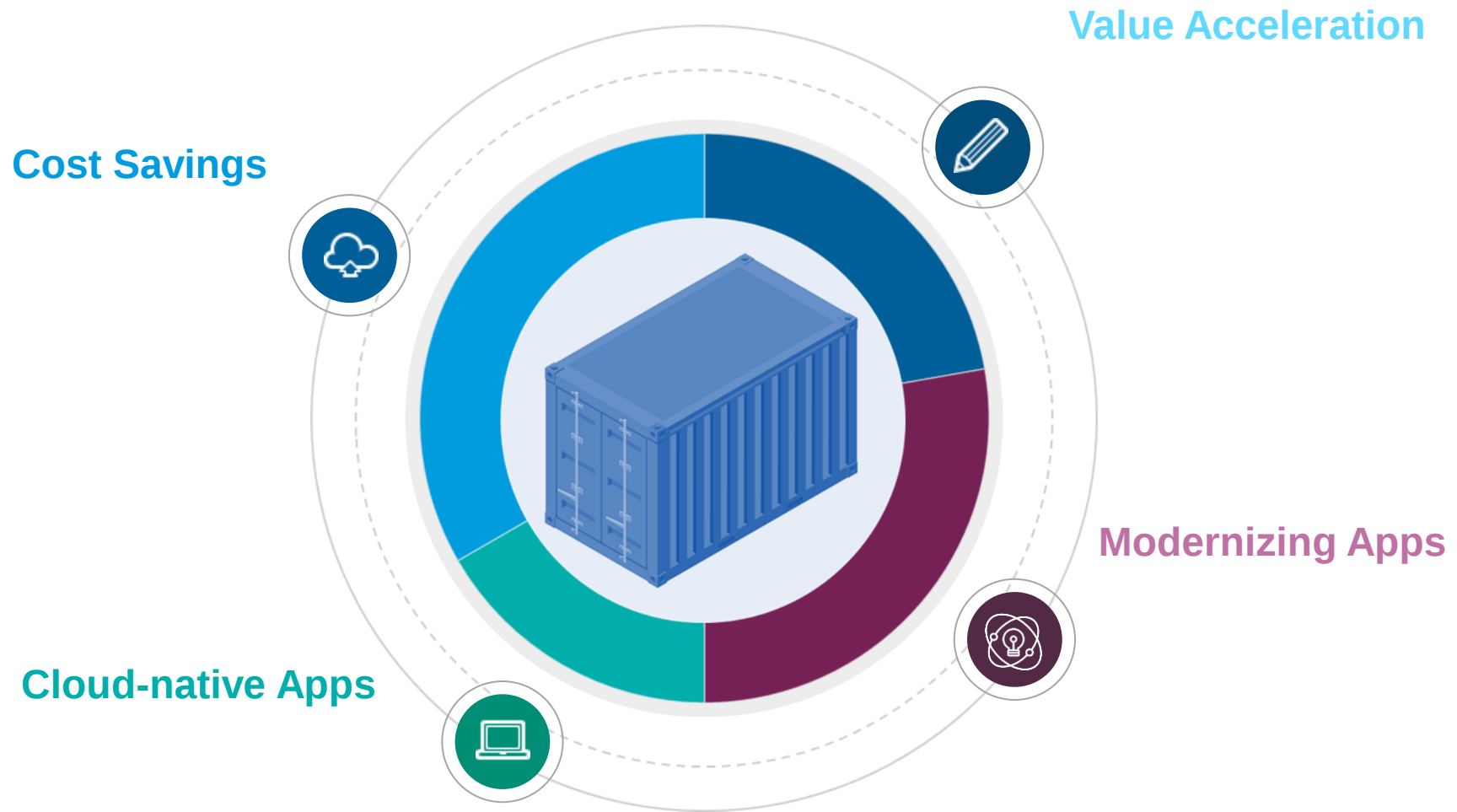
DeepOps

Picture source: NVIDIA POD reference architecture (with NetApp HCI added)

AI/Container Computing 及Storage 特性



為何要用Containers?



