NetApp Solutions for ML/DL/AI

NetApp

David Arnette | david.arnette@netapp.com
Principle Technical Marketing Engineer, NetApp AI Solutions
11 October 2021

© 2021 NetApp, Inc. All rights reserved.

Top challenges of AI today

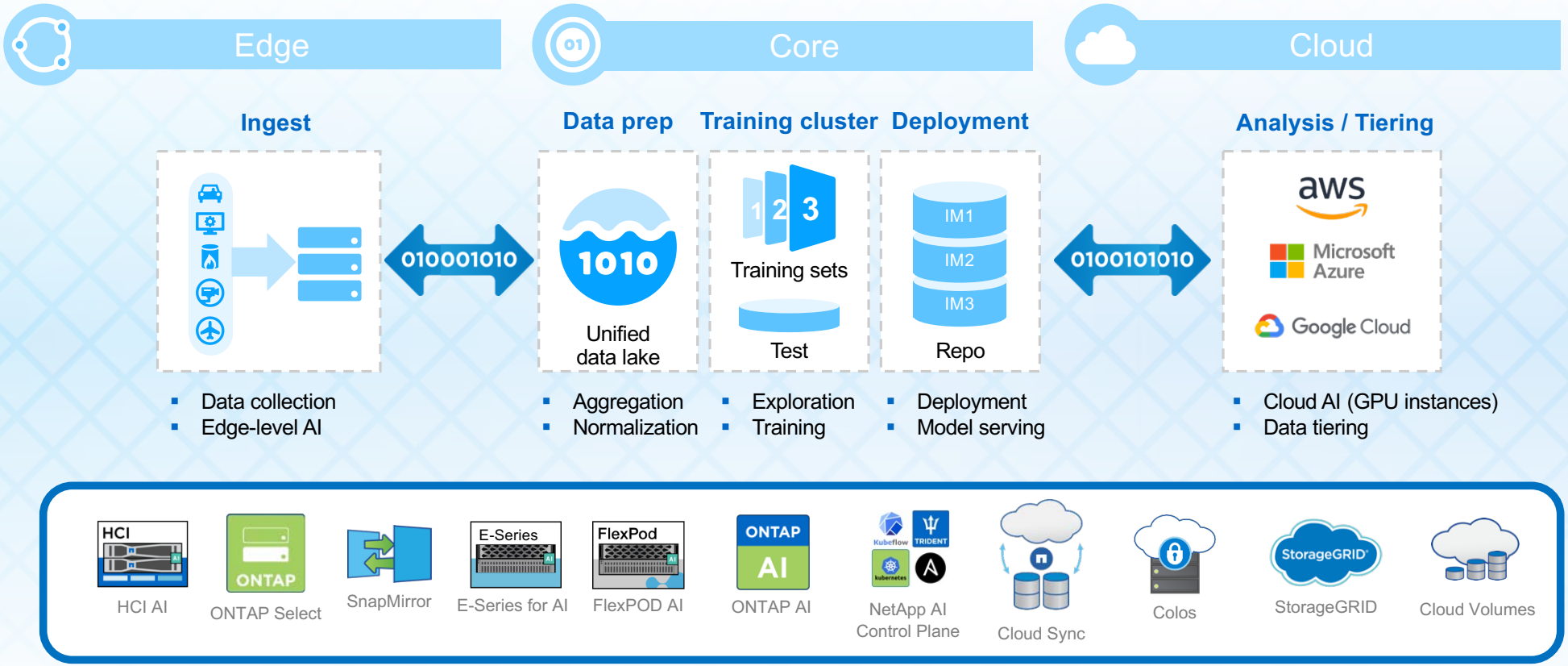
- Expertise
  - Right team, right tools at the right time
  - Maximizing researcher efficiency
- Data
  - Model management and data integrity
  - Data pipeline management
- Technology
  - Building optimized AI PaaS, reporting & tools
  - Performance tuning and managing burst modes
- Infrastructure
  - AI environment configuration
  - Power and Cooling Density

NetApp

© 2021 NetApp, Inc. All rights reserved.

Edge to Core to Cloud Data Pipeline

Strategic Vision: Be the trusted partner and advisor for customers on AI/Data Pipeline

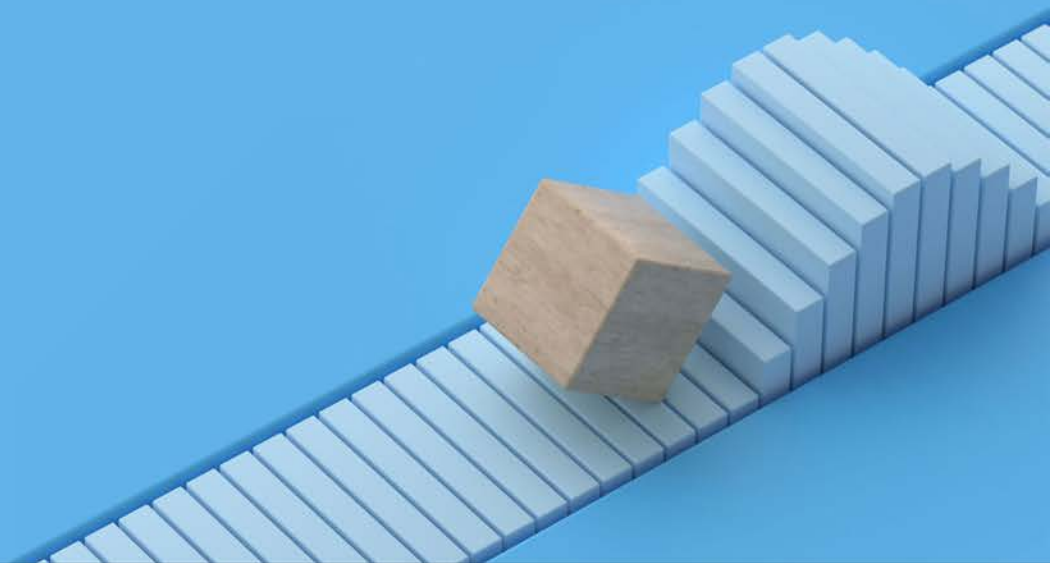


© 2021 NetApp, Inc. All rights reserved.

NetApp

High-performance AI/ML training tier

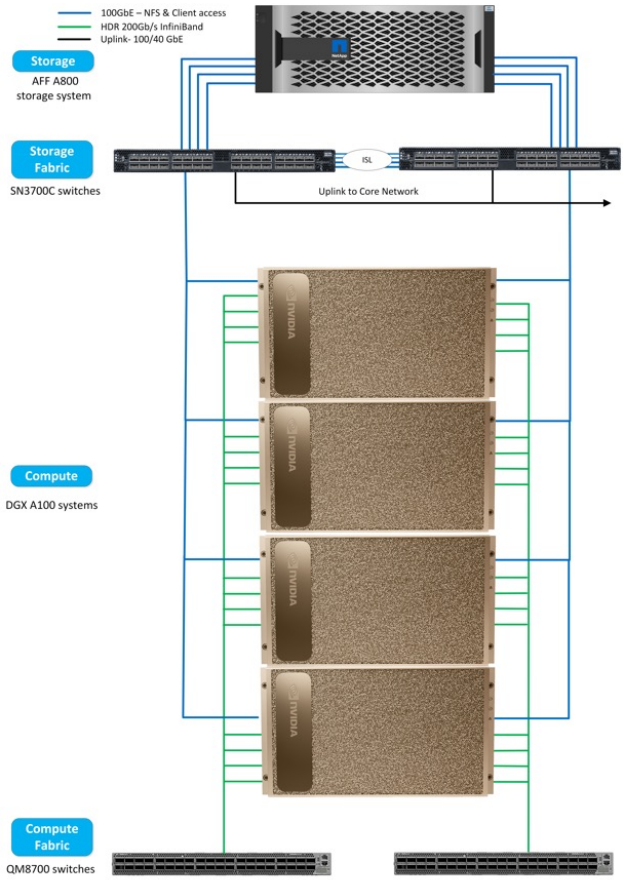
High-performance compute and storage



NetApp

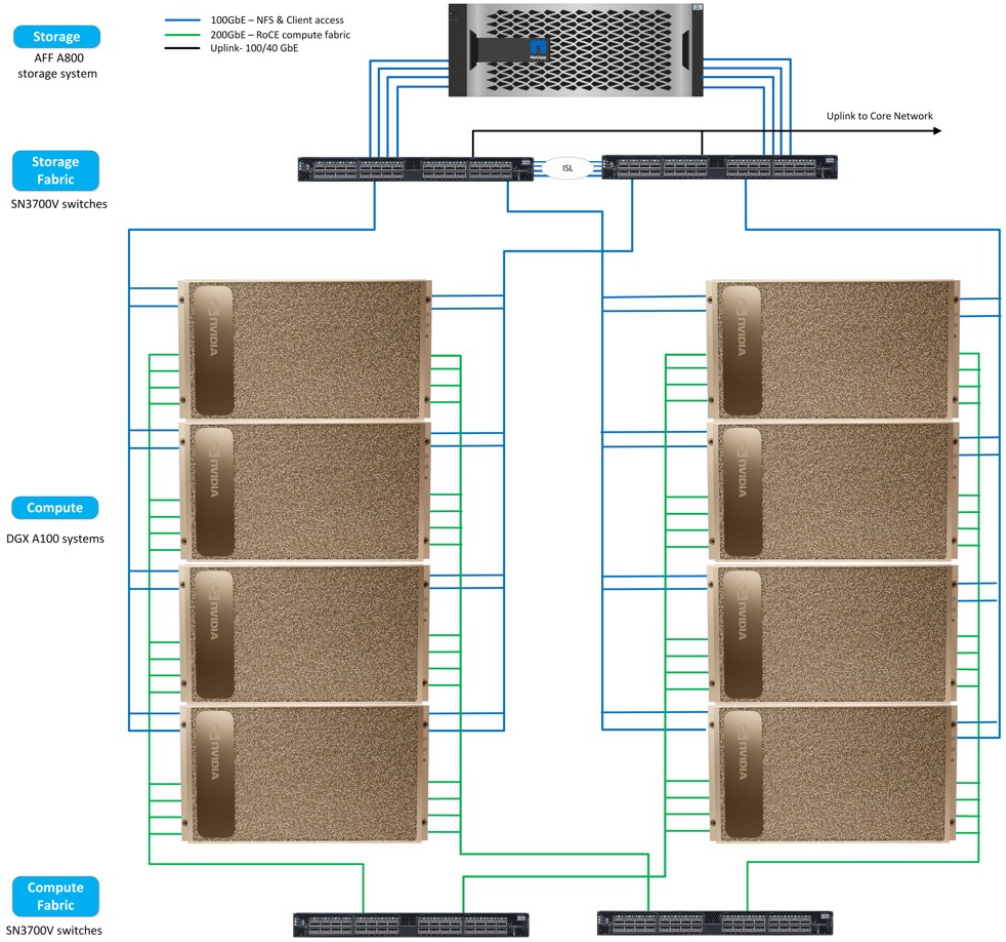
© 2021 NetApp, Inc. All rights reserved.

