Experiences with ESS

Stefan Dietrich, Martin Gasthuber, Jürgen Hannnappel, Janusz Malka (XFEL)
Hamburg, 2023-05-22

HELMHOLTZ

DESY.

Agenda

01 Introduction & reasons for ESS

02 Experiences with ESS over the years

03 Maxwell Compute Cluster & InfiniBand Fabric

04 ESS 3500 Performance

Deutsches Elektronen-Synchrotron DESY

Large-scale facilities for science

One of world's leading accelerator centers

- Scientific main areas
  - Accelerators
  - Photon Science
  - Particle Physics
  - Astroparticle Physics (DESY Zeuthen)
- Approx. 3000 employees, incl. 1300 scientists
- ~3000 guest scientists every year
- Accelerators: X-ray

ESS Deployments

History

- DESY is not a traditional HPC site
  - ...but high demand for HPC-like resources
- Since 2015: continues increasing demand for storage and computational resource
- IBM Spectrum Scale with Native RAID (GNR) for data storage
  - IBM Elastic Storage Server (ESS)
  - No “traditional” GPFS in operation!
- Talk will focus on experiences over the years
  - For details about scientific challenges, check out older GPFSUG presentations

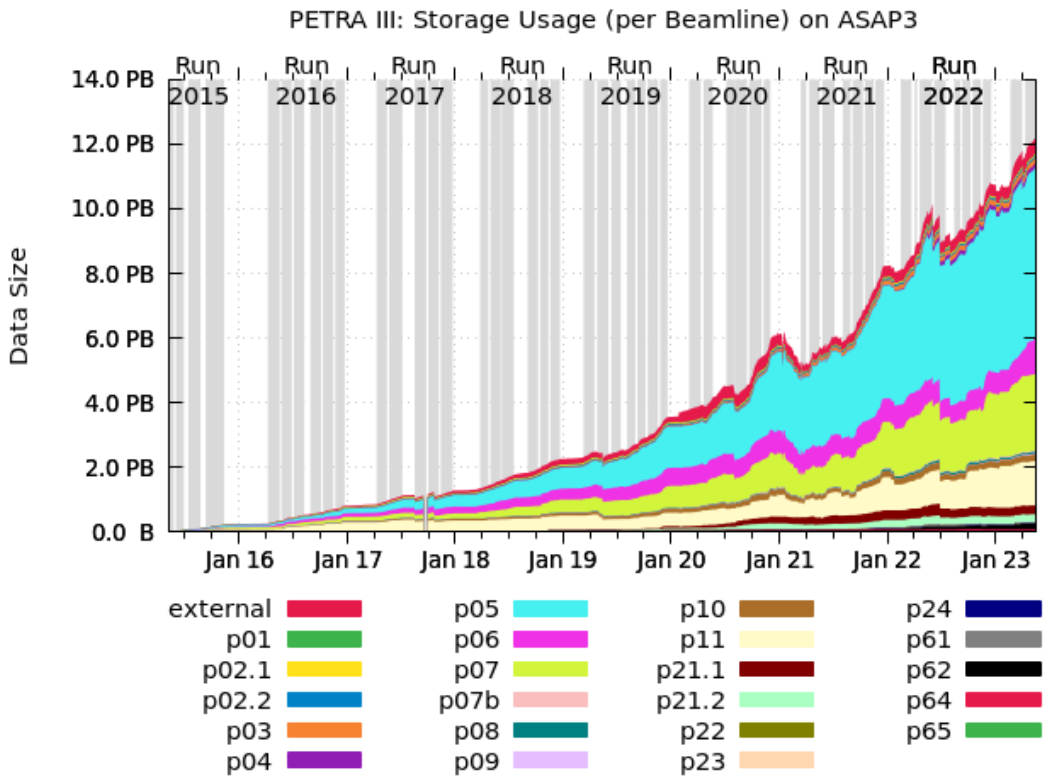
ESS in Operation

- GPFS as initial landing point for Scientific Data
  - Generated by detectors installed at the accelerators
    → PETRA III, FLASH, European XFEL
  - High data rate & volume: ~87 PiB deployed
- Started in 2015 with IBM ESS
- All ESS generations at some point in operation
  - 0th gen: IBM GSS 24
  - 1st gen: GS2, GL4/6
  - 2nd gen: GS4S, GL4/6S
  - 3rd gen: 3000, 3200, 3500, 5000

