

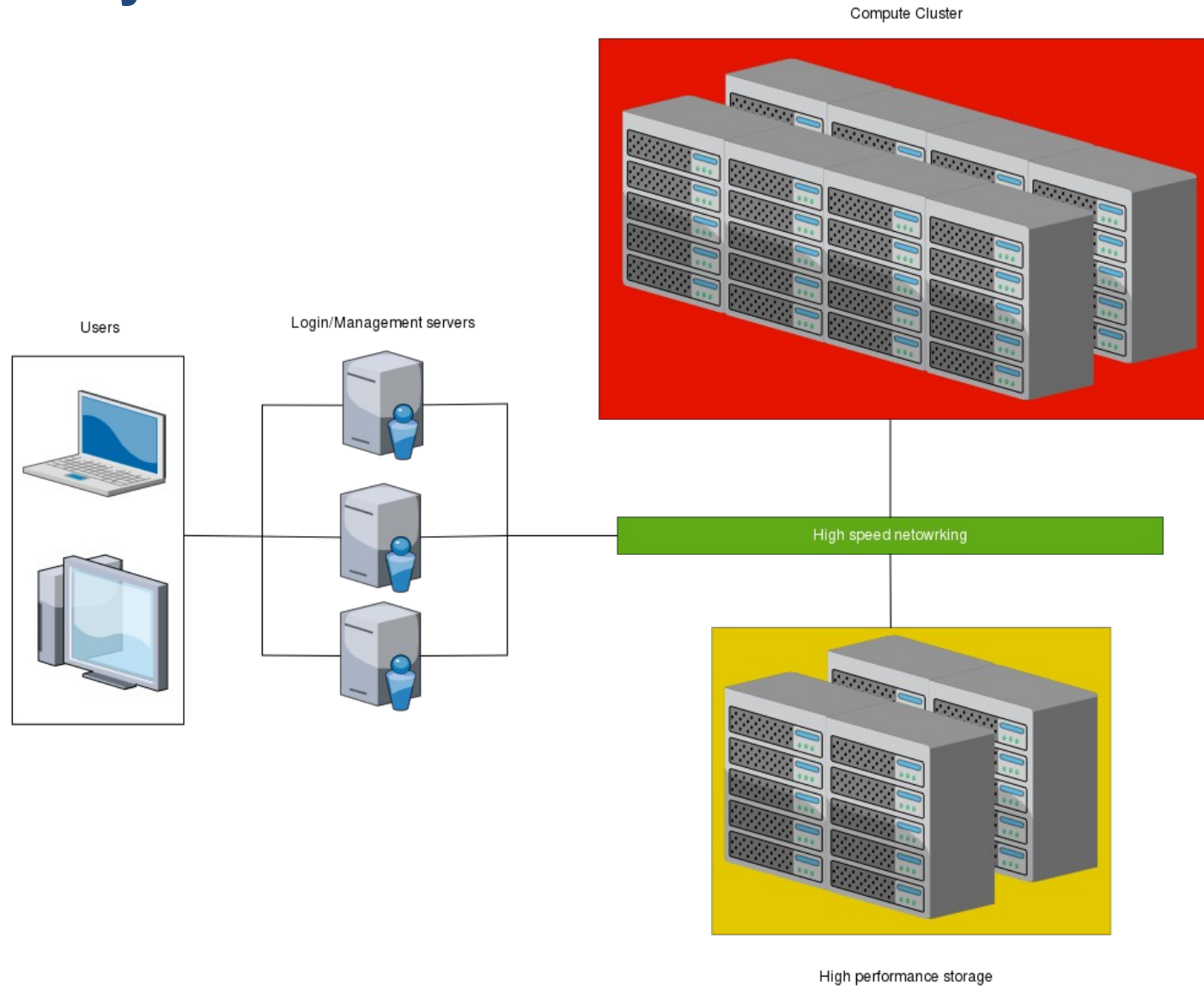


XENON[®]
High Performance Computing

Mastering the Maze

*Simplifying HPC, Kubernetes, and Secure Software
Delivery with DevOps, GitOps, and Platform
Engineering*