ONTAP AI Reference Architecture



InfiniBand Compute Fabric

Ethernet Compute Fabric



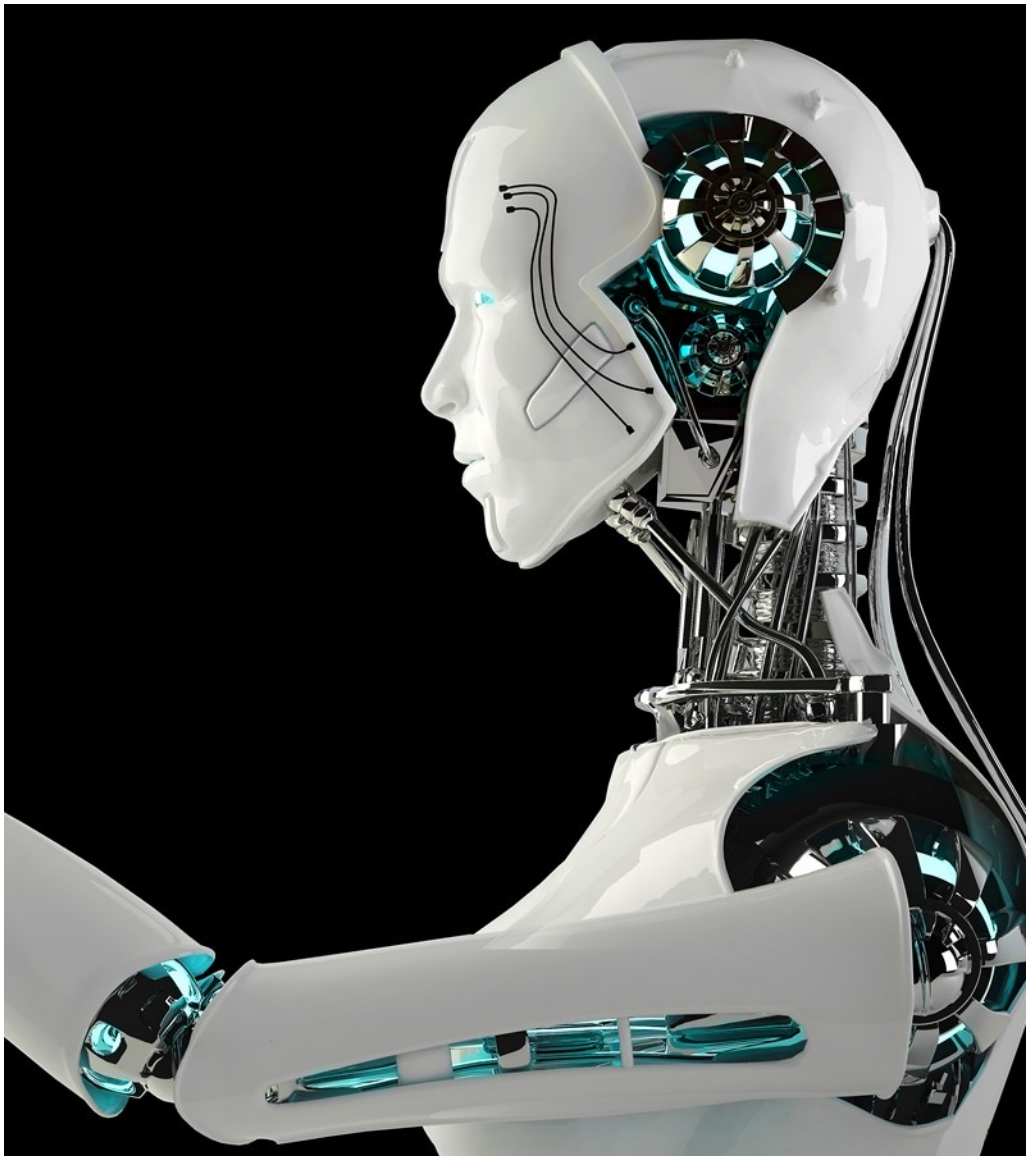
NetApp

© 2021 NetApp, Inc. All rights reserved.

Top reasons to deploy NetApp ONTAP AI

Go faster, spend less, better results…

1) Simplify deployment of AI environments
2) Save up to 70% over cloud-based deployments
3) Solve data center power and cooling challenges
4) BYOT – Bring your own tools
5) 50% Faster than DIY AI



NetApp

© 2021 NetApp, Inc. All rights reserved.

NetApp ONTAP AI integrated solution

Preconfigured offering from NetApp and NVIDIA for optimal performance

Field proven AI
Simplify procurement, configuration, installation, and support with validated best practices.

+

Full configuration
Validated solution from NVIDIA and NetApp® configured and installed by Arrow

+

Effortless support
Complete solution supported by NVIDIA, backed by NetApp

Predefined, configured, delivered, installed, optimized, and supported

Available machine learning operations environment from Domino Data Lab

NetApp

© 2021 NetApp, Inc. All rights reserved.

Three available configurations, each with two capacity options

	Small		Medium		Large	
Compute	2-node NVIDIA DGX A100		4-node NVIDIA DGX A100		8-node NVIDIA DGX A100	
Storage	NetApp® AFF A400		NetApp AFF A700		NetApp AFF A800	
Capacity	65TB	131TB	131TB	263TB	263TB	526TB

NetApp

© 2021 NetApp, Inc. All rights reserved.

EF-Series with BeeGFS and DGX A100

4-Node Configuration

- Validated architecture with building blocks
- Simple to size, setup, and scale up & out
- Decouple compute & storage
- Maximize GPU utilization with GPUDirect Storage
- Support for the whole storage solution

