XENON®
High Performance Computing

Learning from Industry Data Science Platforms

12th Feb 2025

Adrian Torrie – Solution Architect

25+ Years where we come from…

2020

2023
AGRF
Flash storage for genomics

2023
Pawsey
130PB Archive Cluster

2023
PTT Thailand
HPC Consulting

2023
WSU
Neuromorphic AI Cluster

2022
Deakin University
Clustered NVMe software defined storage for AI research

2022
WEHI
Clustered flash storage for genomics, cryo-em

2022
XENON Cloud
HPC-as-a-Service for HPC/AI, private & public cloud

2021
Deakin University
Design & deliver new HPC cluster for shared use across Deakin

2021
Hong Kong Uni HPC
HPC Cluster with XENON Cluster Stack

2021
Murdoch Children's Research Institute
Design and implement backup solution

2021
XENON Cluster Stack
Containerised HPC management solution

2020
CSL Limited
Services for HPC cluster implementation

2010

2020
Monash University
HPC cluster & storage configured as shared cloud node for Australia Research Data Commons

2020
Todd Energy
HPC cluster for oil & gas exploration

2019
Pawsey Supercomputing Centre
New GPU cluster

2018
Harrison.ai
AI solution for IVF, radiology, healthcare

2017
NCI
Supercomputer for scientific research & technology innovation

2017
WEHI
Private cloud for next generation cancer, disease & medical research

2016
Garvan Institute of Medical Research
NVMe solution for medical research

2015
Thales Defence
High fidelity solution for Australian Army Tiger helicopter simulators

2014
RCC
FlashLite HPC cluster

2012
Fujitsu
Infiniband for Raijin Supercomputer

2009
CSIRO
Australia's first GPU Cluster

2007
Victorian Partnership for Advance Computing
HPC Cluster for Advance Computing

2000

2005
Thales Defence
Image Generators for ASLAV simulators

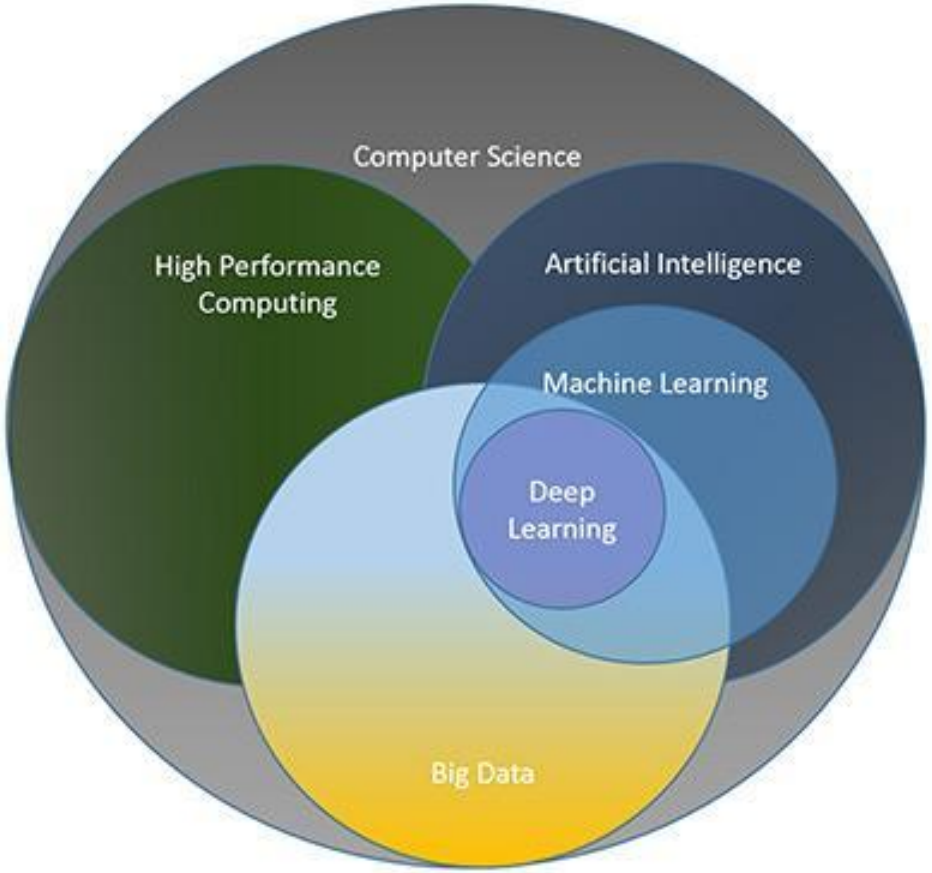
2000
Animal Logic
Render Farm
for "Matrix" movie trilogy

1998
Telstra Digital Video Network
Video Server equipment

Read the Case Studies

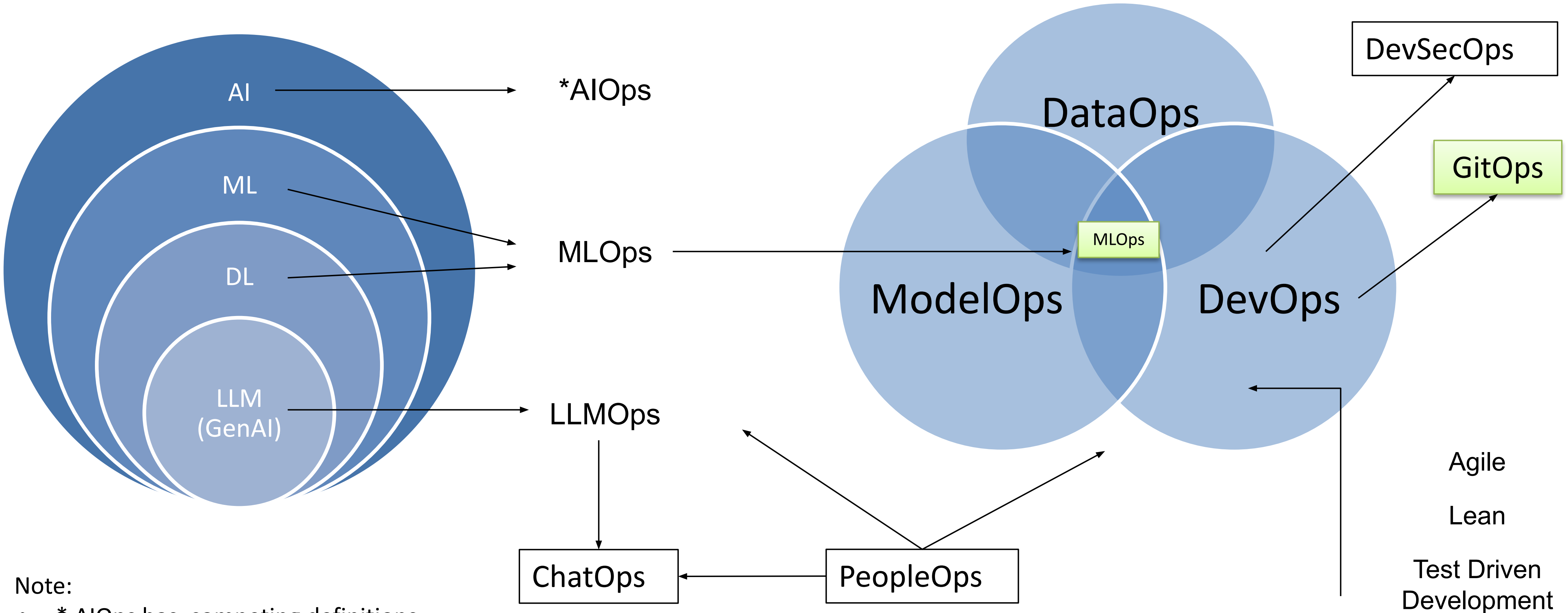
Introduction

- HPC and AI workload convergence
- Demystify and learn from container workflows, DevOps, MLOps, LLMOps, and Platform Engineering
- Can Kubernetes support:
  - Queuing and scheduling typical of HPC?
  - "Lift-and-shift" HPC jobs onto Kubernetes?
  - Ignoring microservices
- How do researchers access storage from Kubernetes?
- How does Kubernetes fit into a hybrid architecture encompassing cloud, on-prem, and IoT (lab devices)?
- How to adopt?



