XENON
High Performance Computing

Securely Share Your Bare-Metal HPC and AI Cluster without Virtualisation: Modern Multi-Tenancy Solutions

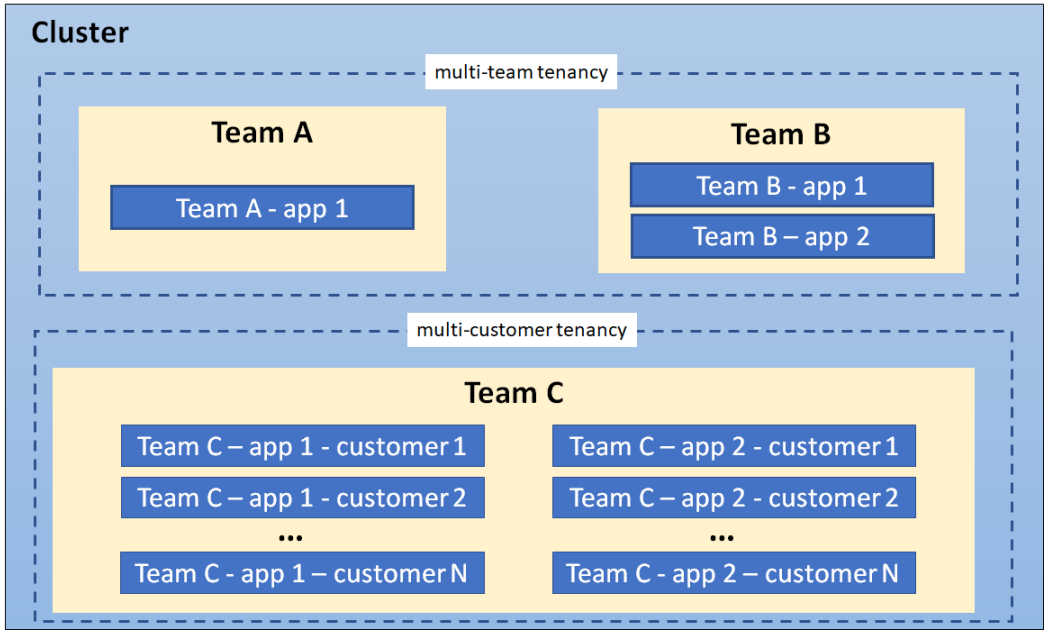
Dr. Werner Scholz, CTO and Head of R&D
XENON Systems

Outline

- What does Multi-Tenancy mean?
- Typical Solutions
- Trade-offs
- From Hard to Soft Multi-Tenancy
- Modern Implementations
- Storage Multi-tenancy
- Challenges

What does Multi-Tenancy mean?

- **Multiple User:**
  - **within the same team** and organisation
  - collaboratively running workloads, which might need to communicate with each other
  - direct access to cluster
- **Multiple Teams:**
  - separate groups **within the same organisation**
  - direct access to cluster
- **Multiple Customers** (“SaaS” multi-tenancy)
  - **strongly isolating** different workloads
  - no access to customer



Typical Solutions

- **Dedicated Hardware:** physically isolated: all components separate and isolated
- **Physically separated networks:** shared compute and storage resources, separate networks
- **Virtualisation:**
  - vmware
  - Xen (Citrix, xcp-ng)
  - KVM (OpenStack, Open Nebula, ProxMox)
- **Containers:**
  - Apache Mesos (retired)
  - docker, docker Swarm
  - Kubernetes

Trade-offs:

- Hardware management and utilisation
- Scheduling, Fairness, QoS
- Cluster management
- Security
- Operating system management, updates
- Peformance: bare metal vs overheads (abstractions)
- Application management: dependencies, software supply chain/asset management

From Hard to Soft Multi-Tenancy

Multi-Tenancy is based on a range of concepts and thus provides a continuum solutions