DGX A100 Servers

8x A100 SXM2 GPUs –

55k CUDA cores, 3456 TensorCores – Up to 5 PFLOPs FP16

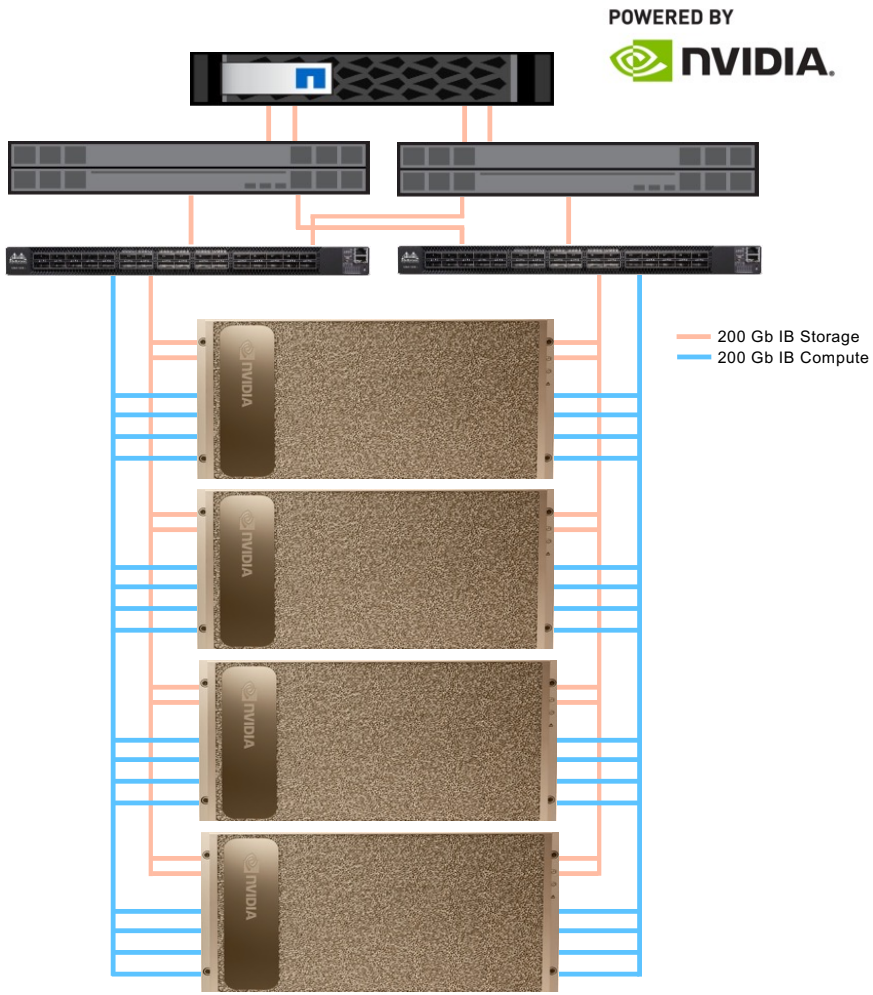
600GB/s NVLink

8x HDR IB ports, 2x 100Gb Eth ports

BeeGFS Building Block

1x HA Pair of x86 servers

2x EF600 200 Gb NVMe/IB, 24 NVMe SSDs

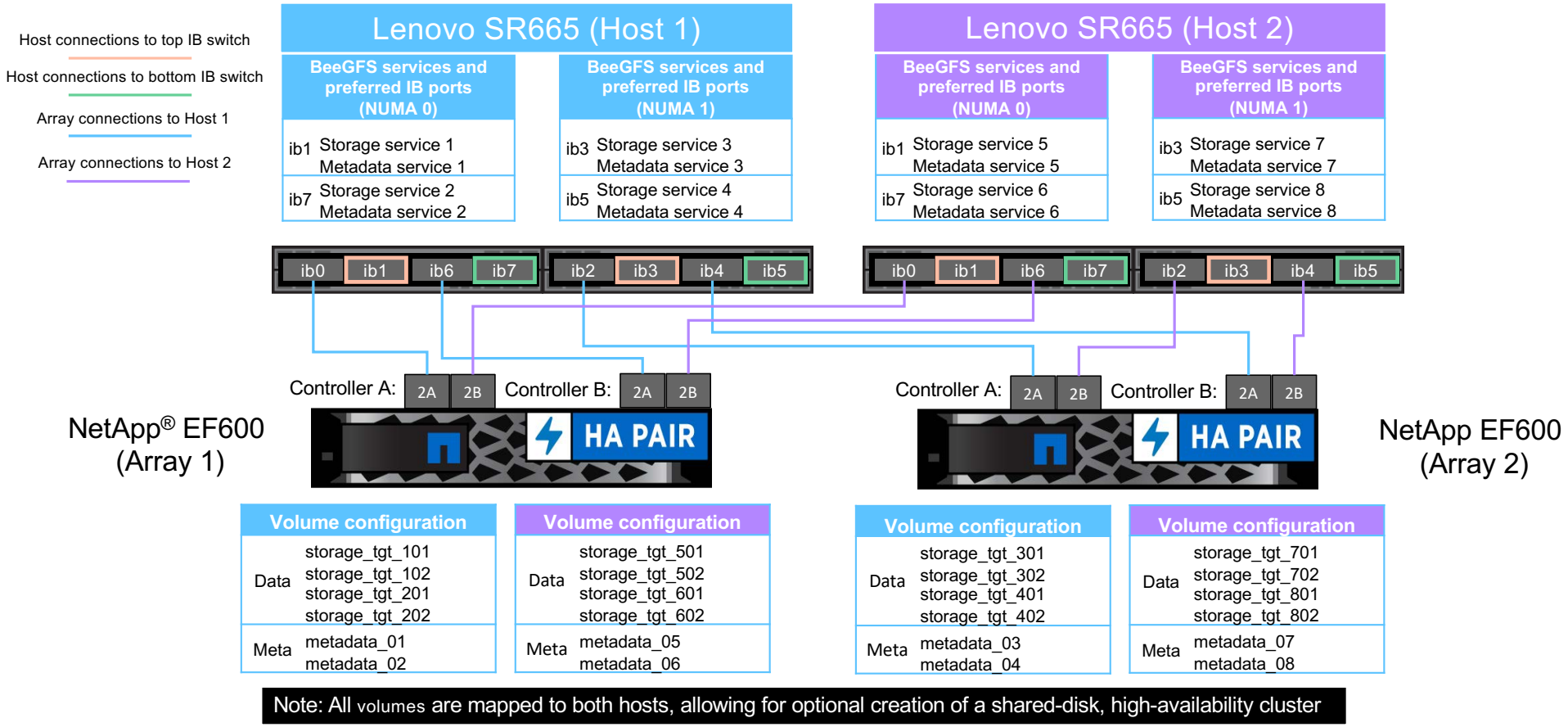


© 2021 NetApp, Inc. All rights reserved.

NetApp

Gen2 Building-block Details

This configuration will be replicated for BeeGFS building block

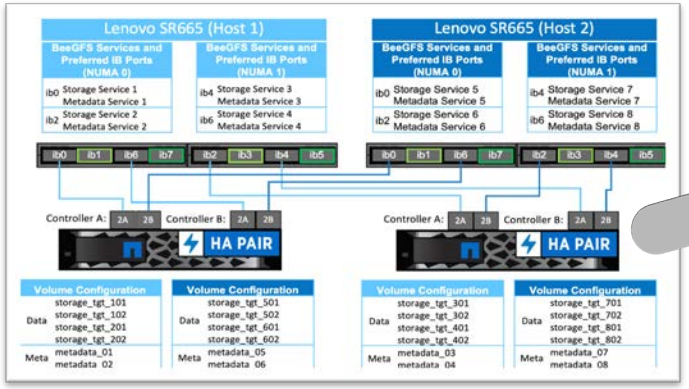


Note: All volumes are mapped to both hosts, allowing for optional creation of a shared-disk, high-availability cluster

 © 2021 NetApp, Inc. All rights reserved. — NETAPP CONFIDENTIAL —

Storage at Massive Scale

NetApp BeeGFS solution with EF-Series

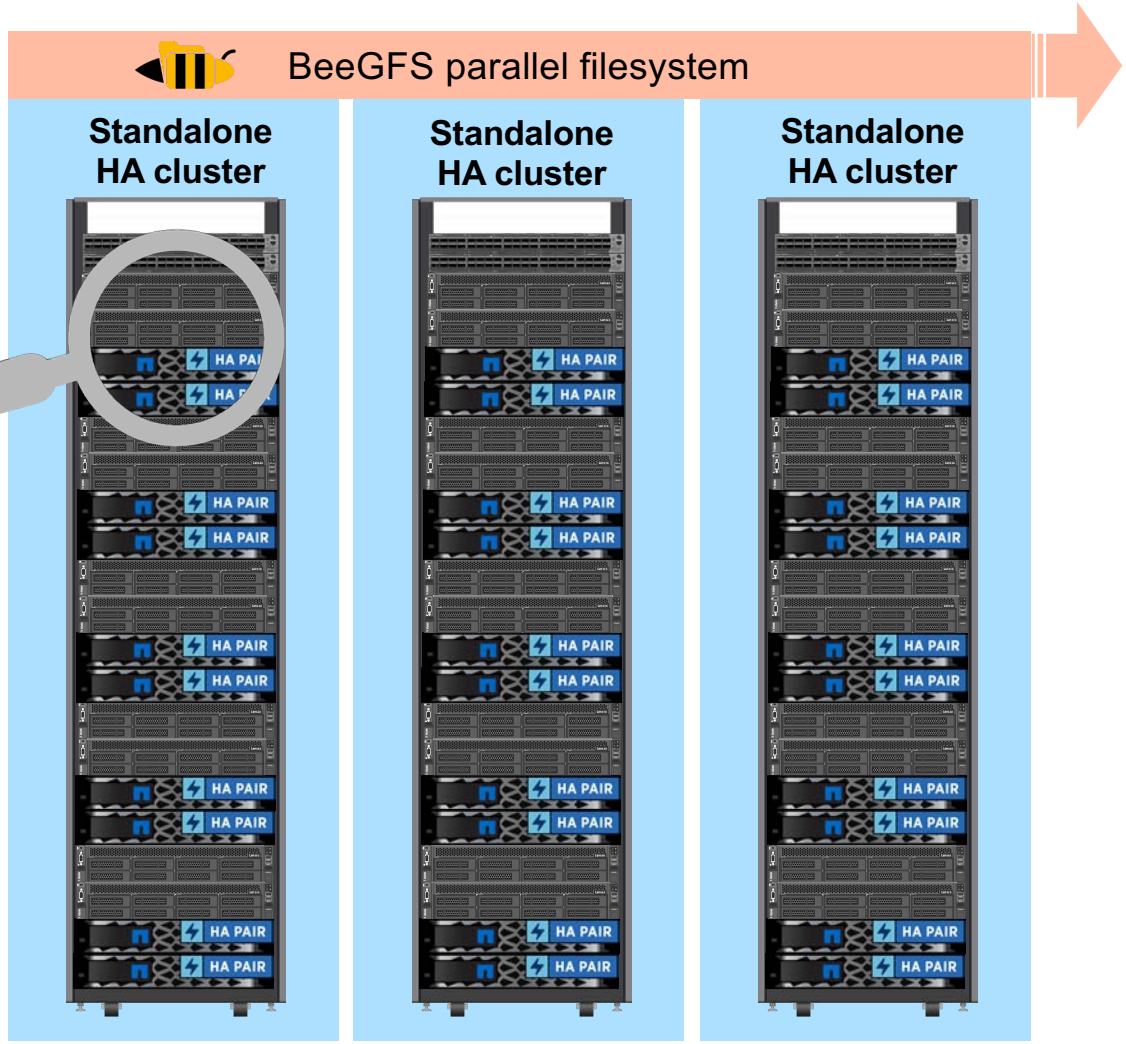


Ansible: To deploy/manage at scale

Grafana/BeeGFS Mon: For monitoring

Pacemaker/Corosync: For high availability

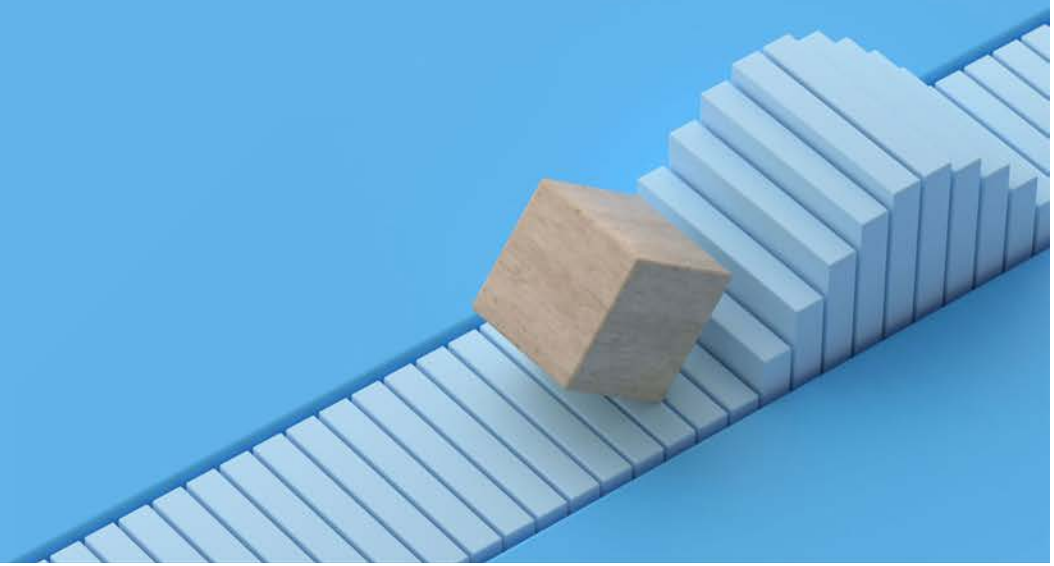
BeeGFS CSI driver: For cloud-native integration



 © 2021 NetApp, Inc. All rights reserved. — NETAPP CONFIDENTIAL —

BaseCommand Platform

An Optimized AI Experience



NetApp

© 2021 NetApp, Inc. All rights reserved.

