HLRS & HLRS Storage Systems

Dr. Thomas Bönisch

Outline

- HLRS Introduction
- Storage Systems@HLRS
- Data Transfer using GridFTP
- Outlook

History of HLRS

• Since 1960 High Performance Computing • 1982 First German Cray System • 1986 Co-Financing of a Cray by Porsche	University Level
– 1995 First NEC System (SX-4) – 1995 Foundation of HWW for co-operation with industry – 1996 First German Federal HPC center	Federal Center and PPP for Industrial Usage
• 1999 NSF/USA Award • 1999 Initiated European Grid pilot project • 2003 Winner of HPC Challenge at SC'03 • 2004 First TFLOP (NEC SX-8)	Growing Science & Education
– 2007 Co-Founder of **Gauss Center for Supercomputer (GCS)** – 2010 ITEA Gold Award for project ParMA – 2011 First PFLOP (Cray XE6) – 2012 European provider for **PRACE** – 2015 Fastest European System HPCG benchmark (#8 TOP500) – 2018 Project lead for European Center of Excellence in Engineering	European Level

Hawk – HPE Apollo 9000

Peak:
25,95PF
HPL:
19,33*PF
HPCG:335TF

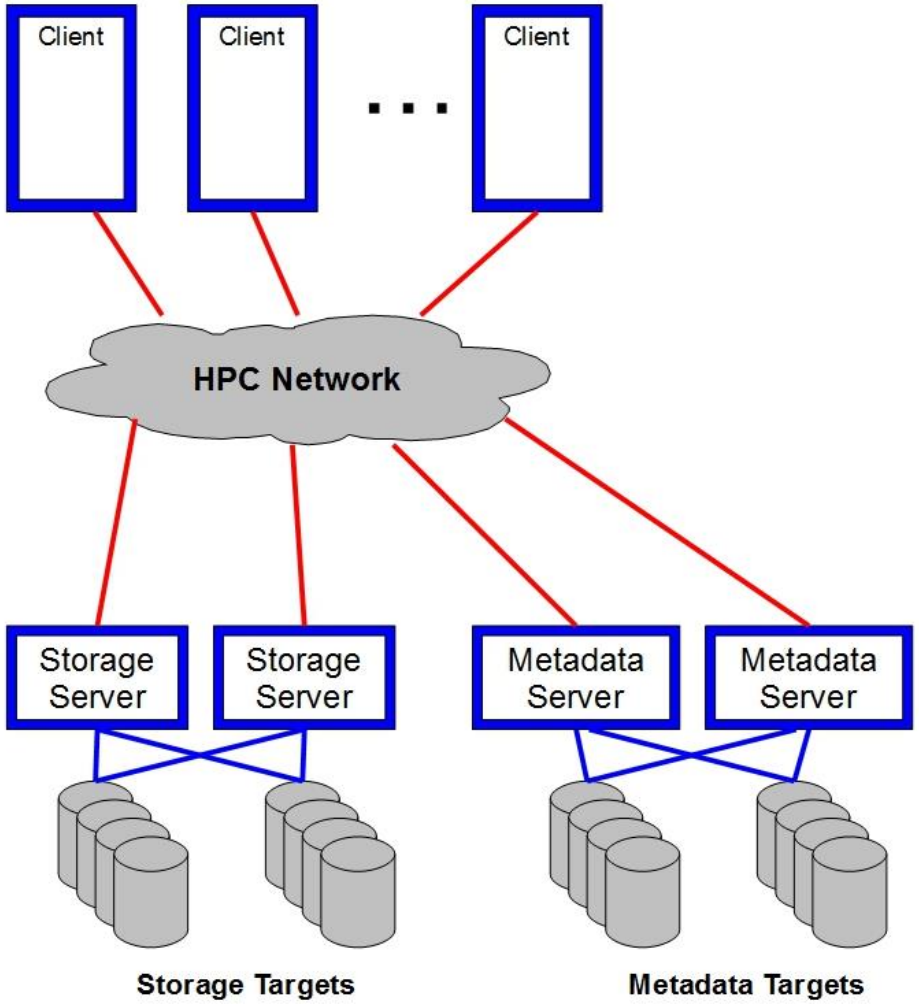
Flagship System

- HPE Apollo System HAWK
  - 720.896 cores AMD Rome 64, 2,25 GHz
  - Racks: 44
  - Nodes: 5.632
  - ~26 PetaFlops Peak
  - Total Memory: ~1,44 PB
  - Infiniband HDR Interconnect
    - 9D Partitial Extended Hypercube Network
  - Pre- and Postprocessing Nodes
  - Average Power Consumption: ~3.2 MW

