



Uwe Ensslin, Birgit Lewendel for DESY-IT

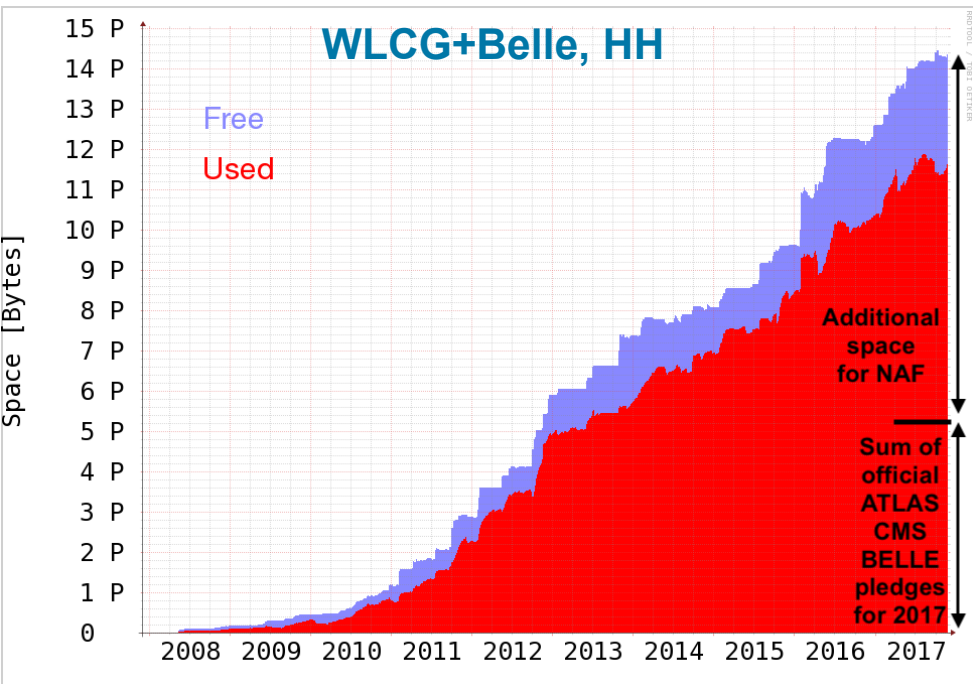
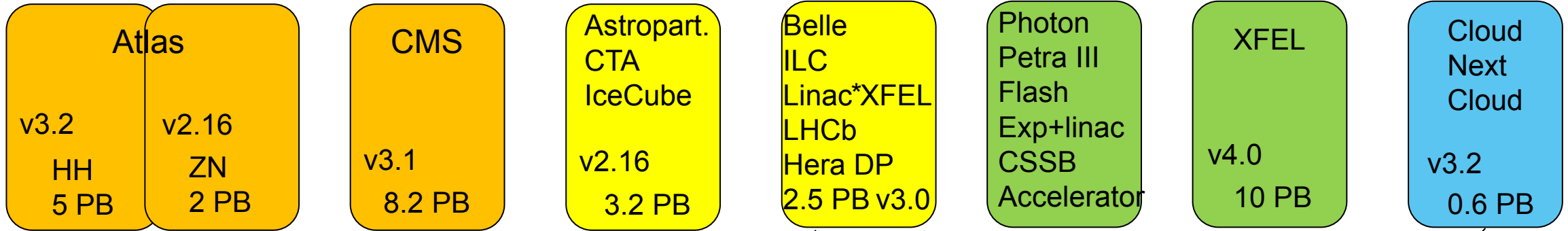
12th dCache WS, 28 – 29 May 2018

dCache for Photon Science, Astroparticle Physics, Accelerator Research, Cloud

HELMHOLTZ RESEARCH FOR
GRAND CHALLENGES



Overview: dCache Instances at DESY



~500 pool nodes
 puppet talk

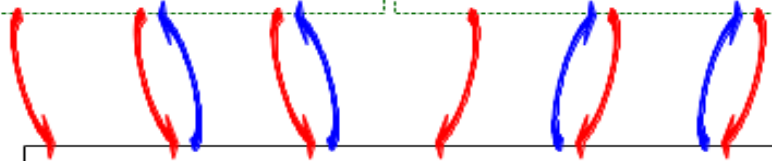
*) XFEL linac sensor data really cached in dCache