Saturn V….

How NVIDIA does AI

- NVIDIA's AI Platform-as-a-Service
- Built By NVIDIA for their AI teams
- Used by over 1200 NVIDIA researchers
- Over 5 years of design & optimization
- Predictable performance at scale
- Optimized environments and validated tools
- Retail, Medical, Genomics, Gaming…

NetApp © 2021 NetApp, Inc. All rights reserved.

What is BaseCommand Platform?

A new partnership between NVIDIA and NetApp

AI Expertise from NVIDIA & NetApp
NVIDIA shares own their AI platform & practices

Your Pods, Our Teams & NVIDIA's managed datacenter
Controlled tenant and prebuilt data center power and cooling

Best of Breed AI Tools and Services
optimized tools for AI models, data management & inferencing

Business Results, Reporting and ROI
Maximize your researchers & lower costs

The leaders in AI will help you deploy the best AIOPS for your business

NetApp © 2021 NetApp, Inc. All rights reserved.

BaseCommand Platform– simplicity, performance and value

Control your environment, know your costs and protect your assets

Faster & Easier than On-Prem

- Prebuilt Power and Cooling for AI environments
- Pre-specified configs accelerate “Time-to-Model”
- Expert deployment team saves time
- Pre-configured PaaS tools and curated tools
- Dedicated AI environment for your business
- Lower operating costs
- Constant performance optimization
- NetApp Data Fabric management services
- NetApp AI Control Plane

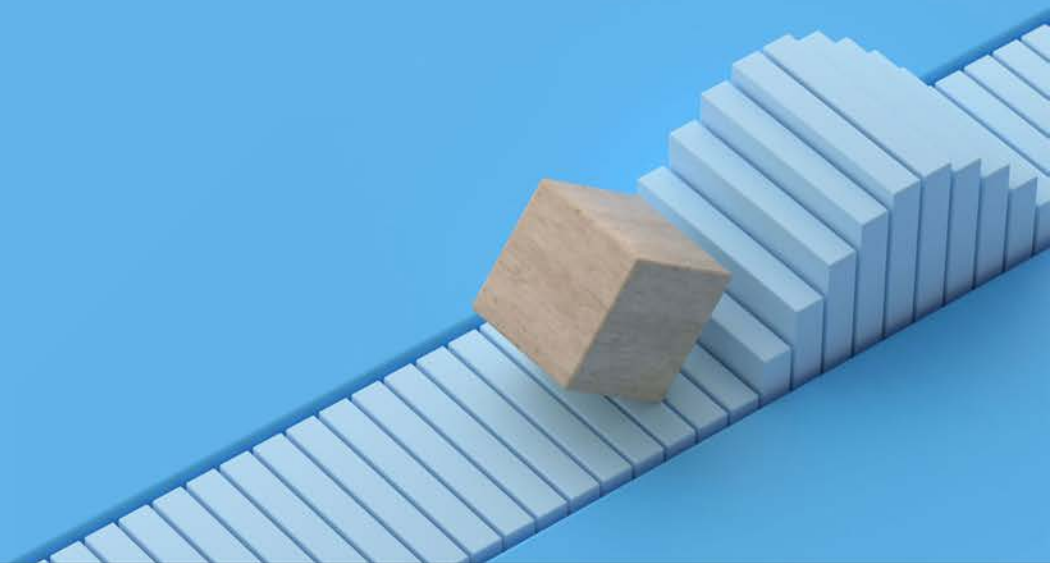
More Cost Effective than Cloud

- Go faster with NVIDIA PaaS and curated tools
- More efficient AIOPS with Multi-Instance GPUs
- Reduce licensing costs
- Eliminates the ”Stew of Services”
- Predictable Infrastructure costs for AI
- More predicable ROI
- No egress fees for storage
- Expert deployment team saves time
- NetApp Data Fabric and AI control plane optimization

NetApp © 2021 NetApp, Inc. All rights reserved.

Object storage data lake

Flexibility and manageability



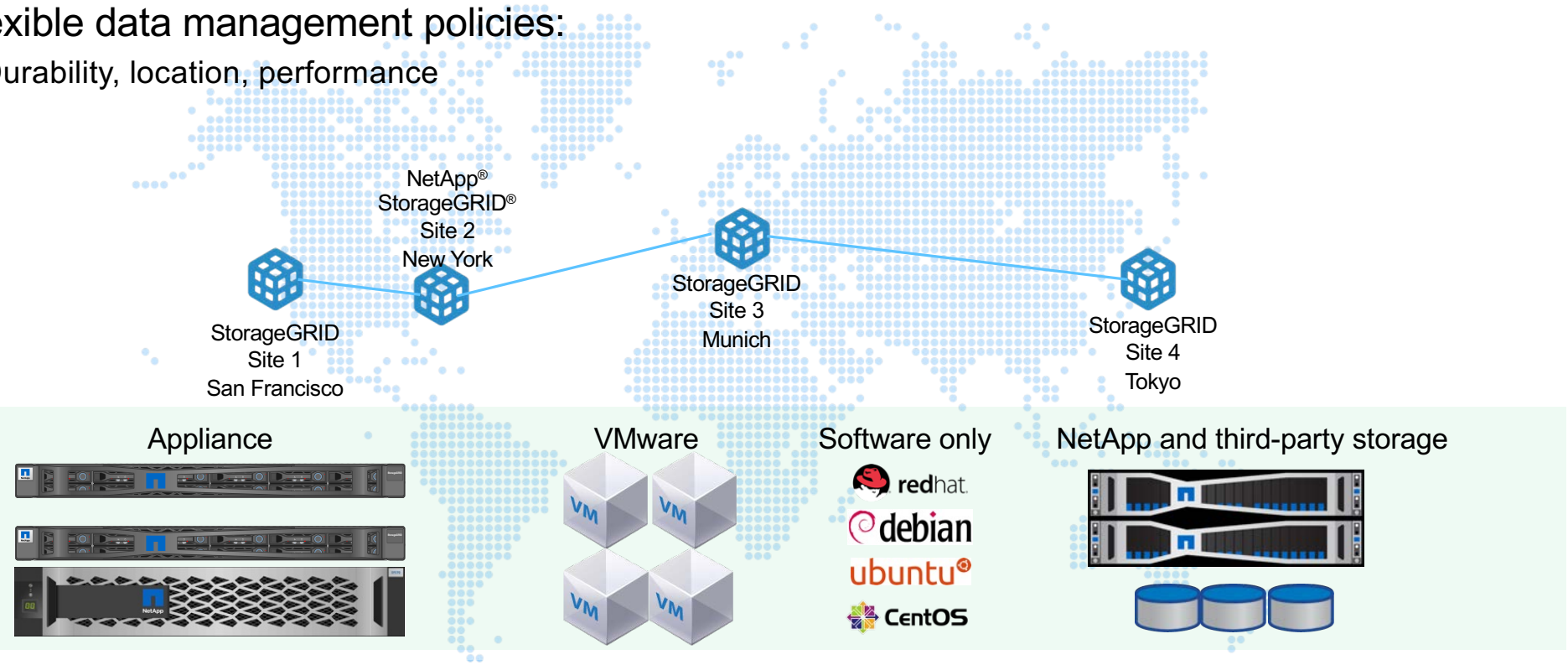
NetApp

© 2021 NetApp, Inc. All rights reserved.

What Is NetApp StorageGRID?

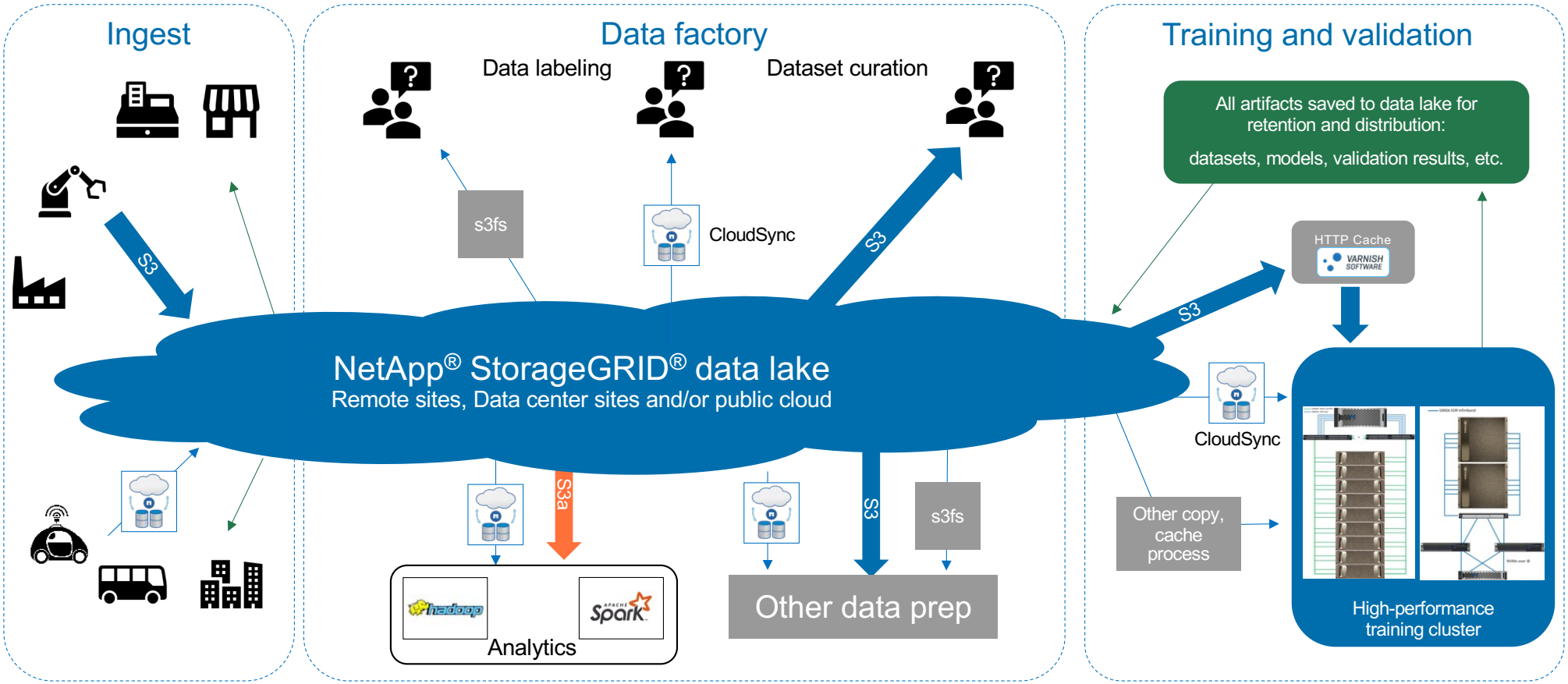
Software-defined data management for unstructured data