Storage Systems @ HLRS

- Shared Home (zhome)
- Lustre (ws10, ws11)
  - Project data
  - Scratch data
- Quobyte (upon request)
  - Mainly für projects which require large file counts
- HPSS

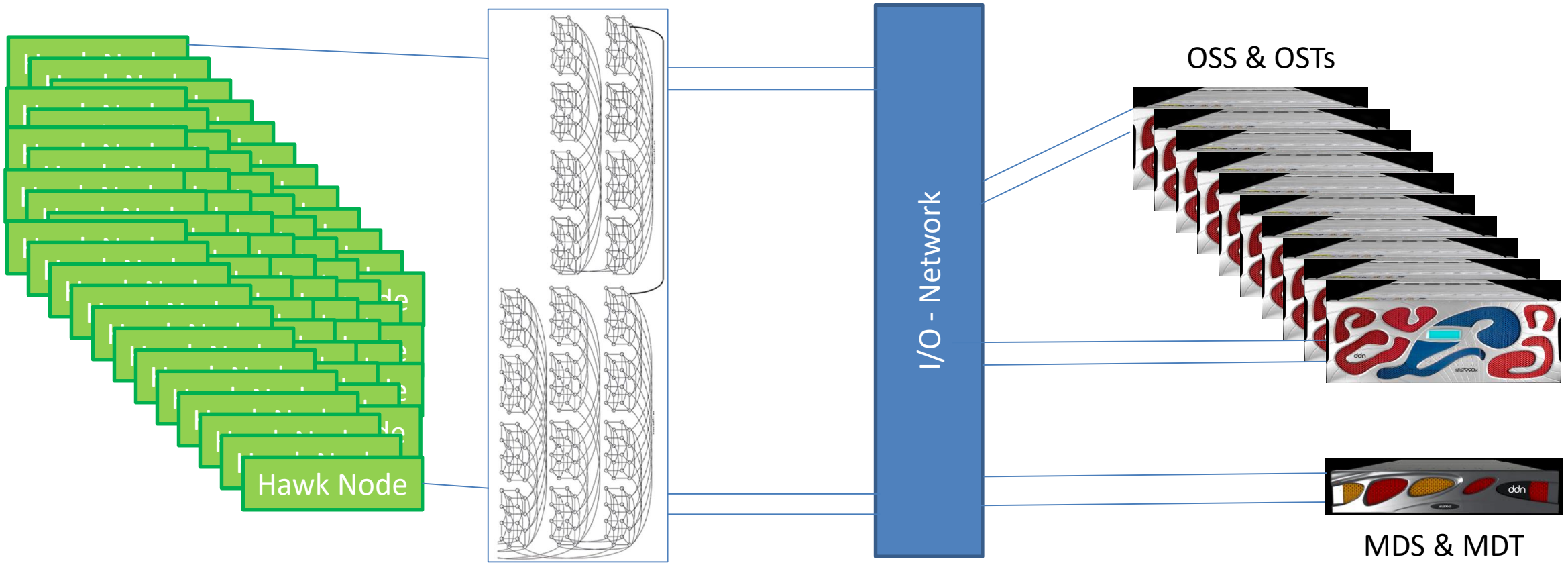
Lustre: Asymmetric Parallel Clustered File System



- Sufficient Number of Servers for high throughput
- Sufficient number of disks for high throughput and high capacity
- Manages access of different nodes to the same storage devices (parallel access)
- Organizes concurrent access
- Guarantees data and metadata consistency

- Pros.:
  - Scalable Bandwidth
  - Serving high number of clients
- Cons.:
  - Metadata Performance sometimes troublesome