- SaaS multi-tenancy: CLI/API access controlled by application or centralised identity management, RBAC
- Workloads/workflows: managed through applications and workflow tools
- Processes: controlled by OS kernel and tools
- Users and groups: typical shared HPC/AI clusters based on multi-user OS
- Control plane isolation
- Containers: OS-level “virtualisation”
- Namespaces: isolating groups of resources
- Logical network isolation: virtual networks, network policies, node isolation
- Hardware based virtualisation
- Physical network isolation
- Physical hardware/cluster separation

Implementing Network Multi-Tenancy

Multi-tenancy at network level:

- Ethernet (high speed and out-of-band networks)
  - Manual network configuration on switches using VLANs
  - Manual configuration using EVPN/VXLAN
  - Semi-automated with configuration management tools (e.g. Ansible)
  - Automated software defined network configuration using e.g. Netris, aarna, Rafay

- Infiniband: Partition Keys:

Manual configuration:

1. Configure base configuration on switches (OOB, MLAG, interface type/speed, NTP, etc)
2. Create a VLAN (or VXLAN) configuration for the new tenant in NVIDIA Air, eg:
3. Copy the config file to /home/cumulus/tenant-vlan.yaml (or paste it in via VI or NANO)
4. nv config patch /home/cumulus/tenant-vlan.yaml
5. nv config apply
6. Confirm the config is running with: nv config show
7. Repeat for all switches as required.

Semi-automated with Ansible:

1. Setup OOB Management and disable ZTP on the switches
2. Enable NVUE API on the switches
3. Add nvidia.nvue module to Ansible environment.
4. Prepare Production Ready Automation
5. Run playbook to load config to switches

