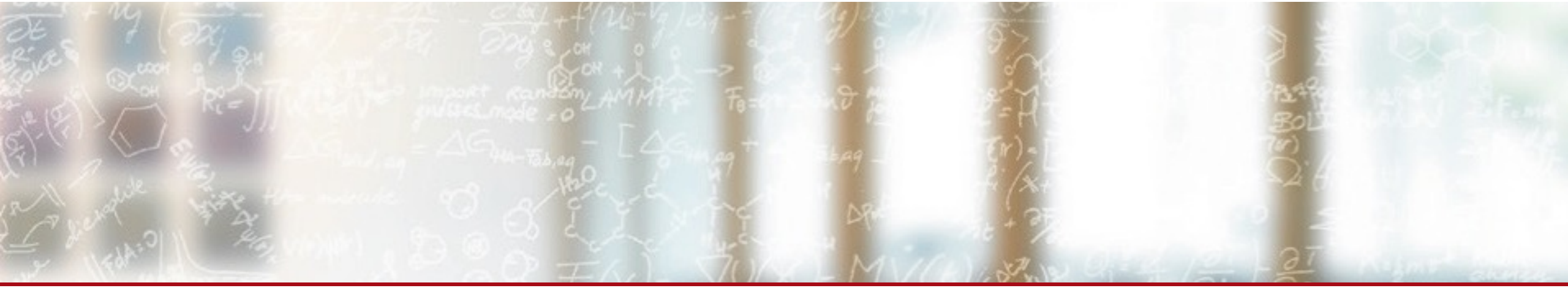
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



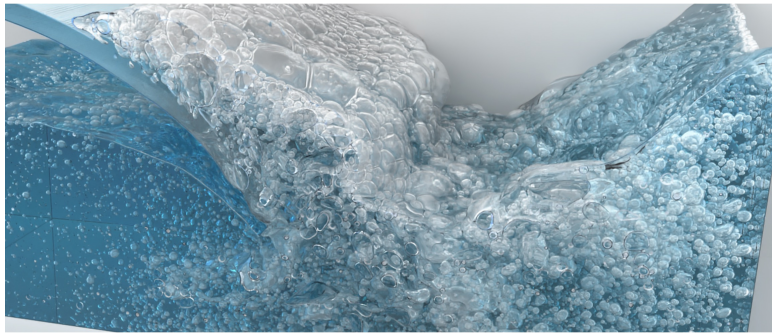
Versatile software-defined cluster
HPC and Cloud convergence

Cray User Group 2024

Ben Cumming, CSCS

HPC and Cloud convergence

HPC design principle – towards science



- Resources – bare metal
  - Compute nodes, high-speed network, high-speed large storage optimized for write
  - Exclusive access to compute resources with accelerators
  - Shared network and storage
- Services
  - Resource manager – node allocations with queues – usually large jobs (nodes and time)
  - Fully vertically integrated software stack from the kernel to the high-level library
  - Pre-defined user environments with limited choices
- Access
  - Design for interactive shell with SSH jumping through a chain of login nodes
  - Posix IAM (limited authentication and authorisation)

