

# 构建高效的AI应用平台

武晨晖 2021 SEP

# 联想凌拓

打造专业的数据管理平台

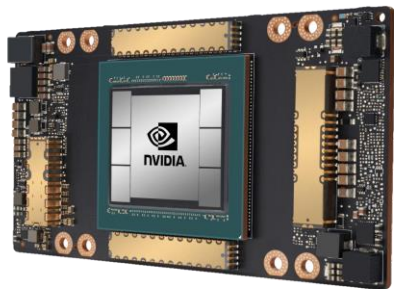
Lenovo



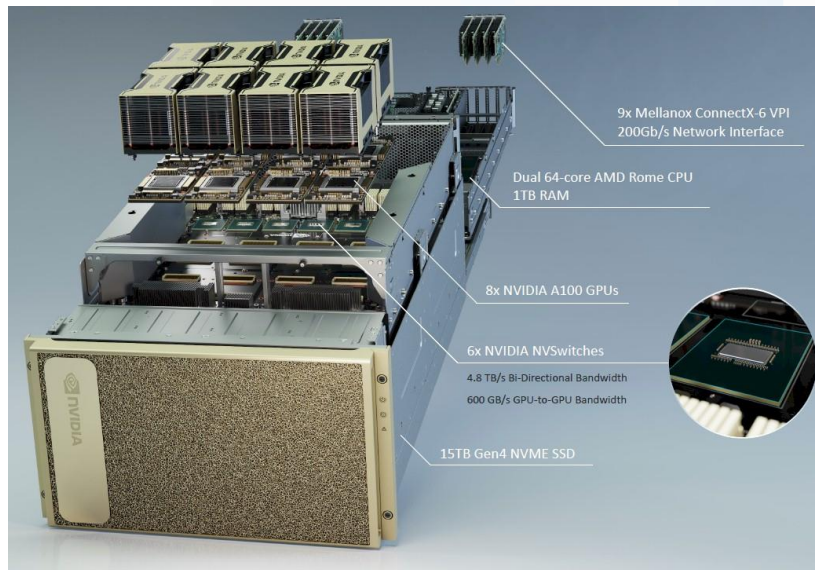
NetApp™



# 强大的算力引擎



NVIDIA A100 产品技术



# NetApp ONTAP AI

一站式的DL平台



## 架构简单

已验证架构，支撑性能已确认，知根知底；  
无自行构建所需的时间和专业技能辅助，快速释放业务价值

## 架构赋能

具备数据管理能力，实现边缘端、核心和云端的数据流动能力  
灵活部署各种的AI框架平台



PYTORCH

Caffe



## 强大能力

性能无限扩展

# ONTAP AI参考架构

采用以太网设备

DGX A100 Servers – Up to 5 PFLOPs FP1

8x A100 SXM2 GPUs –

55k CUDA cores, 3456 TensorCores

600GB/s NVLink

10x 200Gb Ethernet or HDR IB ports

AFF A800 Storage System – 25GB/s read throughput

8x A100 systems per HA pair

48x NVMe SSD

100GbE Network

Split fabric design

200Gb RoCE compute fabric

Mellanox SN3700V

100Gb Ethernet storage & client access

Mellanox SN3700V

<https://www.netapp.com/pdf.html?item=/media/21793-nva-1153-design.pdf>

<https://www.netapp.com/pdf.html?item=/media/21789-nva-1153-deploy.pdf>

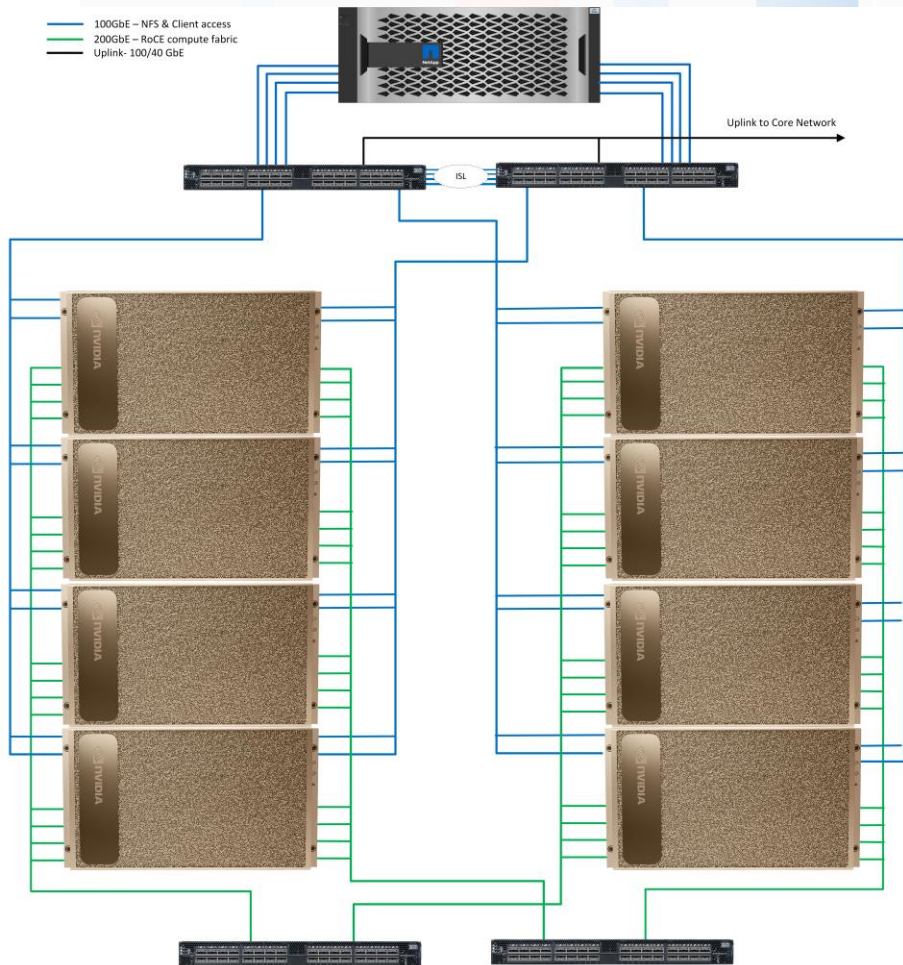


Storage  
AFF A800  
storage system

Storage  
Fabric  
SN3700V switches

Compute  
DGX A100 systems

Compute  
Fabric  
SN3700V switches



# ONTAP AI参考架构

采用IB网络

DGX A100 Servers – Up to 5 PFLOPs FP1

8x A100 SXM2 GPUs –

