HEP/HPC Strategy Meeting: GridKa and HPC

Matthias J. Schnepf | 31. Januar 2025
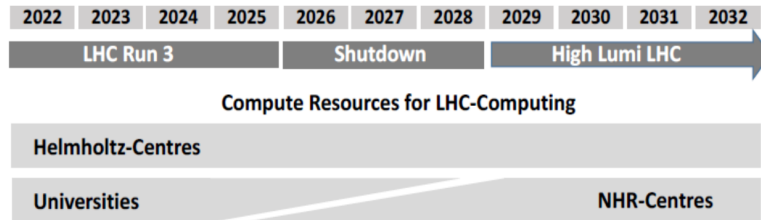
NEMO

- HPC Tier3 in Freiburg
- limited to member of the state of Baden-Württemberg
- $\sim 16,000$ CPU cores (commissioned 2017)
- successor $\sim 32,000$ CPU cores ~ 16 AMD MI300A GPUs, ~ 32 NVIDIA L40s GPUs (will commissioned 2025)
- operation mode
  - fair-share
  - used by local Freiburg and KIT group
  - plan to integrate into Grid for KIT friendly official jobs

NEMO Website
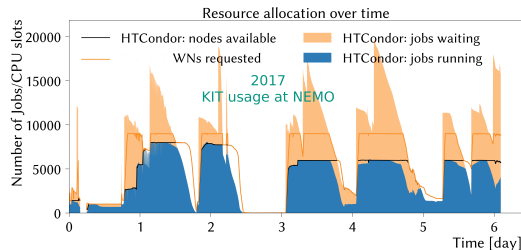
University Tier2 Center in Germany

- NHR (National High Performance Computing) alliance provides computing power to different scientific communities in Germany
- German HEP community decided to move
  - compute power from University Tier2 Centers to NHR HPC centers
  - disk storage to GridKa and DESY-HH
- almost same persons move from maintain University Tier2 to maintain HEP run on HPC
- HEP will be used as other users, but experienced once

| 2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032

| LHC Run 3 | Shutdown | High Lumi LHC |

Compute Resources for LHC-Computing

Helmholtz-Centres

Universities	NHR-Centres

Robin Hofsaess, ACAT 2024

HoreKa

- NHR Tier2 in Karlsruhe (KIT)
- $\sim 60,000$ CPU Cores
- ~ 750 NVIDIA GPUs (A100 and H100)
- HEP usage since burn in phase
- operation mode
  - opportunistic for CMS
    - currently limited set of workflows
    - backfilling (lowest priority)
    - currently only production workloads
    - no preemption
    - max. 21h runtime
    - max. 2300 CPU cores
  - will be pledged for ATLAS

HoreKa Website
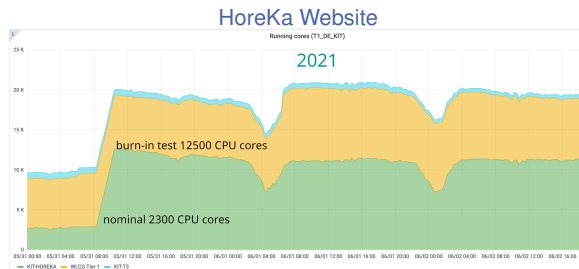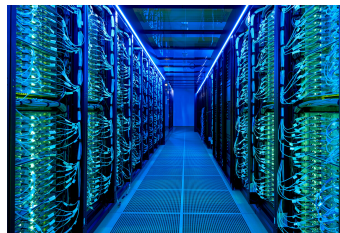
# What have the HPC Cluster in this Talk in Common?

- nearby WLCG Site
  - Aachen (CLAIX) CMS Tier2
  - Freiburg (NEMO) ATLAS Tier2
  - Göttingen (EMMY) ATLAS Tier2
  - Karlsruhe (HoreKa) Tier1 for all LHC VOs
- HEP friendly
  - allowed to use via one HPC user account
  - worker nodes have connection to the outside
  - provide container technology (apptainer)
  - support CVMFS infrastructure
    - CVMFS is installed
    - cvmfs_exec and alien cache on share file system
    - squid proxy near by
  - currently only CPU resources are used
 - usage of GPUs is possible
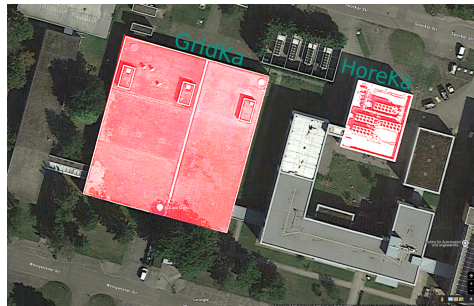
Google Maps

