GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

# Storage systems at GWDG

and how to use them

Sebastian Krey

08.10.2025

# Outline

# Unified storage concept for NHR/SCC/KISSKI

- Central home storage and software installation for all users
- HDD based capacity storage via central `ceph-hdd`
- Central medium performance SSD storage via `ceph-ssd`
- Compute island specific high performance storage, all flash (Lustre, VAST or BeeGFS)
- Tape based storage for inactive data which is required again in a later phase of the compute project
- Data projects which outlive the compute projects will be possible in future for providing long term storage.
- Access to campus home directory (StorNext) only via data mover nodes

# Storage Systems: Homestorage

- Unified home storage for all user groups
- 600 TiB of all flash storage
- Mounted via NFS on all compute nodes
- Also provides the central software installation
- Strict volume quota, relaxed inode quota
- Daily snapshots and offsite backup

# Storage Systems: Coldstorage

Hardware:

- 53 Servers, 21 PB HDD, 3.5 PB NVME
- HDD Cluster with 45 Servers (`ceph-hdd`):
  - 24x 20TB HDD, 4x 7.68 NVME
  - 2x24 Core Sapphire Rapids CPUs, 512 GB memory
  - 2x25G Ethernet
- NVME Cluster with 8 Servers (`ceph-ssd`)
  - 20x 15.36TB NVME
  - 2x32 Core Milan CPUs, 512GB memory
  - 100G Ethernet
- HDD cluster capacity optimized $\rightarrow$ Erasure Coding
- NVME cluster performance optimized $\rightarrow$ Replication

# Storage Systems: High Performance storage

- Two Lustre based filesystems (general availability) and one VAST storage system (KISSKI)
- `lustre-mdc` 1.6PB, `lustre-rzg` 510TB, `vast-kisski` 550TB
- Usage limited to specific compute island to ensure high performance
- Strict volume and inode quota
- All flash filesystems to allow best performance in all workload types

# Storage assignment

- Every user gets home directories for their project specific user accounts
- Every project gets capacity storage in the central coldstorage (ceph-hdd)
- Every project has $PR0JECT for software installations, central configuration files, etc.
- Default quota are applied, which can be increased based on the specifications in the project application
- Every NHR project gets archive storage based on requirements
- Permanent directories on the high-performance filesystems only on request with proper reasoning (e.g. data every user of the project requires)
- In RZGÖ assignment of high performance storage based on I/O requirements (Lustre or VAST depending on read/write mix)

# Storage usage model

**Central storage**

- Home storage: configuration files, software and stuff that needs backup
- Coldstorage `ceph-hdd`: Central capacity storage (HDDs) available on all compute islands.
- `ceph-ssd`: SSD based central storage for medium performance available on all compute islands.

**High performance storage**

- `lustre-mdc, lustre-rzg, scratch-scc` and the institute specific high performance filesystems
- Available only on the designated compute islands
- Mounted via RDMA capable fabrics for lowest latencies
- Switching to HPC Workspaces based usage model
- Project specific permanent directories only on request

# Workspaces

- Filesystems `ceph-ssd`, `lustre-mdc`, `lustre-rzg` (formerly know as `scratch-grete`) usage primarily via HPC Workspaces
- Every user can allocate a space for a limited period of time (30 days with extension option of 2 times 30 days, max 90 days)
- Aim: automatic cleanup after defined period of time
- Expired workspaces can be recovered for a limited time
- Documentation `https://docs.hpc.gwdg.de/how_to_use/storage_systems/workspaces/index.html`
- `ceph-hdd` provides longer running workspaces (6 times 60 days) for test accounts, which do not have a project (can also be used by projects for special purposes)

# Summary

- New larger and more performant filesystems are installed
- All user groups can use the same storage systems
- Unified operational concept will allow easier migration from Tier 3 (SCC) to Tier 2 (NHR) usage for university users
- Easier maintenance and documentation will allow a better user experience, performance and availability
- Hopefully less wasted space on high performance storage due to “forgotten” data ensuring consistent performance