Source (Feb 2025):
https://www.hpcadvisorycouncil.com/subgroups_hpc_ai.php

xOps – Buzzword Explosion



AI

ML

DL

LLM (GenAI)

*AIOps

MLOps

LLMOps

ChatOps

PeopleOps

DataOps

ModelOps

DevOps

MLOps

DevSecOps

GitOps

Agile

Lean

Test Driven Development

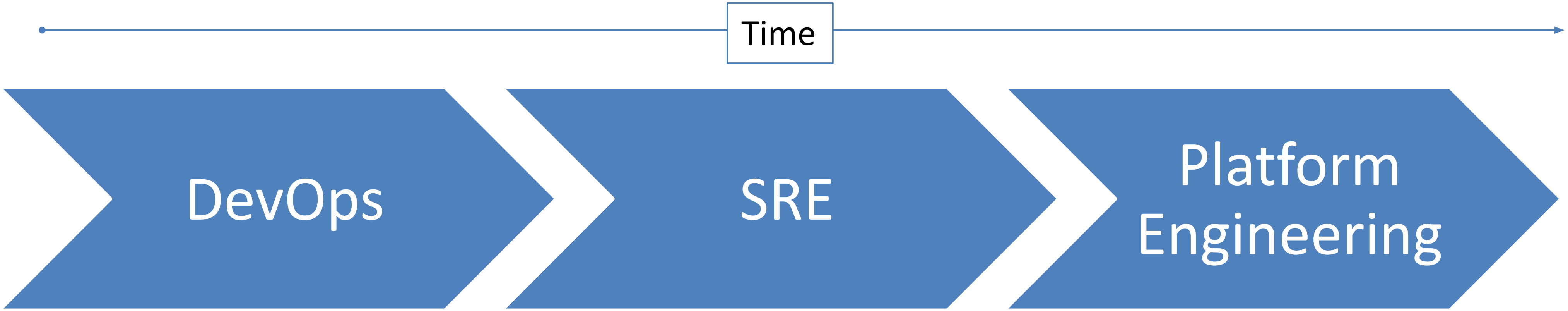
ITIL

Asset Lifecycle Management

Note:

- * AIOps has competing definitions
- Depending on an organisation’s or individual’s (x)Ops maturity, culture, communication style …
- … some may lump LLMOps under MLOps, or AIOps, or use all 3 interchangeably
- … similar interchangeable usage may occur with DevOps, DevSecOps, CI/CD
- … some just refer to all the concepts here as simply **DevOps**

Evolution of Automation



- Culture
- Practices
- Tools
- Shift-left / “Fail-fast”

➢ Tools are chosen with automation and security in mind

You build it, we run it

- Toil reduction
- Error budgets
- SLOs and SLIs
- Design for failure

➢ Think of SRE as an operating model for DevOps

We build it, we run it

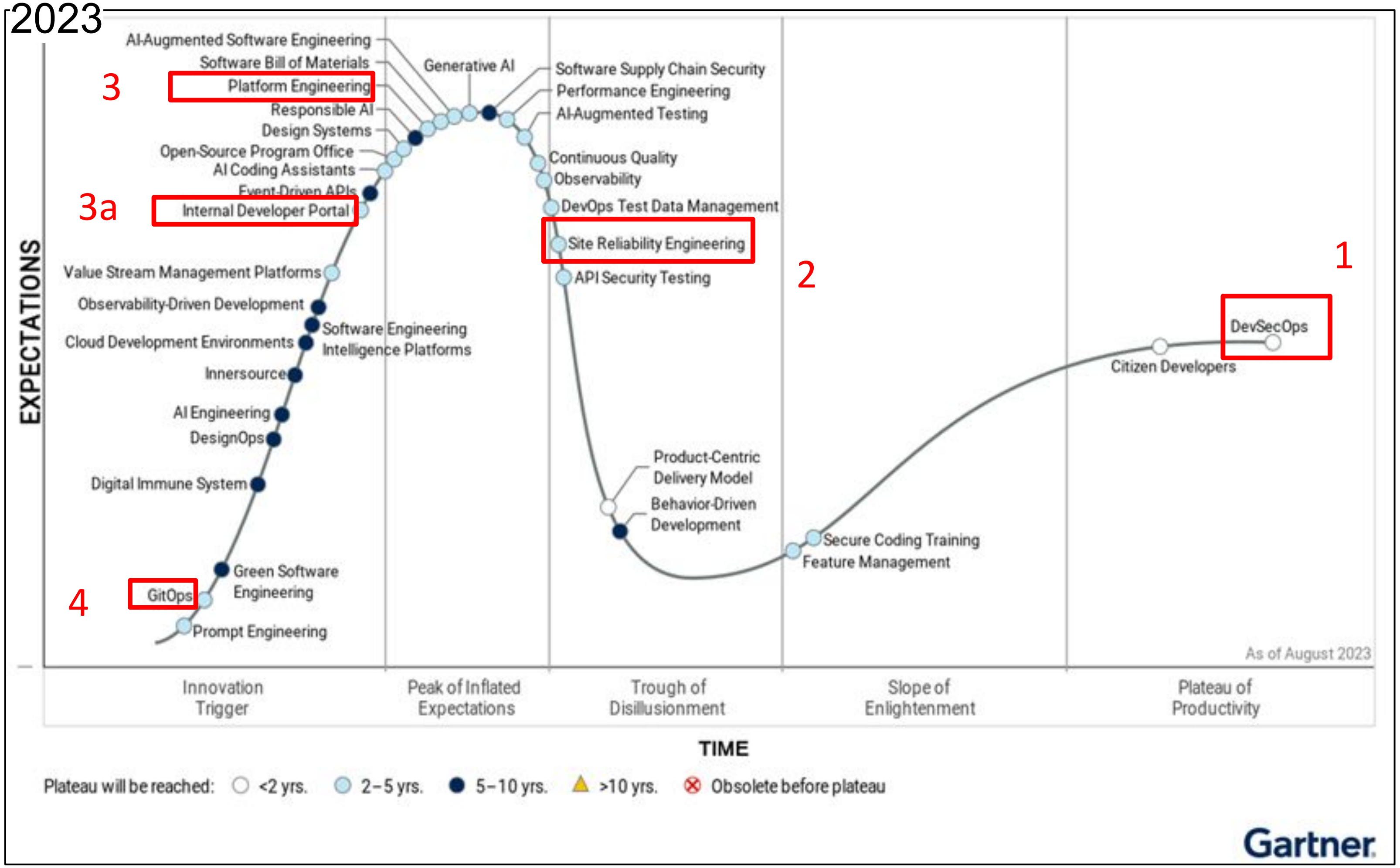
- Reduce cognitive load
- “Golden path”
- Citizen Data Analyst
- Citizen Data Scientist

➢ Self-service delivery model of organisation capabilities

You build it, you run it

DevOps vs SRE vs Platform Engineering

Gartner Hype Cycle for Software Engineering, November 2023



Source (Feb 2025): https://www.gartner.com/

1. DevSecOps
2. Site Reliability Engineering (SRE)
3. Platform Engineering
   a. Internal Developer Portal (IDP)
4. GitOps

What's the difference?