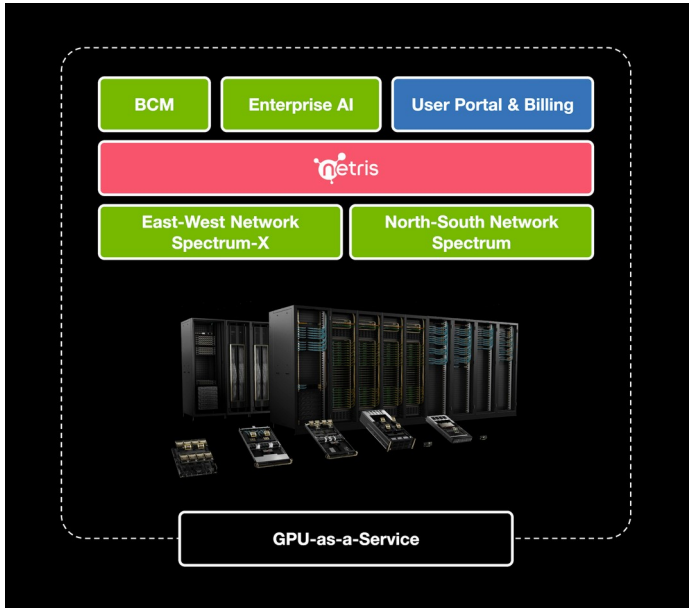
Modern Implementations

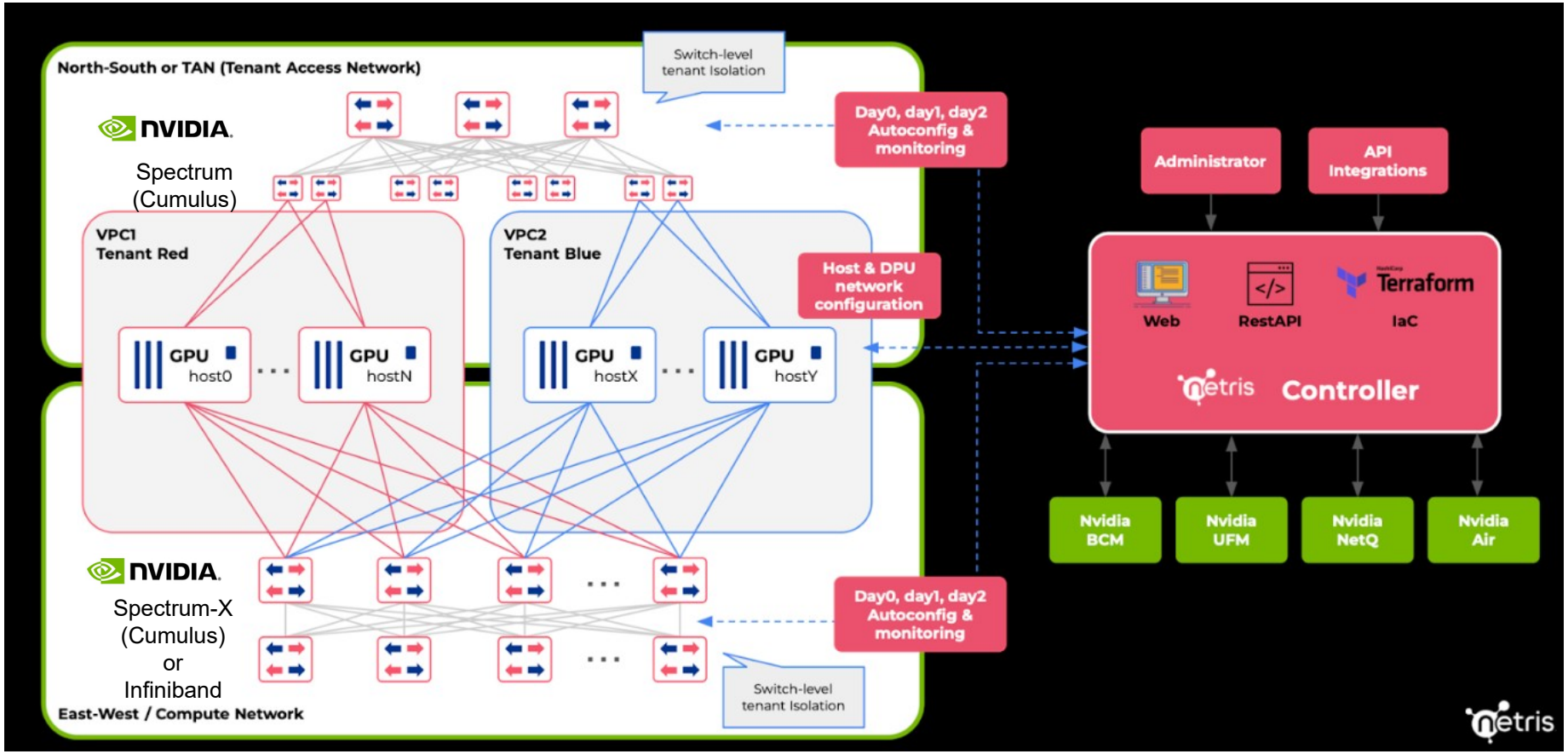
Solutions:

- Netris: software defined networking
- Kubernetes: many implementations
- Rafay
- aarna.ml
- etc.