- Flexible data management policies:
  - Durability, location, performance



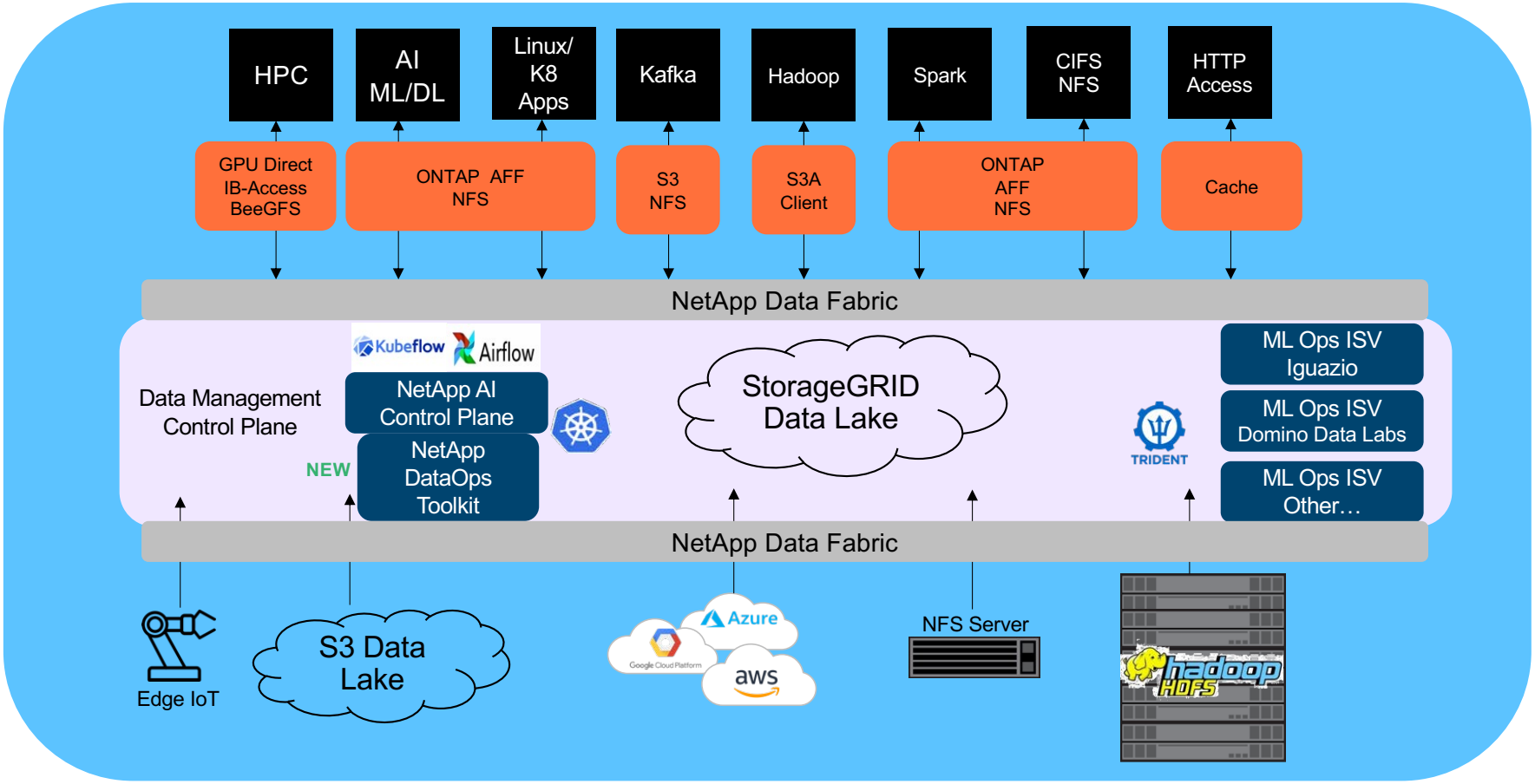
© 2021 NetApp, Inc. All rights reserved.

Many Workflows, Many Workloads



NetApp © 2021 NetApp, Inc. All rights reserved.

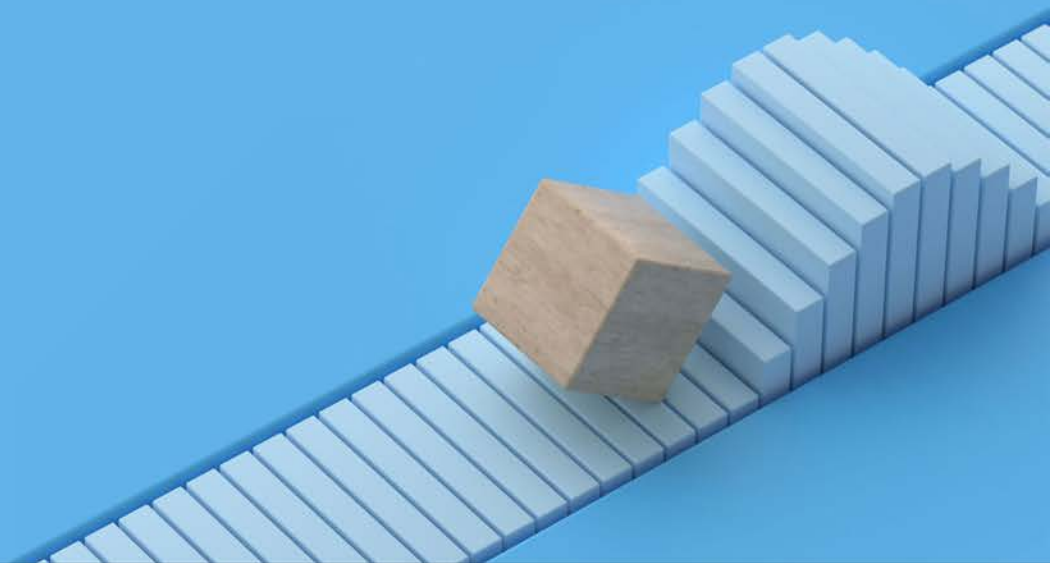
Data Modernization with NetApp Data Management Control Plane



NetApp © 2021 NetApp, Inc. All rights reserved.

Moving the data

Build your data pipeline

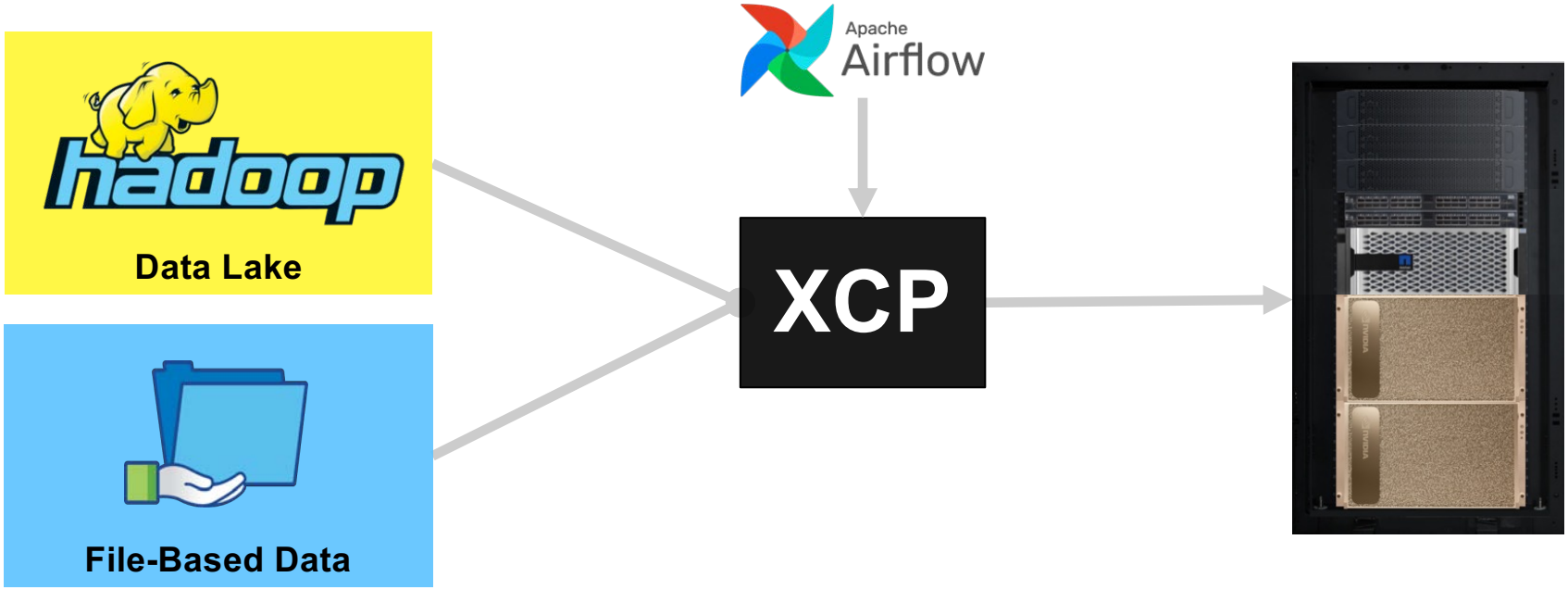


NetApp

© 2021 NetApp, Inc. All rights reserved.