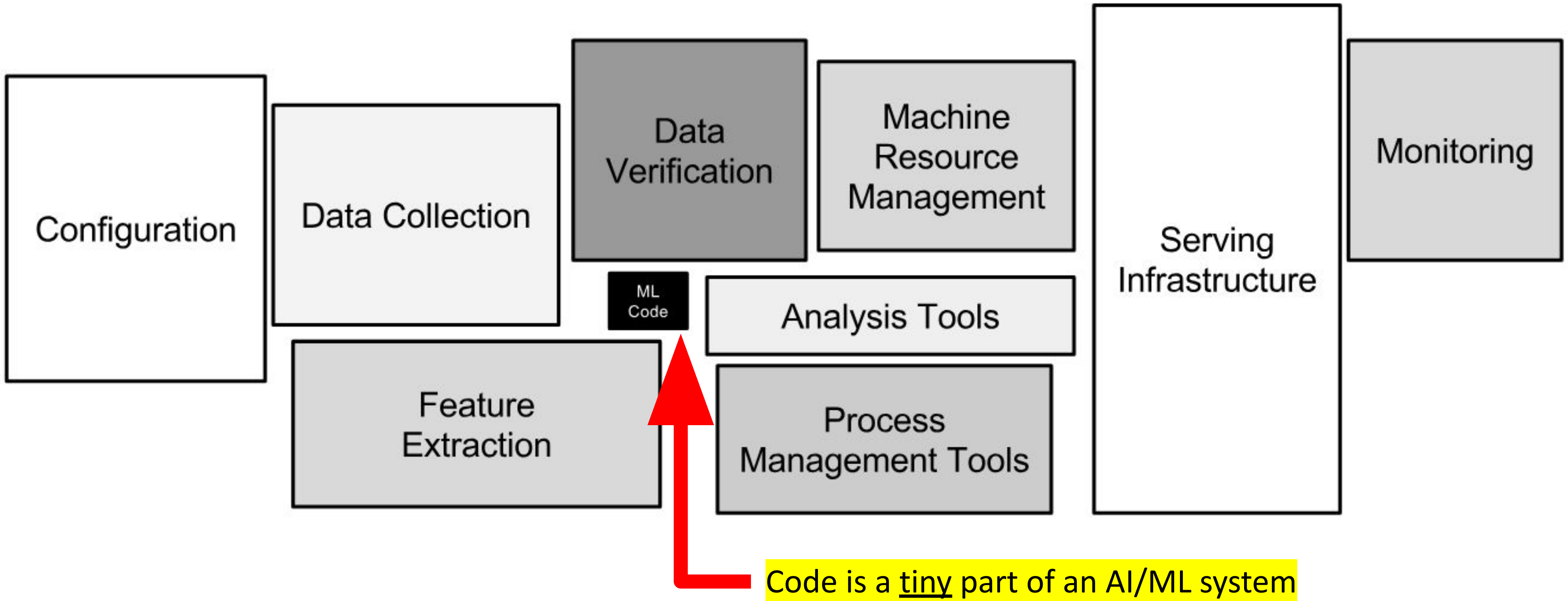
Choose 1 only?

Choose all 4?

They're complimentary, use all 4.

Why Platform Engineering is Important

Note: The size of each box shows the respective size and importance for a functional MLOps platform

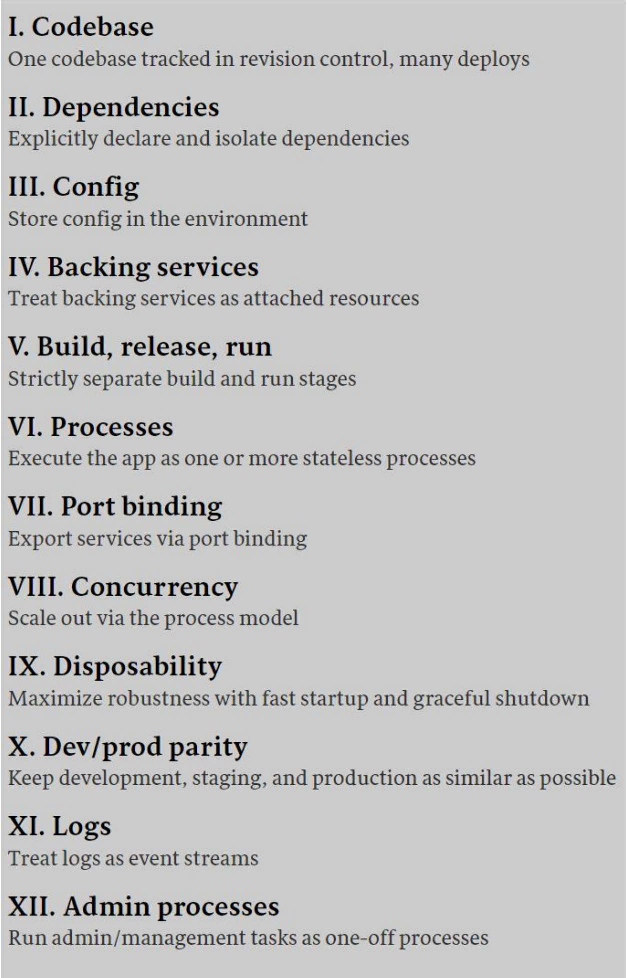


Code is a tiny part of an AI/ML system

Source (Feb 2025): Hidden Technical Debt in Machine Learning Systems

Principles and Standards

1. The Twelve-Factor App



I. Codebase
One codebase tracked in revision control, many deploys

II. Dependencies
Explicitly declare and isolate dependencies

III. Config
Store config in the environment

IV. Backing services
Treat backing services as attached resources

V. Build, release, run
Strictly separate build and run stages

VI. Processes
Execute the app as one or more stateless processes

VII. Port binding
Export services via port binding

VIII. Concurrency
Scale out via the process model

IX. Disposability
Maximize robustness with fast startup and graceful shutdown

X. Dev/prod parity
Keep development, staging, and production as similar as possible

XI. Logs
Treat logs as event streams

XII. Admin processes
Run admin/management tasks as one-off processes

2. Semantic Versioning 2.0.0 (SEMVER)

MAJOR.MINOR.PATCH, e.g. v3.2.8

1. MAJOR version when you make incompatible API changes
2. MINOR version when you add functionality in a backward compatible manner
3. PATCH version when you make backward compatible bug fixes

Additional labels for pre-release and build metadata are available as extensions to the MAJOR.MINOR.PATCH format.

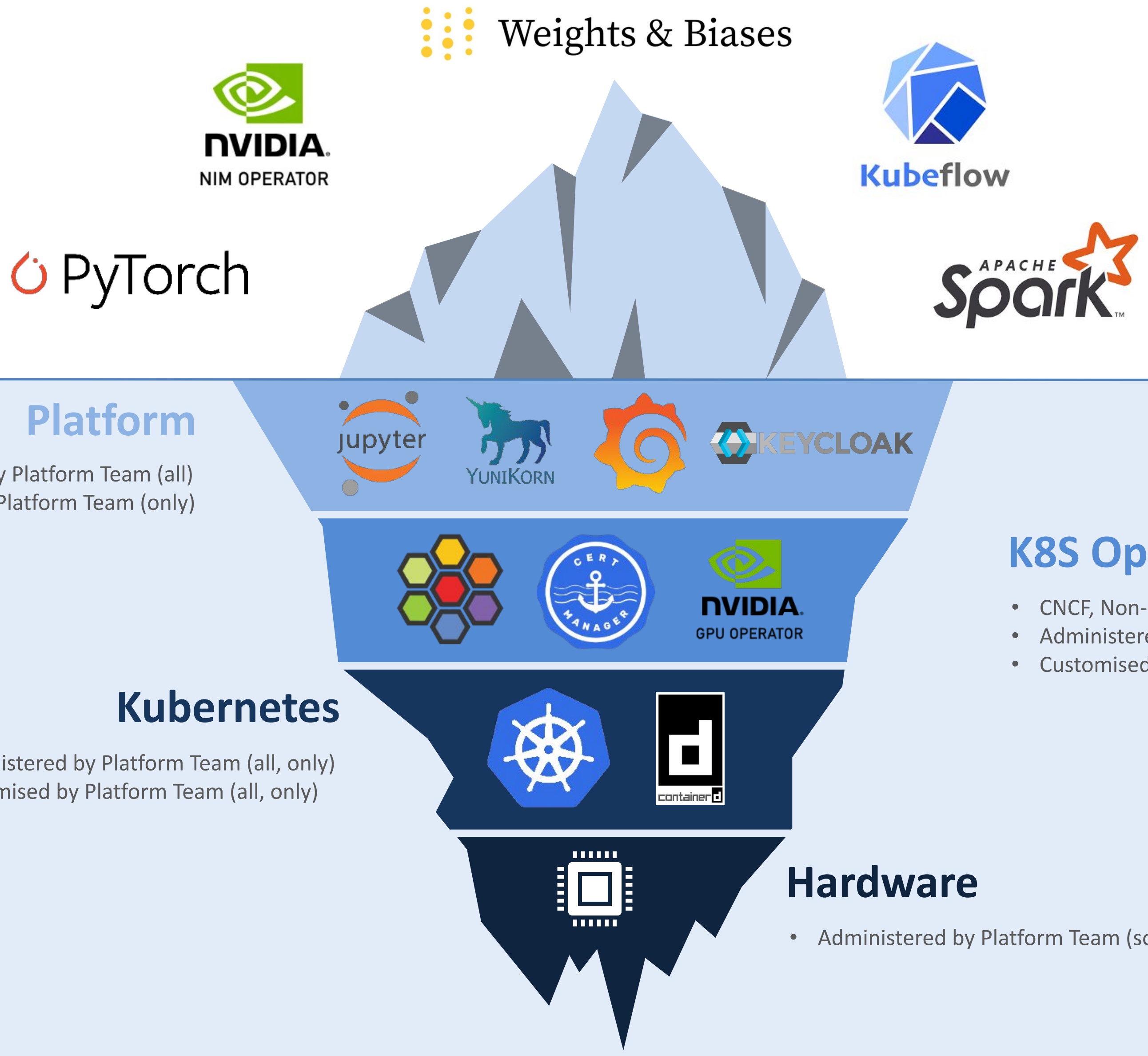
3. Open Container Initiative (OCI)