55k CUDA cores, 3456 TensorCores

600GB/s NVLink

10x 200Gb Ethernet or HDR IB ports

AFF A800 Storage System – 25GB/s read throughput

8x A100 systems per HA pair

48x NVMe SSD

100GbE Network

Split fabric design

200Gb HDR InfiniBand compute fabric

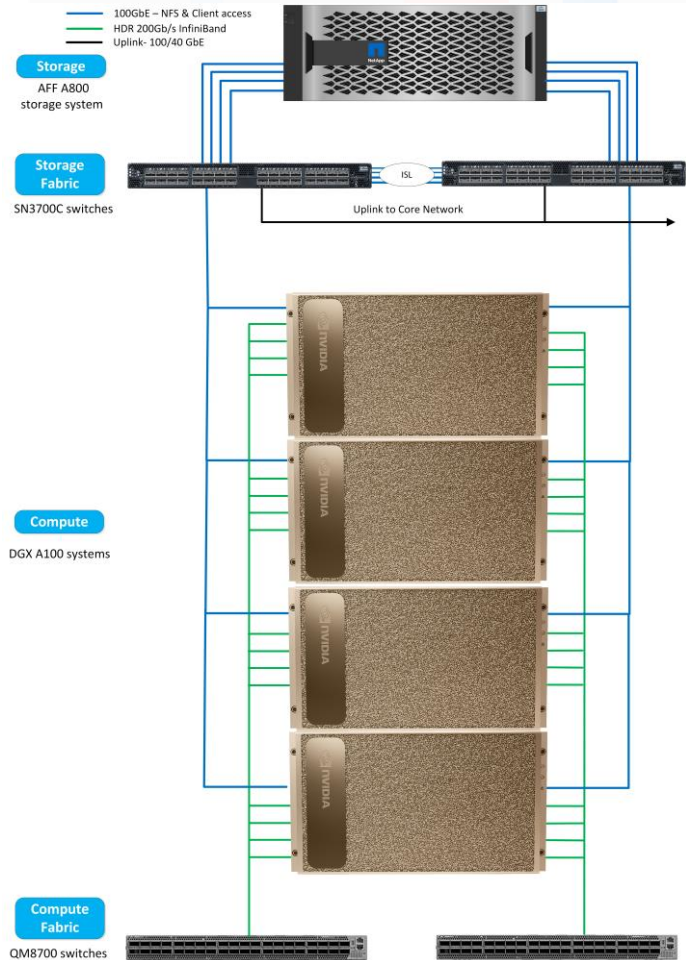
Mellanox QM8700

100Gb Ethernet storage & client access

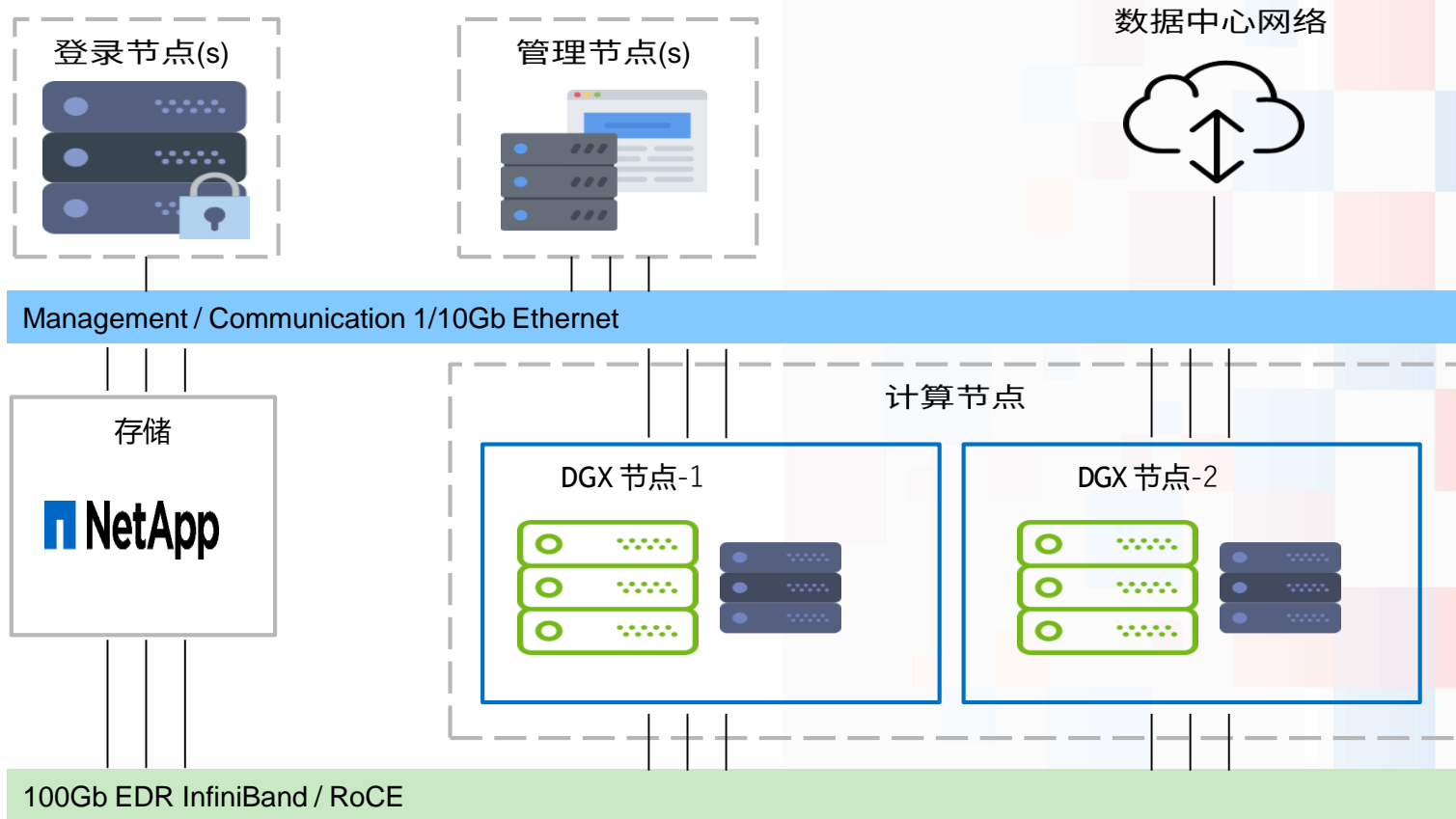
Mellanox SN3700V

<http://www.netapp.com/us/media/nva-1151-design.pdf>

<https://www.netapp.com/media/20708-nva-1151-deploy.pdf>



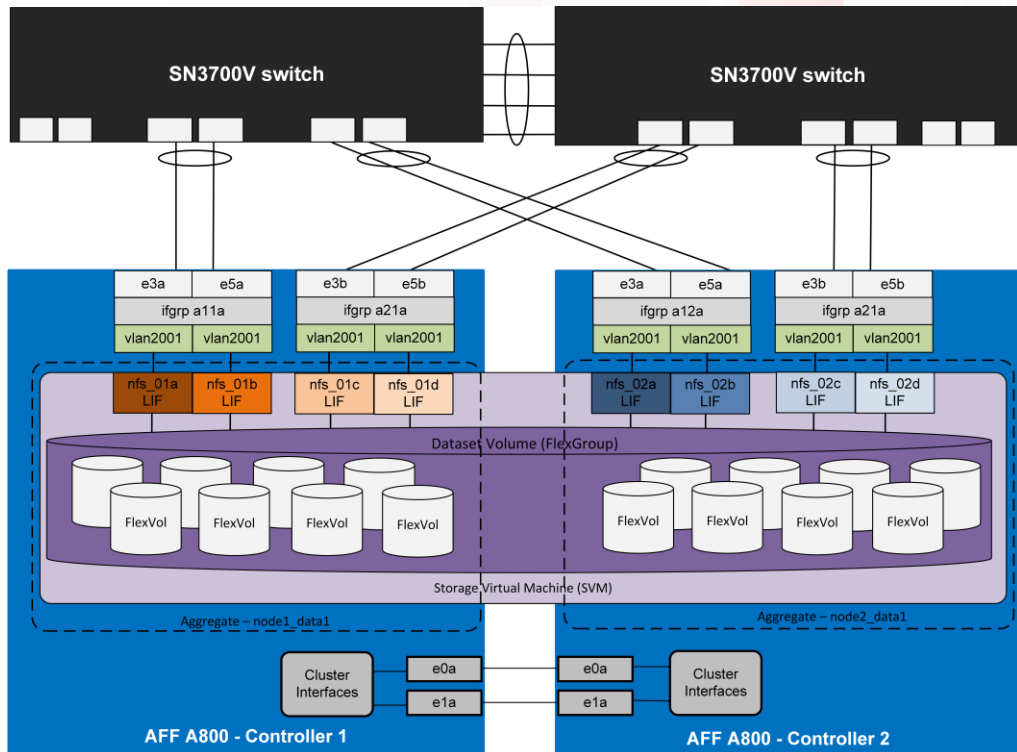
# ONTAP AI 拓扑图



# ONTAP AI参考架构

## 配置说明

- nConnect NFS option 在统一的NFS挂载点的方式下会有更好的I/O带宽
  - Only available with Linux Kernel 5.3 or higher, DGX OS 5.0 is based on Ubuntu 20.04 LTS
- 验证的主机 mount 命令参数
  - wsize=262144,rsize=262144 (256k transfer size)
  - Nconnect=16
- 为每台DGX配置独立存储访问端口
  - Not required but helps w/ LACP hashing
- **Single FlexGroup** (单一存储空间)
  - 16 constituents, 8 per node
- 存储采用以太网连接,DGX采用以太网/IB网连接



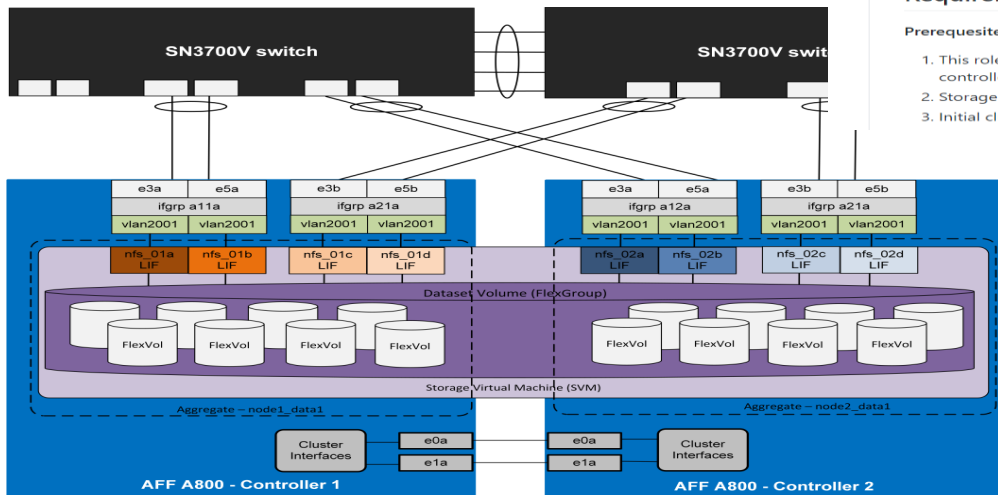


# ONTAP AI with Ansible



## 自动化部署

- 加速部署工作，减少配置错误
- 防止配置的不一致
- 存储配置自动化
  - 网络、存储空间、文件协议及访问策略



