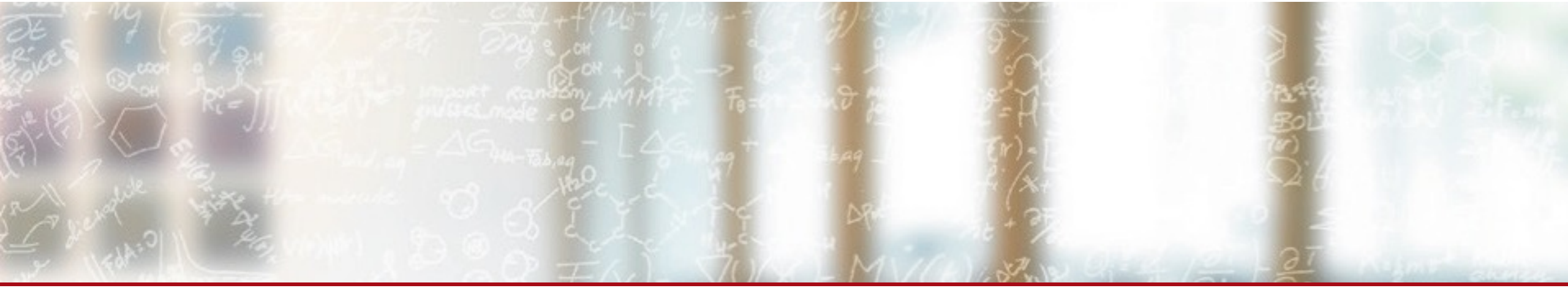
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



Alps technology

CSCS User Lab Day 2024

Maxime Martinasso, CSCS

Alps Technology in a nutshell

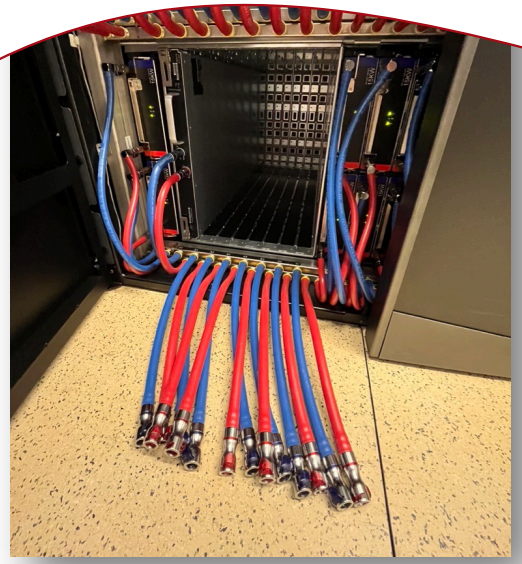
- Architectural concept: network end points for resources
- Heterogeneous infrastructure (Nvidia GPU, AMD GPU, x86, ARM,…)
- Managed by a micro service architecture control plane (CSM/OpenCHAMI)
- Slingshot network: performance and zone segregation
- Distributed Alps (multiple geo-distributed infrastructure)
- Versatile software-defined Cluster (vCluster) technology
  - Convergence Cloud and HPC
- Multitenant infrastructure
- Science as a Service concept with innovative resource access