PETRA III, FLASH

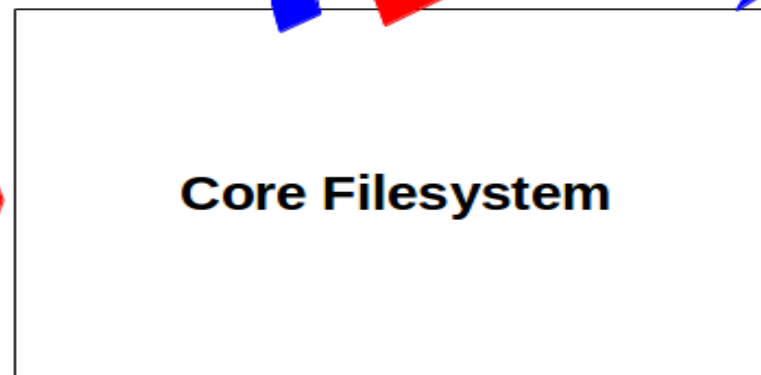
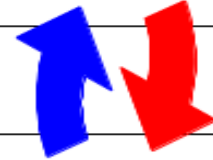
Logical Dataflow



Sandbox per Beamline



- Low latency
- Low capacity
- Host-based authentication



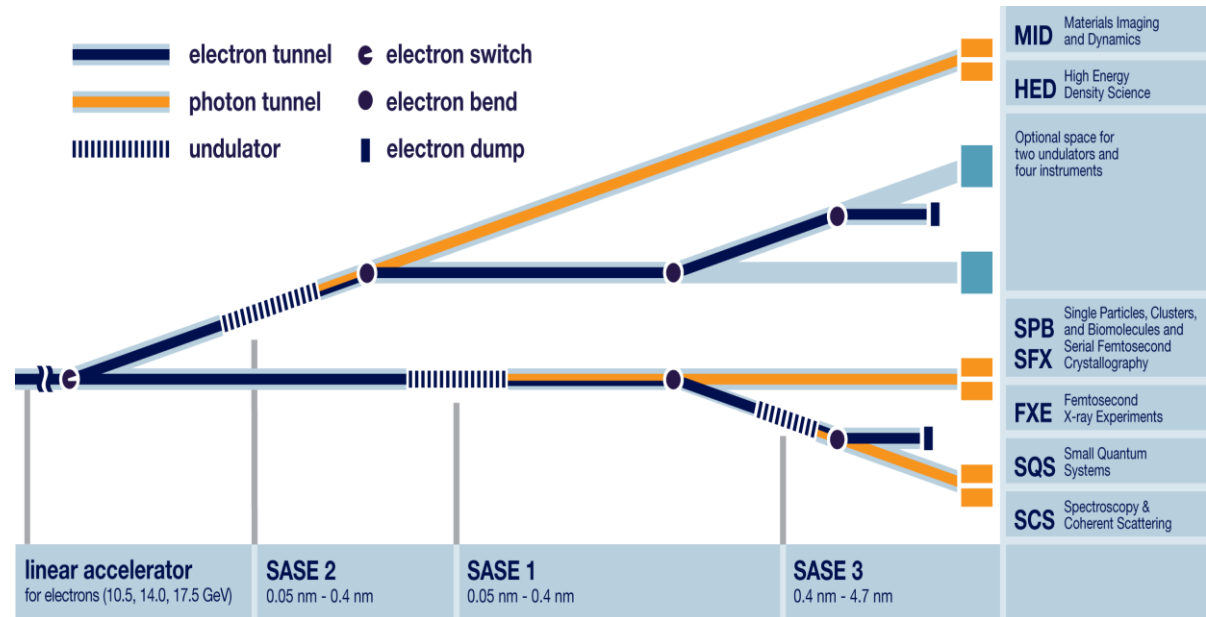
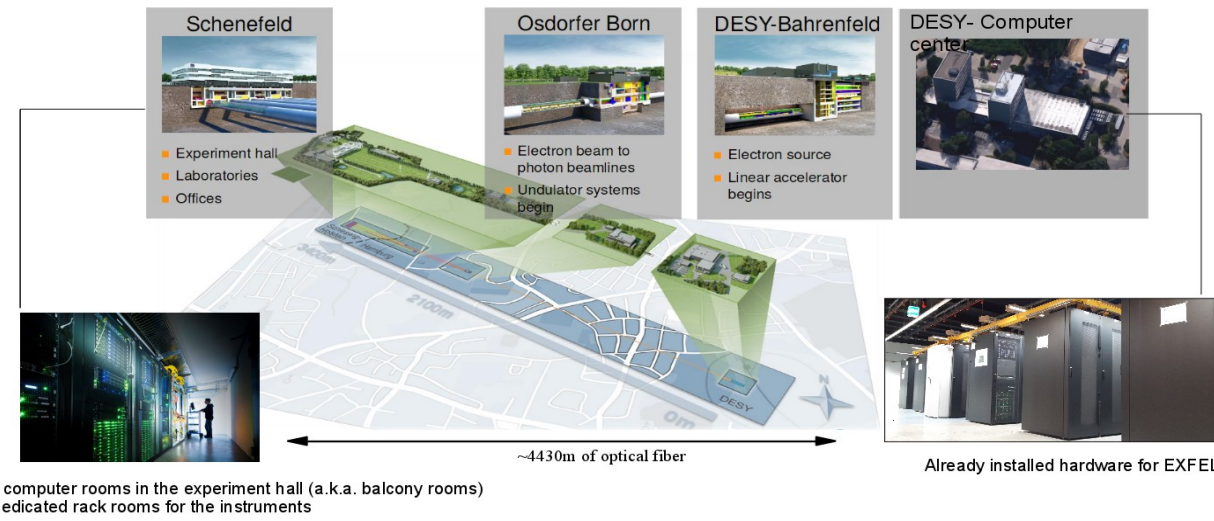
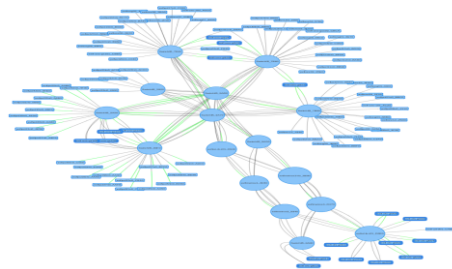
- 4 min latency
- High capacity
- Full user authentication