The screenshot shows a GitHub repository for 'na\_ontap\_ai\_aff\_deploy'. The repository has 1 branch and 0 tags. The commit history shows a push on 6 Jun with 2 commits. The README.md file is visible, containing the following information:

### na\_ontap\_ai\_aff\_deploy

Ansible role that can be used to deploy/provision a new NetApp AFF/FAS cluster as part of an ONTAP AI converged infrastructure pod.

#### Requirements

**Prerequisites:**

1. This role assumes the storage system is an AFF A800 with X1146 or X1148 network cards in slots 3 and 5 in each controller (2x cards/controller).
2. Storage system nodes should be cabled as specified in NVA-1151.
3. Initial cluster setup/creation must have already been completed.

[https://github.com/NetApp-Automation/na\\_ontap\\_ai\\_aff\\_deploy](https://github.com/NetApp-Automation/na_ontap_ai_aff_deploy)

# NetApp--NFS协议主导者

NFS是一种网络文件系统，从1985年推出至今，共发布了3个版本：NFSv2、NFSv3、NFSv4，NFSv4包含两个次版本NFSv4.0和NFSv4.1。经过20多年发展，NFS发生了非常大的变化，最大的变化就是推动者从Sun变成了NetApp，NFSv2和NFSv3基本上是Sun起草的，NetApp从NFSv4.0参与进来，并且主导了NFSv4.1标准的制定过程，而Sun已经被Oracle收购了。

编号	版本	RFC	时间	页数
1	NFSv2	rfc1094	1989年3月	27页
2	NFSv3	rfc1813	1995年6月	126页
3	NFSv4.0	rfc3530	2003年4月	275页
4	NFSv4.1	rfc5661	2010年1月	617页