Hawk High Performance File System



5632 x Dateisystem Klient

HPC Netzwerk

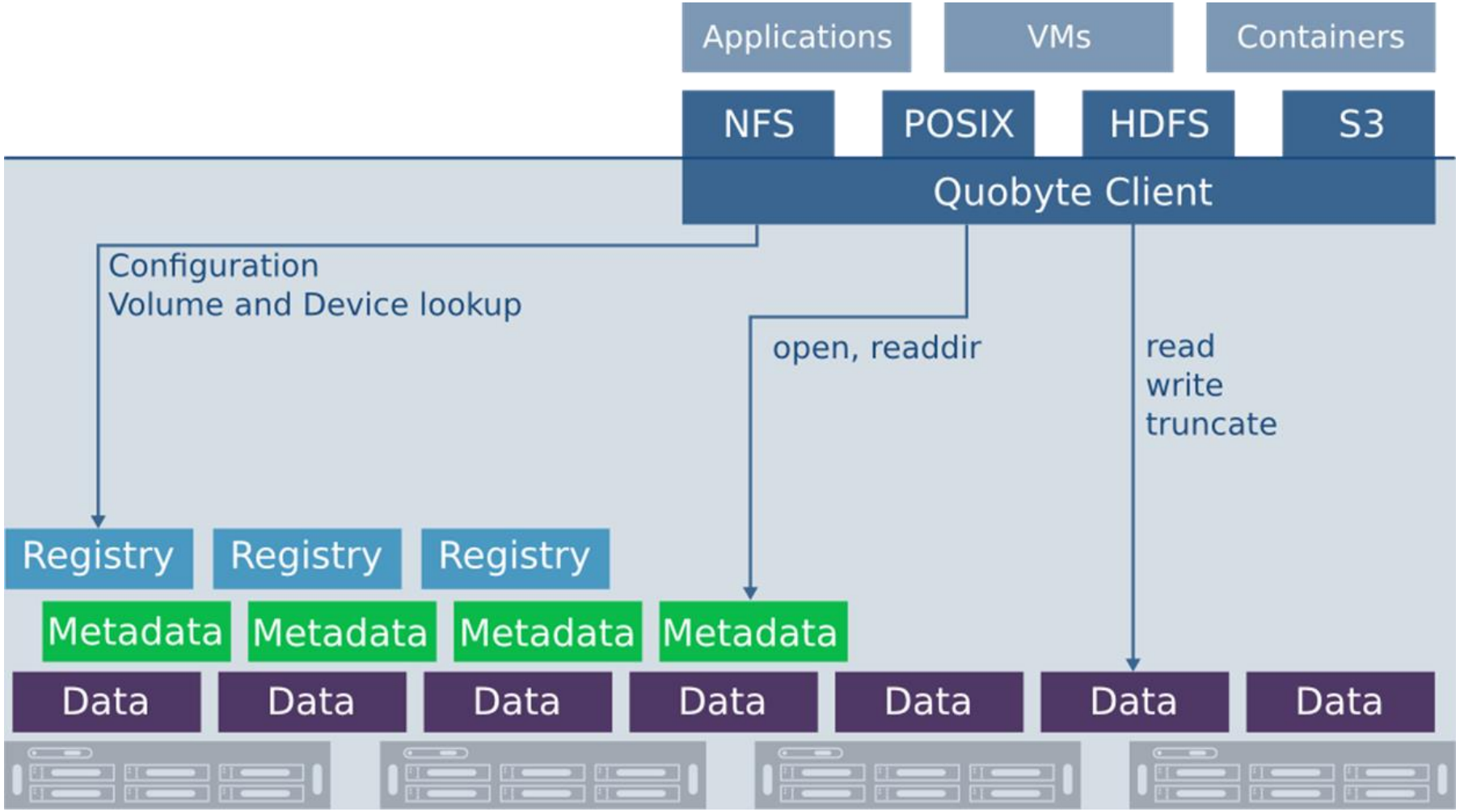
Work Space Mechanism

- A directory in the project file system is created upon request with a user defined name
- The directory is available for 60 days
- The directory life time can be extended 3 times by 60 days
- At the end of life, the directory with its content!!! is automatically deleted
- There are tools for
  - finding available workspaces
  - Releasing workspaces
  - Setting a reminder in calender tools
- Quota is enabled

Quobyte - ObjectStore

- Based on Quobyte Data Center File System
  - Storage as a Service Technology
  - Software-Only-Solution
    - Builds a horizontal storage infrastructure from heterogeneous server hardware
    - avoids vendor lock-in
- Goal
  - Provide Storage Space for special requirements
    - Requirements not feasible for Lustre
      - E.g. large file counts
  - File System Access to the outside (to come)
  - Other special requirements