Netris

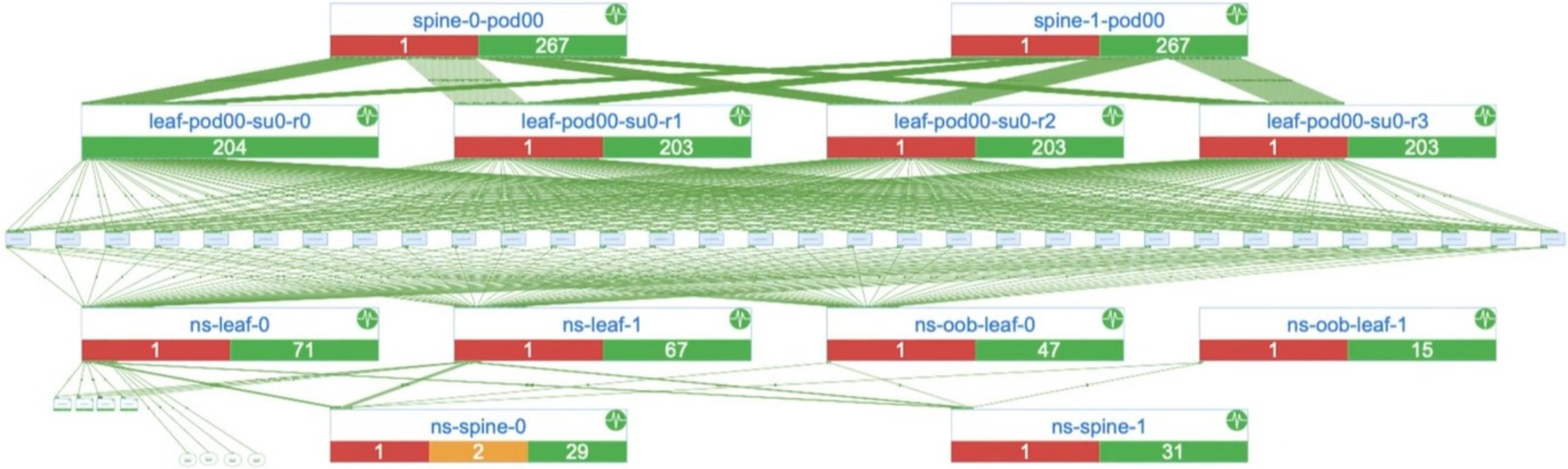
Cloud provider-style network automation and abstraction suitable for both multi-tenant public cloud providers and private cloud environments





Netris

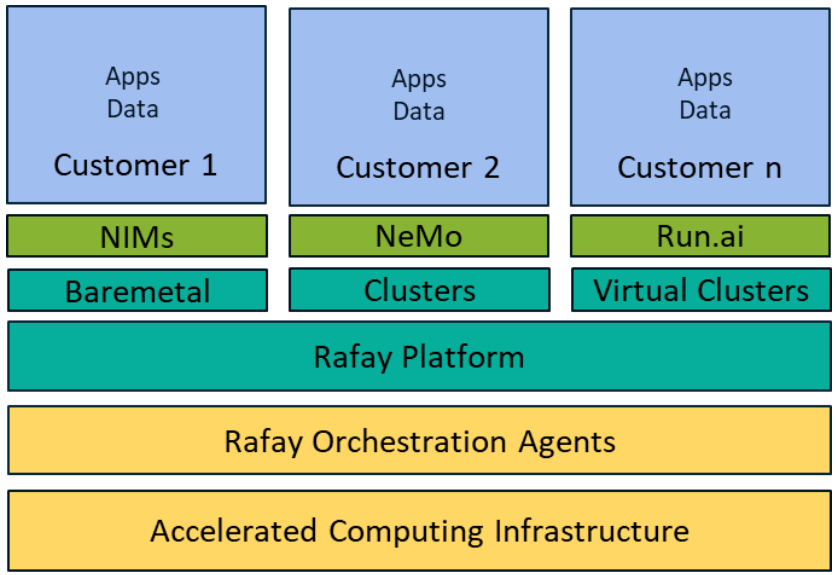
- Dynamically assigns infrastructure components to various tenants
- Managed through Netris web console or by integrating Netris APIs.
- Automatically configures the appropriate VRFs, VXLANs, and all other low-level networking details across every network switch and managed gateway
- Delivers highly scalable multi-tenancy and network-level isolation.