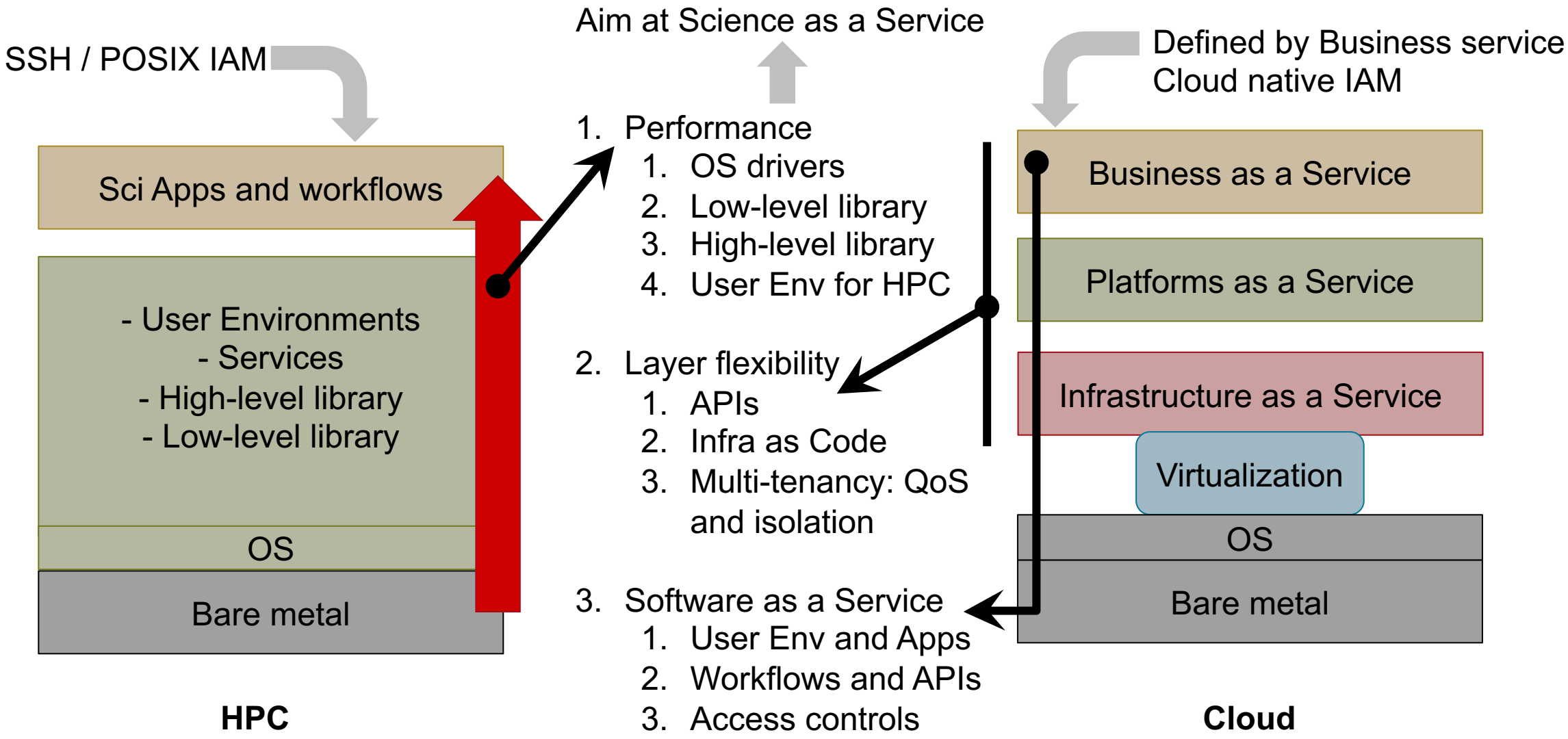
High performance ➔ fully integration of stack ➔ limited set of services

Cloud design principle – towards enterprise

- Resources – virtualisation
  - vResouces: vCPU, vGPU – uses virtual images/containers
  - Bare metal nodes runs many different vResources
  - Multi tenancy for compute, network and storage
  - Price model cheap for vResource to very expensive for bare metal
- Services
  - DIY infrastructure - define the configuration of your infrastructure
  - Bring your own service
  - Extensive offering of pre-configured services
- Access
  - UI in your browser, interactive shell in your browser or with SSH
  - Web IAM: complex protocol for authentication and authorisation