The HPC world today



The HPC world today

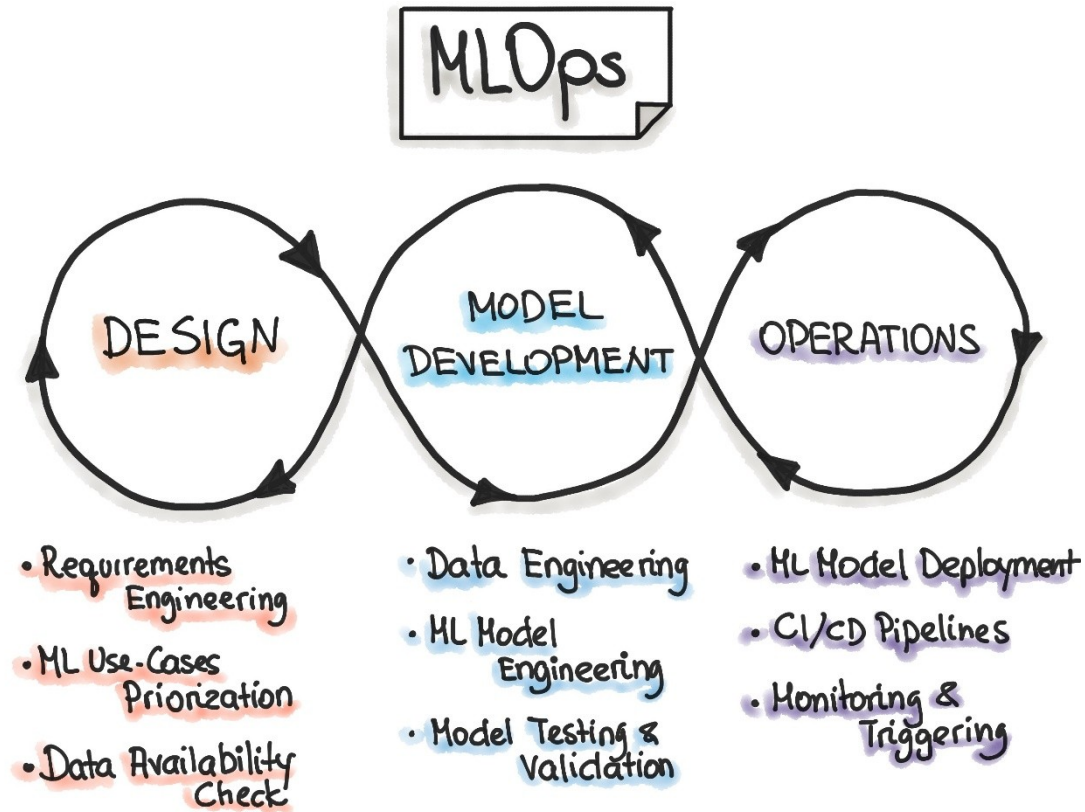
- Users access the machines via login nodes
- They submit jobs via a job scheduler
- Jobs are generally one application
- Jobs use files from the high performance storage directly
- Jobs are generally “stand alone” and don’t interact with other running jobs
- Libraries and other resources are curated via “Modules” or similar

The emerging ML/AI landscape

- Multi-component workflows (RAG, multiple stages, multiple models)
- Heavy use of docker/containers
- Interactivity often required/wanted for some stages of work
- Rapid pace of change in regards to libraries and supporting tech
- More suited to “On-Demand” than a “batch scheduled” model
- Differing requirements depending on stage of work (Training vs running of models)
- More use of “Cloud” resources