Resource Management via COBalD/TARDIS

- resources are integrated into an Overlay Batch System (OBS)
- OBS WN runs as container matterminers/wlcg-wn via CVMFS
- COBalD/TARDIS is booking and integrating resources based on usage
- resources are available via CEs
- COBalD/TARDIS interacts via SSH with HPC and supports SLURM and MOAB
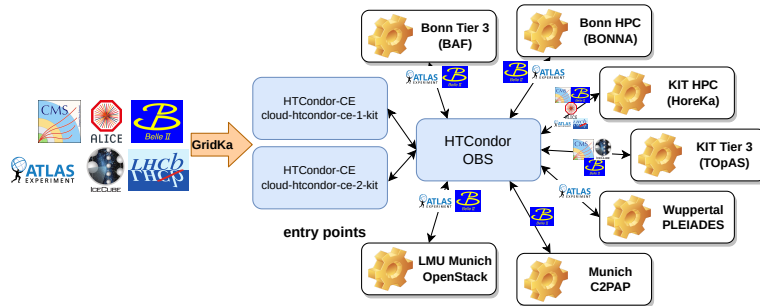
Accounting via Auditor

- designed for accounting of opportunistic resources
- able to account subsites and different scores
- GridKa uses Auditor for EGI accounting for opportunistic and Tier1 resources
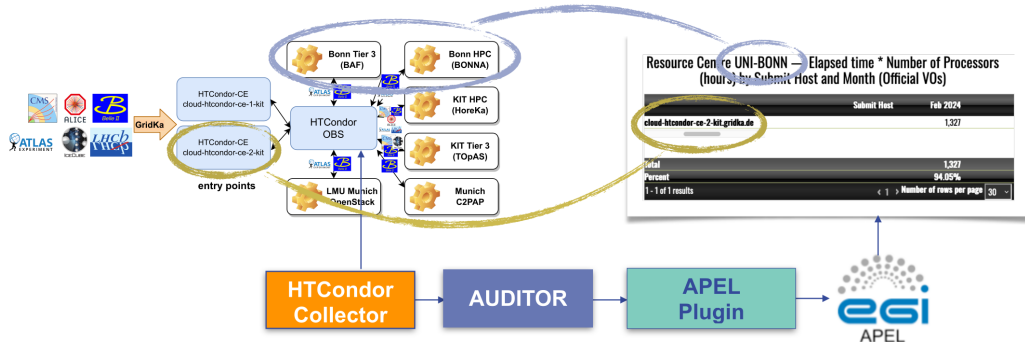
AUDITOR: An Accounting tool for Grid Sites and Opportunistic Resources; HEPiX Fall 2024 Workshop

Improve External Network Usage via XRootD Proxy

- XRootD is a widely use file transfer protocol used by several VOs
- HPC centers have high bandwidth connection within their cluster
- many HPC centers have a insufficient external bandwidth for HEP, e.g. HoreKa
- proxy improve bandwidth usage
  - proxy node to increase external bandwidth
  - caching within the HPC cluster
- option: use an dedicated routing host for HEP
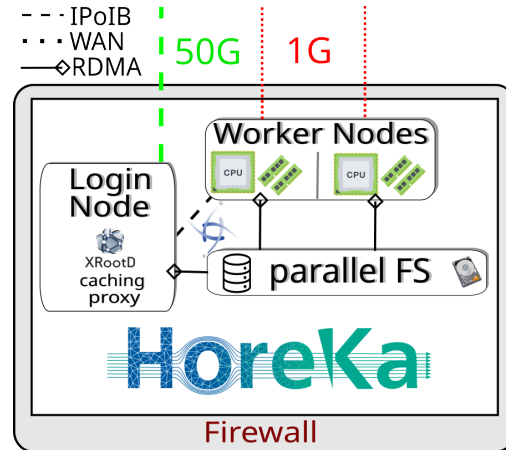

Robin Hofsaess, ACAT 2024

Conclusion

- computing at Germany university Tier2 centers will replaced by HPC centers and storage move to GridKa and DESY-HH
- HEP at KIT has long term experience with HPC centers
- working with HEP friendly HPC next door works perfectly fine
- also small HPC centers can provide thousands of CPU cores to HEP
- software and technology to use and improve usage
  - resource management COBalD/TARDIS
  - accounting Auditor
  - XRootD Proxy to improve usage of external network connection

Backup
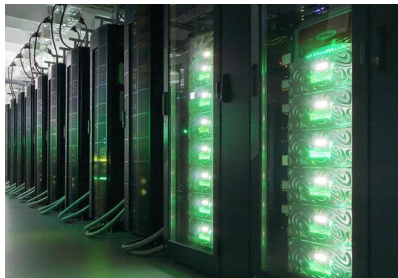
CLAIX

- NHR Tier-2 in Aachen
- $\sim 65,000$ CPU Cores
- ~ 208 NVIDIA H100 GPUs
- operation mode
  - fair-share
  - monthly pledge
  - possibility to use unused CPU-time the next month

EMMY
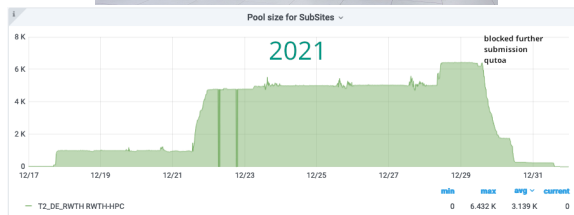
- NHR Tier-2 in Göttingen
- $\sim 111,000$ CPU Cores
- ~ 12 NVIDIA H100 GPUs
- operation mode
  - fair-share

GWDG

COBalD/TARDIS