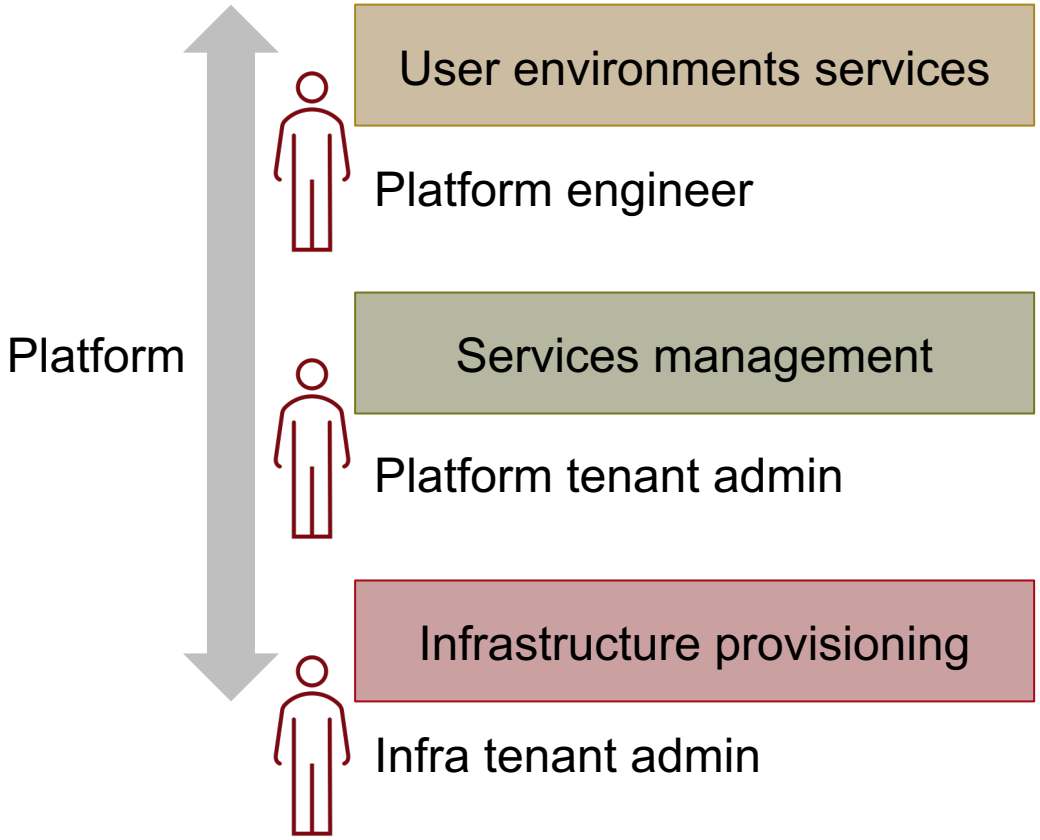
Virtualization at scale ➔ high flexibility ➔ limited performance

HPC and Cloud concepts to enable Science



Versatile software-defined cluster (vCluster)