Enter MLOps



<https://ml-ops.org/content/mlops-principles>

MLOps platforms

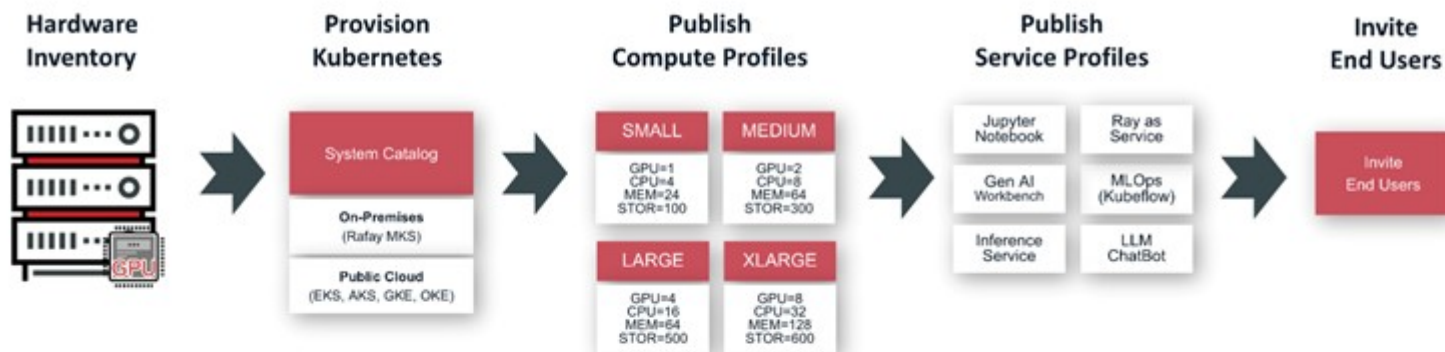
- Cloud-based
 - Amazon SageMaker
 - Azure ML
 - Google Vertex AI
 - Databricks
- Open source
 - MLflow
 - Kubeflow
 - Metaflow
 - Ray

The basic requirements

- Authentication and permissions control
- Local container registry
- User management
- Workflow design and observation
- “Chart” repository
- “Billing” and/or quotas
- Network access control
- Storage interfaces for multiple kinds of storage (Objects, POSIX, databases)
- Security scanning and alerting