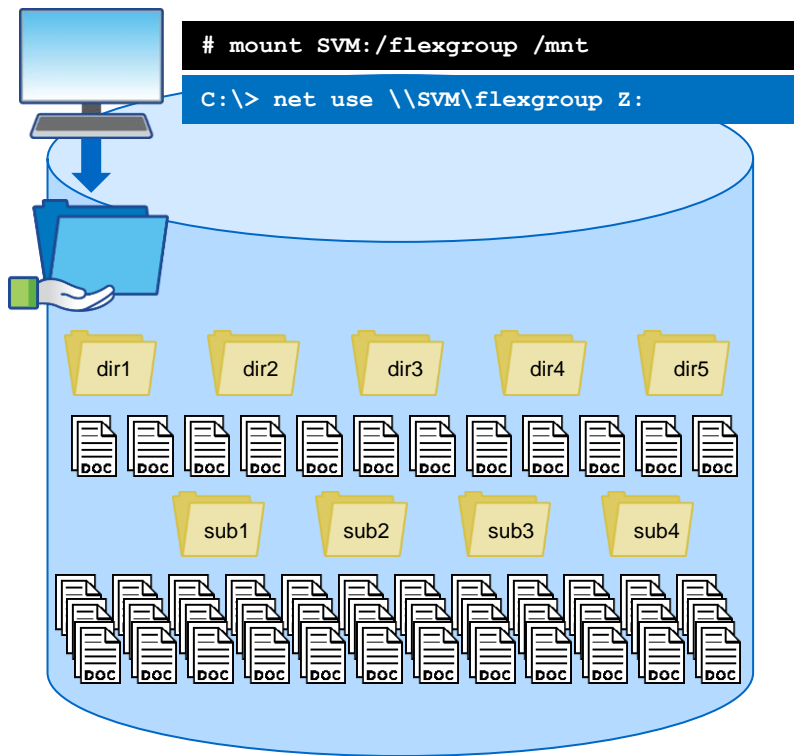


Alex McDonald  
Office of the CTO  
NetApp

Alex McDonald joined NetApp in 2005, after more than 30 years in a variety of roles with some of the best known names in the software industry .

With a background in software development, support, sales and a period as an independent consultant, Alex is now part of NetApp's Office of the CTO that supports industry activities and promotes technology & standards based solutions, and is co-chair of the SNIA NFS Special Interest Group.

# NetApp--擅长海量小文件场景



## FlexGroup Volumes 统一集群文件系统

- 横跨存储多节点方式部署，弹性扩展
- NFS or SMB 访问协议，标准网络接口
- 标准用户认证方式：AD域/LDAP
- 4KB文件块、分布式metadata存储及文件系统一致性
- 文件空间大于**20+PB**
- 文件数量大于 **4000亿**

# NETAPP产品方案



AFF A800

AFF A700

AFF A400

AFF A250

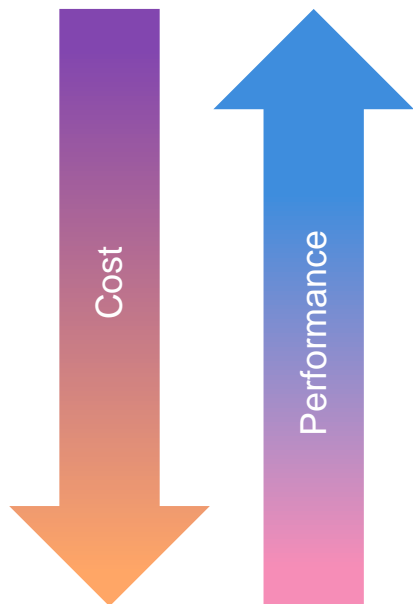
AFF A220



- NetApp 自研
- 文件存储行业第一名
- 丰富的存储功能
- NFS / SMB / S3

# NetApp存储性能规格

按需配置



		Throughput**	Raw Capacity (typical / max)	Connectivity	# NVIDIA A100 GPUs supported****
AFF A800	1 HA pair *	25GB/s	368TB / 3.6PB	100 GbE	08 - 64
	4 HA pairs	100GB/s	1.47PB / 14.4PB		256
	12 HA pairs	300GB/s	4.4PB / <b>43.2PB</b>		<b>768</b>
AFF A400	1 HA pair	11GB/s	182TB / 14.7PB	40/100 GbE	06 - 16
	4 HA pairs	44GB/s	728TB / 58.8PB		64
	12 HA pairs	132GB/s	2.1PB / <b>176.4PB</b>		<b>192</b>
AFF A250	1 HA pair	7.4GB/s	91.2TB / 4.4PB	25 GbE	04 - 06
	4 HA pairs	29.6GB/s	364.8TB / 17.6PB		16 - 24
AFF A220	1 HA pair	5.1GB/s	91.2TB / 4.4PB	10 GbE	02 - 04
	4 HA pairs	20.4GB/s	364.8TB / 17.6PB		08 - 16