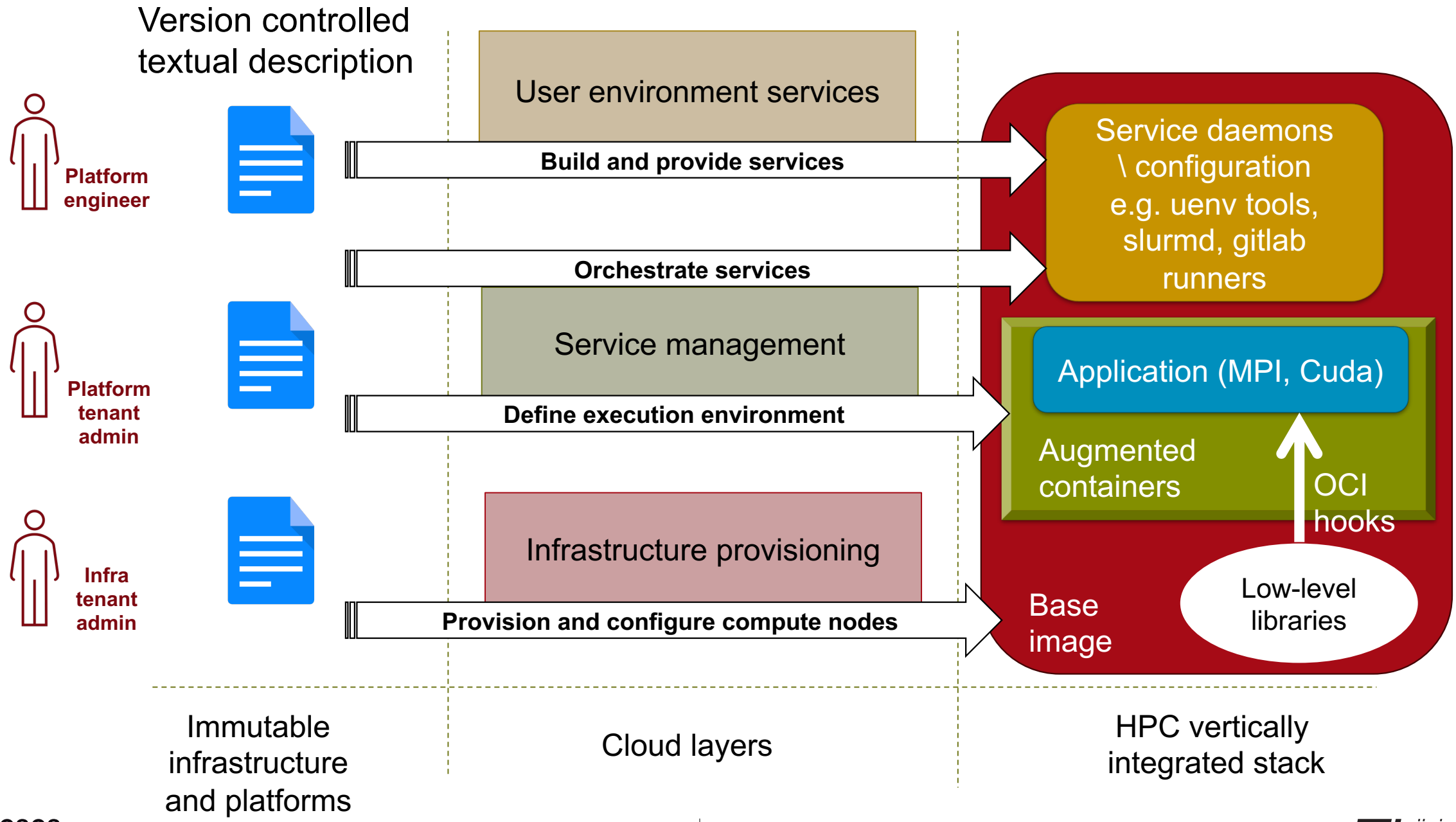
vCluster layers and tenant concept



- User tailored environments
- Programmable resource access
- Scientific application build services

- Orchestration of platform services
- Execution environments
- Soft tenant (resource labelling)

- Interface to the management plane
- Hard tenant (network segregation)


Version controlled
textual description
Platform
engineer
Platform
tenant
admin
Infra
tenant
admin
User environment services
Build and provide services
Orchestrate services
Service management
Define execution environment
Infrastructure provisioning
Provision and configure compute nodes
Service daemons
\ configuration
e.g. uenv tools,
slurmd, gitlab
runners
Application (MPI, Cuda)
Augmented
containers
OCI
hooks
Low-level
libraries
Base
image
Immutable
infrastructure
and platforms
Cloud layers
HPC vertically
integrated stack

User environment services

- Programming environment management and deployment
  - Uenv: self-contained programming environments and applications
  - Spack configurator
  - Containerized applications

- Resource access
  - Programmable resource access with FirecREST a Restful API
  - Interactive resource access: Jupyter, web interface, shell
  - Batch scheduler, orchestrator, or other scheduler types

- Sustainable scientific development
  - External CI/CD
  - Reframe: testing framework

Service management

- Service orchestration
  - Manage services via Nomad orchestration
- Soft tenant: label resources

- Execution environment
  - Job running inside containers
  - OCI hooks augment containers with HPC capability