Data Volume

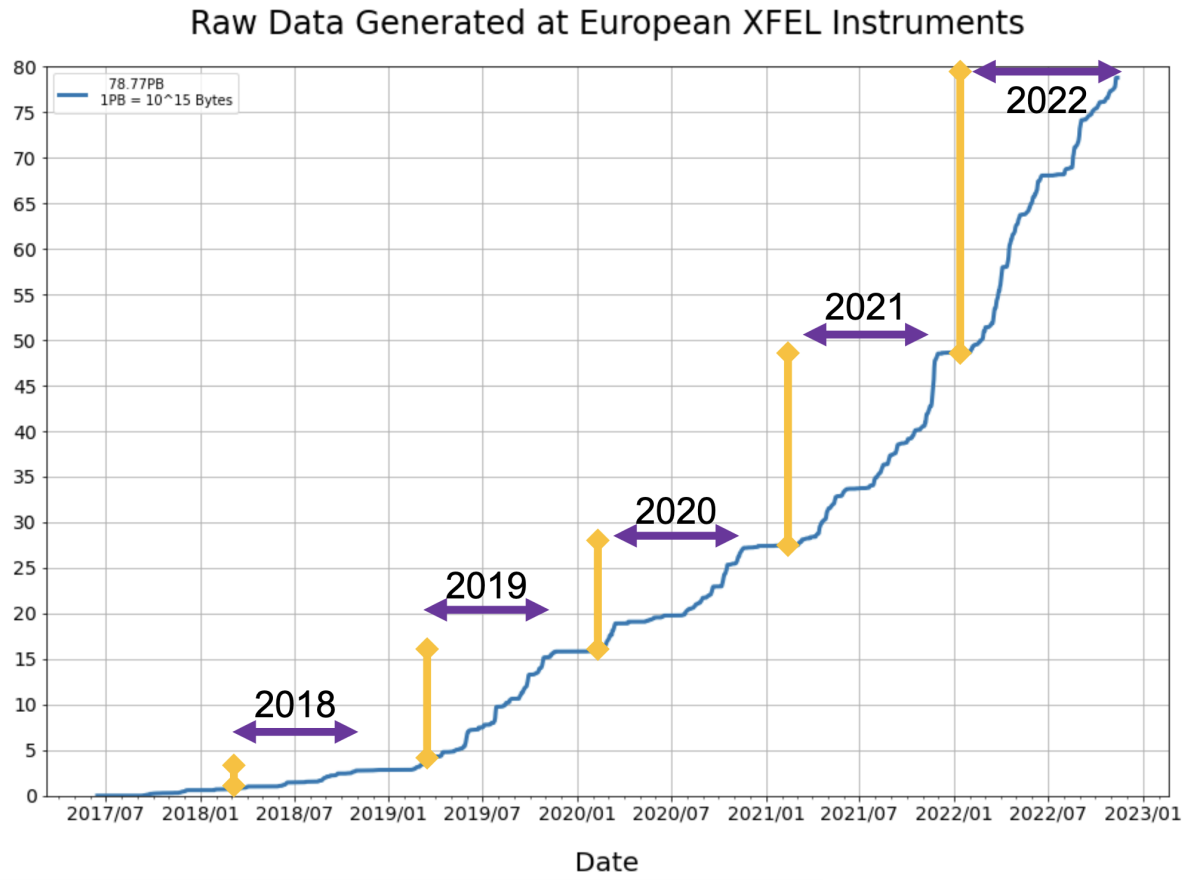
DESY – PETRA III

- Active capacity: ~15 PiB



European XFEL

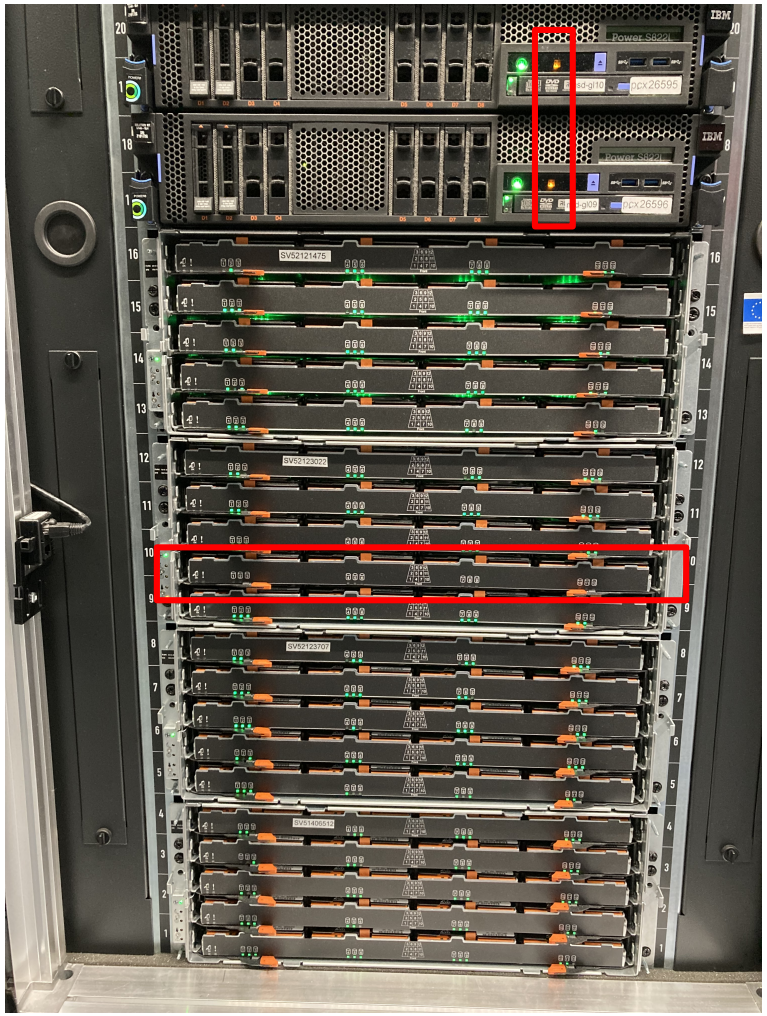
- Active capacity: ~45 PiB, Plot includes archive



Why GNR?

Why Spectrum Scale with Native RAID?

- Software RAID – better rebuild times
  - Slow and long rebuilds for RAID6
    → increased chance for data loss
  - No huge performance degradation during rebuilds
- End-to-end checksum for data integrity
- Worked well for us over the last ~8 years
  - No (known) data loss due to faulty hardware, bit rot etc.
  - 8+3p for HDD for enclosure fault tolerance, 8+2p for SSD/NVMe
  - Similar hardware without GNR: Experienced data loss incidents
- An extreme example: out of warranty GL4 in (non-productive) operation



ESS Deployments over the years

Continues growth and removal of old hardware

- Storage expansion: Remove & add new NSDs to existing filesystems
  - Usually in operation during low activity, without QoS to migrate as fast as possible, full restripe afterwards
  - except: on-disk format migration between GPFS 4.2.3 → 5.0
- Handling different performance characteristics
  - e.g. NSD from ESS 5000 faster than GL6S in single pool
  - Adjusting # of NSDs – more NSDs for newer generations
- Interoperability between ESS generations important
  - Worked well for us with 2nd/3rd generation
  - Problem with 1st gen: discontinued Big Endian support
    → Cluster stuck on old ESS release, no new features