Rafay MLOps-Kubeflow

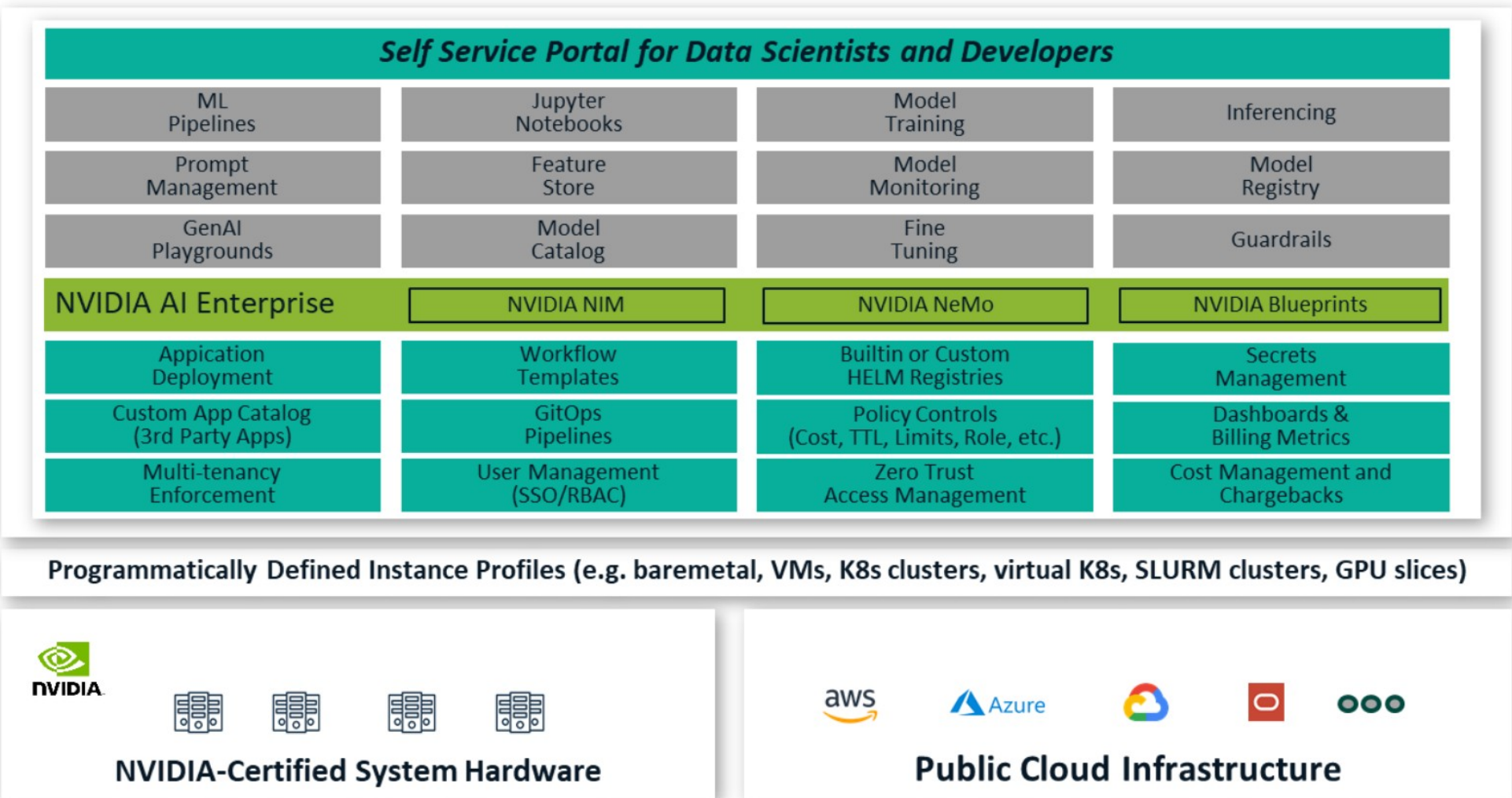


- ➔ Simplified process with Rafay Blueprints



Training	Monitoring	Integrations
Distributed Training Operator	TensorBoard	arize EVIDENTLY AI
Serving	Auto ML	Feature Store
KServe	Katib	FEAST
Notebooks	Pipelines	Model Registry
jupyter	Kubeflow	mlflow
Authentication	Service Mesh	Kubernetes
dex	Istio	Rafay MKS