- **Image** Specification**:** Defines the structure and format of container images (e.g. Dockerfile).
- **Runtime** Specification**:** Specifies how a container should be executed and managed.
- **Distribution** Specification**:** Outlines standards for distributing container images.

➡ Think in terms of OCI compliant images for “lift and shift” enablement

AI Platform

"Above the Line"



Platform Customers

- Accessible by Customer Users (some)
- Accessible by Customer Admins (all)
- Administered by Customer Admins (all)
- Customised by Customer Admin (all)

Platform

- Administered by Platform Team (all)
- Customised by Platform Team (only)

K8S Operators

- CNCF, Non-CNCF, Apache projects
- Administered by Platform Team (all, only)
- Customised by Platform Team (all, only)

Kubernetes

- Administered by Platform Team (all, only)
- Customised by Platform Team (all, only)

Hardware

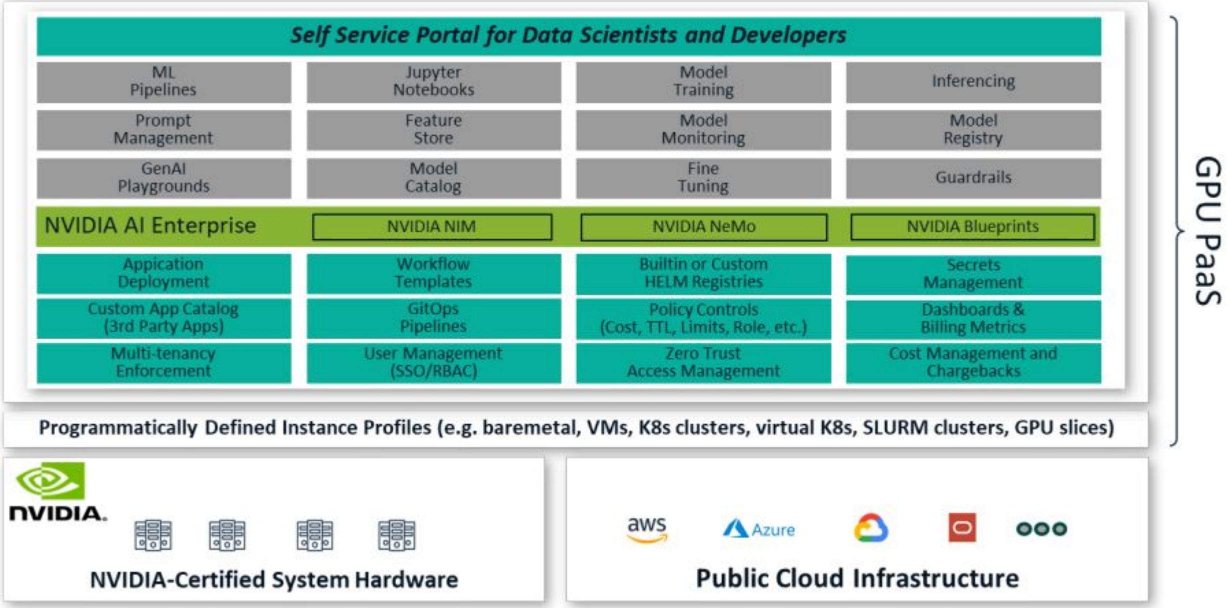
- Administered by Platform Team (some/all/only)

"Below the Line"

PaaS (I)

PaaS solutions can be used as a shortcut to building your platform

Rafay



Self Service Portal for Data Scientists and Developers

ML Pipelines	Jupyter Notebooks	Model Training	Inferencing
Prompt Management	Feature Store	Model Monitoring	Model Registry
GenAI Playgrounds	Model Catalog	Fine Tuning	Guardrails
NVIDIA AI Enterprise	NVIDIA NIM	NVIDIA NeMo	NVIDIA Blueprints
Appication Deployment	Workflow Templates	Builtin or Custom HELM Registries	Secrets Management
Custom App Catalog (3rd Party Apps)	GitOps Pipelines	Policy Controls (Cost, TTL, Limits, Role, etc.)	Dashboards & Billing Metrics
Multi-tenancy Enforcement	User Management (SSO/RBAC)	Zero Trust Access Management	Cost Management and Chargebacks

Programmatically Defined Instance Profiles (e.g. baremetal, VMs, K8s clusters, virtual K8s, SLURM clusters, GPU slices)

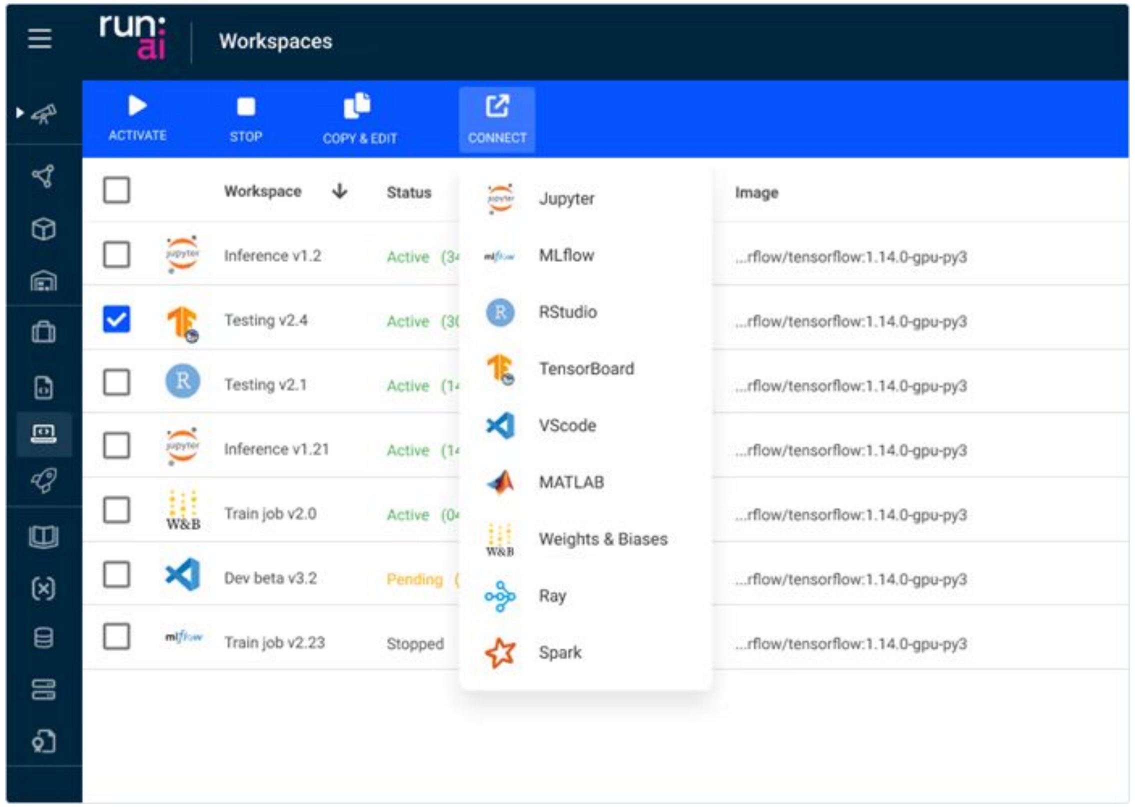
GPU PaaS

NVIDIA-Certified System Hardware

Public Cloud Infrastructure

PaaS Reference Architecture for GPU Clouds

Run:AI



	Workspace	Status	Image
□	Inference v1.2	Active (3	...rflow/tensorflow:1.14.0-gpu-py3
☑	Testing v2.4	Active (3	...rflow/tensorflow:1.14.0-gpu-py3
□	Testing v2.1	Active (1	...rflow/tensorflow:1.14.0-gpu-py3
□	Inference v1.21	Active (1	...rflow/tensorflow:1.14.0-gpu-py3
□	Train job v2.0	Active (0	...rflow/tensorflow:1.14.0-gpu-py3
□	Dev beta v3.2	Pending (	...rflow/tensorflow:1.14.0-gpu-py3
□	Train job v2.23	Stopped	...rflow/tensorflow:1.14.0-gpu-py3