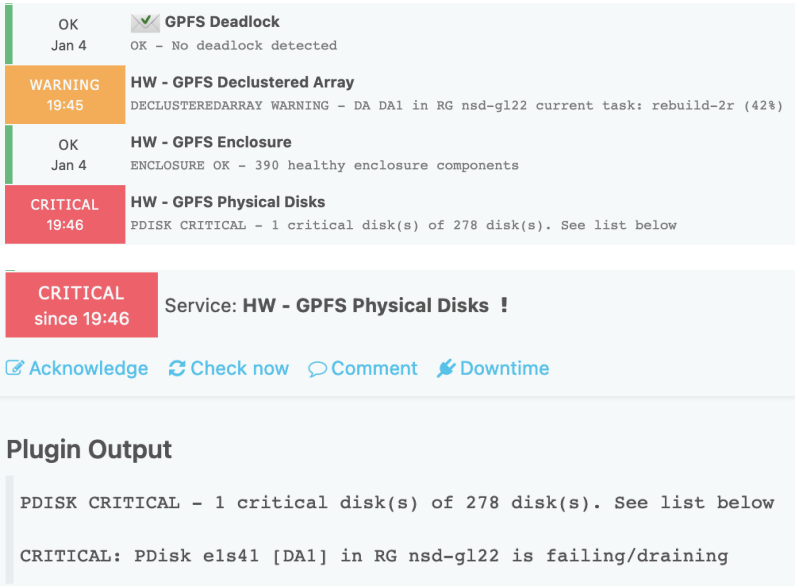
```
# mmlsfs all -V | grep Original
15.01 (4.2.0.0)
17.00 (4.2.3.0)
20.01 (5.0.2.0)
20.01 (5.0.2.0)
21.00 (5.0.3.0)
23.00 (5.0.5.0)
27.00 (5.1.3.0)
27.00 (5.1.3.0)
27.00 (5.1.3.0)
```

```
# mmvdisk vs list
fs0_d8    fs0 BB013, BB014, BB015, BB016
fs0_d0    fs0 BB017, BB018
fs0_d10   fs0 BB017, BB018
fs0_d2    fs0 BB019, BB020, BB021, BB022
fs0_d3    fs0 BB019, BB020, BB021, BB022
fs0_d11   fs0 BB023, BB024, BB025, BB026
fs0_d1    fs0 BB023, BB024, BB025, BB027
fs0_d5    fs0 BB031-32
fs0_d6    fs0 BB031-32
fs0_d7    fs0 BB031-32
fs0_d9    fs0 BB031-32
fs0_d12_1 fs0 BB033-34
fs0_d12_2 fs0 BB033-34
fs0_d13_1 fs0 BB033-34
fs0_d13_2 fs0 BB033-34
```

Site Integration

Integration into existing site infrastructure

- No green field – existing network infrastructure
  - ESS deployment sometimes cumbersome, dedicated networks for BMC & deployment
- xCAT → Ansible for ESS updates: Less manual work, partially buggy
  - XFEL: 1 EMS node to deploy 4 clusters
- Integration into Icinga for central service- and hardware monitoring
  - Running NRPE/Icinga Agent on ESS systems for monitoring
  - RHEL Slim ISO includes relevant dependencies
    → I guess we are not the only site installing software on ESS ;-)
- Extensive use of Spectrum Scale Bridge for Grafana
  - Visualization of GPFS performance metrics



Maxwell Compute Cluster

HPC-like Platform for Photon Science Data Analysis

DESY is not a traditional HPC side

- Maxwell Computer Cluster
  - ~1.000 compute nodes, 544 TB RAM, ~42.700 physical cores, ~200 GPU nodes
- HPC characteristics
  - Fast Cluster Filesystems: GPFS & BeeGFS
  - Fast Interconnect: InfiniBand
  - Vast compute resources: Big CPU & GPU resources
- Variety of Photon Science communities, e.g.
  - Data analysis of PETRA III, FLASH, XFEL
  - Accelerator & PETRA IV simulation
  - Plasma Wakefield simulation