Rafay Platform overview



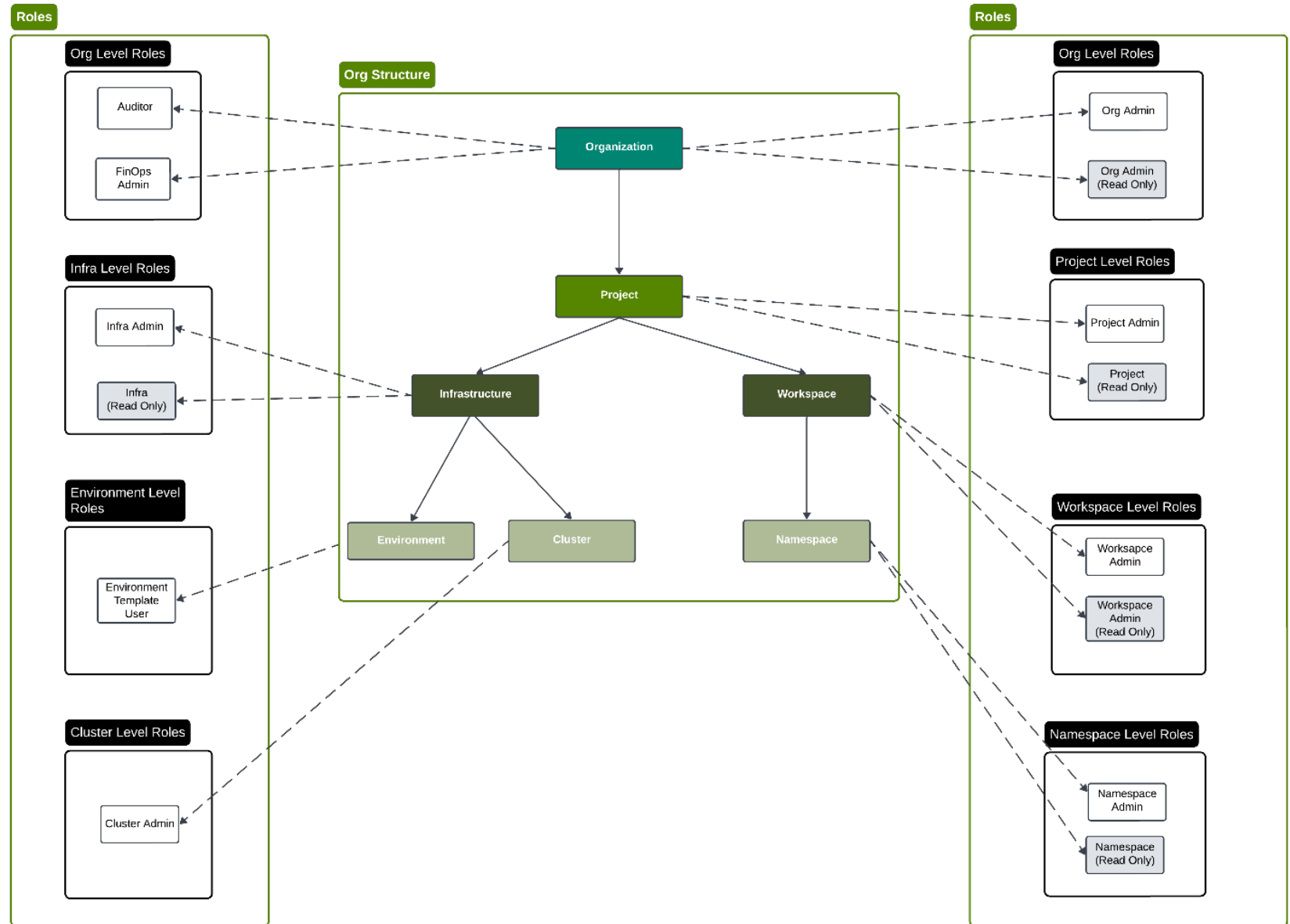
Rafay Functional Overview

Rafay is a Kubernetes Operations Platform (KOP) that simplifies the management, deployment, and scaling of Kubernetes clusters across hybrid and multi-cloud environments. It provides automation, security, and governance for Kubernetes workloads.

- **Backup:** Ensures the protection and recoverability of Kubernetes cluster data, safeguarding against data loss.
- **Blueprints:** Provides templates for consistent and repeatable deployment of applications and infrastructure across clusters.
- **Catalogue:** Offers a repository of pre-built applications and services, facilitating quick deployment and standardisation.
- **Cost:** Delivers insights into cloud expenditures related to Kubernetes, aiding in budgeting and cost optimisation.
- **GitOps:** Enables declarative deployment and management of applications using Git repositories as the source of truth.
- **Policy Management:** Allows the definition and enforcement of policies across clusters, ensuring compliance and governance.
- **Multi-Cluster Management:** Simplifies the administration of multiple Kubernetes clusters, providing centralised control and visibility. Pre-existing clusters can also be “imported” for deep visibility and workload management of the imported clusters.
- **Network Policy:** Facilitates the creation and enforcement of network security policies within Kubernetes environments.
- **Service Mesh:** Manages microservices communication within Kubernetes, enhancing observability, security, and reliability.
- **Zero Trust:** Implements security models with secure, zero-trust access, ensuring secure access within Kubernetes environments with seamless access through both browser-based and KubeCTL CLI interfaces, eliminating the need for special client software installations.