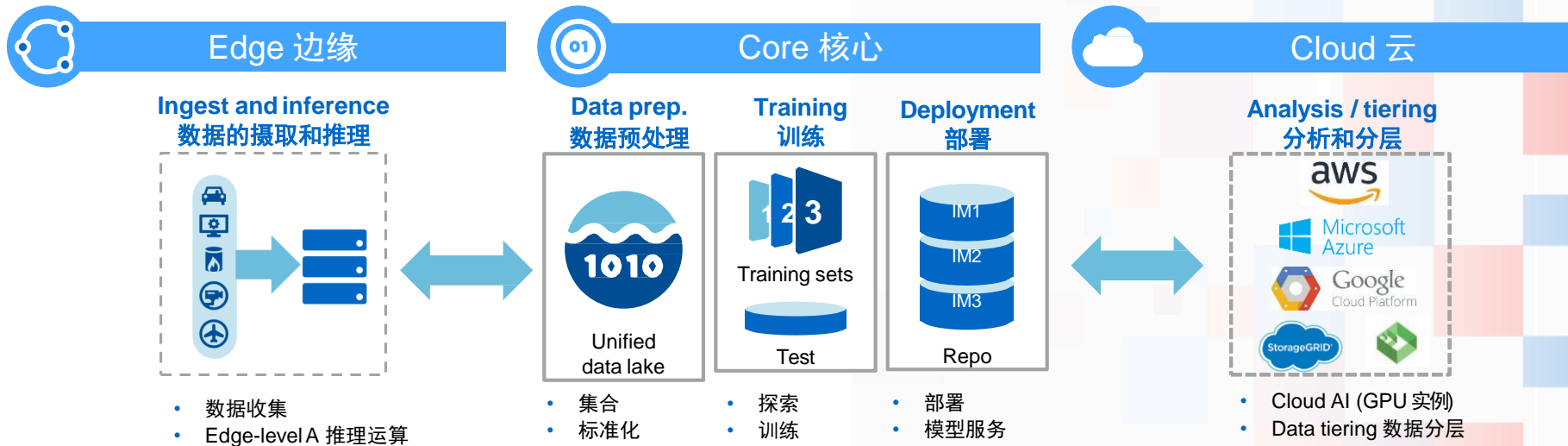
\* 1 AFF = 1 HA pair = 2 Nodes

\*\* 100% Sequential Read

\*\*\* Estimates based on workload testing in NVA-1153

# 从边缘到核心到云

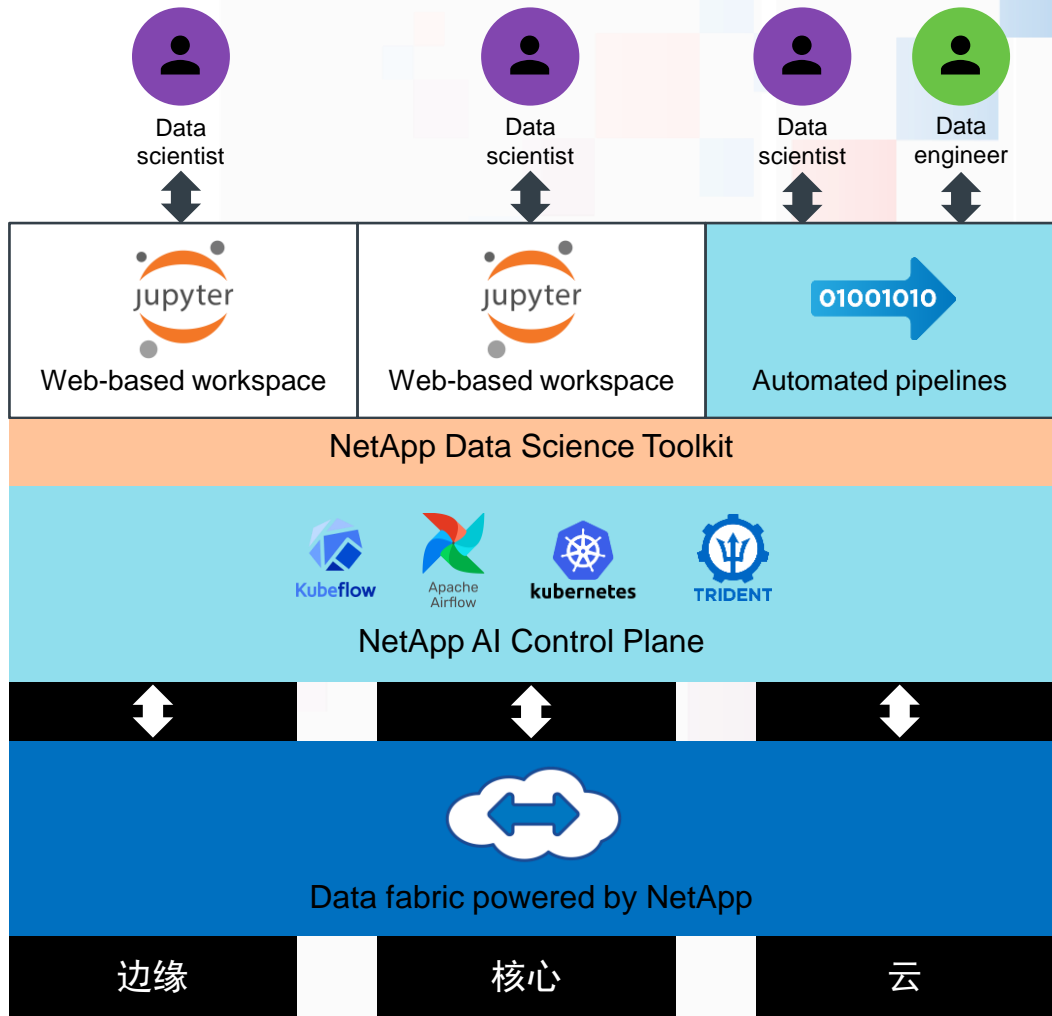
Seamless data management 无缝的数据管理



# NetApp Data Science Tools

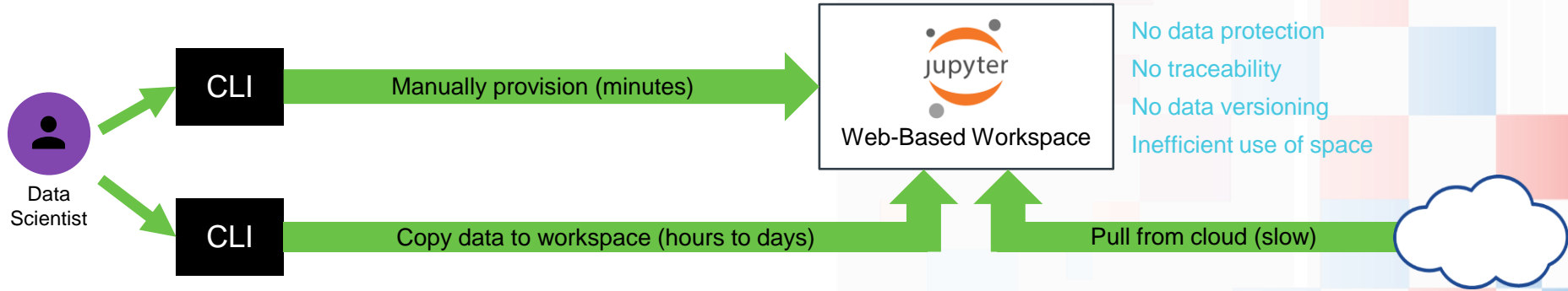
为数据科学家增加价值

- 灵活选择的 Jupyter 工作空间
  - Full access to production datasets
- 流程自动化:数据抽取、数据准备、模型训练和模型部署
- 工作空间和工作负载可横跨边缘、核心和云
  - Choice of any compute and/or cloud
  - Cross-site Data Scientist collaboration
- 嵌入版本管理功能
  - Full dataset to model traceability
  - Seamlessly switch between model versions for dev/test, A/B testing, etc.