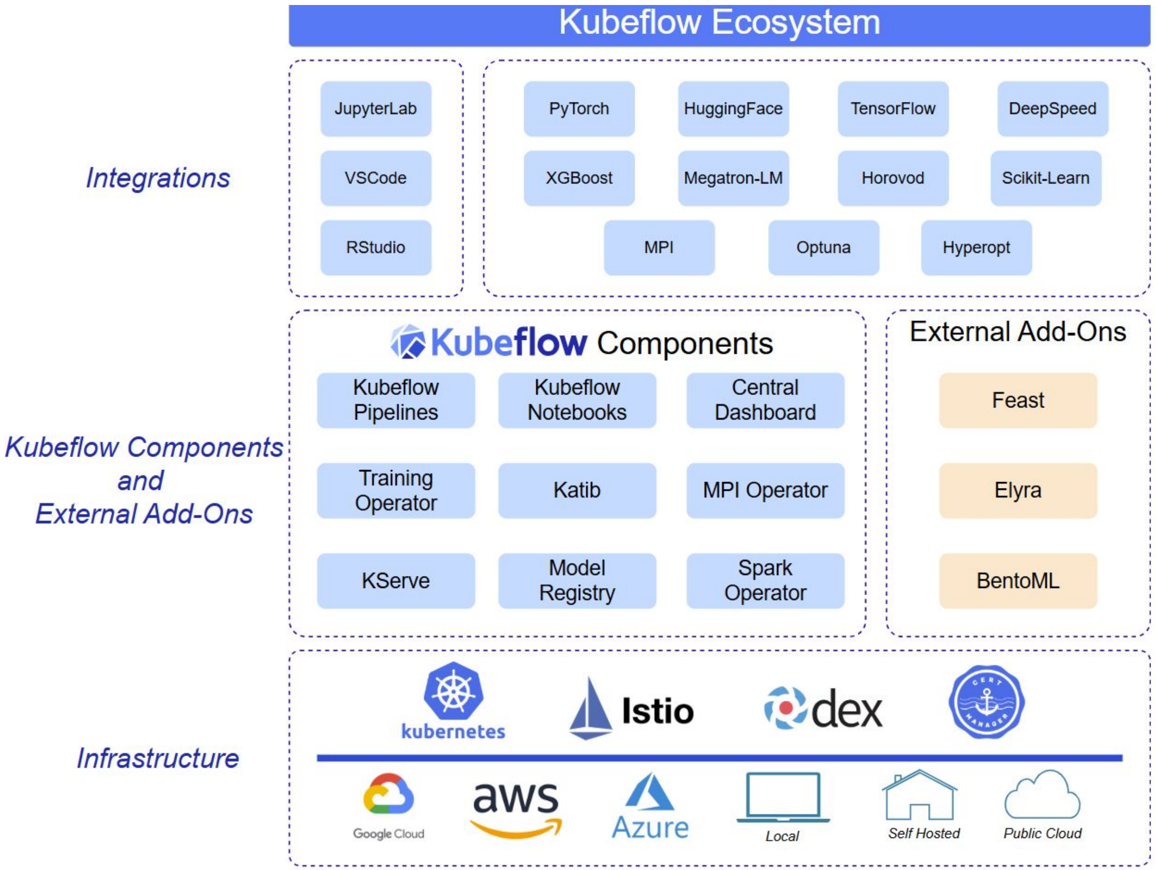
Source (Feb 2025): Rafay Platform - GPU PaaS Reference Architecture for Nvidia Cloud Partners & Enterprises

Source (Feb 2025): https://www.run.ai/

PaaS (II)

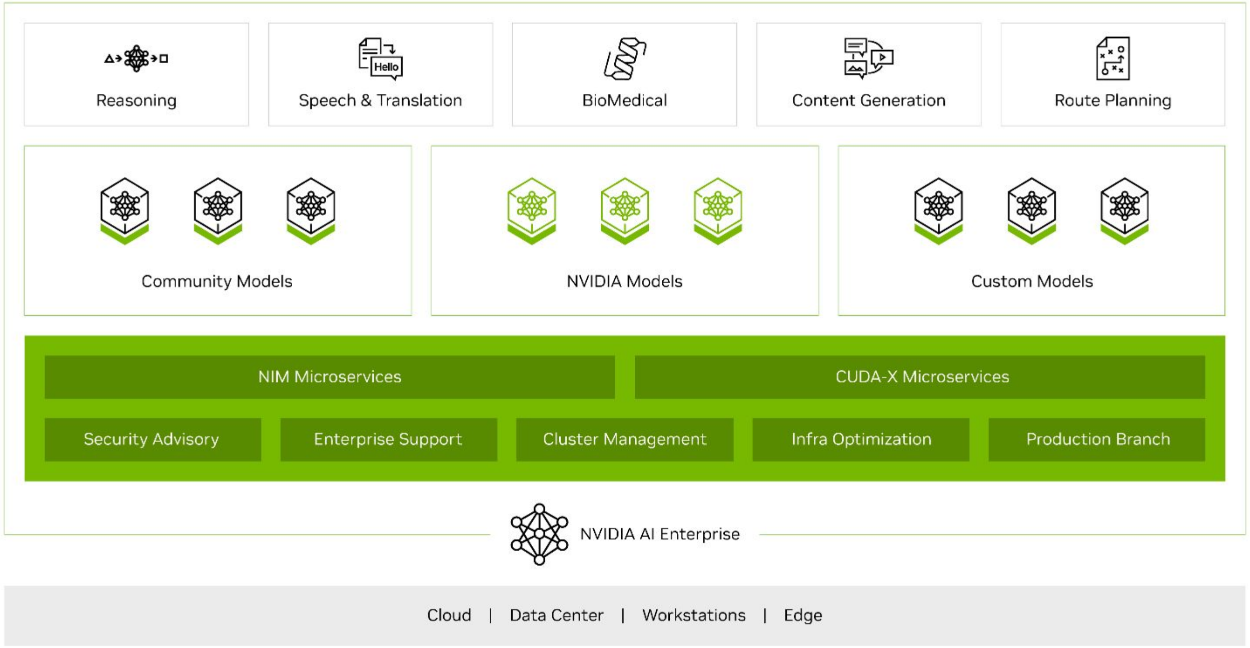
PaaS solutions can be used as a shortcut to building your platform