Ultra-high-performance Data Replication

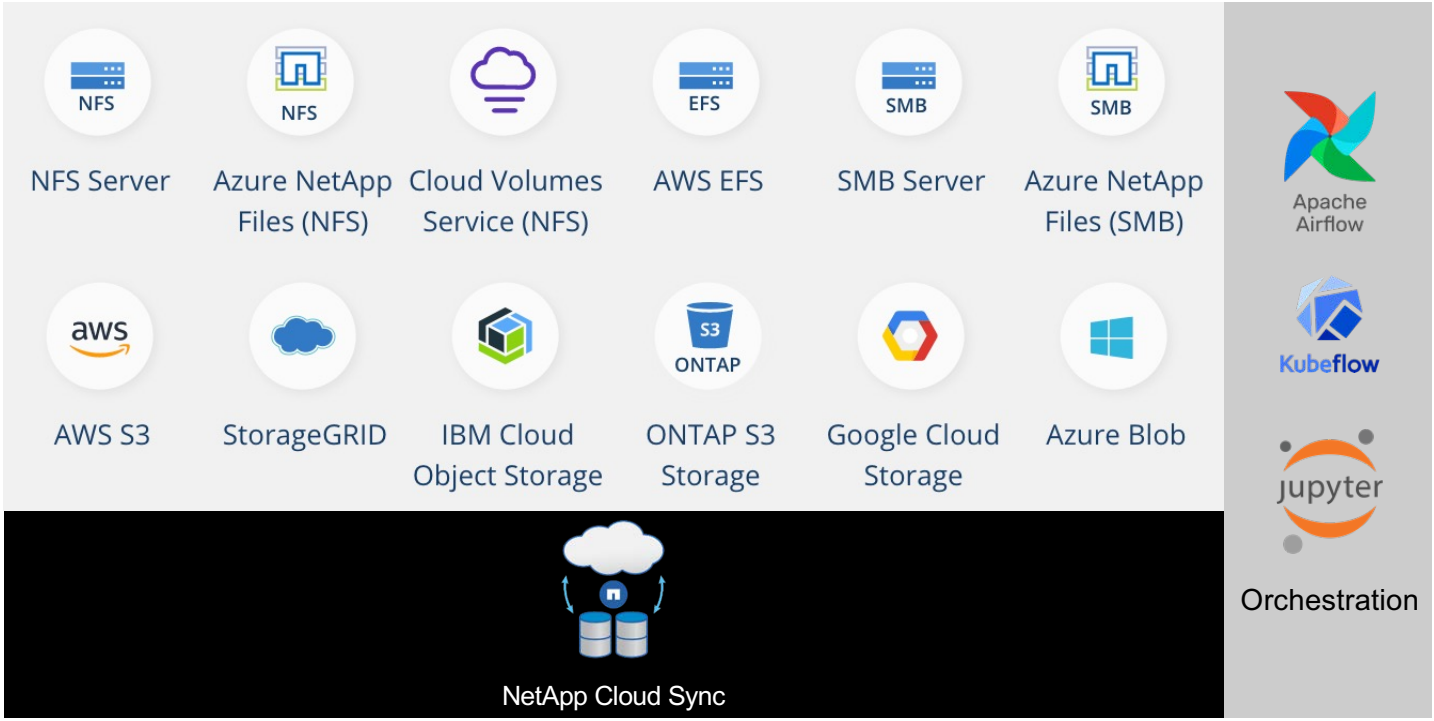
NetApp XCP



NetApp © 2021 NetApp, Inc. All rights reserved.

Unified Data Pipeline

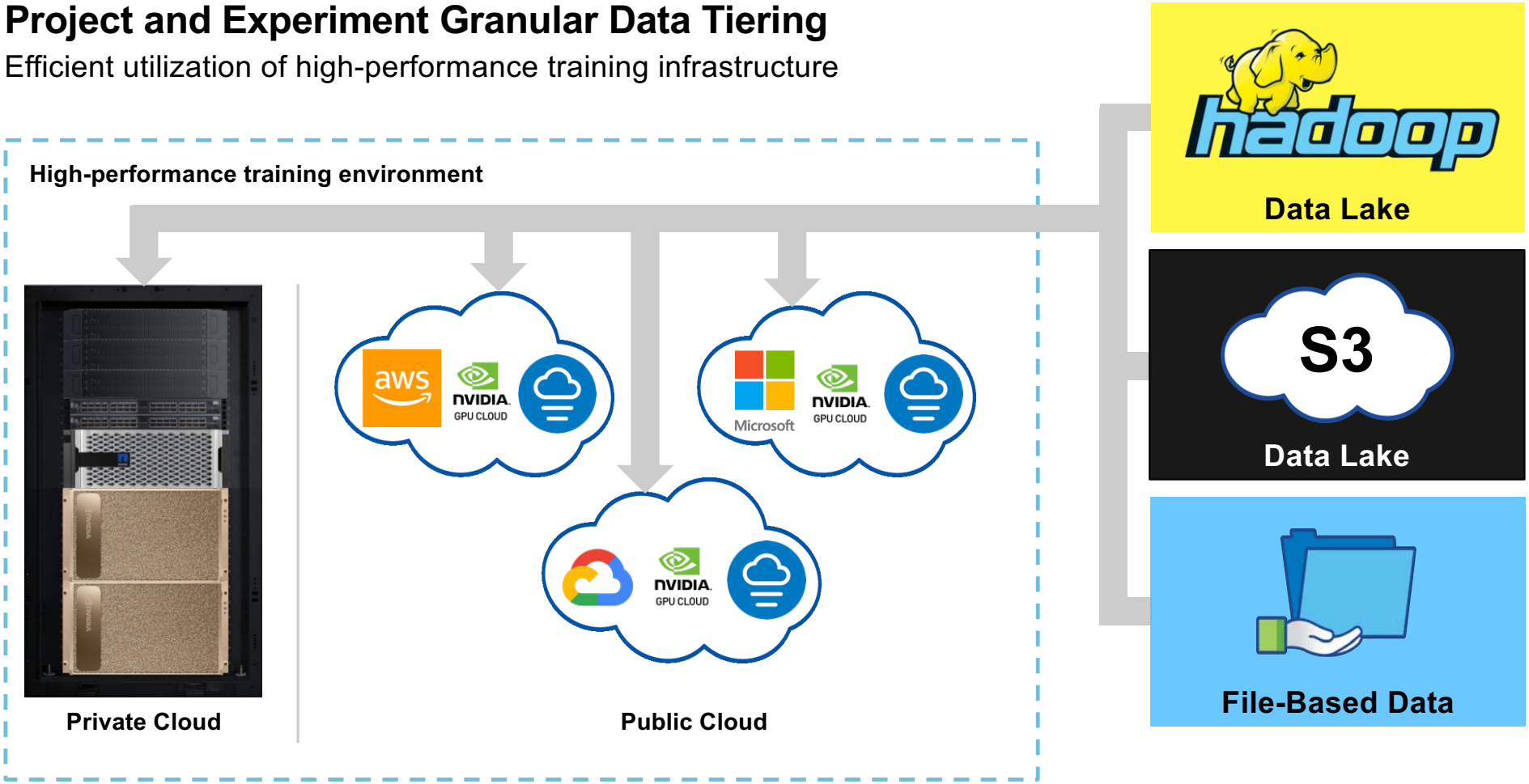
Integrate heterogeneous data sources, centrally orchestrated



© 2021 NetApp, Inc. All rights reserved.

Project and Experiment Granular Data Tiering

Efficient utilization of high-performance training infrastructure



NetApp © 2021 NetApp, Inc. All rights reserved.

NetApp DataOps Toolkit

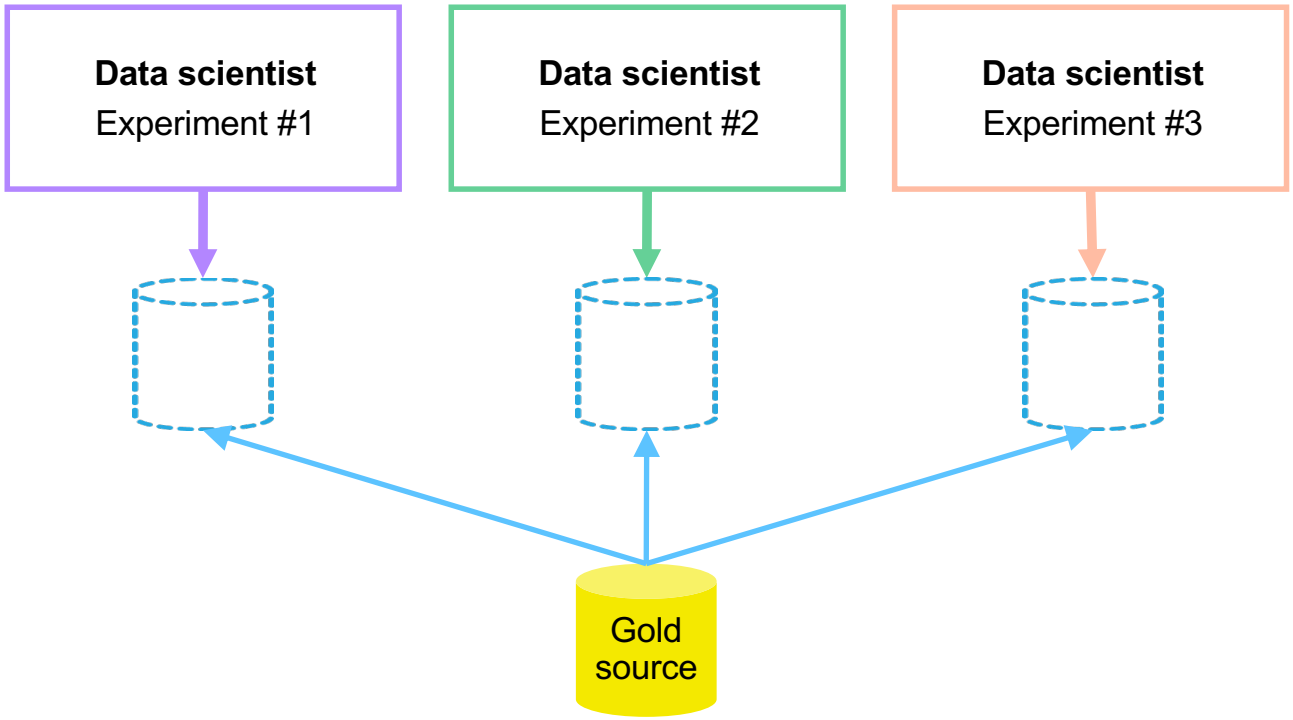
Simplifying AI data management

NetApp

© 2021 NetApp, Inc. All rights reserved.