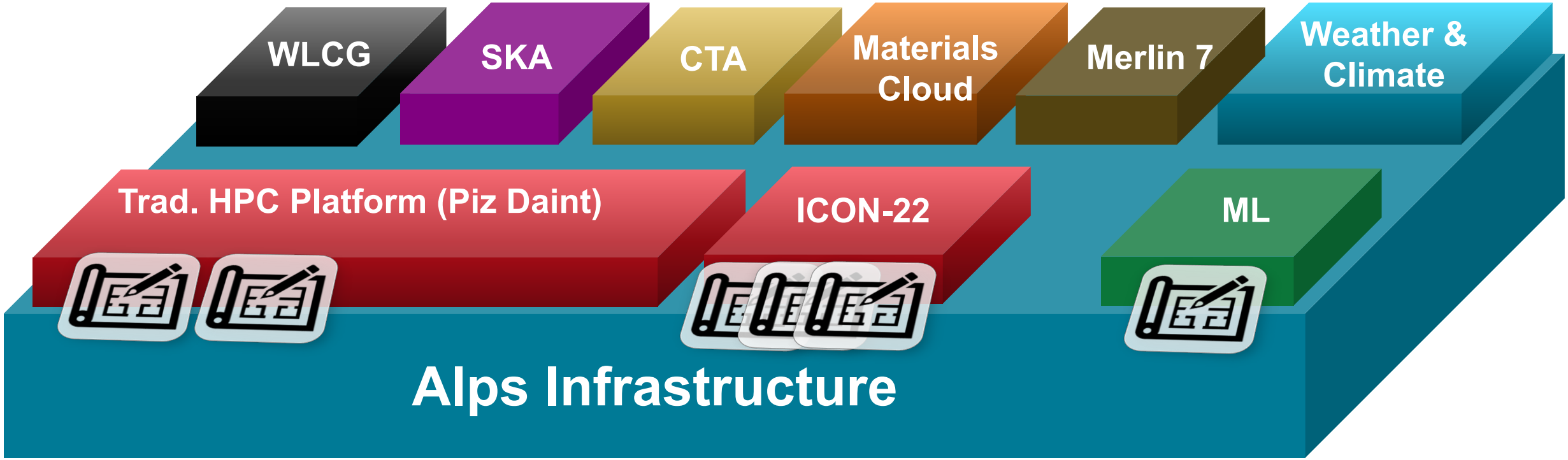
Infrastructure provisioning

- API interaction with management plane
  - Manta to manage and provisioned resources
- Hard tenant: network segregation

