CSCS site report

LAD 2024, Paris
Marco Passerini, CSCS

CSCS, the Swiss National Supercomputing Centre

- Founded in 1991

- "We develop and operate a High Performance Computing and data research infrastructure that supports world-class science in Switzerland."

- 141 employees
    - ~100 in Lugano
   - ~40 in Zurich
    - 22 nationalities

Alps technology in a nutshell

- **Science as a Service** concept with innovative resource access

- Architectural concept: **network end points** for resources

- Versatile software-defined Cluster (**vCluster**) technology
  - Convergence Cloud and HPC

- **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)

- **Multitenant** infrastructure

cscs

ETH zürich

Alps Research Infrastructure

- Alps is an **HPE Cray EX** supercomputer being our new flagship infrastructure
  - Multi-phase installation started in 2020
- Specs:
  - ***2688 Grace-Hopper nodes***
    - ***~10752 GPUs***
  - 1024 AMD Rome-7742 nodes 256/512GB
  - 144 Nvidia A100 GPU nodes
  - 24 AMD MI250x GPU nodes (LUMI1 type)
  - 128 AMD MI300A GPU nodes
  - Slingshot network (200 Gbps injection)
    - (GH nodes have 4 NICs)
  - Two availability zones (HA, non-HA)
  - 100% liquid cooled
  - ~10 MW (envelope for power and cooling)
- Performance
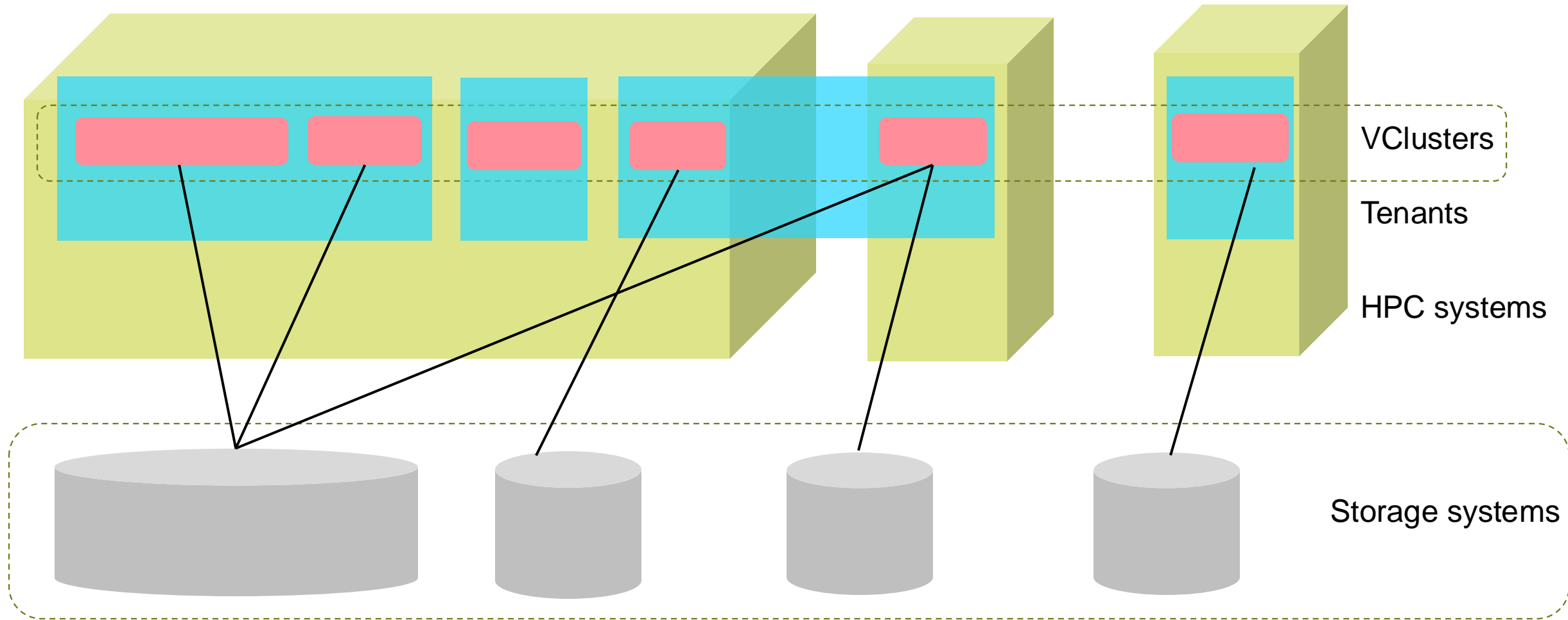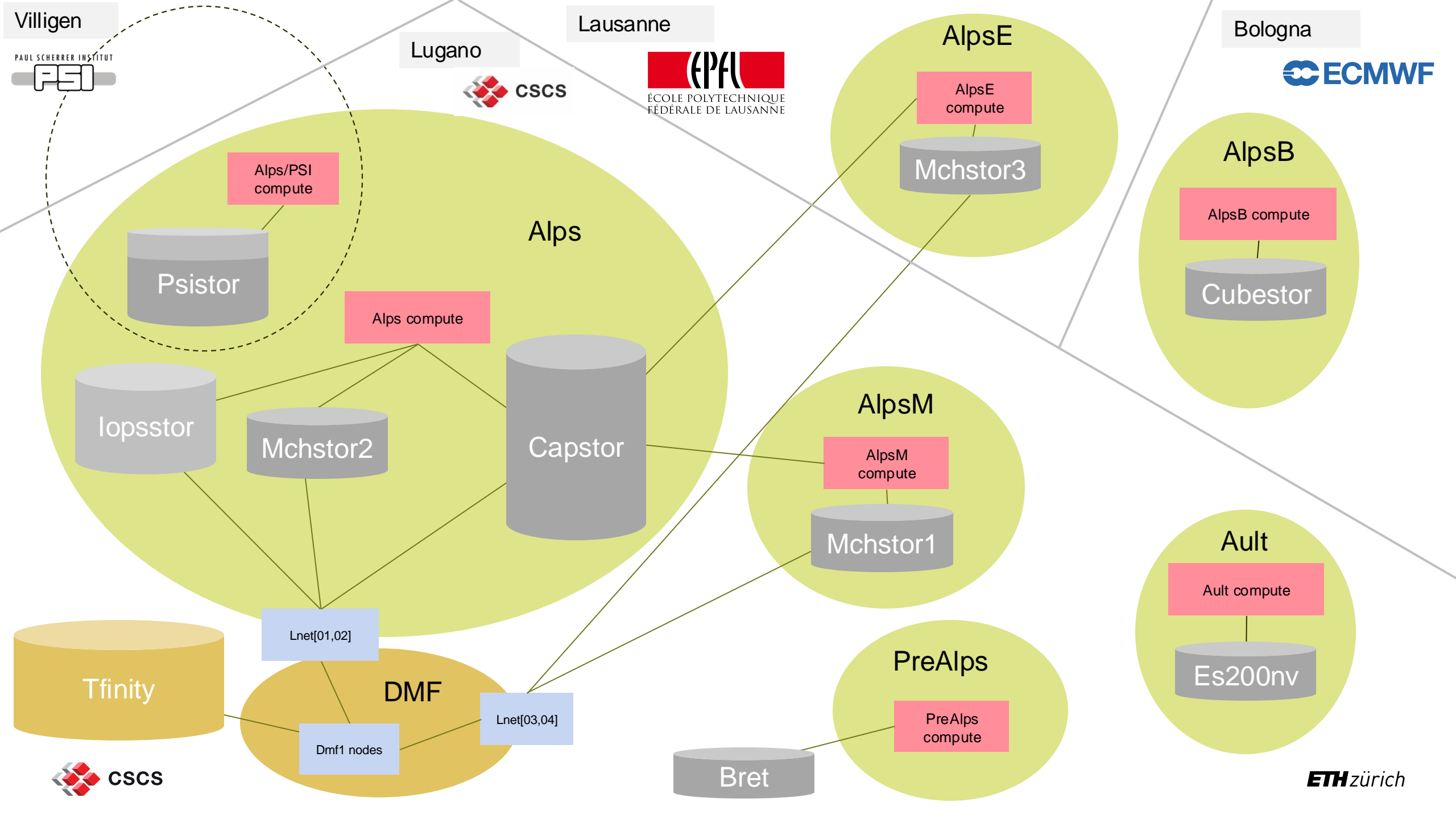 - **~353 PF/s Rpeak** (Top500 #6, 06/2024)
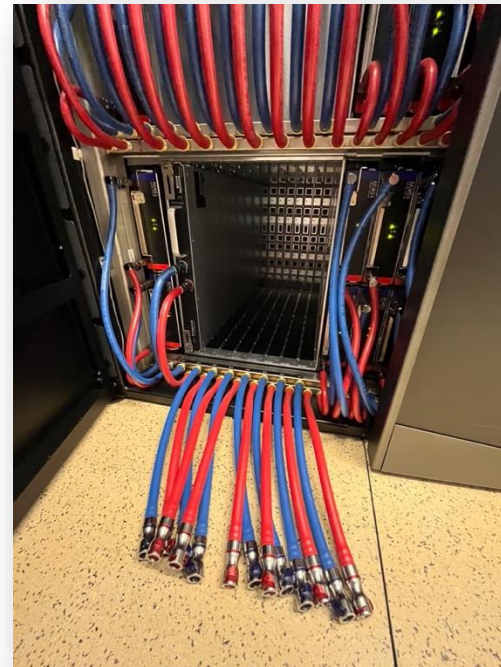
Water cooled blades