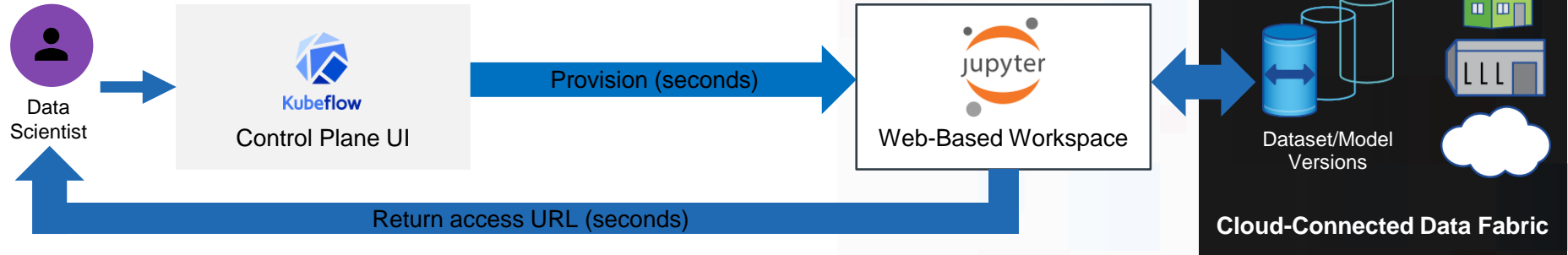


# Data Science Workspace Creation

## Common DIY method

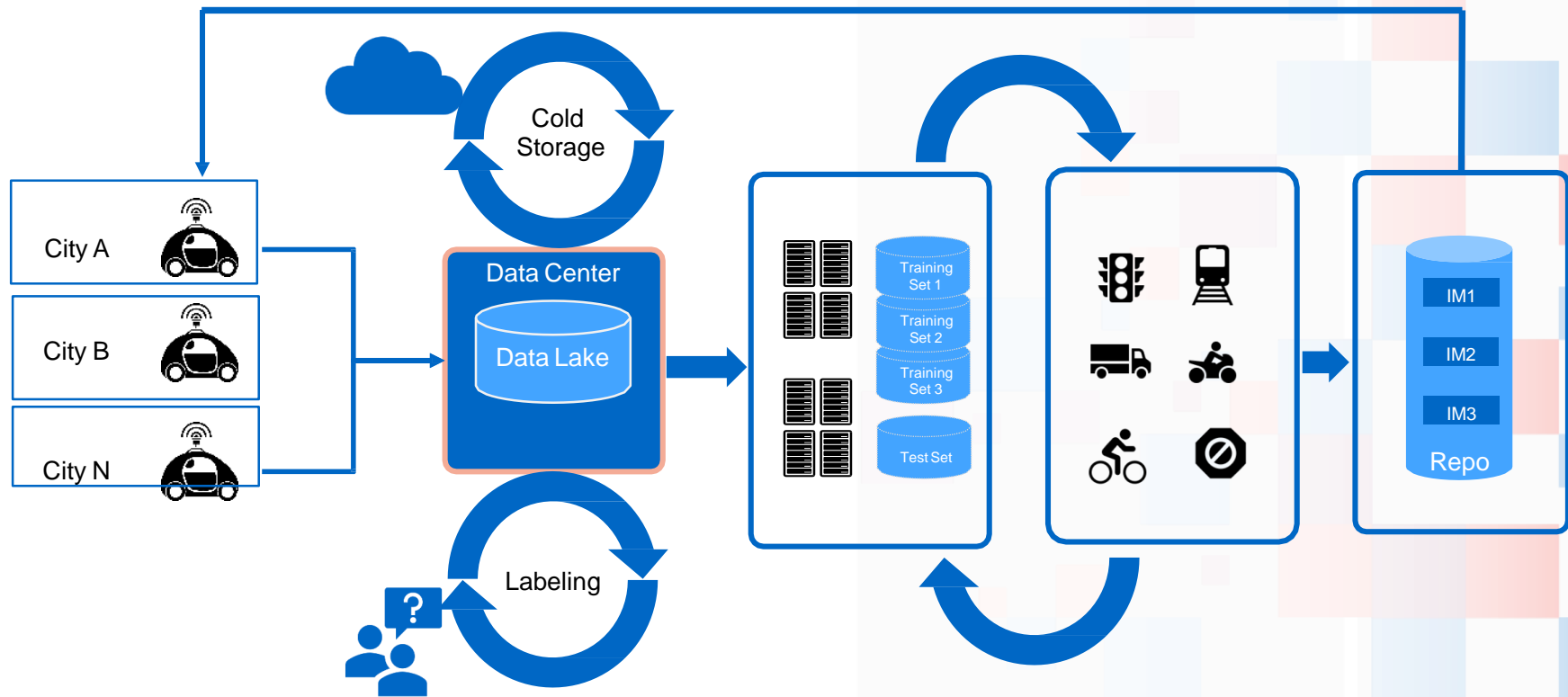


## NetApp AI Control Plane

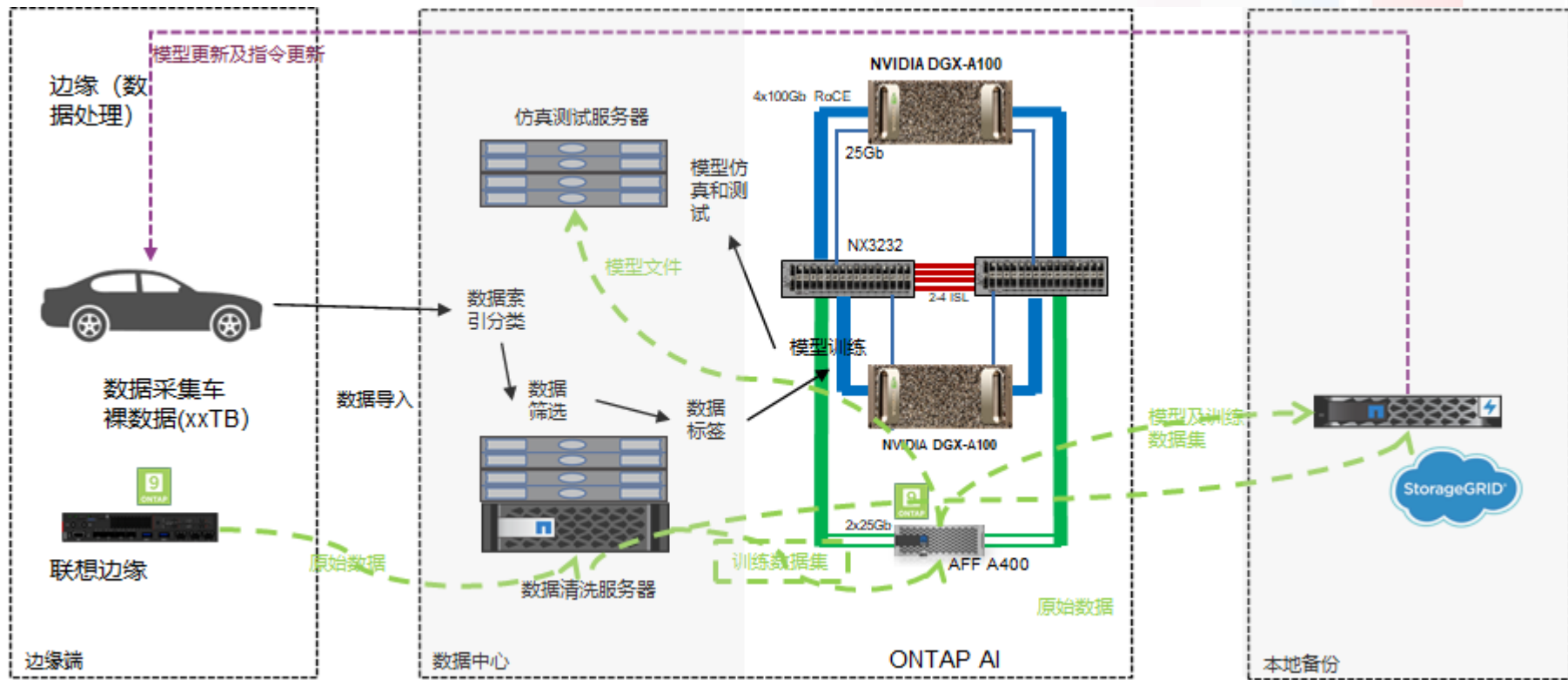




# 无人驾驶数据流



# 某汽车公司案例



# 联想的AI战略

**\$1.2 Billion**

**AI 投资**

“AI 改变世界，规模公司的发展依赖于大数据分析。”



杨元庆  
CEO - Lenovo



## AI Research

提供工具和专家来加速AI创新.