KubeFlow



Source (Feb 2025): https://www.kubeflow.org/docs/started/architecture/

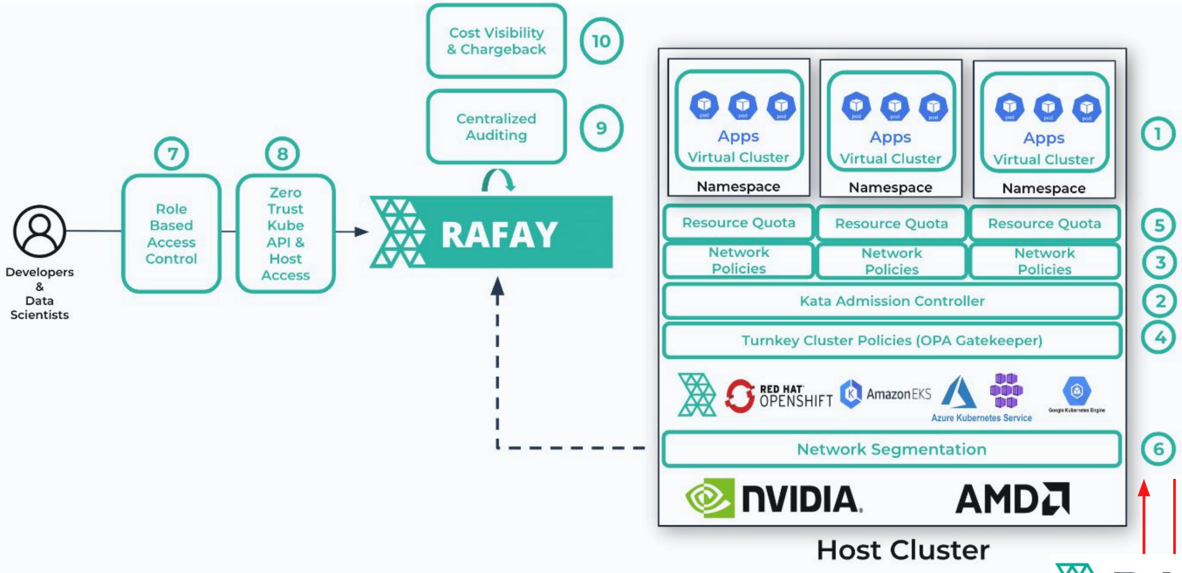
NVIDIA AI Enterprise



Source (Feb 2025): https://docs.nvidia.com/ai-enterprise/overview/latest/platform-overview.html

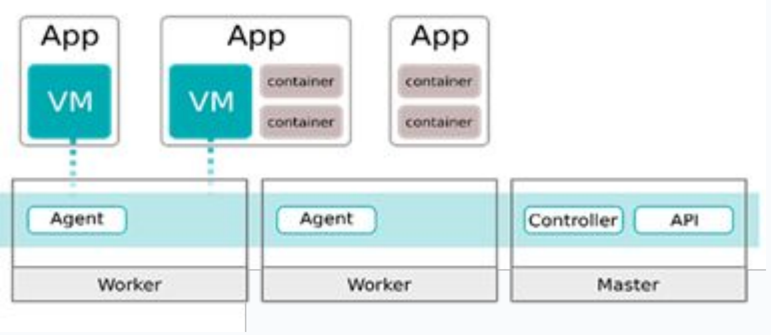
Next-gen Data Centre Management

Soft Multi-tenancy

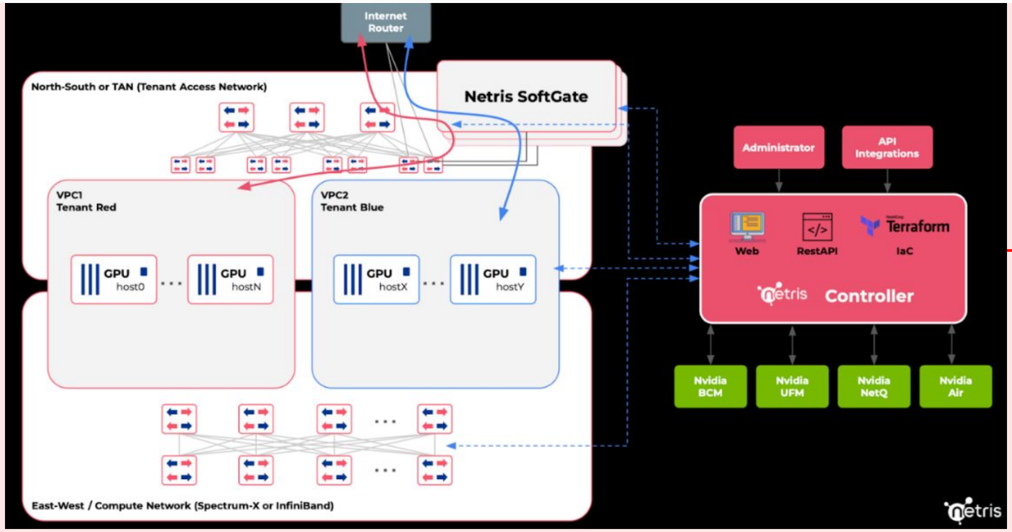


vCluster

KubeVirt

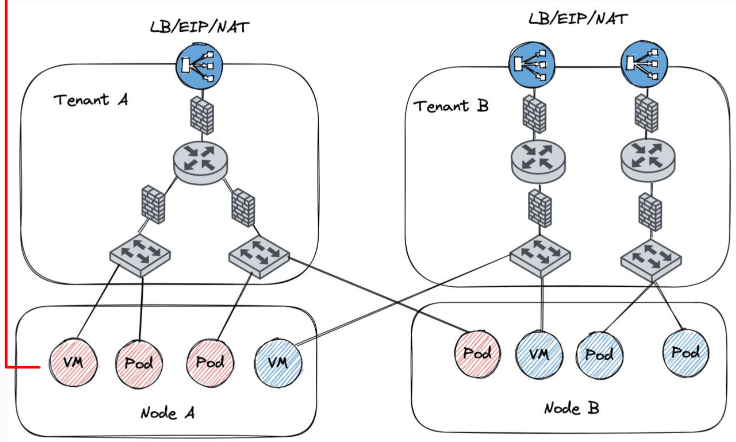


RAFAY



netris

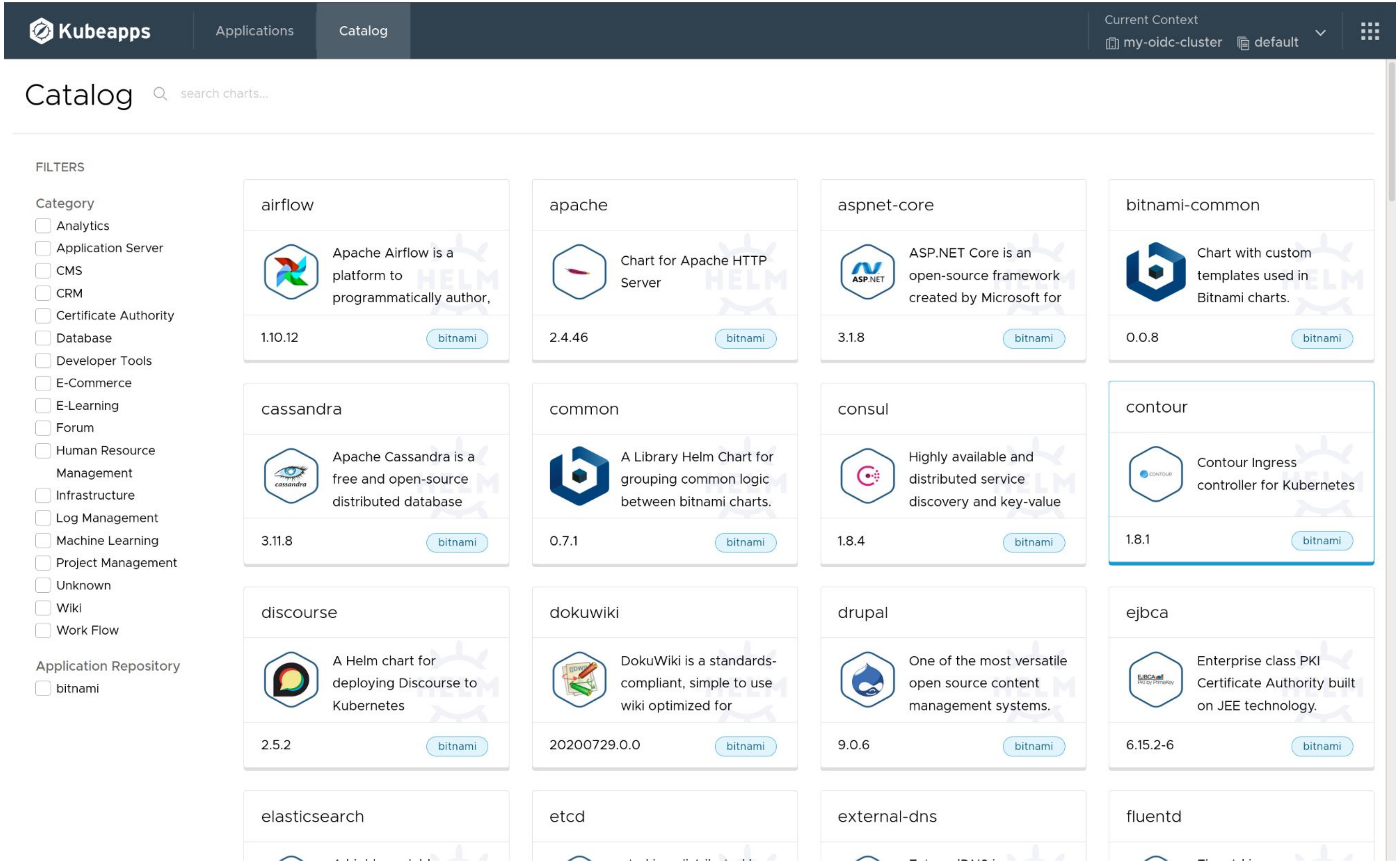
Kube-OVN



Hard Multi-tenancy

XENON
High Performance Computing

App Catalogue



Catalog

search charts...