為何需要永久性儲存？



Containers are ephemeral — data is not

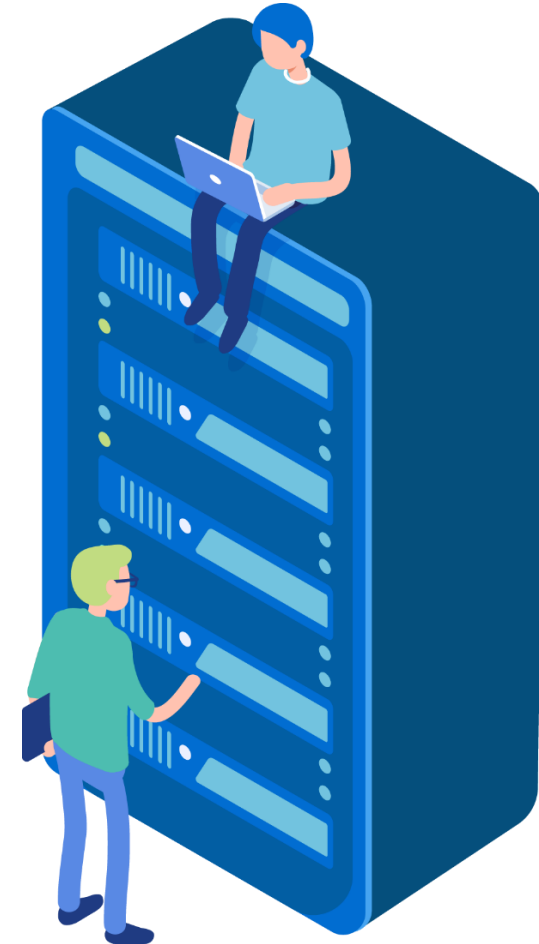


There are no real stateless applications

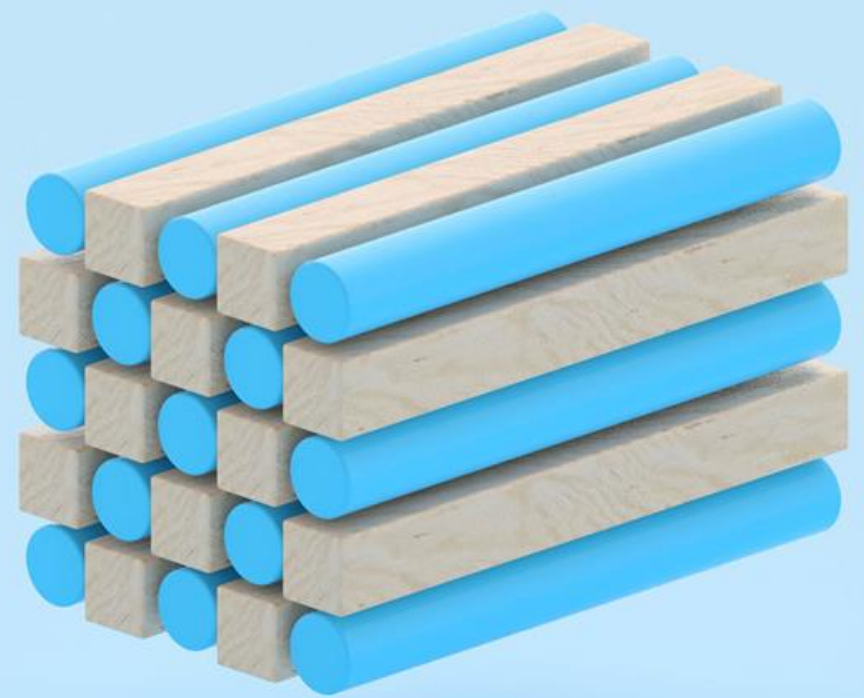


Consuming storage should be easy