## Enterprise AI Solutions

授权各个组织启动他们的人工智能计划.



## AI Exploration

普及人工智能.

100+ 数据科学家和开发者

AI 研究和创新中心

AI 硬件和软件平台

End-to-end Solutions



IoT

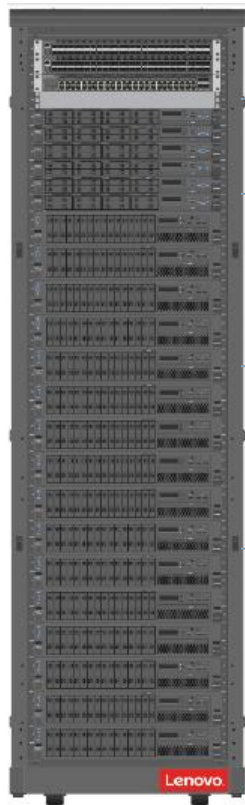


Big Data



AI

# LENOVO AI



网络

管理节点

推理节点



训练节点



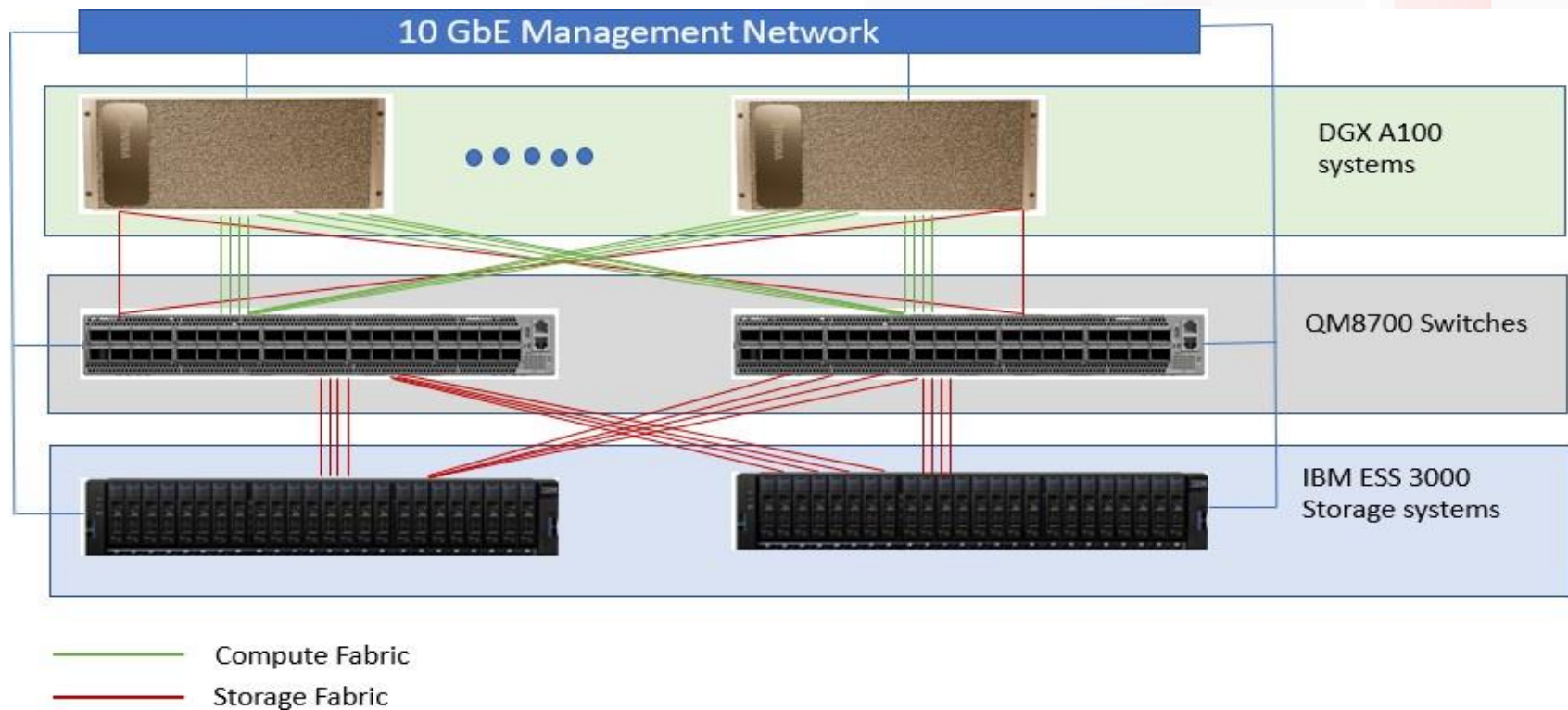
数据节点



DM / DE



# GPFS存储方案



# 专业AI应用平台--炼AI大师



AI专家



数据科学家

视觉

语音

自然语言处理

表格数据

...