FILTERS

Category

- □ Analytics
- □ Application Server
- □ CMS
- □ CRM
- □ Certificate Authority
- □ Database
- □ Developer Tools
- □ E-Commerce
- □ E-Learning
- □ Forum
- □ Human Resource Management
- □ Infrastructure
- □ Log Management
- □ Machine Learning
- □ Project Management
- □ Unknown
- □ Wiki
- □ Work Flow

Application Repository

- □ bitnami

airflow
Apache Airflow is a platform to programmatically author,
1.10.12 bitnami

apache
Chart for Apache HTTP Server
2.4.46 bitnami

aspnet-core
ASP.NET Core is an open-source framework created by Microsoft for
3.1.8 bitnami

bitnami-common
Chart with custom templates used in Bitnami charts.
0.0.8 bitnami

cassandra
Apache Cassandra is a free and open-source distributed database
3.11.8 bitnami

common
A Library Helm Chart for grouping common logic between bitnami charts.
0.7.1 bitnami

consul
Highly available and distributed service discovery and key-value
1.8.4 bitnami

contour
Contour Ingress controller for Kubernetes
1.8.1 bitnami

discourse
A Helm chart for deploying Discourse to Kubernetes
2.5.2 bitnami

dokuwiki
DokuWiki is a standards-compliant, simple to use wiki optimized for
20200729.0.0 bitnami

drupal
One of the most versatile open source content management systems.
9.0.6 bitnami

ejbca
Enterprise class PKI Certificate Authority built on JEE technology.
6.15.2-6 bitnami

elasticsearch

etcd

external-dns

fluentd

XENON®
High Performance Computing

Batch Schedulers

YuniKorn

A light-weight universal resource scheduler for container orchestrator systems.

- App-aware scheduling
- Hierarchical queues
- Gang scheduling
- Job ordering & queuing
- Resource fairness
- Resource reservation
- Preemption
- Max application enforcement

VOLCANO

Kueue

slinky

Slinky = Slurm on Kubernetes

- Offering flexibility and ease of use for both HPC and cloud-native users
- Run and manage Slurm clusters on Kubernetes
- Manages the scaling of Slurm nodes within Kubernetes
- Job allocation/accounting/dependencies
- Fair-share, and priority scheduling
- “Lift-and-shift” potential for unified infrastructure

Batch Workloads

YuniKorn

+

Kubeflow

Training Operator

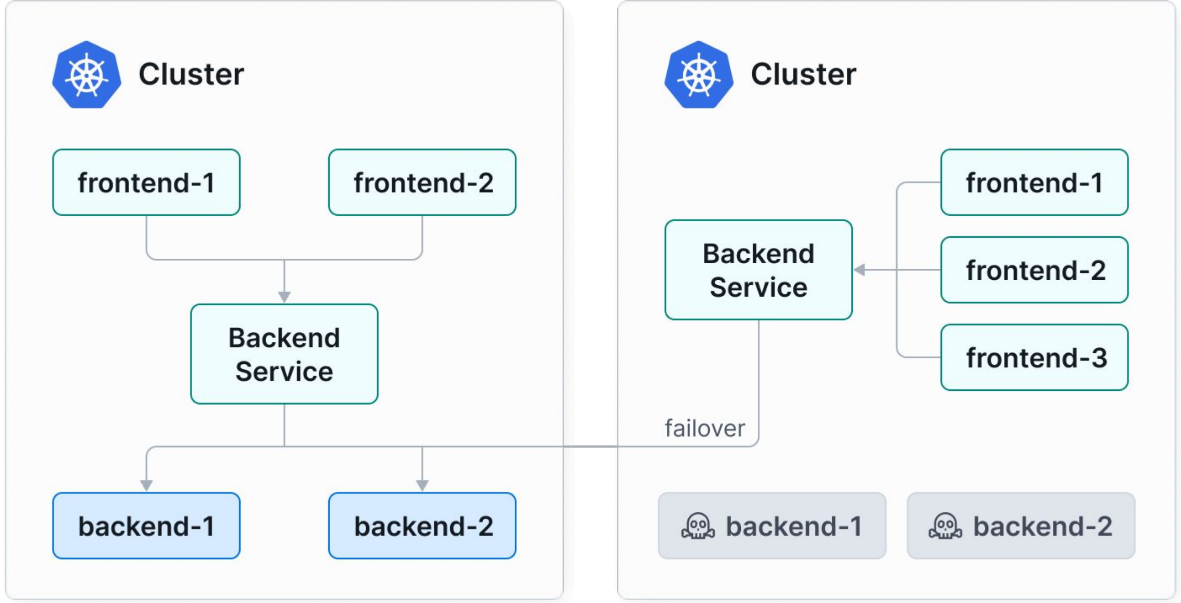
- MPI
- Pytorch
- Tensorflow
- Ray Clusters / Jobs / Services
- Spark
- Flink
- PaddlePaddle
- XGBoost
- JAX

Cluster Mesh – Example Scenarios

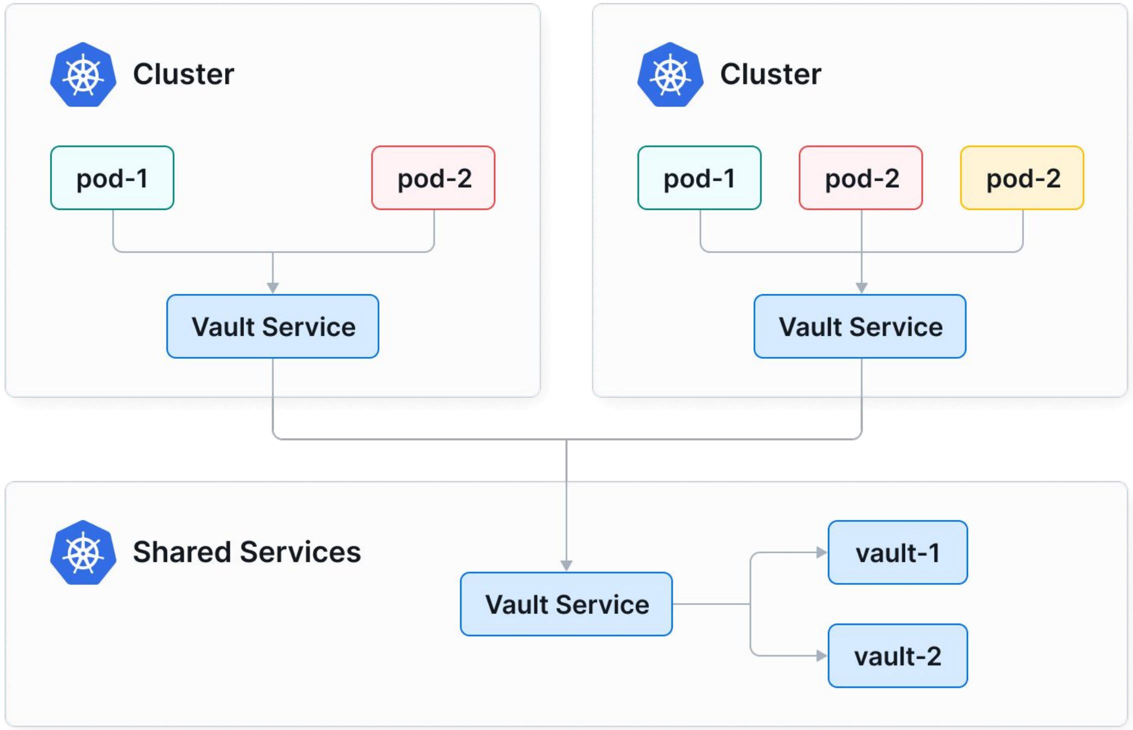
cilium

…"an open source, cloud native solution for providing, securing, and observing network connectivity between workloads, fueled by the revolutionary Kernel technology eBPF."