Data persistence requirements in the enterprise have not changed



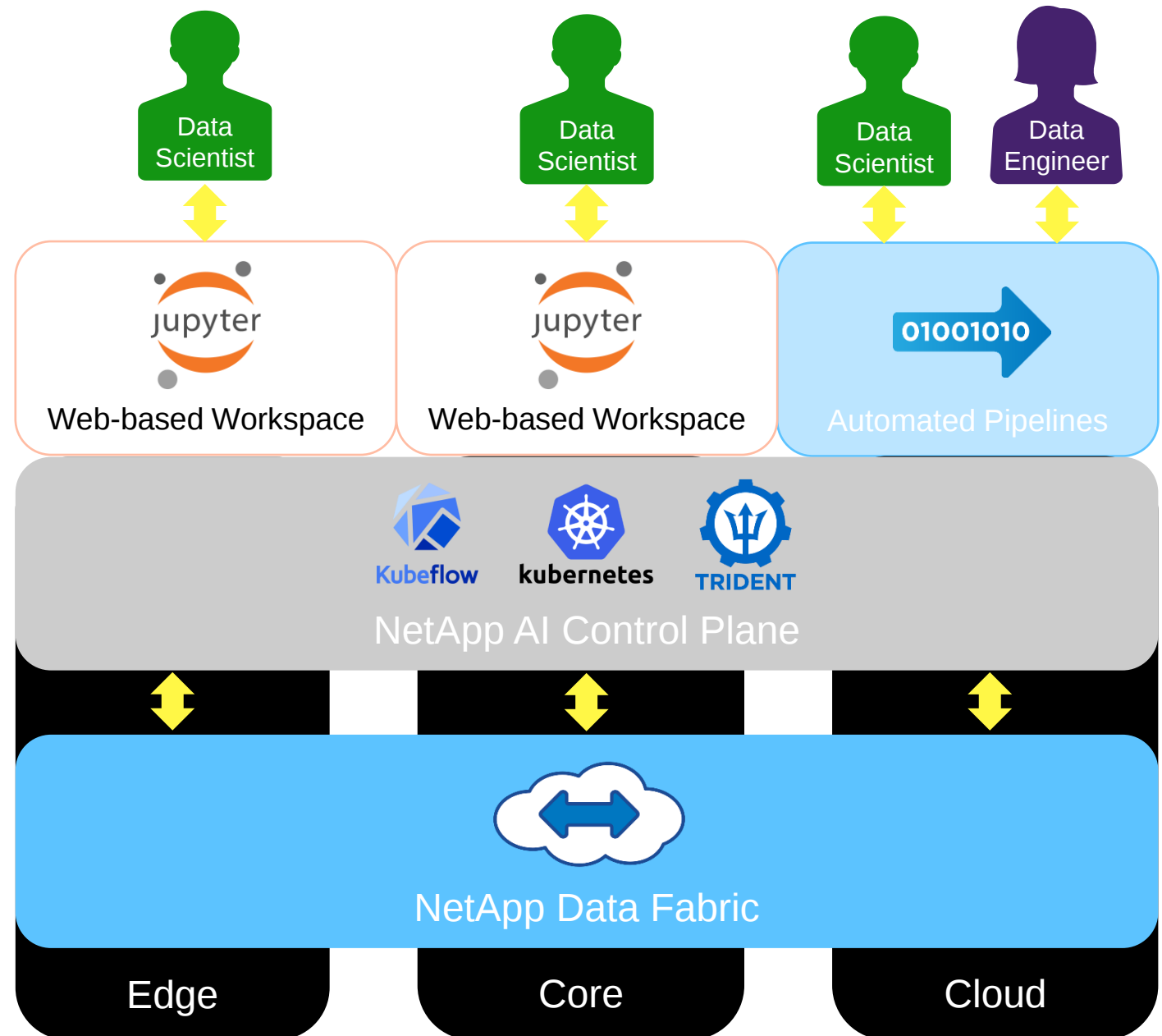
NetApp 產品特色



NetApp AI Control Plane

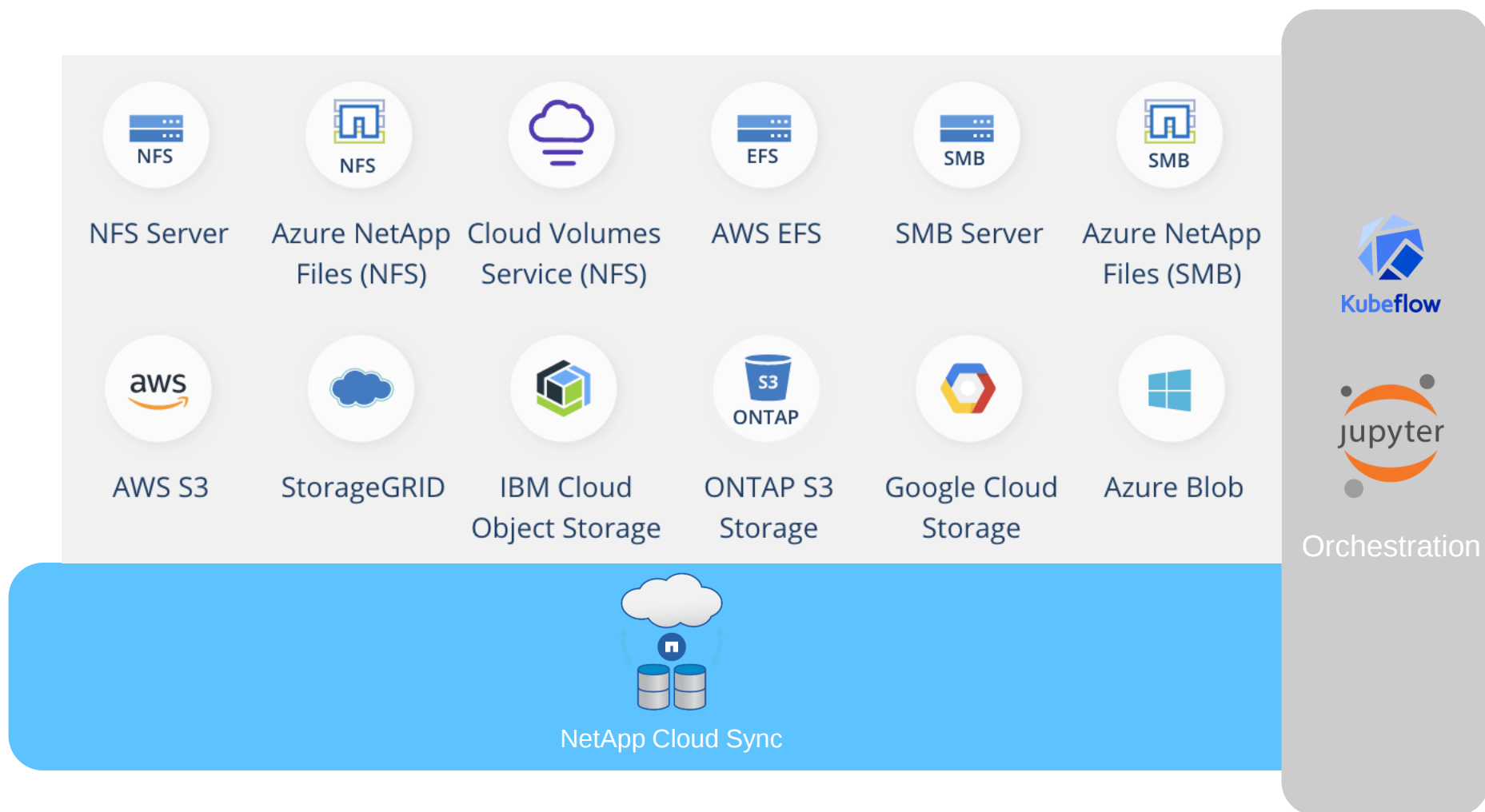
Full-stack AI data and workload management

- On-demand Jupyter Workspaces
- 自動化數據準備，培訓和部署工作流程
- 工作空間和工作負載可以跨越邊緣，核心和雲
 - Choice of any compute and/or cloud
 - Cross-site Data Scientist collaboration
- 內建版本控制
 - Full dataset to model traceability
 - Seamlessly switch between model versions for dev/test, A/B testing, etc.