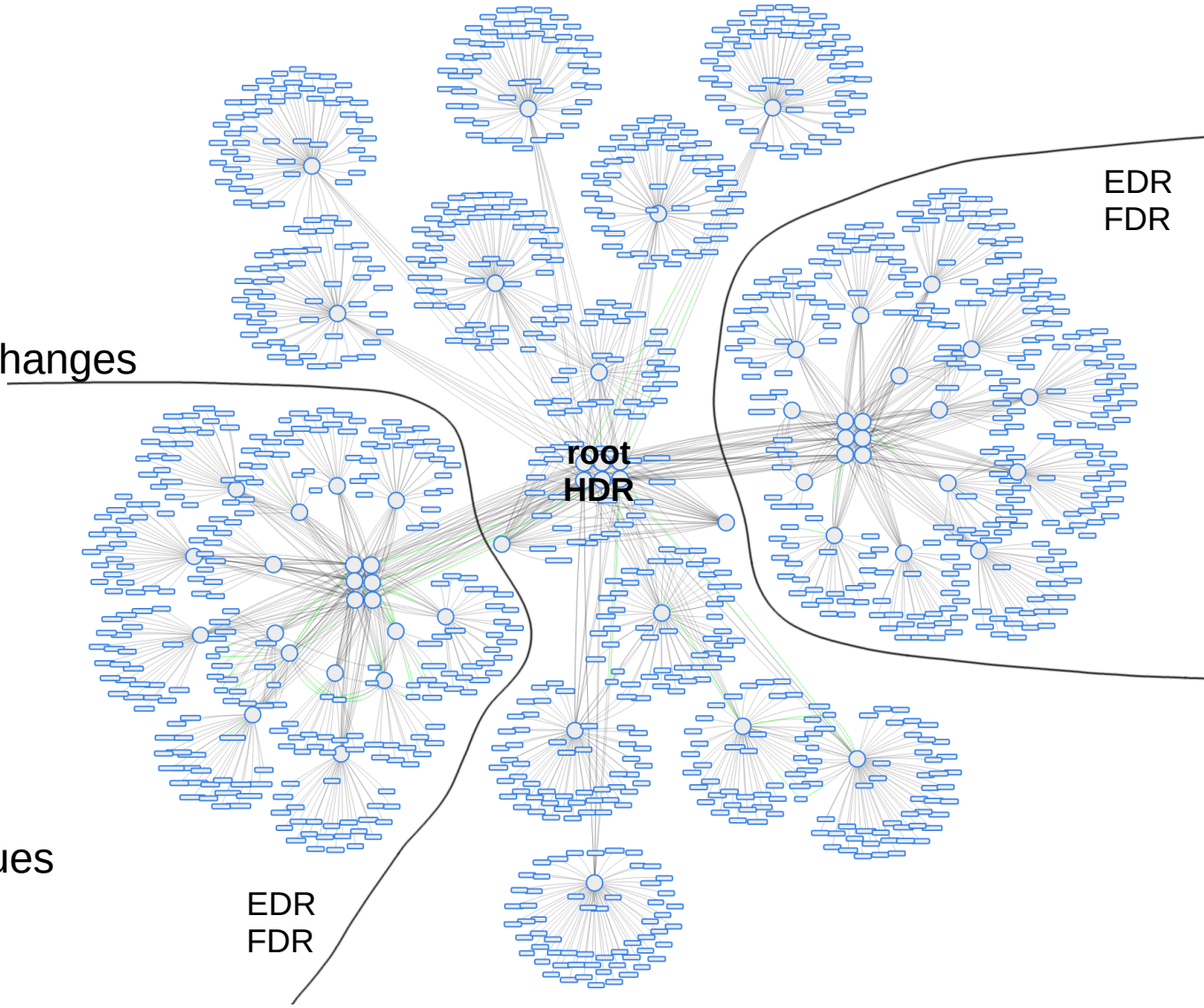
Scheduling

- Multiple interaction possibilities for users
  - Interactive login nodes for users
  - Slurm for batch jobs
  - Jupyterhub
- Core- and group specific compute resources
  - Implemented as partitions in Slurm
  - Buy-in model for groups
  - User can run jobs on group specific resources, but might be preempted

Maxwell InfiniBand Fabric

Growing and upgrading through the years

- All compute- and storage nodes use the Maxwell InfiniBand fabric
- Continues growth and new technology required changes over the years
  - Transition: 2 layer FDR → 3 layer FDR/EDR → 2 layer HDR
- Slow phase out of FDR/EDR with removal of old compute/storage nodes
- Challenging – multiple recablings required
  - Aging FDR cables causing performance issues
- Upgrade to NDR with new OSFP connectors?