Alps Research Infrastructure

- Alps is an HPE Cray EX supercomputer being our new flagship infrastructure

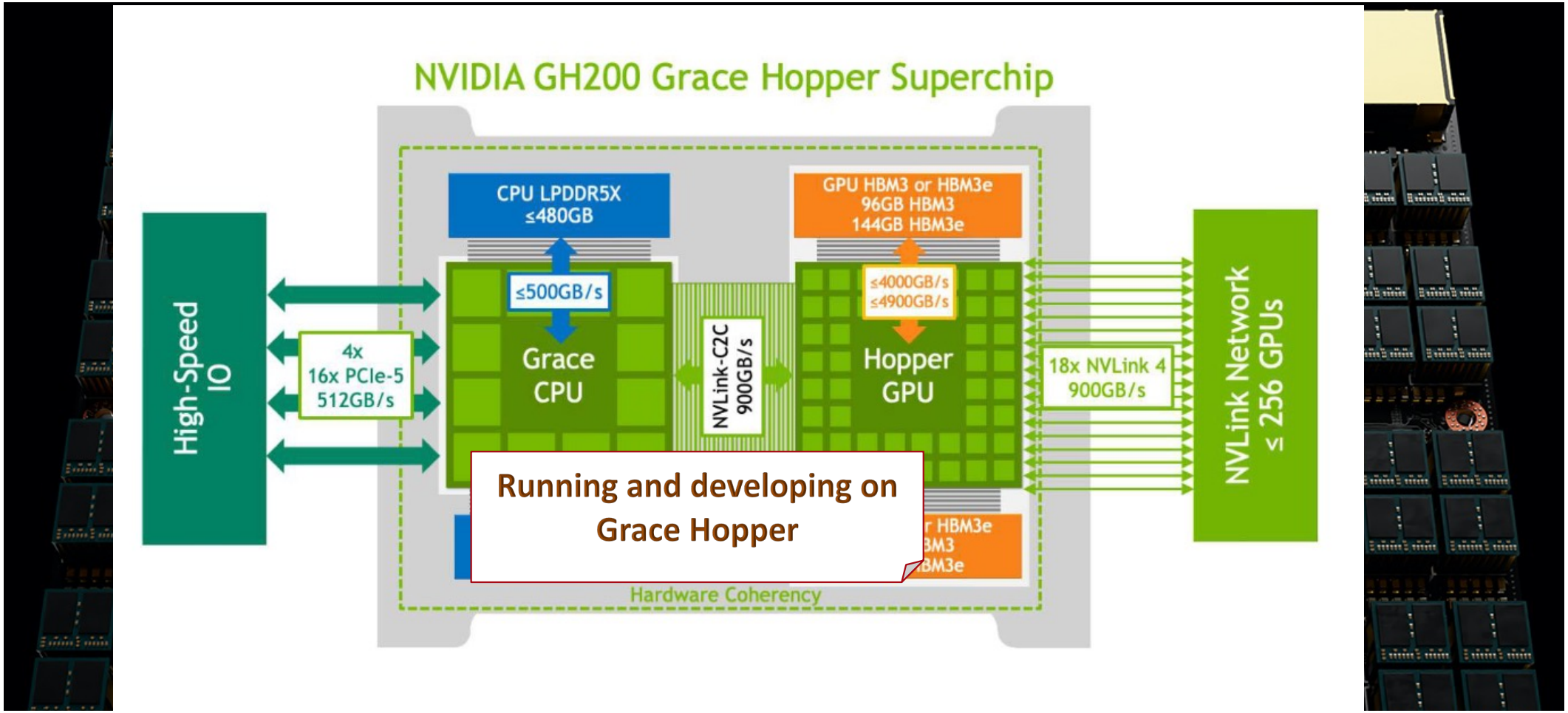
- Some specs
  - 1024 AMD Rome-7742 nodes 256/512GB
  - 144 Nvidia A100 GPU nodes
  - 24 AMD MI250x GPU nodes (LUMI1 type)
  - 128 AMD MI300A GPU nodes (24Q4)
  - ***2688 Grace-Hopper nodes***
  - Slingshot network (200 Gbps injection)
  - Two availability zones (HA, non-HA)
  - 100% liquid cooled
  - 100+10 PiB HDD
  - 5+1 PiB SSD (RAID10)
  - 100s of PiB tape library
  - ~10 MW (envelope for power and cooling)

Inauguration of "Alps" – Open Day
Saturday 14th of September 2-5pm
LUGANO



Water cooled blades

Grace-Hopper superchip (GH200)



Problem statements – more than an HPC infra

One-size-fits-all approach for HPC

- HPC systems provide a vertically integrated stack

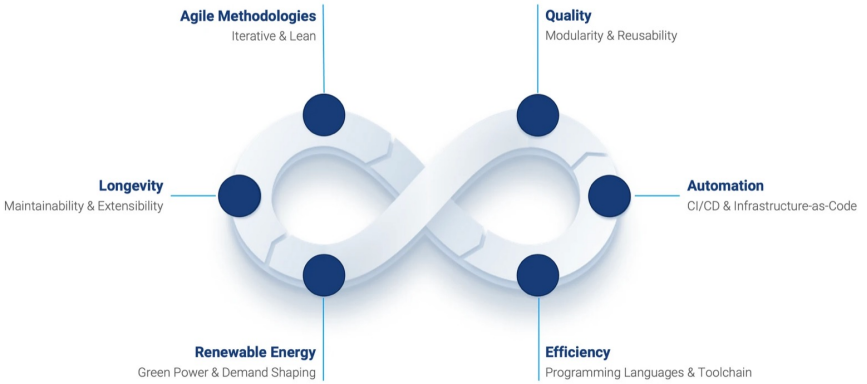
- User Environments
- Services
- High-level library
- Low-level library
- OS