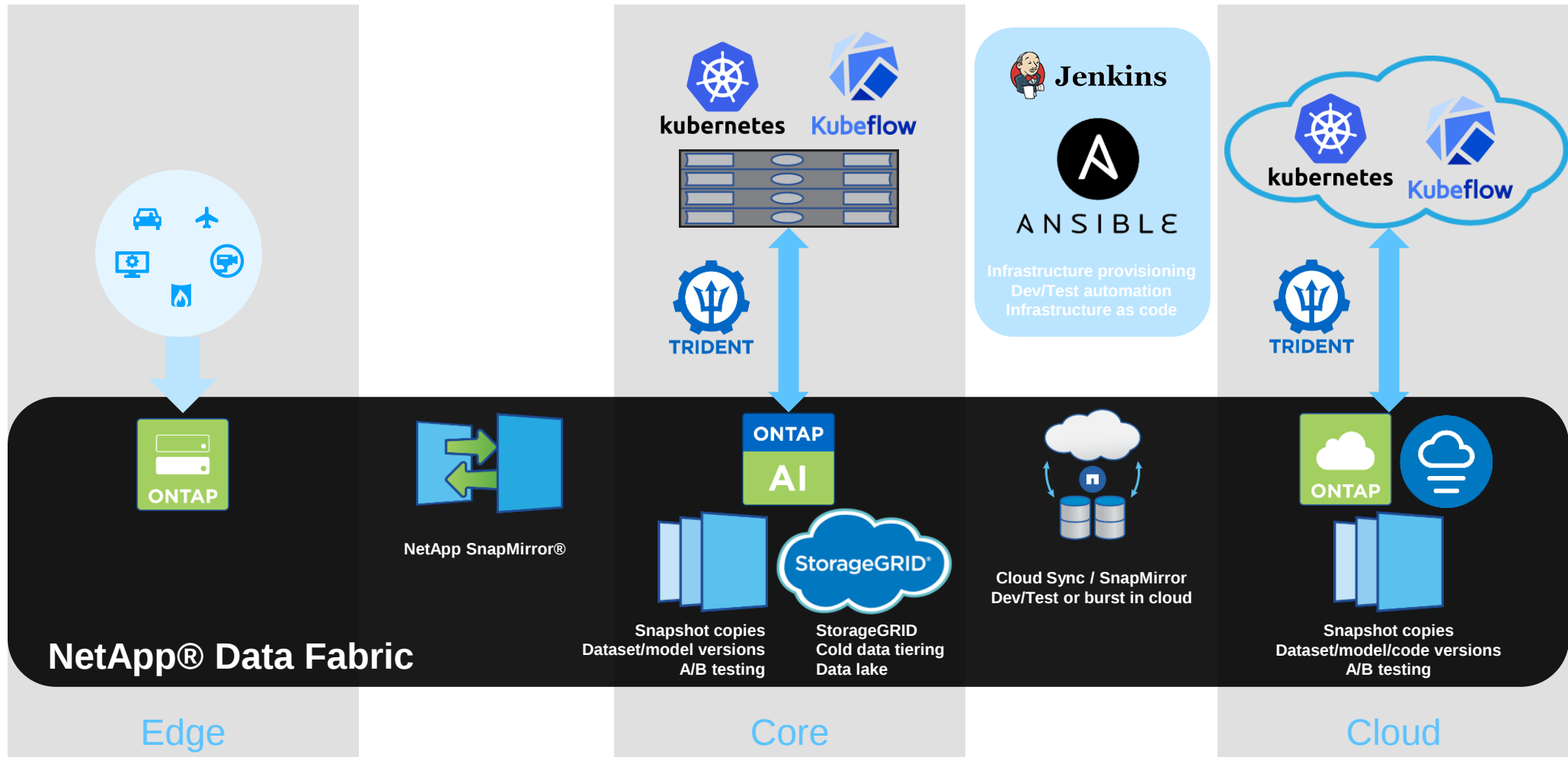
統一數據管道

Integrate heterogeneous data sources; centrally orchestrated



靈活的架構

Data and workloads are available whenever and wherever they are needed



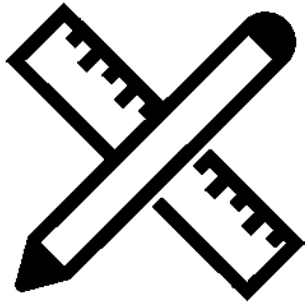
合作加速AI數據發展



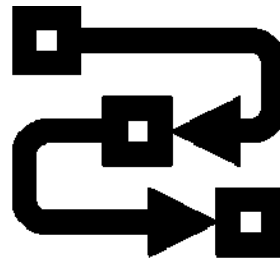
在本地擴展AI面臨的挑戰

Addressing design, deployment and operations bottlenecks

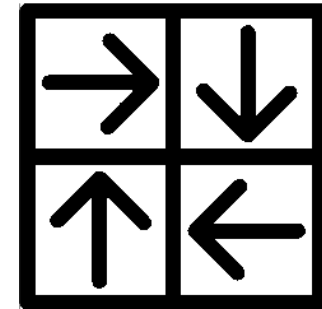
DESIGN
GUESSWORK



DEPLOYMENT
COMPLEXITY



MULTIPLE POINTS
OF SUPPORT



NetApp AI Family

Start small, scale big

HALF RACK



1x DGX-1
1x AFF A220
1x 5624Q

HALF RACK



2x DGX-1
1x AFF A300
2x 5624Q

HALF RACK



4x DGX-1
1x AFF A700
2x 5624Q

FULL RACK



9x DGX-1
1x AFF A800
2x 3232C

FULL RACK



3x DGX-2
1x AFF A800
2x 3232C

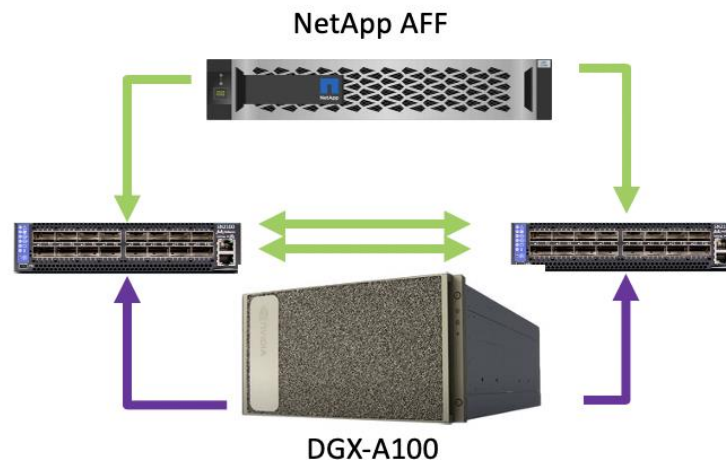
FLEXPOD AI



1x UCS 480 ML M5
1x UCS 5108
1x AFF A800
2x 9336C
2x 6454 FI

泰瑩科技與 NetApp & NVIDIA 建構 無阻礙的 AI 教學與研究平台

NetApp ONTAP AI Platforms are DGX A100 Ready