Rafay IAM

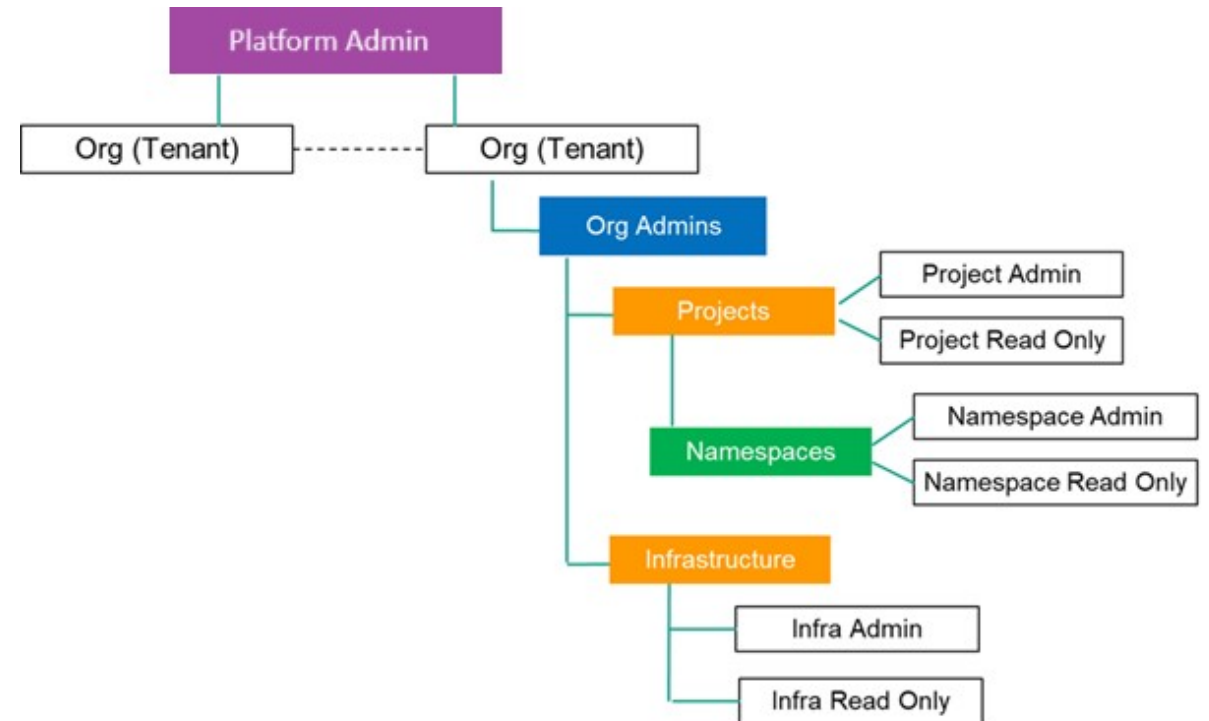
- **Role-based Access Control (RBAC)**
Helps define roles and permissions to ensure secure access management for users and groups.
- **Multi-cluster Support**
Facilitates centralised IAM policies across multiple clusters, enhancing control over diverse environments.
- **Integration with External Identity Providers**
Supports integration with services like Okta, LDAP, and Microsoft Entra to name a few, streamlining user management and authentication processes.
- **Audit and Compliance**
Provides visibility into IAM activities, ensuring compliance and aiding troubleshooting efforts.
- **Granular Permissions**
Allows detailed permissions at the project, cluster, and namespace levels to meet specific access needs.