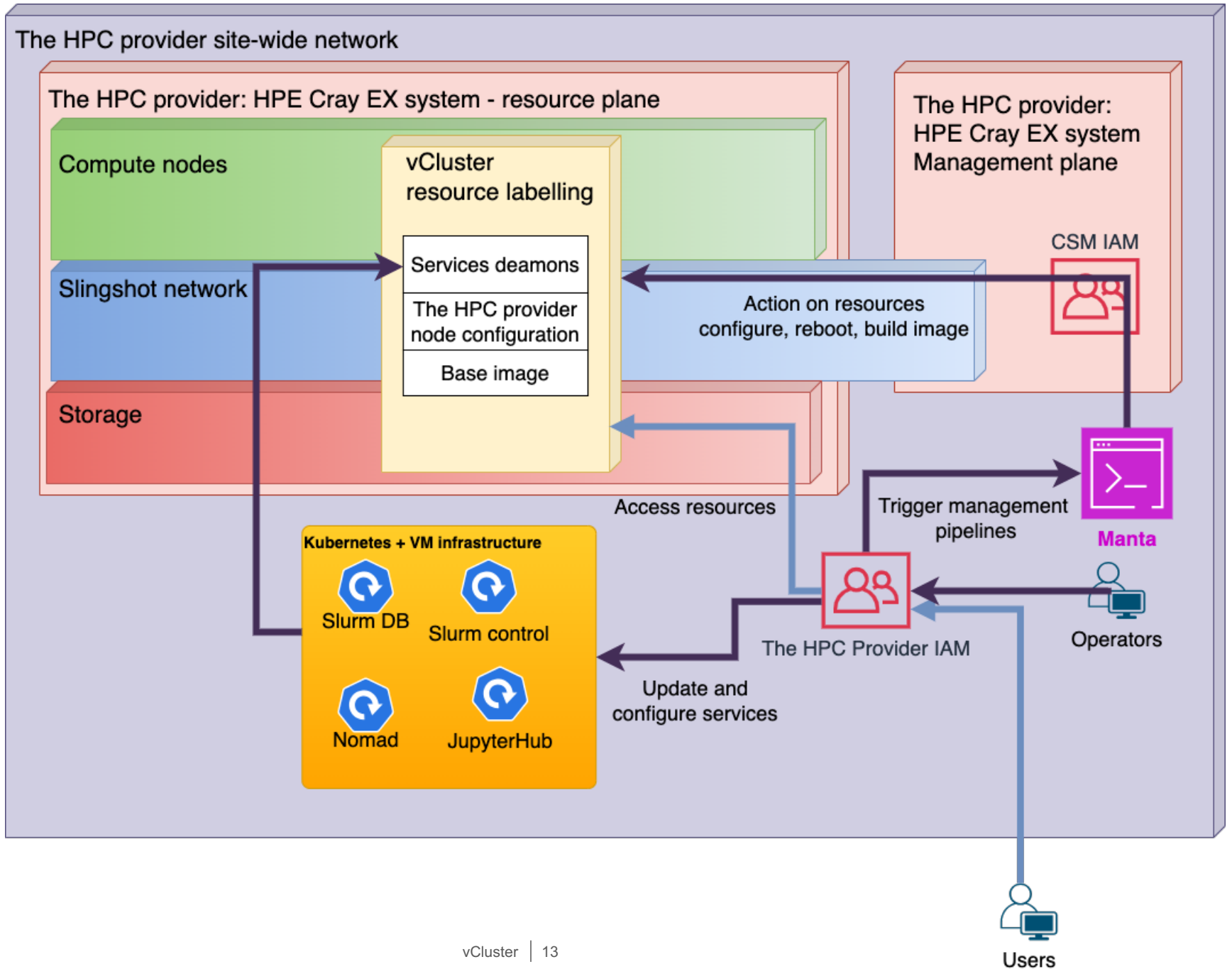
vCluster on Alps

Tenants, Platforms and vClusters

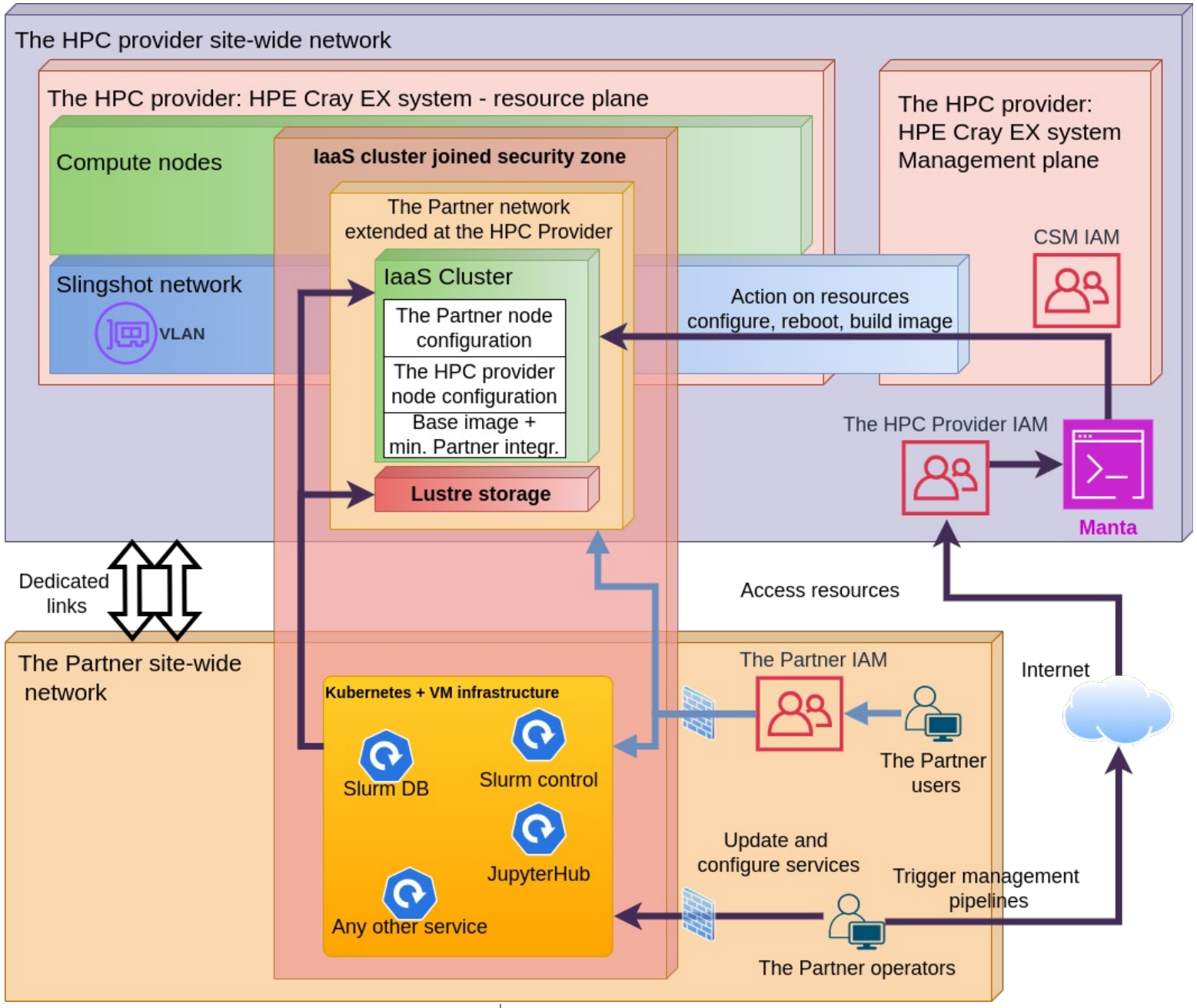
- Infra tenant: institution having access to the management plane
- Platform tenant: institution, project, community that manages the services on a set of platforms
- Platform: a set of vClusters that answer a business/scientific need
- vCluster: a set of services