NetApp和NVIDIA正在密切合作，提出共同解決方案，將DGX A100的優點與NetApp經過驗證的高性能儲存與數據管理結合為一，提供了創新、先進、與靈活的AI平台。

- NetApp ONTAP AI 提供機器學習及深度學習從資料萃取、訓練到推理，提供無與倫比的加速快感。
- NetApp AFF系統利用端到端的高速儲存技術，是業界最快與最靈活的ALL FLASH架構。
- 經NetApp認證的架構與NVIDIA DGX A100 與超高速乙太網路交換機結合在一起，提供快速與穩定的高速網路資料中心。
- 搭配NVIDIA NGC平台，提供預先訓練的模型、私人專屬容器映像倉庫、映像版本管控及專案分享，讓你的AI專案事半功倍。
- 使用新的多實例GPU技術，通過將DGX A100劃分為每個系統多達56個實例，可以加快多個較小的工作負載。

Resources

- **Videos:**

- **Overview:** [NetApp AI Control Plane video](#)
- **Demo:** [Provision a Jupyter Workspace with NetApp AI Control Plane](#)

- **Blog Posts:**

- **Overview:** [How to Effectively Integrate Data Management into Your MLOps Processes with NetApp AI Control Plane](#)
- **Technical How-to:** [AI/ML/DL Dataset and Model Versioning from Within a Jupyter Notebook](#)

- **Solution Brief:** [NetApp AI Control Plane solution brief](#)

- **Technical Paper:** [NetApp AI Control Plane technical report](#)