Storage technologies

- Lustre
- HPE DMF, ZWS
- Spectra Tfinity, IBM TS4500
- Spectra Blackpearl
- Vast
- CEPH
- DCache
- Globus online
- Spectrum Scale
- Spectrum protect
- …

VClusters and tenants

VClusters

Tenants

HPC systems

Storage systems

cscs

ETH *zürich*

Lustre systems (1/3)

- **Capstor**
    - HPE ClusterStor E1000D
    - **129 PiB** raw GridRAID
    - 8480 **HDDs** (16 TB)
        - 80 disk enclosures
    - 11 racks
    - 6 MDS, 80 OSS
    - Slingshot 11
    - Performance:
        - **~1.09 TB/s write, 1.19 TB/s read**
            - (ARM different page size! > tunings)
        - 1.5M IOPS
        - 373K file creates/sec
    - Neo 6.6-010, lustre 2.15

- **Iopsstor**
    - HPE ClusterStor E1000F
    - **7.2 PiB** raw RAID 10
    - 240 **NVMe** devices (30TB each)
    - 1 rack
    - 2 MDS, 20 OSS
    - Slingshot 11
    - Performance:
        - **393 GB/s Write | 782 GB/s Read**
        - **24M write, 8.6M read IOPS**
        - 214K file creates/sec

CSCS

ETH zürich

Lustre systems (2/3)

- **Mchstor1, Mchstor2, Mchstor3**
  - Each:
    - HPE ClusterStor E1000D
    - 2 MDS, 2 OSS
    - **~1.2PB** raw (6TB HDDs) GridRAID
    - Performance
      - 21 GB/s write
      - 26 GB/s read
      - 150K file creates/sec

- **Psistor**
  - HPE ClusterStor E1000D/F
  - 2 MDS, 12 OSS
  - ~9.2 PB
    - **~300 TB SSD** raw RAID10
    - **~13.5 PB HDD** raw GridRAID
  - Performance:
    - HDD 120 GB/s write 110 GB/s read
    - SSD 47 GB/s write, 120 GB/s read
      - 4.6M write, 1,6M read IOPS

CSCS

ETH zürich

Lustre systems (3/3)

- **Cubestor**
  - HPE Clusterstor E1000D
  - 2 MDS, 2 OSS
  - **~2PB** raw GridRAID
  - 16 TB HDD drives

- **Bret**
  - HPE Clusterstor L300
  - 2MDS, 2 OSS
  - **~224TB HDD**

- **DDN ES200NV**
  - 2 MDS, 2 OSS
  - **~80TB SSD**

- Currently decommissioning:
  - Snx3000 (Piz Daint storage)
  - Snx3000tds
  - snxarolla
 - snxtsa
  - DDN ES7K

ETH zürich

HPE DMF

- Use cases:
  - **Backups** of /users and /store
  - **HSM** for /archive
  - Scratch **cleanup scripts**