Architecture



General Purpose Storage System

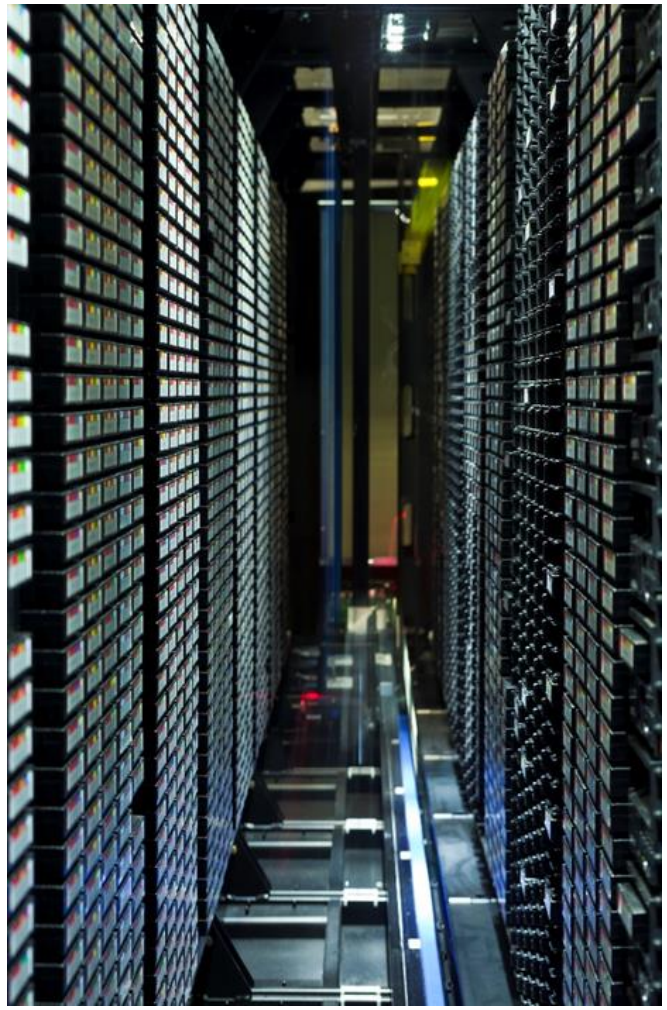
- Architecture
  - distributed parallel file system
- POSIX file storage
  - NFS, Unix apps
- High-performance block storage
  - VMs, databases
- S3 object storage (to come)
- Parallel file IO
  - HPC applications
  - HDFS connector
- Unified
  - copy data in via S3, run it through Hadoop and post-process it via NFS

HLRS Installation

- 20 storage units
  - One server
    - 128 GB RAM
    - 2x25 Gbit network connectivity
  - One JBOD
    - 60*10 TB gross capacity each
- 400 Gbit connectivity
- ~12 PB total raw capacity
  - Usable capacity depends on the redundancy schemes in use
- Directly accessible from all Hawk nodes

HSM at HLRS (and Backup of Stuttgart University)

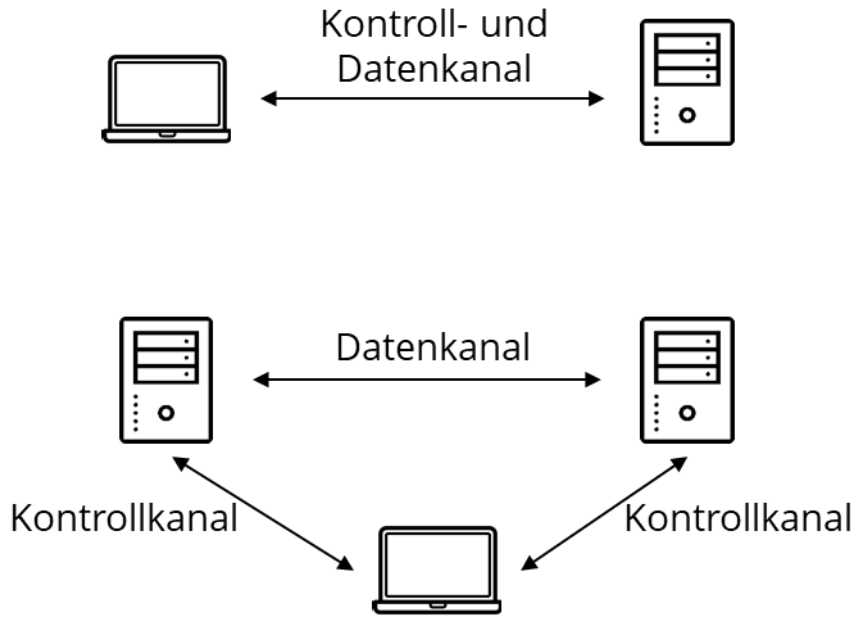
- Data Backend using tape technology
- Using HPSS software
- Two copies per file
- Redundant setup



Data Transfer Using GridFTP

- Data Transfer Software developed by Globus project
- Software Stack maintained by Grid Community Forum
- Features
  - Acess via certificates
  - Split Setup (frontend backend) for improved security
  - features for high speed data transfer like
    - Parallel data connections
    - Parallel data streams

- Third party transfer possible



Thank you