Training accurate models

Rapid experimentation is necessary



NetApp © 2021 NetApp, Inc. All rights reserved.

2 seconds

vs. 17 hours

Accelerating the AI training workflow, efficiently

- Copying a dataset takes seconds instead of hours/days
  - Ex. 10 TB dataset – 2 seconds vs. up to 17 hours
- Copies are stored efficiently
  - Ex. 10 copies of each dataset – up to 90% reduction in storage space

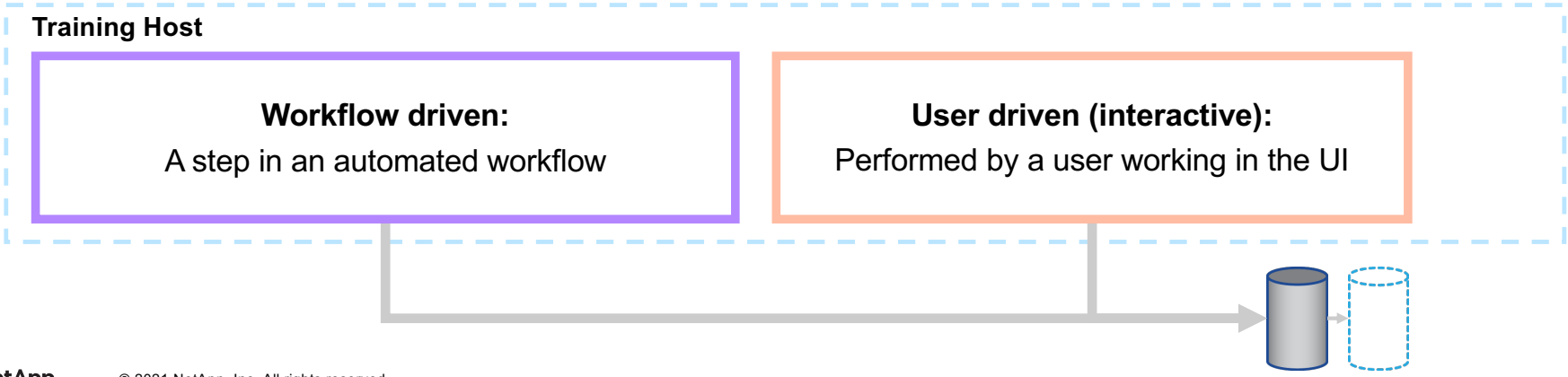
NetApp

© 2021 NetApp, Inc. All rights reserved.

Explainable AI

Traceability is mandatory

- The challenge:
  - As AI moves into highly regulated/sensitive industries – dataset-to-model traceability and reproducibility is a ***mandatory requirement***
  - Traditional versioning tools (e.g. Git) are designed for code, not data
- The solution:
  - Utilize existing, proven tech: ***volume snapshots***–long and proven history of being used to ***quickly and efficiently version datasets***
  - Snapshot name can be saved in code or model repository to implement traceability

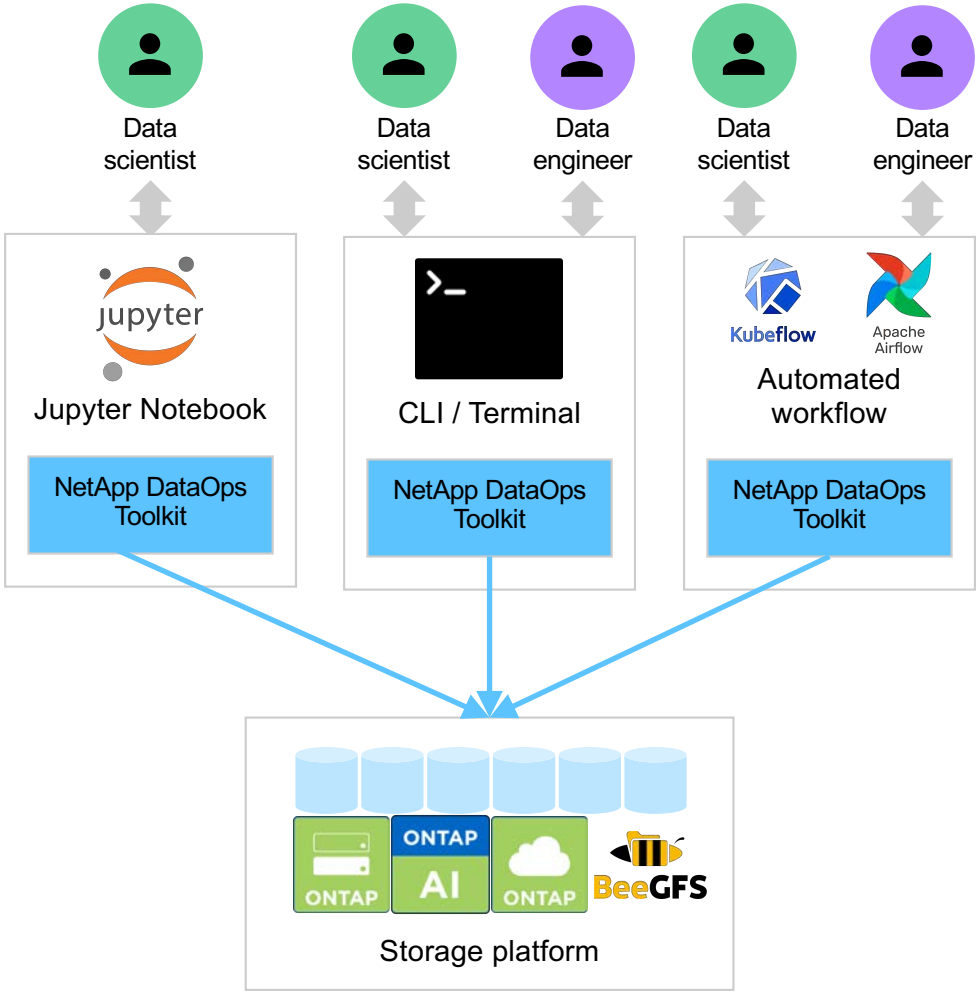


NetApp © 2021 NetApp, Inc. All rights reserved.

NetApp DataOps Toolkit

Simplifying AI data management

- NetApp's industry-leading, multitenant data management capabilities
  - Traditional toolkit: Supports ONTAP (AFF, FAS, Cloud, Select)
  - Kubernetes toolkit: Supports ONTAP (AFF, FAS, Cloud, Select), Azure NetApp Files (ANF), Cloud Volumes Service (CVS), and BeeGFS (limited)
- Simple, easy-to-use interfaces; designed for data scientists and data engineers
  - CLI utility
  - Importable library of Python functions
- Provides access to advanced features that would normally require help from storage admin
- Key capabilities
  - Rapidly provision a new data volume
  - Near-instantaneously clone a data volume
  - Snapshot a data volume for traceability/versioning
  - Trigger data sync



NetApp © 2021 NetApp, Inc. All rights reserved.

NetApp DataOps Toolkit for Kubernetes

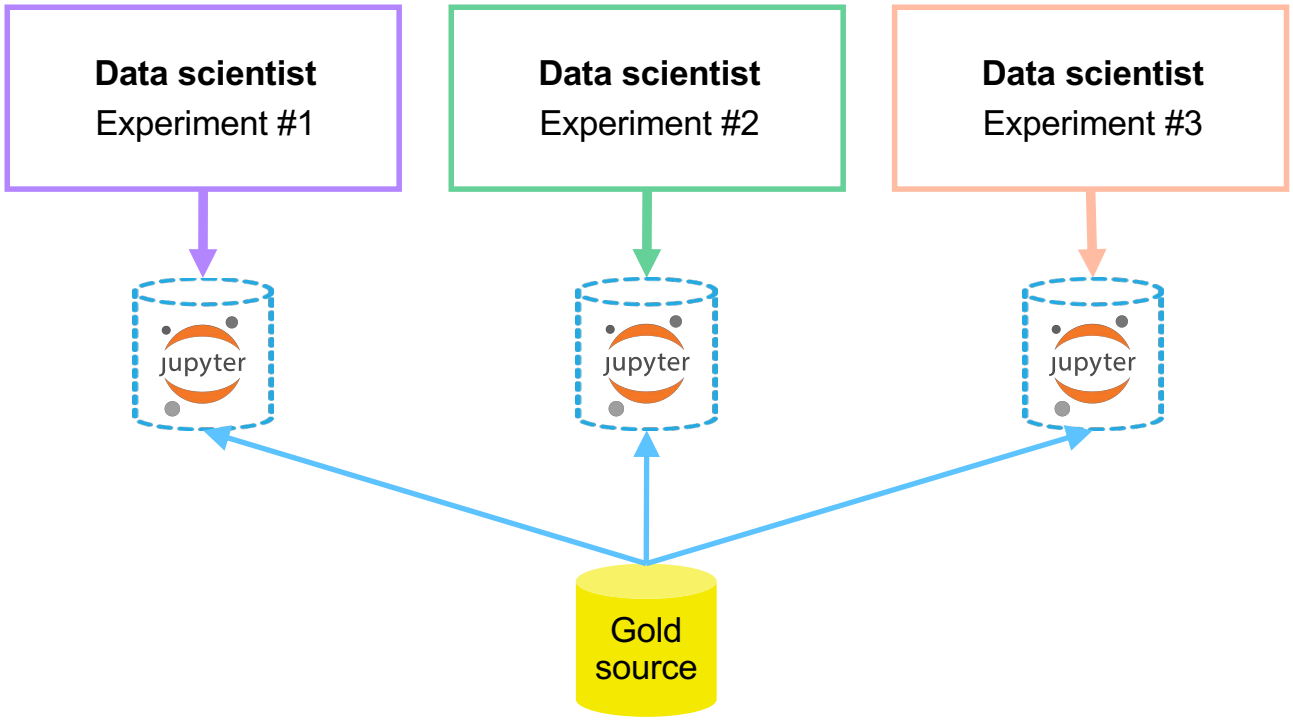
Manage workspaces, not data volumes

- Kubernetes-native version of toolkit…
  - Supports upstream Kubernetes and OpenShift
- Provision/clone/snapshot/delete JupyterLab workspaces
  - Use cases: Rapidly provision/clone/delete data scientist workspaces; rapidly snapshot data scientist workspaces for backup/traceability
- Provision/clone/snapshot/delete Kubernetes PVCs
  - Use cases: Rapidly provision/clone/delete Kubernetes-native persistent data volumes; rapidly snapshot Kubernetes-native persistent data volumes for backup/traceability

NetApp © 2021 NetApp, Inc. All rights reserved.