- **Architecture**
  - Servers:
    - 1x management node
    - 5x DB nodes
    - 2x datamovers for tape
    - 2x datamovers for ZWS
  - Storage:
    - Spectra Tfinity
      - ~40 PB
      - 2230x LTO9 tapes 18TB
      - 20 drives assigned
    - ZWS
      - 3.4 PB, 212x 16TB HDD

User environment

- Automatic **folder creation** and quota assignments using Waldur and internal accounting system
- Folder structures for "**multitenancy**"
  - */fs_name/function/tentant/(customer)/user_or_project*
- **Load distribution**:
  - 2x MDS for /users and /store
  - 4x MDS for /scratch
  - DNE1 with round robin assignments of folders (alphanumeric)
- **ACLs**
  - Let user-support team access user files, preserving UID/GID for users
- **Project quotas** soft/hard
  - Custom scripts and DB to manage them
  - Grace time
- **Changelogs** enabled for DMF
- **Cleanup policies**
  - Custom scripts based on DMF DB
- Metadata **bottlenecks**
  - Compilation on Tempfs
 - Squashfs
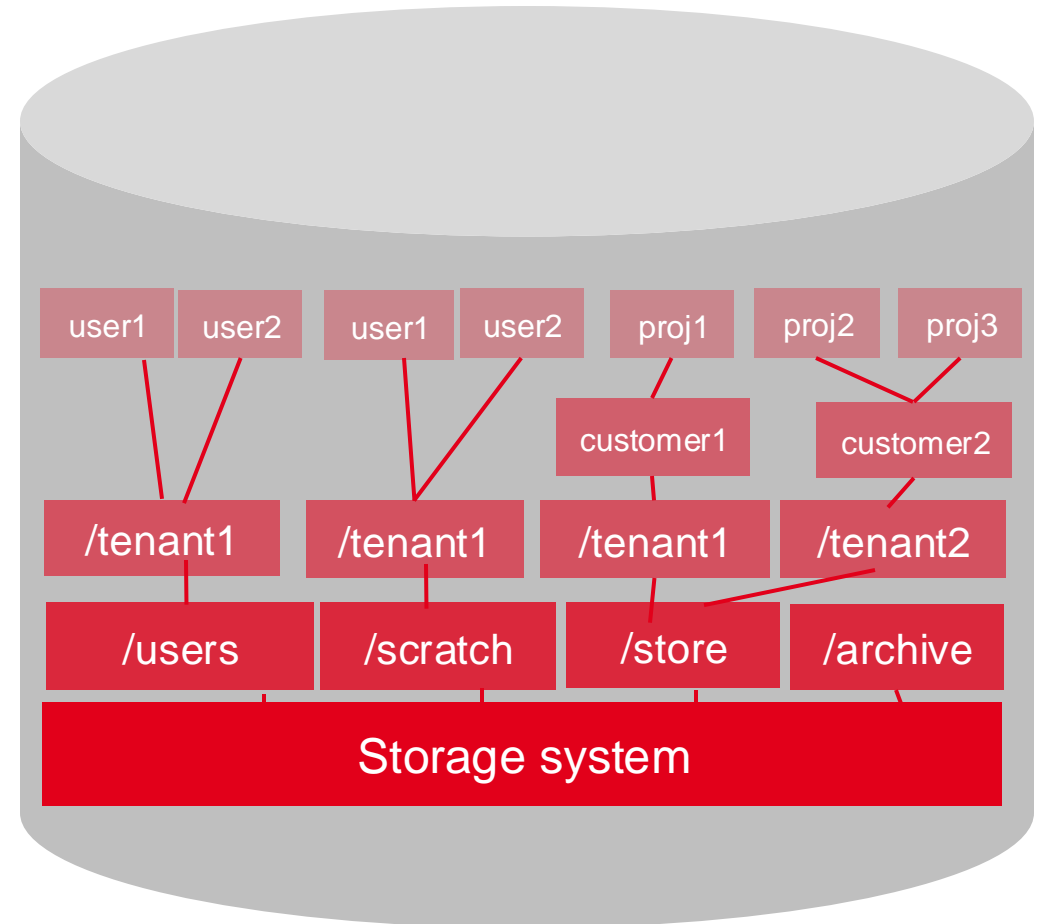
Lustre multitenancy