- **On-demand Webcast:** [AI Control Plane: Full Stack AI Data Management for Hybrid Cloud](#)

netapp.com/ai

NetApp 近兩年亮眼成績

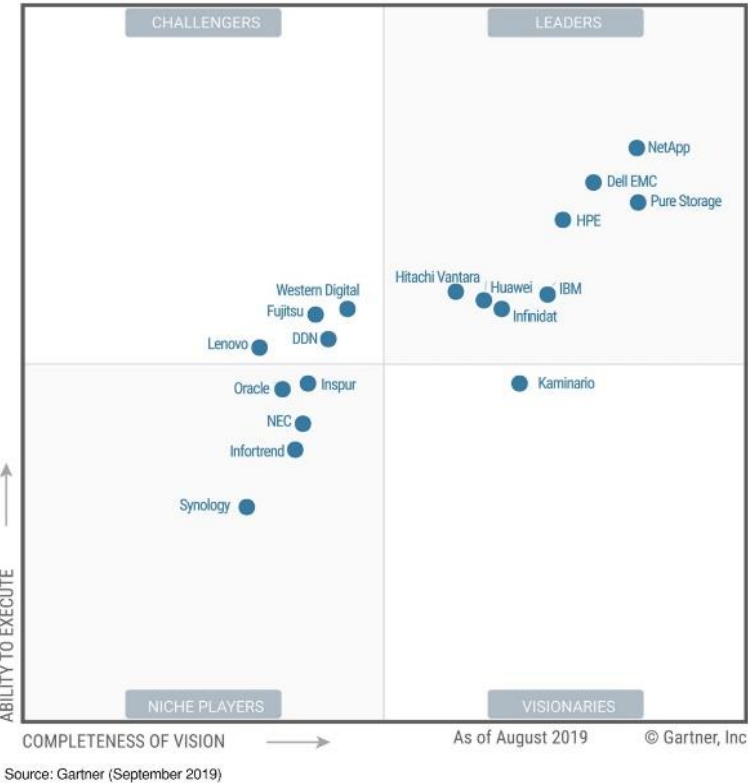
Gartner / IDC Reports

Figure 1. Magic Quadrant for General-Purpose Disk Arrays



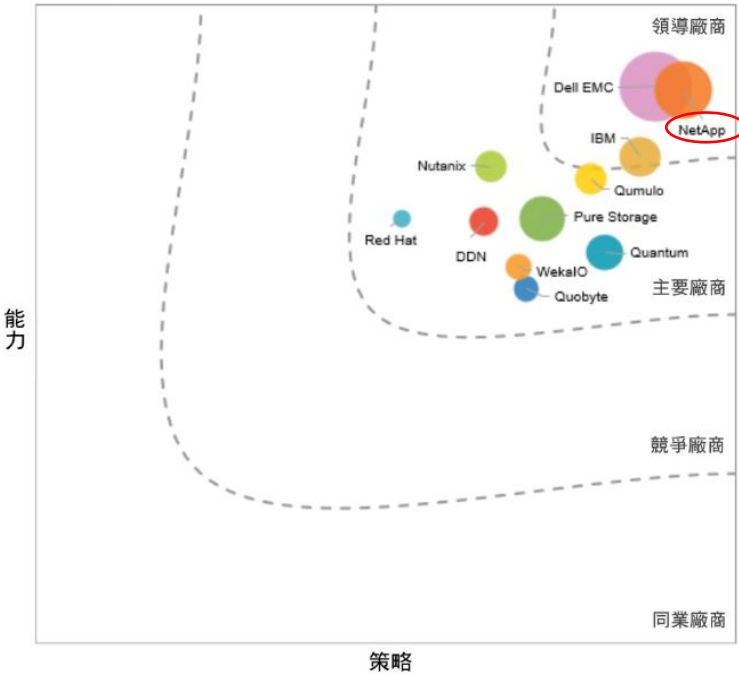
Magic Quadrant for General-Purpose Disk Arrays – Gartner 2018

Figure 1. Magic Quadrant for Primary Storage



A Leader in Primary Storage – Gartner 2019

IDC MarketScape 全球橫向擴充檔案型儲存設備



資料來源：2019 年 IDC

MarketScape: Worldwide Scale-out File-Based Storage - IDC2019

如何更進一步？

Thank You