Rafay Multi Tenancy

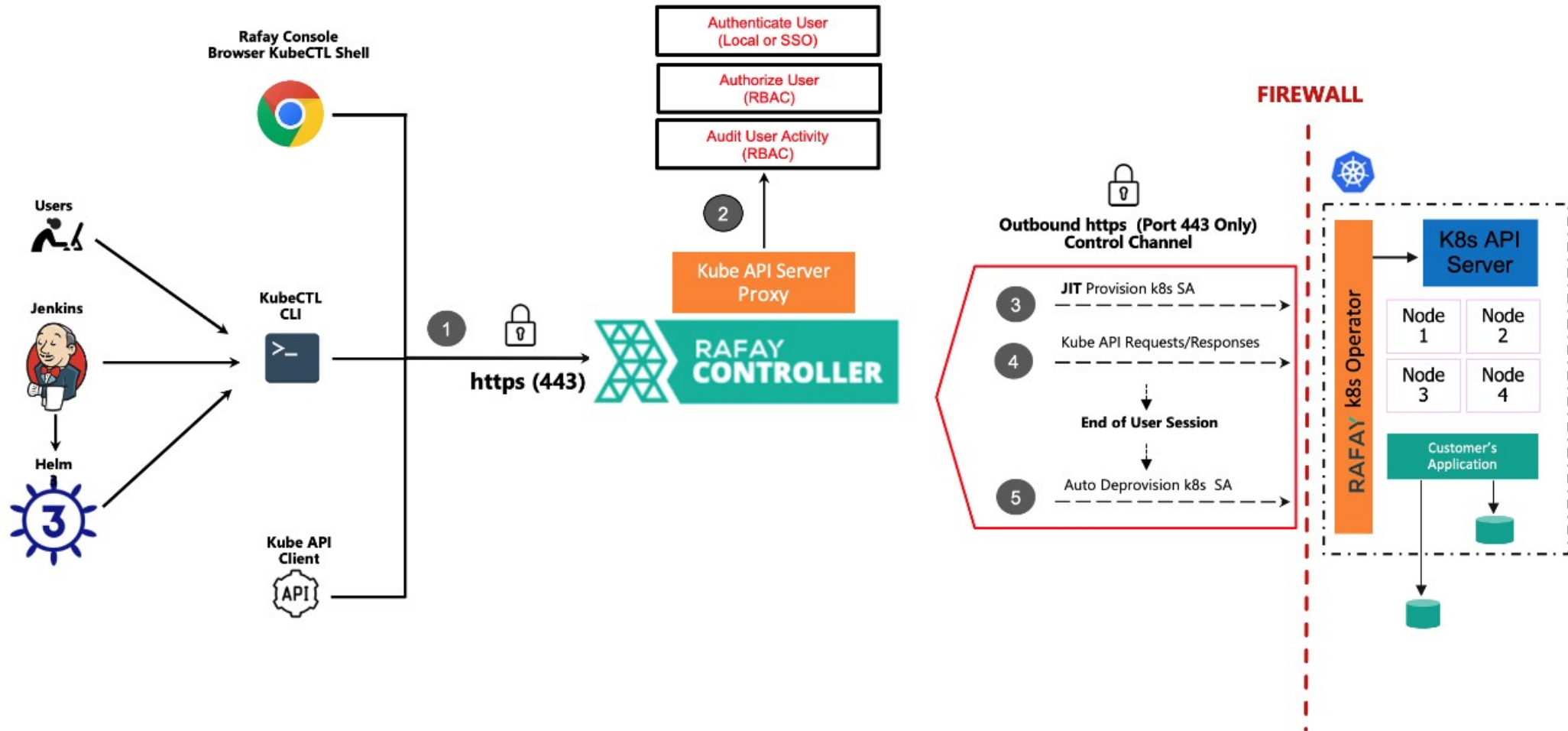
Multiple Levels of Multi Tenancy

Rafay's platform supports multi-tenant organisations (Org) with role-based access control (RBAC), allowing administrators to manage multiple Orgs through a unified interface.