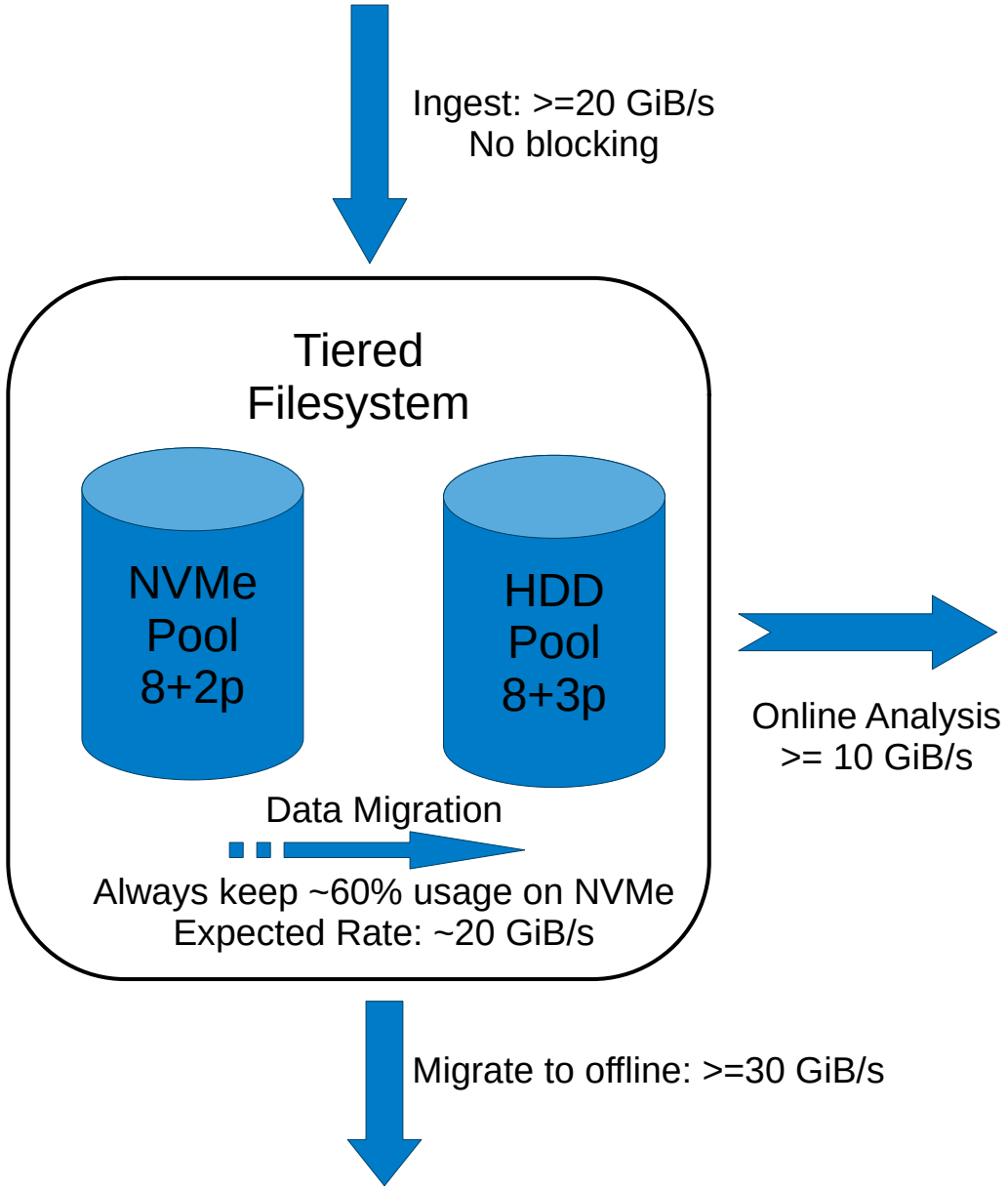


ESS 3500 Performance

Benchmarking FLASH systems is hard

New Burst Buffer for European XFEL

- 2xESS 3500
  - Performance Model: 24x 30 TiB NVMe
  - Capacity Model: 408x 10TB HDD
  - 2x Dual Port HDR HCA per canister
- Burst buffer as tiered filesystem
  - Initial placement on NVMe pool
  - Migration between tiers should be fast and not interfere with data acquisition
- Initial performance impression
  - 50 GB/s read – far away from 90 GB/s+ read
  - Erratic, results not reproducible



ESS 3500 Performance

Benchmarking FLASH systems is hard

Tuning and debugging

- ESS 3500 Performance was too low and erratic
  - ~2.5 days of tuning with Olaf Weiser from IBM
- NVMe → HDD Drain Tuning
  - Additional client tuning (workerThreads, pitWorkerThreadsPerNode)
- NVMe Tuning
  - Erratic behavior: Same test, different results
  - TRIM: No change, NVMe busy after TRIM finished
  - Preconditioning drives to get into “steady state”
  - Upgrade from 2 → 4 HCAs, each with 1 port connected → overcome PCIe 4 x16 limit of ~30 GiB/s
- End result: Workload now running fine

NVMe Read

HDD Write

Summary

- Using ESS worked well for us in the last ~8 years
  - Would most likely choose a GNR system again, if started today
  - Deployment needs more bug fixing
- Faster systems & networks, but harder to get performance
  - FLASH controllers add another layer of in-transparency
- Important to have good support

Thank you

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

Stefan Dietrich
IT
stefan.dietrich@desy.de