1. Flexibility of the programming env. is minimal
2. Composability of services is limited to few options
3. Upgrading means service disruption and forcing the rebuild of the entire upper stack

How to use and build HPC software at CSCS: a practical introduction

→ Separate community of users and provide them with custom services

Sustainable software development



- Scientific software have a longer lifespan than supercomputers

1. Code will be refactored to use latest hardware (accelerators) leading to costly scientific validation of outputs
2. Hardware heterogeneity + new programing env. lead to combinatorial number of tests

Continuous Integration and Development for Scientific Applications

→ Adapt supercomputer services to application sustainability needs

Flexible scientific workflows

- Simplify access to HPC resources for workflow to increase researcher efficiency

1. Need programmable interfaces to HPC resources
2. Bring your own software stack or user environments (example ML) without compromising on performance

→ Use REST API and containers to facilitate scientific workflows

Automate your HPC workflows with FirecREST API and Sarus

ML and PyTorch in containers

Alps technology - vCluster

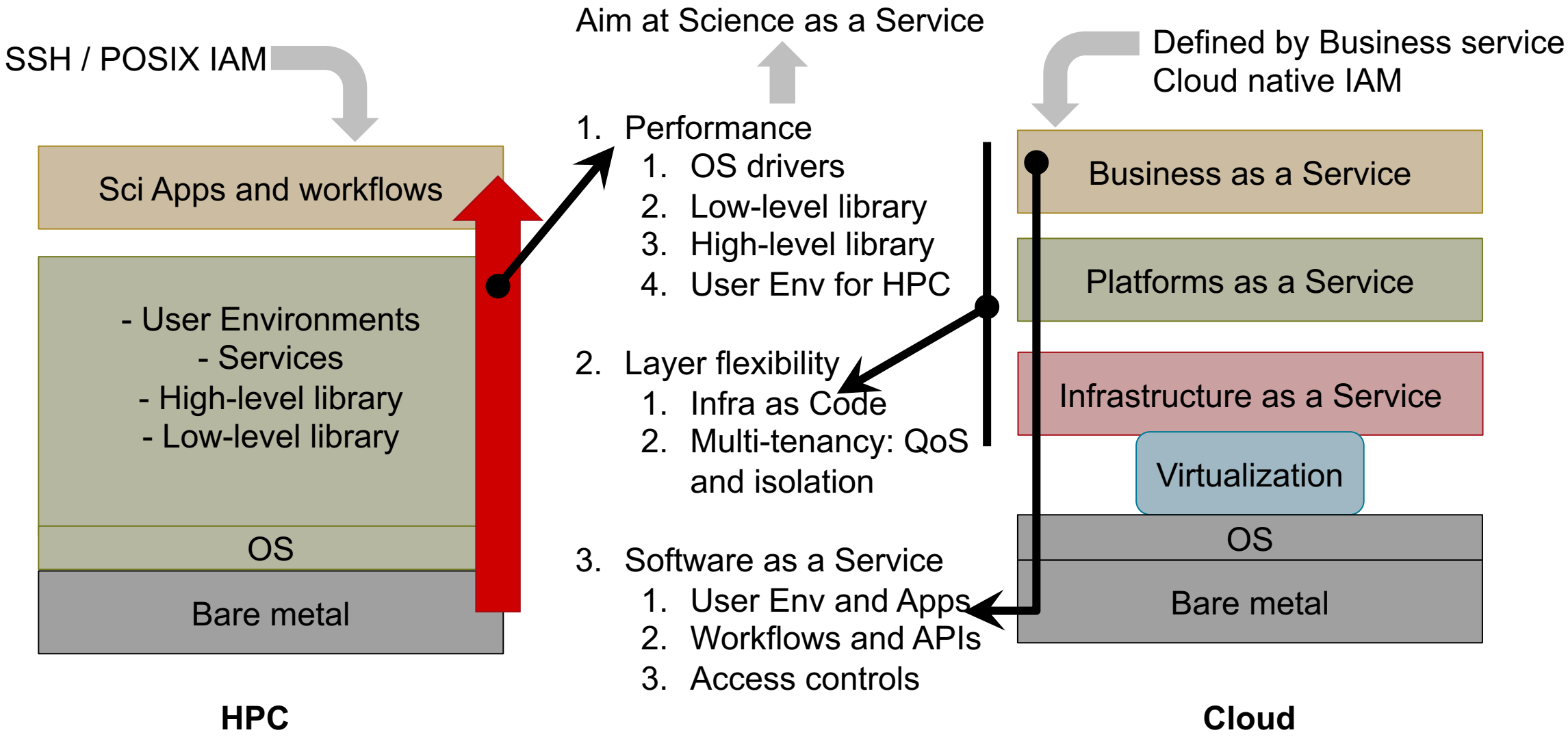
Versatile software-defined cluster
vCluster technology