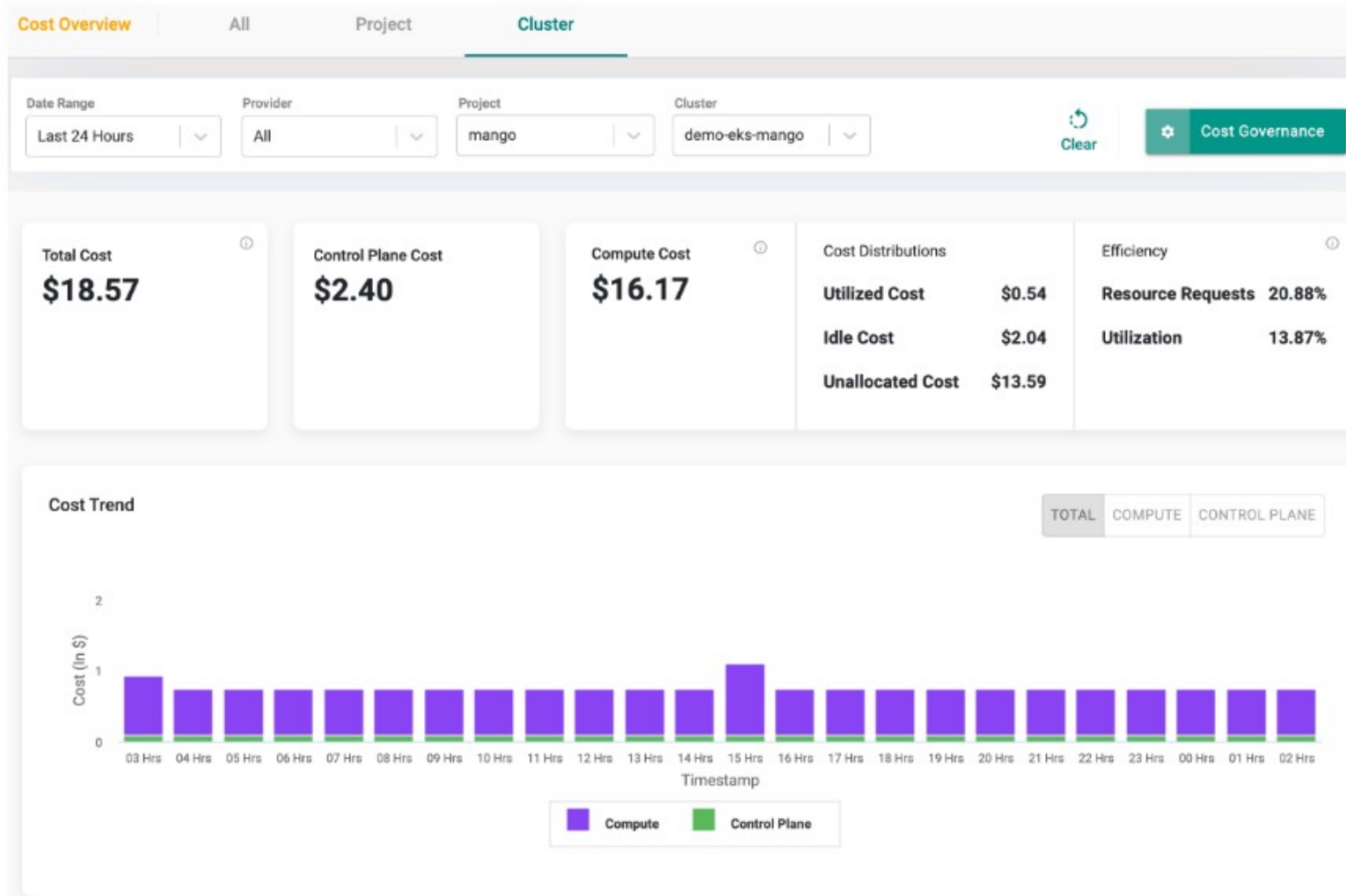


Rafay Zero-trust

➔ The Rafay Kube API Server Proxy authorises all clients



➔ Built-in capabilities for chargeback and cost allocation purposes.



At scale example



Project Ohm Tech stack



Pulumi at a glance

- Pulumi allows true IaC (Python, GO, Node.js, .NET, Java)
- It handles Cloud APIs natively (AWS, Google, Azure, Kubernetes)
- Pulumi Handles bare metal via Terraform providers:
 - Netris
 - Rafay
 - MAAS
 - iDRAC
 - iLO
- It also natively handles Virtualization platforms (Vmware, Proxmox)

Pulumi + Rafay + Netris

- Pulumi handles all the infrastructure, be it virtual, cloud, or bare metal.
- With Netris, all the underlying switch fabric is abstracted away and then configured using Pulumi
- The deployment of Rafay on Rafay's MKS distribution of K8s is also possible via Pulumi
- Rafay then handles all of your MLOps-Kubeflow needs

Back to Project Ohm





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



Thank You

Malcolm Haak
Solutions Architect

P +61 3 9549 1111
M +61 488 700 496
E malcolmh@xenon.com.au

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