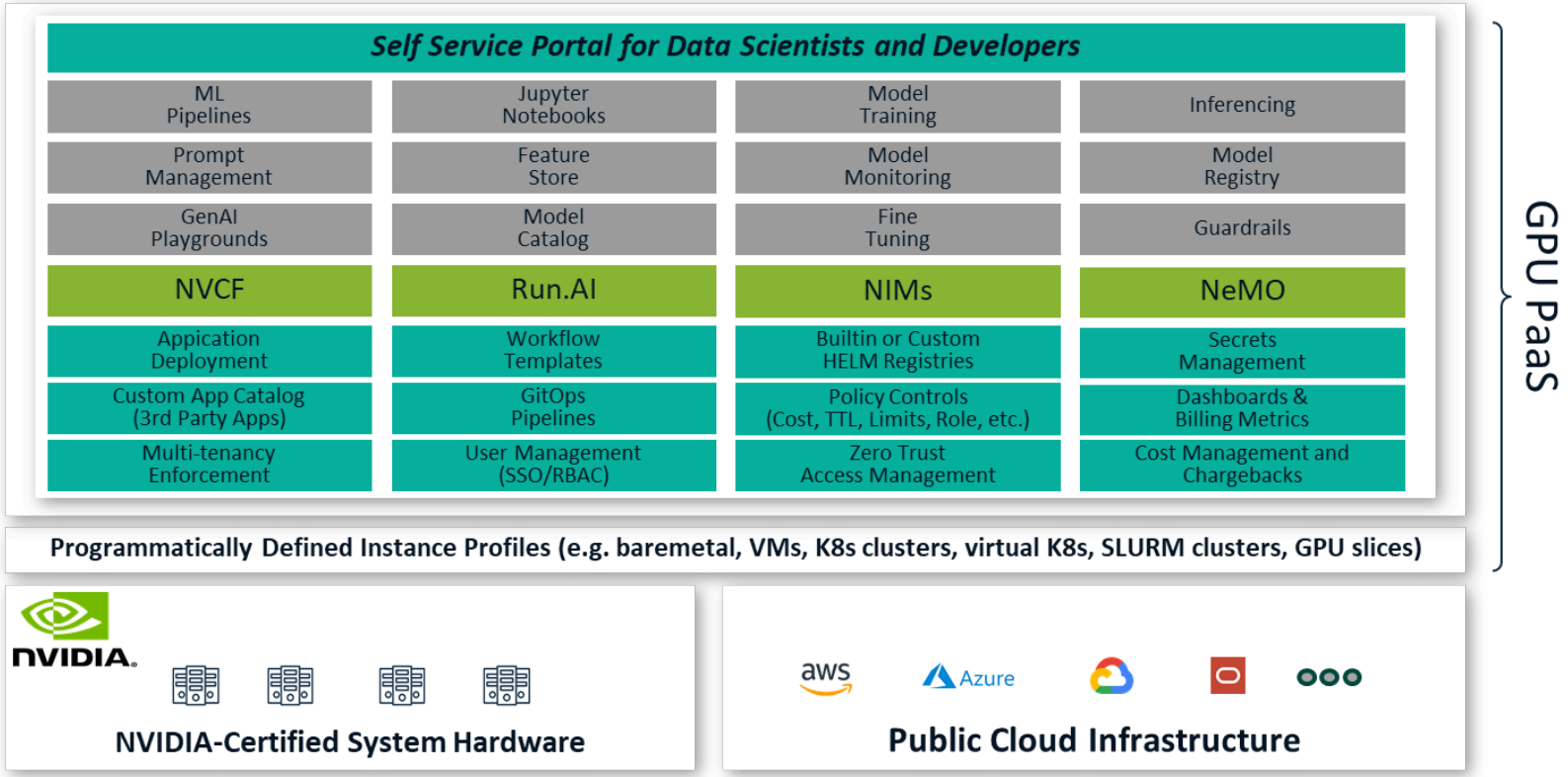
Infiniband multi-tenancy: access controlled with partition keys (PKEYs)

Rafay

Platform-as-a-Service for GPU & CPU Deployments
Infrastructure Orchestration and Workflow Automation for Cloud-Native and AI Innovations