Soft tenancy



Hard tenancy



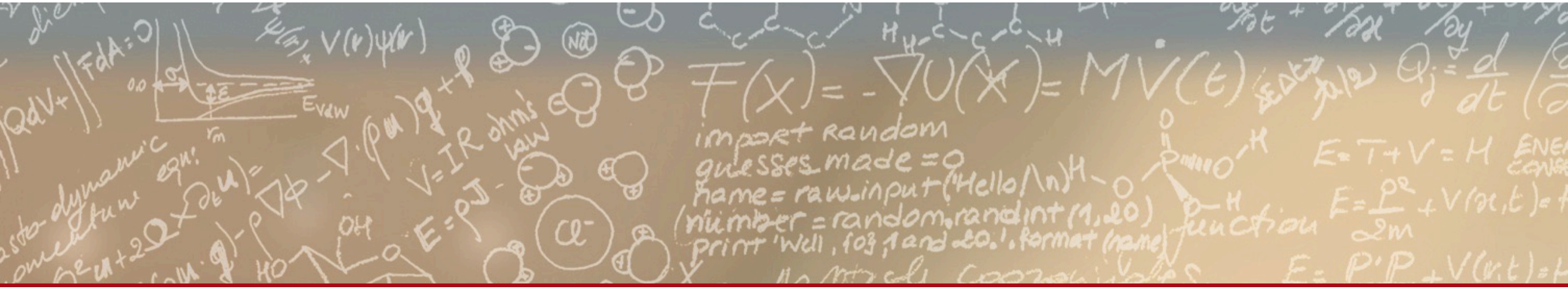
vCluster Mobility

- On Cloud providers for company business scalability
  - Need to understand infrastructure provisioning
  - Common abstraction layer
- On redundant infrastructure at the same site for ensuring service availability during maintenance
- On redundant infrastructure at different sites for geo-redundancy service
  - Synchronisation of control planes across sites

AlpsE at EPFL

Wrap-up

- vCluster is a set of technology
  - User environments
  - Service management
  - Infrastructure provisioning
- Enable abstractions with performance
  - Soft and Hard tenancy
  - Augmented containers using OCI Hooks
- Offers multi-tenancy on a single large HPC infrastructure
  - HPC and Cloud convergence



Thank you for your attention.