Driving rapid experimentation

Clone workspaces and data together

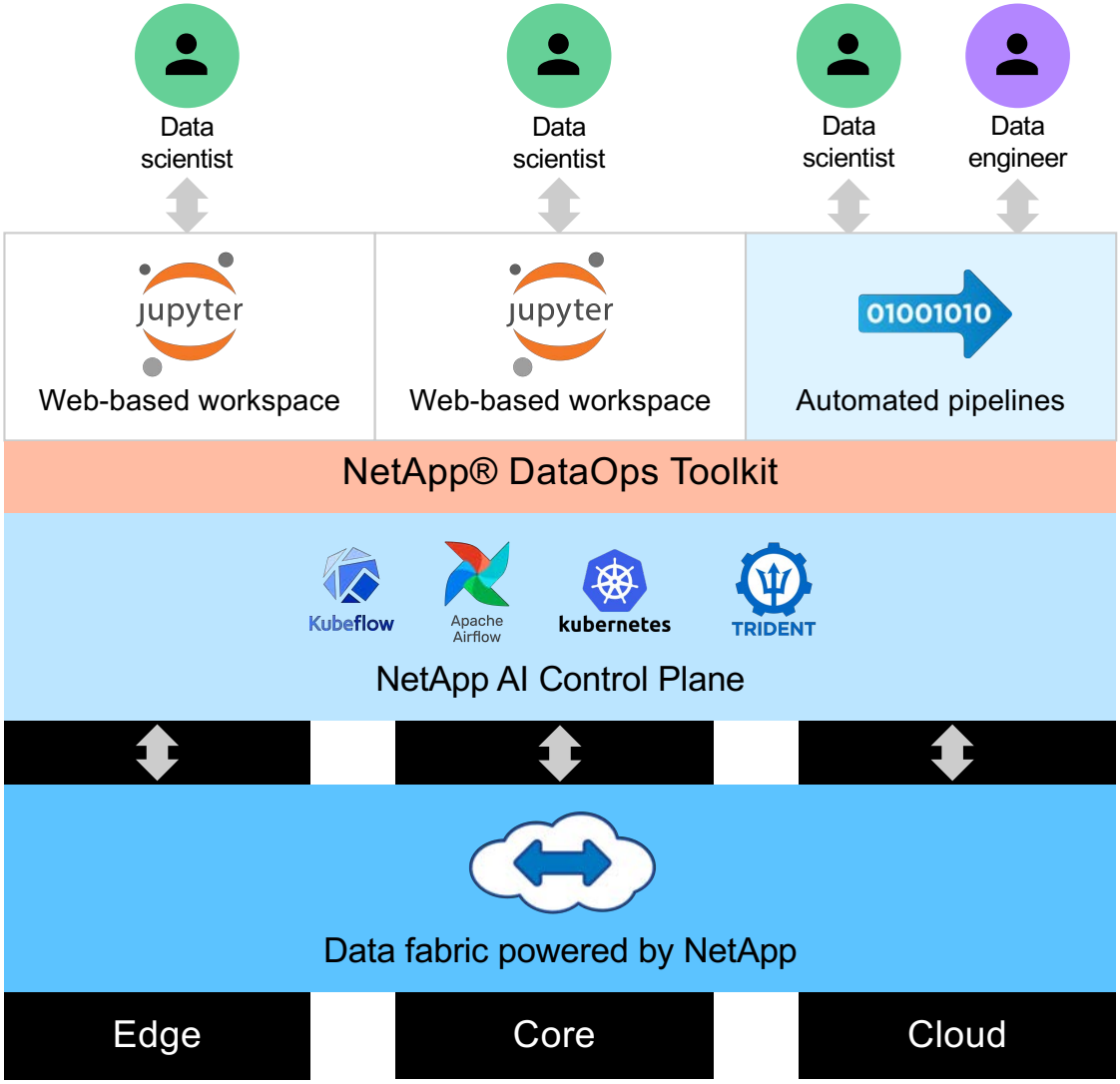


NetApp © 2021 NetApp, Inc. All rights reserved.

Use cases

Multiple deployment options

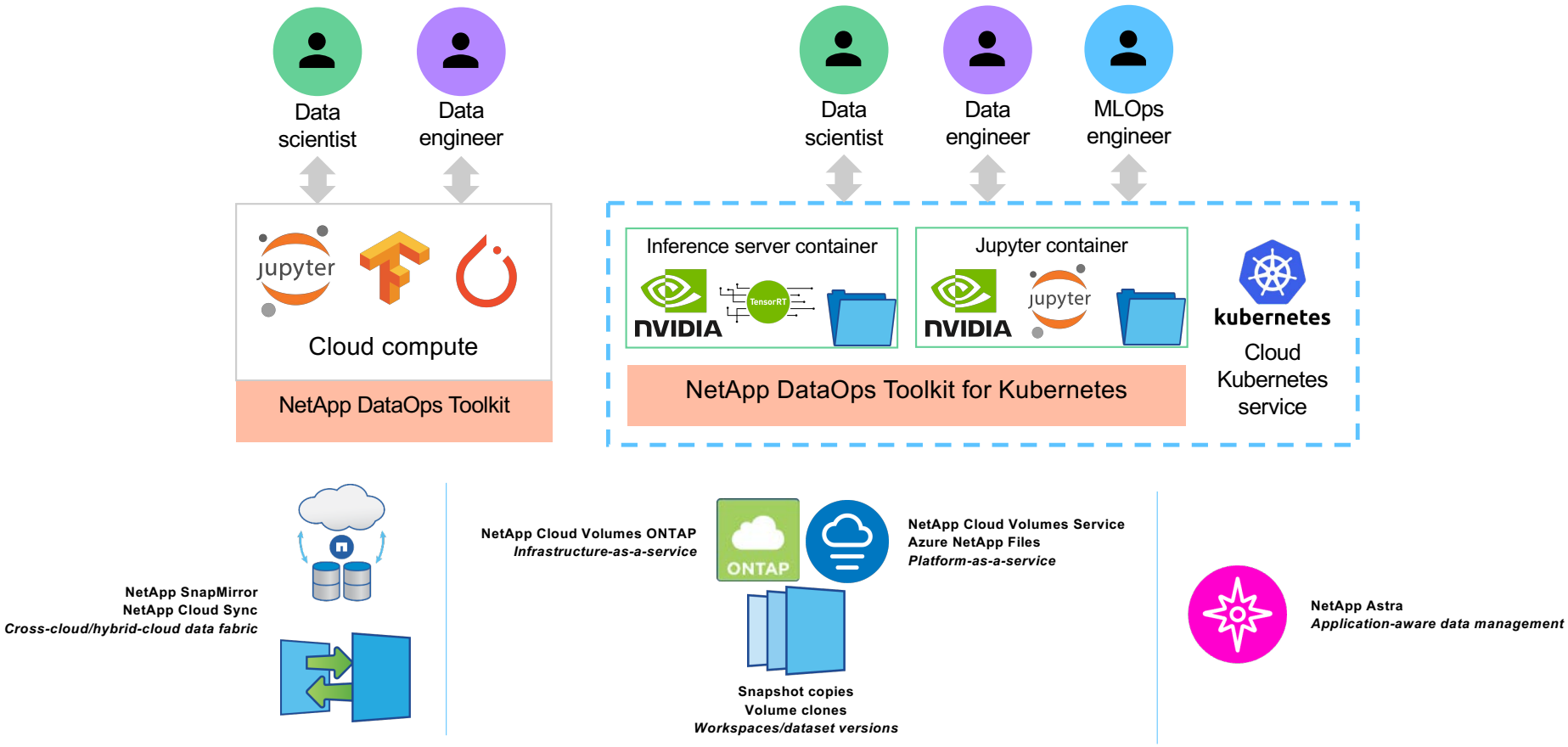
- Standalone AI data management solution for smaller teams
  - Easy-to-use, simple-to-manage
  - No MLOps or storage expertise required
- Use with NetApp AI Control Plane
  - Full-stack MLOps reference architecture built around popular open-source tools
  - Simplifies invoking of data management operations from within Jupyter notebooks, Kubeflow pipelines, and/or Apache Airflow workflows
- Use with other MLOps platforms
  - Incorporate advanced NetApp data management capabilities into any MLOps platform
  - For example: proprietary, custom, and homegrown MLOps platforms



NetApp © 2021 NetApp, Inc. All rights reserved.

NetApp AI in the cloud

Cloud-based AI accelerated by the NetApp



NetApp

© 2021 NetApp, Inc. All rights reserved.

For more information, visit…

netapp.com/ai

NetApp

© 2021 NetApp, Inc. All rights reserved.

Thank you

© 2021 NetApp, Inc. All rights reserved.

NetApp