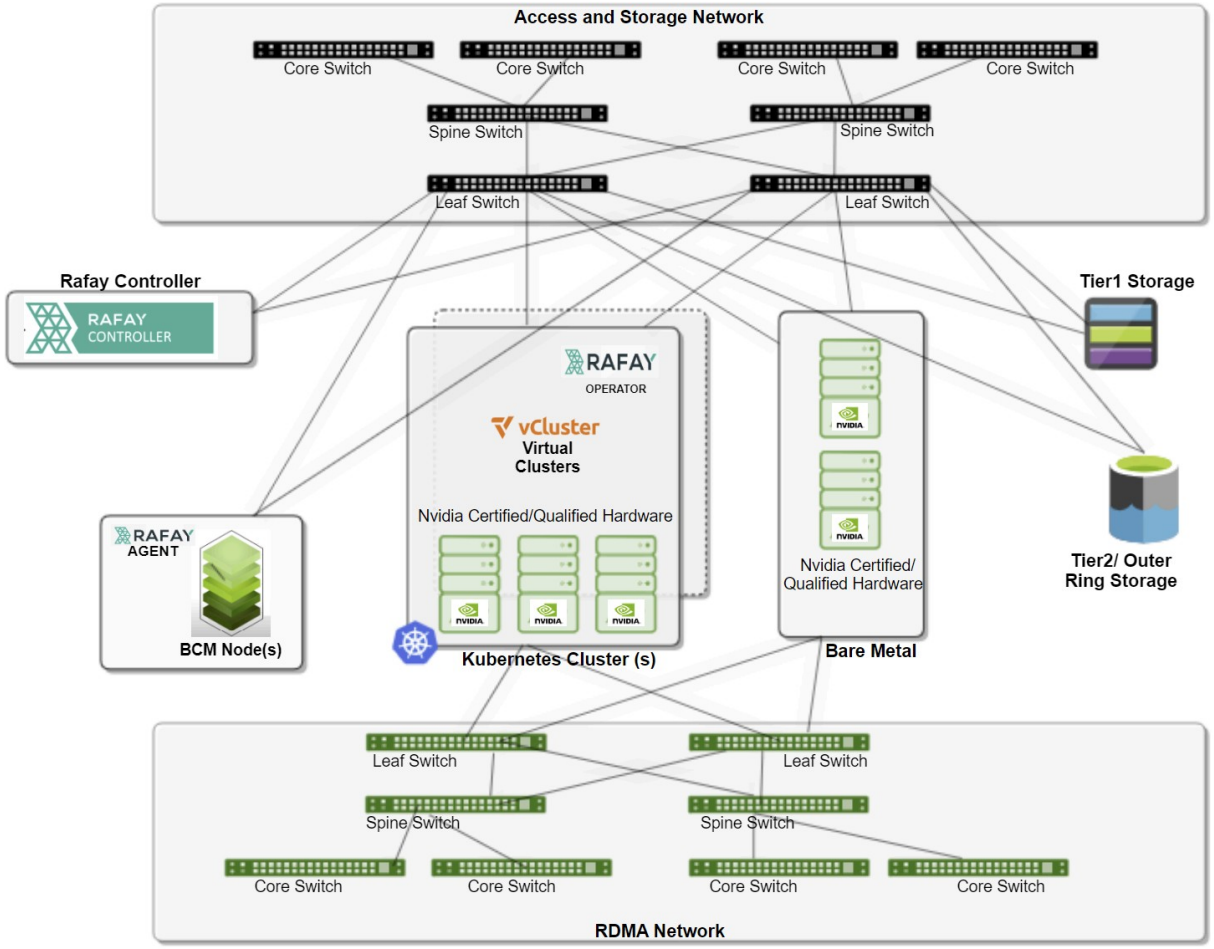


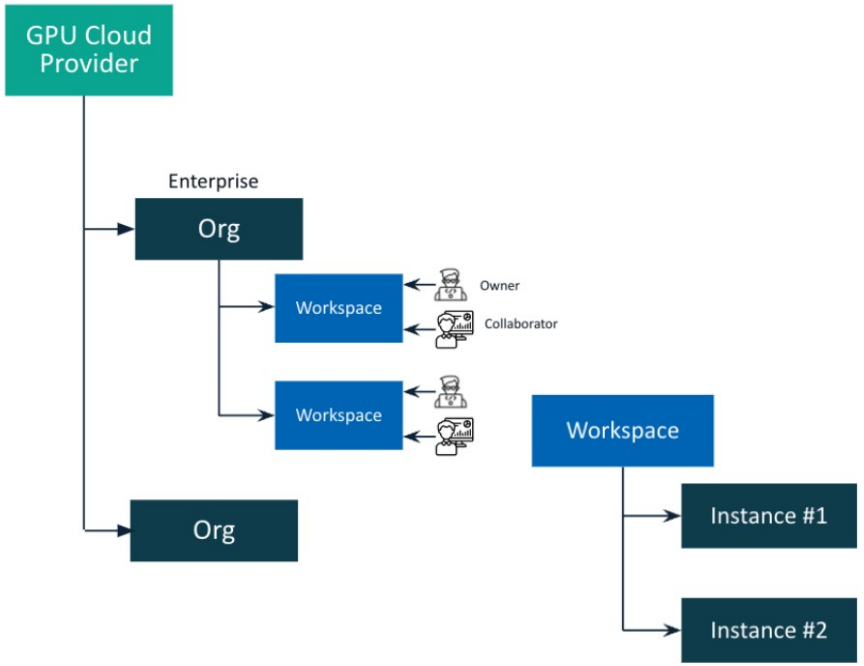
Rafay Operating Model



Rafay

Platform-as-a-Service for GPU & CPU Deployments

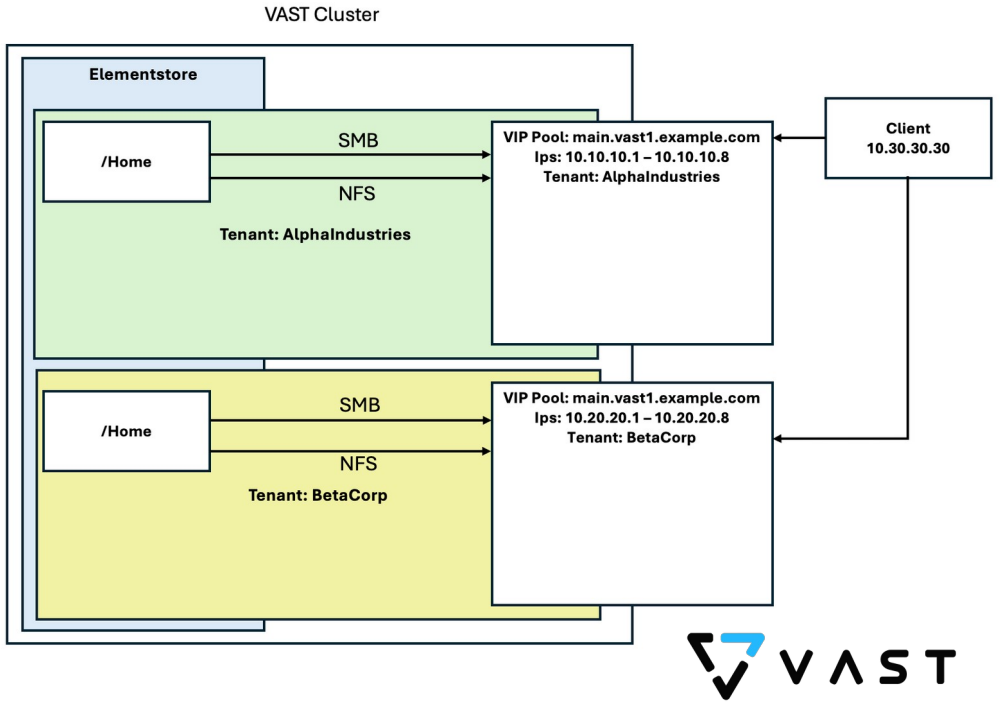


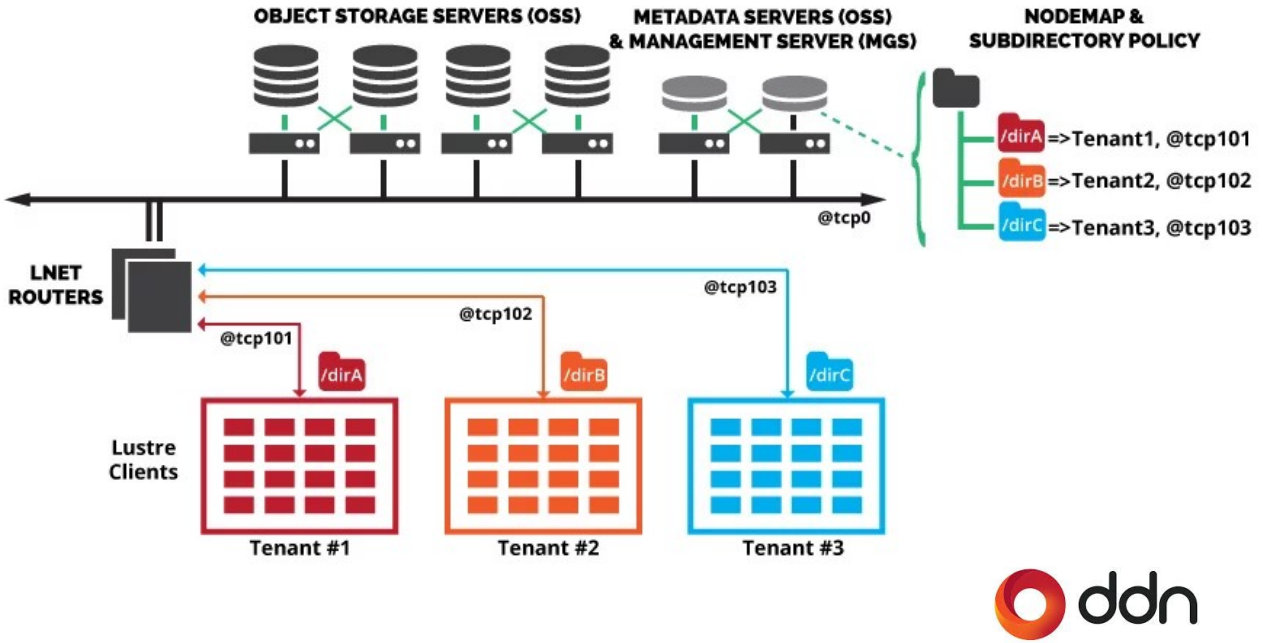


Storage Multi-Tenancy

Storage systems also need to support multi-tenancy:

- Separate storage areas for different tenants
- Integration with (different) identity providers: local, AD, LDAP, or NIS
- Unique encryption keys, quotas, and quality-of-service (QoS) settings
- Networking rules (restricting access between tenants/servers/clients/applications and storage)





Solutions available from various vendors (incl. NVIDIA certifications): VAST, DDN, Weka, IBM, etc.

Conclusion

Challenges

- Hard multi-tenacy is *hard*
- Soft multi-tenancy is less hard
- Manual network configuration becomes unsustainable very quickly
- Switching to a network automation solution requires a complete cluster reinstall

Learnings

- Hard vs soft multi-tenancy is a continuum of solutions
- Carefully assess requirements against this continuum
- Plan for future requirements (many customers moving towards “harder” multi-tenancy)
- Implement automation from the start
- Pick the right solution (and partner) for the implementation

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
E info@xenon.com.au

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