- What we evaluated:
  - **Multiple VLANs** on same storage (POC with DDN)
    - Limitations in maximum amount of VLANs (32)
    - Patch to dynamically adding VLANs
  - **QOS**
    - **Dedicated MDTs**
      - Can be restricted to root
    - **OST pools**
      - Users with normal permissions can override
  - **Authentication**
    - Tenants with their own LDAP
    - Identities must be replicated to the Lustre backend, UID/GID remapping
  - **Quotas**
    - Anybody with root can change them
- What we implemented:
  - **Managed** vclusters (tenants don't have root)
    - Subdirectory mounts to vclusters
  - **Unmanaged** vclusters (tenants with root access)
    - Dedicated storage systems, on dedicated VLANs

Lustre network configuration

- **TCP/KFI** UDSP dual mode
  - Routed outside SS network
  - no dedicated LNET servers
  - **Priorities:**
    - *lnetctl udsp add --src kfi --priority 0*
    - *lnetctl udsp add --src tcp --priority -1*
    - *lnetctl net show -v 4 | grep -P 'nid|priority'*
    - *lnetctl net show -v | grep -P 'nid|send_count|recv_count'*

Lustre long distance mounts

- Mchstor3, **Lausanne, ~300KM distance**
  - Over the internet, encrypted
    - 2x 100Gbps links
  - Performance:
    - **16.6 GB/s write**
    - **11.4 GB/s read**
    - **3.5 ms ping latency**
  - MTU size 9000
  - No special tunings
    - Mixed user workload
  - DMF integration also works
- Cubestor, Bologna
  - Work in progress
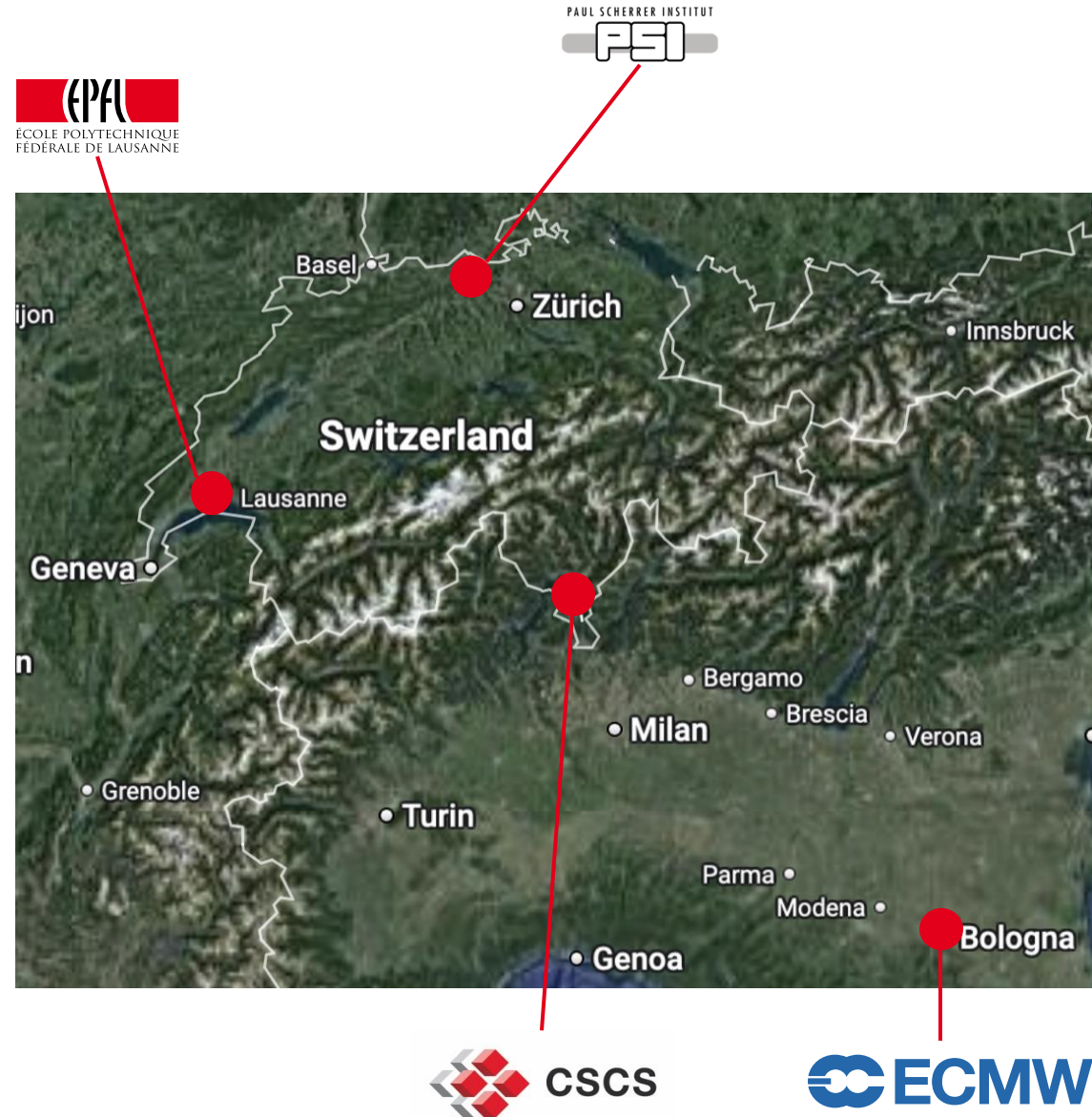
Monitoring (1/3)