- Concept of Cloud and HPC convergence

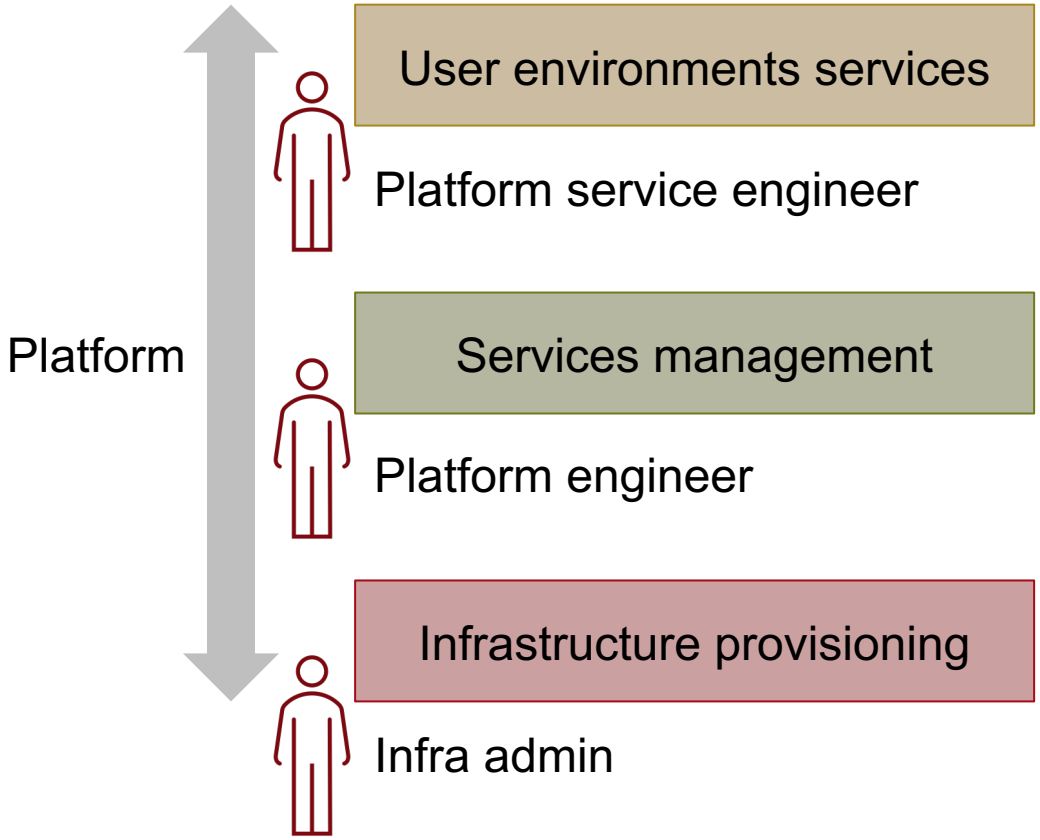
1. HPC: High performance ➔ vertically integrated stack ➔ limited set of services
2. Cloud: Virtualization at scale ➔ high flexibility ➔ limited performance