工作空间

批处理任务

MLOps流程

用户管理

配额管理

监控告警

任务调度

计费系统

分布式存储



GPU服务器



GPU服务器



CPU服务器



边缘服务器



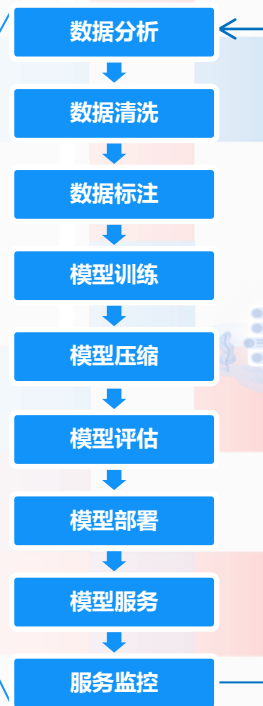
嵌入式设备



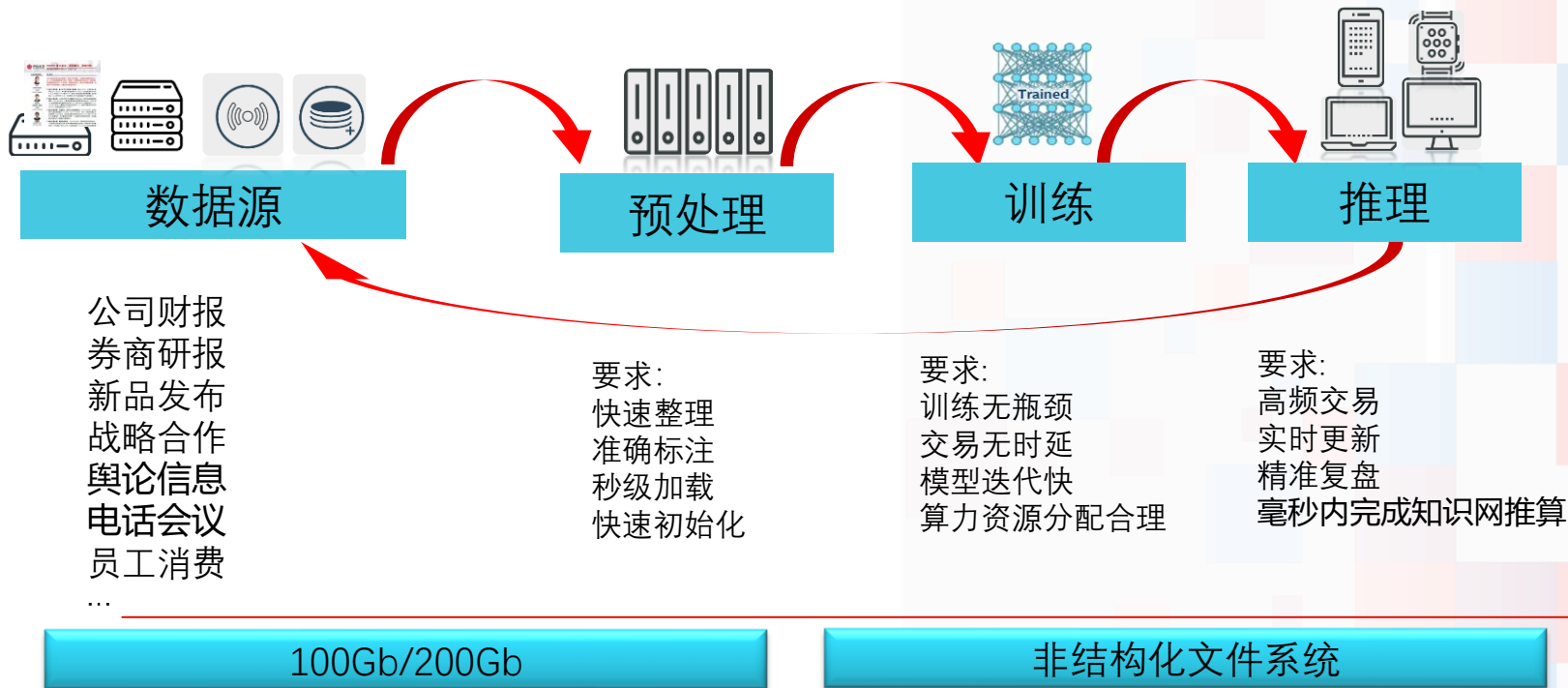
存储设备



网络设备

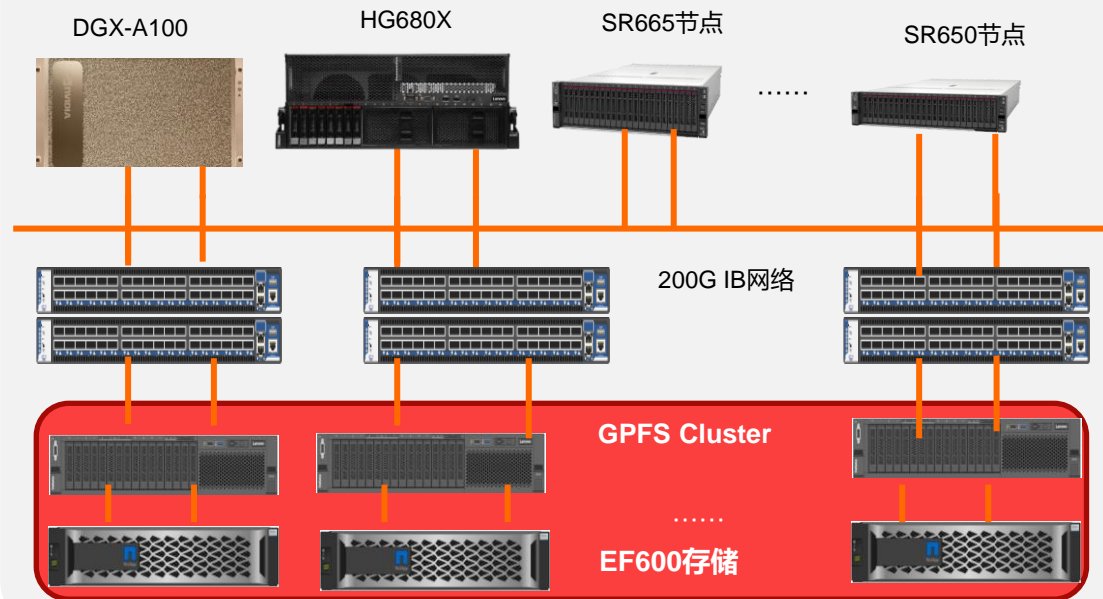


# 量化交易AI workflow



# 某私募公司AI量化平台案例

## “联想大脑” AI调度管理平台



业界第一，单节点带宽44GB/S，横向扩展！

- 构建异构GPU集群，提高算力效率
- 构建无阻塞、低时延网络
- 海量数据秒级读取，横向扩展，
- 日间数据瞬间复盘
- 统一算力资源AI调度平台
- 毫秒内完成知识网推算
- 分层数据存储





# 谢谢!

智慧数据构建智能世界

