PAUL SCHERRER INSTITUT

PSI

Marc Caubet Serrabou

HPCE AWI Group Update 2024-02-05

HPCE News

Merlin6

- **Batch system updates:**
  - **Slurm**
    - v23.02.7
    - **Slurmrestd** service deployed for merlin5, merlin6, gmerlin6 → first use case will be BIO
    - **Recommended OpenMPI software versions:** openmpi/4.1.5_slurm with ucx/1.14.1_slurm → Newer will come soon!
    - **Temporary fixed issues** with nodes not being able to run jobs from different partitions → SchedMD Bug #18616
      - Will improve usage of the cluster
  - **Other updates:** HP SPP software: 2022.09 → 2023.09
  - **GPU cluster:**
    - **Public nodes** based on cuda/12.1
    - **Gwendolen** based on cuda/12.3
    - cuda/12.2 must be avoided (HPCLOCAL-721)

BIO

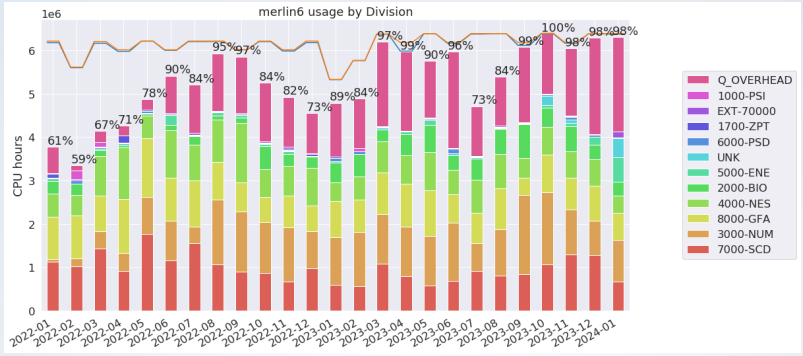
- **New BIO division head:** Michel Steinmetz
- **New EM @ A. Wanner (BIO group)** - needs 3-5 PB storage in the next 3 yrs
  - Probably RA (Leo involved), comp. needs (heavy! 10M CPU/h 1M GPU/h for ~40 days) at the moment @ Google clouds for free: needs to be discussed for the time after Google clouds (in 2 years time)
- **MX modules** will/are supposed to be taken over by **MX** (a few modules by Greta) - tutorial in two weeks

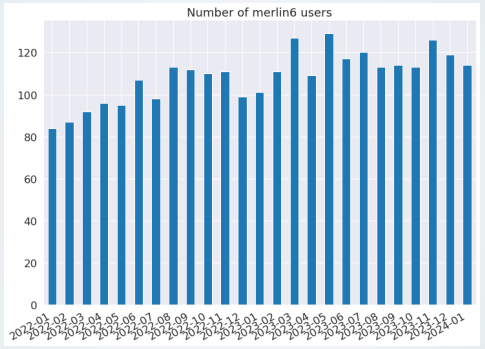
LHC Computing

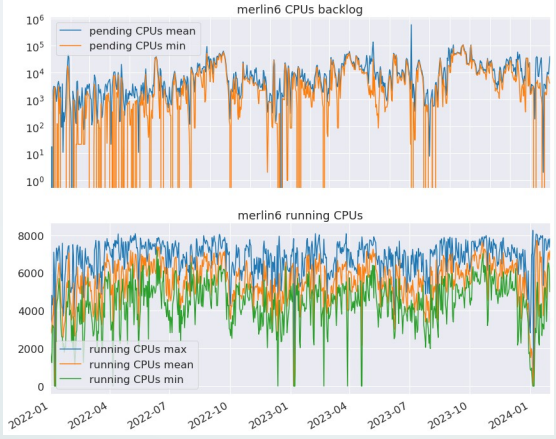
- **Tier-2 on CSCS Alps:** stable in January after ~6 months of frequent problems and downtimes (mostly metadata/IOPS storage related)

HPCE News – Merlin6 usage

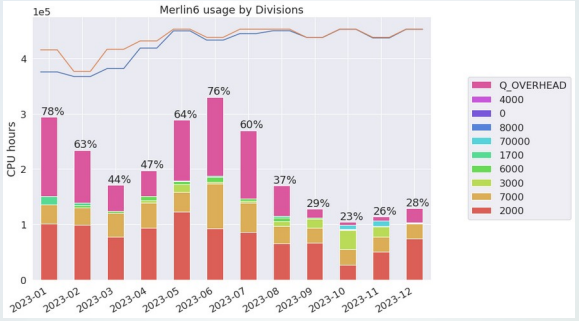
CPU







GPU



PSI core services for ALPS

- The old **Morgana (K8s)** cluster being migrated to the new VMWare infrastructure
  - **Rancher** instance working with vSphere, still missing some issues and configurations (CA issues)
    - GUI + multi-cluster orchestration (instead of KubeSpray), integration with vSphere
  - Deploying Slurm core services based on K8s as main option
- Firsts **Virtual Machines** already installed (i.e. Vault service, Admin and Deployment nodes) and being configured.
  - **Based on RHEL8**, installed via PSI central infrastructure, running Ansible on top
  - **OpenSuse 15.5** for some specific VMs related to the computing cluster (i.e. alternative Slurm core services)
- **CI/CD** adapted for VM core services

ALPS News

- **ALPS network migrated** from 172.30.100*/24* to **172.30.136*/22***. After that change, cluster unavailable for ~1 week:
  - Necessary changes in the deployment layers (controlled by CSCS)
  - Deep changes on the Cray supplied COS layers, which required configuration updates by CSCS
- **Migration of provisioning configurations** from CSCS to PSI layers will be held in the next days with CSCS:
  - Some changes implemented after the network updated → provisioning problems due to changes on upper layers
  - **Goal –** minimal image provided by HPE/CSCS and PSI provisioning main configurations
- **Production Merlin7 storage**:
  - Storage acceptance → **moved to the PSI subnet**
    - *Pending sessions with CSCS* for allowing initial mounting and initial access by PSI admins
  - Concerns regarding performance values given in acceptance test page
    - HDD IOPS can seemingly be saturated by a single node, running 48 tasks.
    - **No real isolation of IOPS**, however exotic workarounds are possible (i.e. file loopback mounts) to mitigate problems.
      - SSD and HDD appliances merged → shared MDS instead of independent metadata ops
      - No clear options to mitigate IO-related problems (Network Request Scheduler?)