- Nagios, Grafana
- Prometheus exporter (python) > Metricbeat > Kafka > ELK stack

- System stats
    - CPU Load
    - MDT IOPS
    - MDT IOPS/type
    - Bandwidth
    - Bandwidth/OST
    - Write IOPS/OST
    - Read IOPS/OST
    - OST RPCs
    - Capacity
    - Number of files

Monitoring (2/3)

- Lustre **job stats**
  - *lctl get_param *.*.job_stats*
  - Python script reads every minute the values through all OSSs, picking the max snapshot time
    - aggregate MD operations
    - aggregate write/reads from each OSS
  - We measure:
    - Job IOPS
    - Job Write Bytes/s
    - Job Read Bytes/s
  - Fine grained selection by
    - Storage system
    - Vcluster
    - User
    - Node
    - Job ID
  - Job name:
    - vcluster_jobid
    - vcluster_username_node
      - Set **$LUSTRE_JOB_ID** on a profile script

Monitoring (3/3)

- **Quota** reports

[DWDI] Lustre Scratch

Username	Folder	Files Used	Space Used
am░░░░░░░	/mnt/capstor/scratch/cscs/ar░░░░░░░	23,448,475	14,346,789,847,040
am░░░░░░	/mnt/capstor/scratch/cscs/am░░░░░	23,433,499	14,335,961,202,688
ah░░░░░░	/mnt/capstor/scratch/cscs/ah░░░░░	1,220,832	48,508,946,087,936
ah░░░░░	/mnt/capstor/scratch/cscs/ah░░░░	1,220,832	48,508,946,087,936
ah░░░░░	/mnt/capstor/scratch/cscs/ah░░░░	656,235	373,584,166,912

```
# lsattr -p -d /capstor/users/cscs/user_1
 2029 -----------------P- /capstor/users/cscs/user_1

# lfs quota -p 2029 /capstor/ -h
Disk quotas for prj 2029 (pid 2029):
    Filesystem   used   quota  limit  grace  files  quota   limit   grace
    /capstor/    29.4G   50G    50G     -      44  500000  500000      -
```

Data movement

- **Internal data transfer**
  - Xfer Slurm queue
  - Dedicated nodes for internal transfers between filesystems
  - 4 servers
- **External data transfer**
  - Globus
  - Gateway to I/O data from outside CSCS to our storage systems
  - 3 servers

Future work

- Quotas
  - **Notifications** for users
- QOS
  - **Lustre NRS TBF**
- **VAST** for /users
  - Snapshots
  - Metadata performance
  - Multitenancy built in
- **Blackpearl** S3
- **NVMEoF**
- **FirecREST** for data movement

CSCS

ETH zürich

Thank you for your attention.