→ vCluster is a set of technologies to enable service flexibility on top of HPC

HPC and Cloud concepts to enable Science



vCluster layers and tenant concept

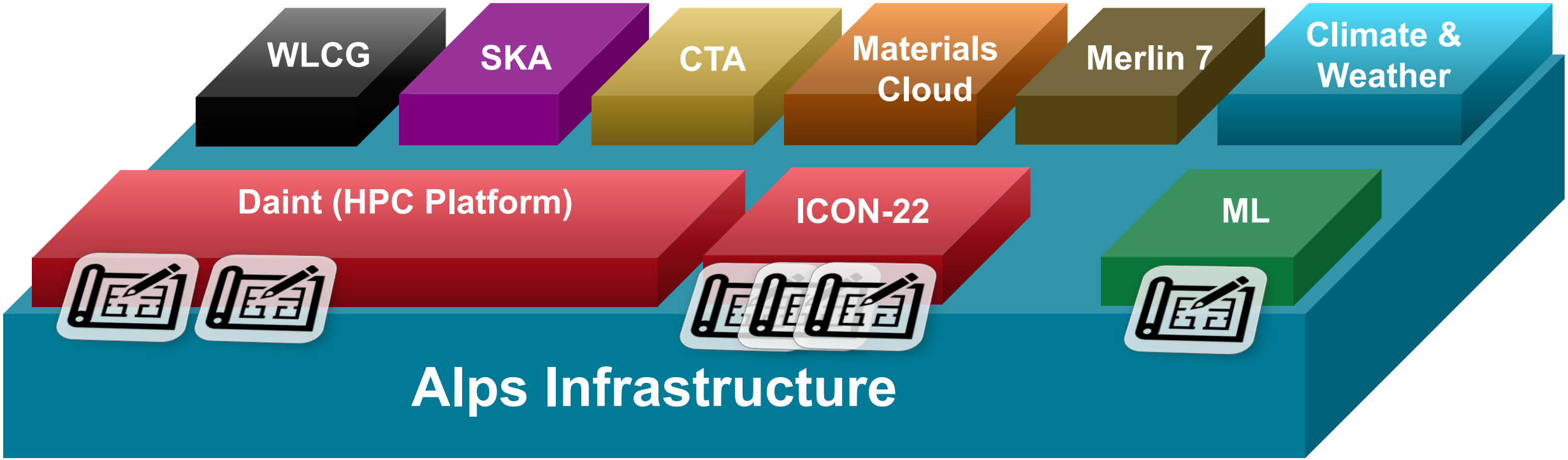


- User tailored environments
- Programmable resource access
- Scientific application build services

- Orchestration of platform services
- Execution environments

- Interface to the management plane
- Network segregation

Platforms and vClusters



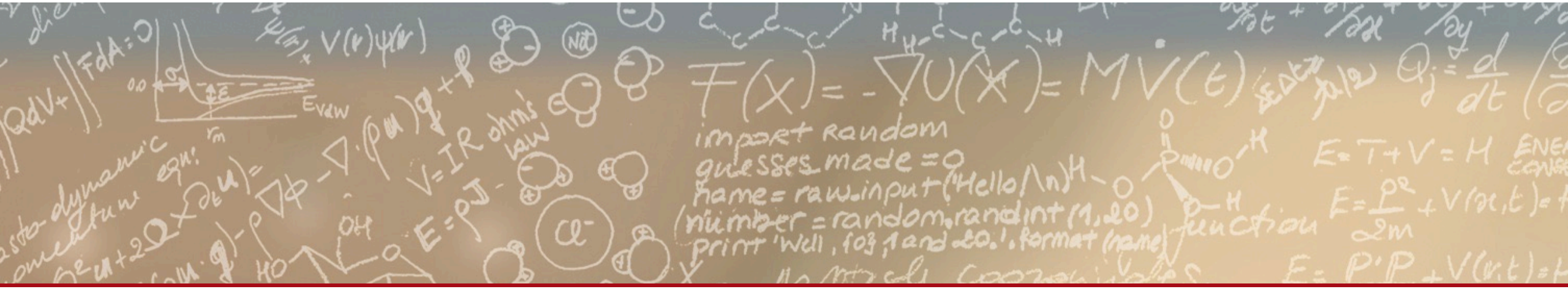
General information for working on Alps

On-going and future technology developments

- vCluster and Alps in practice
  - On-going work to mature the technology
  - Multiple platform developments: HPC, ML, C&W,…
  - IaaS use cases in production
  - Data-bridges access and usage

- Develop and increase adoptions of APIs for resource access and configuration
  - FirecREST, API Gateway, Sarus, Container engine
  - CI/CD pipelines, user environments

- Identify new technology opportunities to enhance our services
  - vCluster elasticity, on-demand storage, multi-interface data managers, no login nodes, power-aware scheduling, zero-trust architecture, domain specific language and intermediate representation, DPU on network cards, code identification, LLM bots and user tickets,…



Thank you for your attention.