High Availability and Fault Tolerance

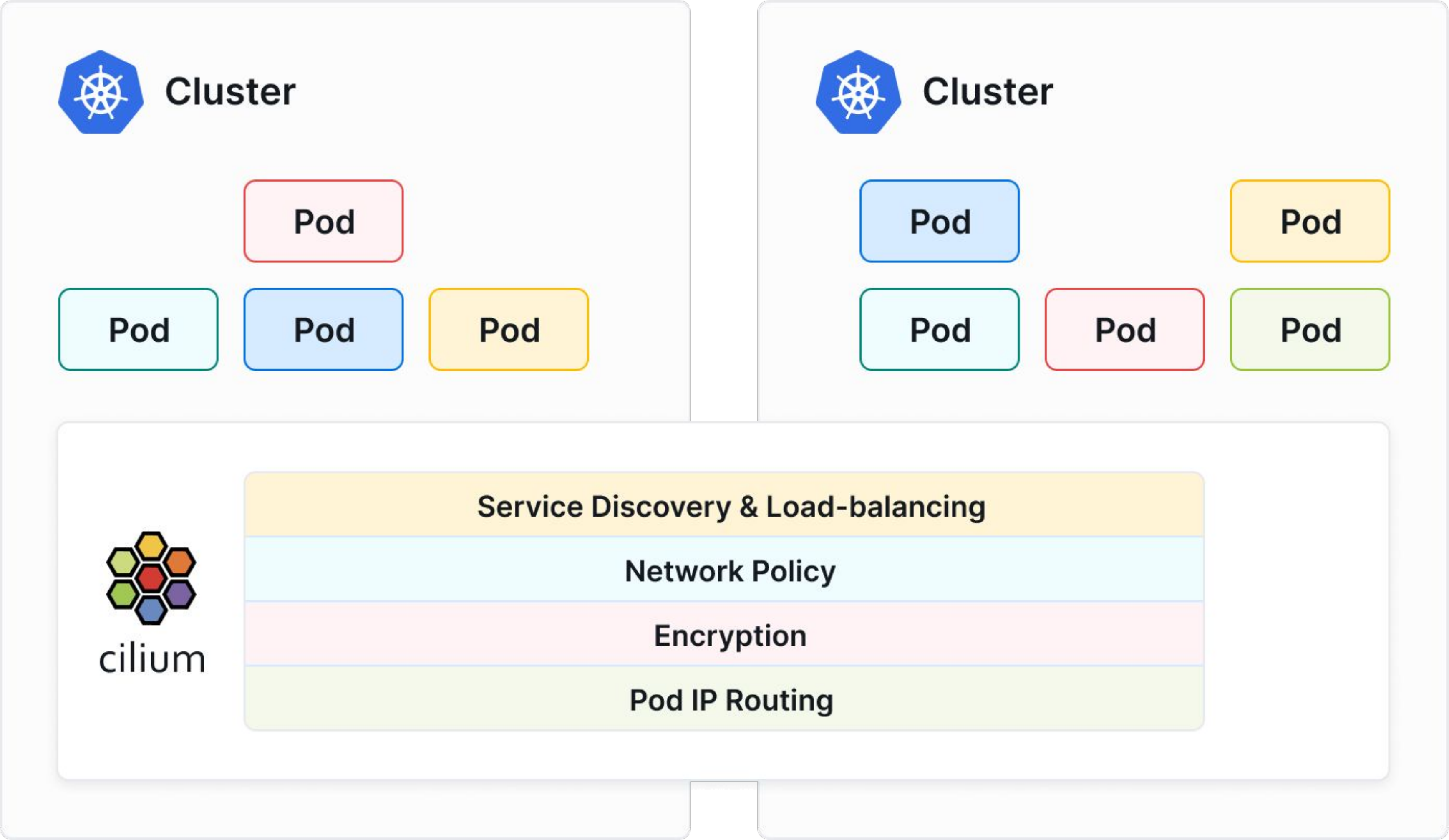


Shared Services Across Clusters



XENON
High Performance Computing

Cluster Mesh – Uniform Network Policy Enforcement



IoT Edge Gateways and Devices

KubeEdge

Provides fundamental infrastructure support for network, app deployment and metadata synchronization between cloud and edge.

□ Good for edge gateways

K3S

Kubernetes distribution built for
IoT & Edge computing

□ Good for lightweight vClusters
□ Good for devices

Note: Cilium works on both distributions for cluster mesh scenarios

Storage

CSI Drivers

- NFS
- SMB
- Local file (on-node)
- Cloud hyperscalers

CSI Drivers for native integration with Vendor storage

HAMMERSPACE

DELL Technologies

WEKA

ddn

VAST

Simple Adoption Idea – 1. POC

Objective

- Prove the technology, i.e. batch jobs as containers on Kubernetes

Step 1

NVIDIA GPU OPERATOR

- ✔ Driver management
- ✔ GPU in containers

Kubeflow Training Operator

- ✔ MPI Jobs
- ✔ Pytorch Jobs

Step 2 (Optional)

NVIDIA NETWORK OPERATOR

- ✔ GPU Direct Storage
- ✔ RDMA

Tips:

- Use a single node Kubernetes “cluster” (no VM) to simplify and focus on the objective
- Focus on a single Kubeflow Training Operator workload
- Test the Nvidia Network Operator once you have the job running

Simple Adoption Idea – 2. MVP

Objective

- Prove scheduling workflows, understand additional job types

YUNIKORN

- ✔ Batch Scheduling
- ✔ Gang scheduling
- ✔ Fair use
- ✔ Queues

Kubeflow

Training Operator

- ✔ MPI Jobs
- ✔ Pytorch Jobs
- ✔ Tensorflow Jobs
- ✔ Spark Jobs

NVIDIA.

GPU OPERATOR

- ✔ Driver management
- ✔ GPU in containers

NVIDIA.

NETWORK OPERATOR

- ✔ GPU Direct Storage
- ✔ RDMA

Simple Adoption Idea – 3. Scale-out Production

Production considerations

- HA/DR
- Tighter user access controls
- Secure UIs
- Certificates for securing traffic
- Multi-node
- Platform services

Other Tips

1. Leaky Tap Strategy
2. Blueprints and Deployment stamps
    1. Blueprints tend to refer to software stacks (Rafay and Nvidia use this term)
    2. Deployment stamps refer to Infrastructure as Code (IaC) – Ansible and Pulumi
    3. Sometimes the lines are blurred between the two in cloud-native approaches
    4. Both are essentially templates that rely on configuration to stand-up resources (use config as code!)
3. Scaffold repos
    1. 1x functional repo per use case
    2. Standards embedded in repo design
    3. Configuration locations pre-defined and required to trigger DevOps processes
    4. Clone from to execute and learn, and modify to start new projects in a new repo
    5. Good to teach new users
    6. Fast way to learn for those new to Platform Engineering (even if you're a senior)
4. vClusters
    1. Give you access to blue-green deployment patterns for whole clusters
5. GitOps
    1. Tools provide access to blue-green deployment patterns for cluster resources
6. Reference architectures
    1. Azure MLOps v2, AWS, Nvidia, Xenon AI Sandpit

Connect with Xenon

Thank You

XENON®

High Performance Computing

XENON Systems Pty Ltd
10 Westall Road, Springvale, Victoria 3171, Australia

www.xenon.com.au

P +61 3 9549 1111
F +61 3 9549 1199
E info@xenon.com.au

A member of the XENON Technology Group
www.xtg.com.au