- Very high capacity, tape
- Full user authentication

European XFEL

- Data taking off-campus in Schenefeld, storage & analysis on the DESY campus
- User runs in November 2017, March & May 2018
 - More continuous operation starting in summer
- About 1 PiB of data stored already, up to 50 TiB per day
- Calibration and analysis using the Maxwell HPC cluster at DESY
- One SASE active, two more starting this year
- Infiniband monitoring developed by IT
- PETRAIII data taking used as blue print



dCache use in Photon Science at DESY

Different requirements

PETRA-III and FLASH

- Many (millions..) of small (4k, 8M) files
 - Larger containers, too
- Interesting character sets in file names
- dCache as gateway to tape archive
- Preserve ACLs when copying to dCache
- Re-staging of complete experiments
- Challenge:
 - Efficiently copy 100k to several millions of files *per experiment* to dCache and further to tape

European XFEL

- Large files (4GiB to 16 GiB)
 - Some small (~ MiB) ones, too
- Well behaved file names
- More active use of dCache
 - ACL protected repository for raw data and calibration constants with tape backend
 - Calibration pipeline reads from dCache
 - Data export via dCache to associated institutes
- Challenge:
 - Handle the data volume coming from the experiments (up to 5GiB/s for 12h)

dCache use in Photon Science at DESY

Getting the data to dCache: A common, automated setup

Triggers

- Place an entry into a queue.
Examples:
- cron: Mark data to be copied 7 days after close of the experiment (PETRA-III, FLASH)
- Request tracker (should be Gamma Portal) for re-staging (PETRA-III, FLASH)
- External: User pushes a button in the metadata catalog (XFEL)
- Internal: A copy job submits a job that feeds back *locality* information to the metadata catalog (XFEL)

Queues

- Implicitly define a workflow (example: PETRA-III/FLASH)
 - Check UTF-8 compliance of paths
 - Execute a GPFS policy run to efficiently get a url-encoded list of files to copy
 - Start the copy process (archival)
- Per queue and global run limits provide basic resource scheduling

Parallel transfer

- The Archiver: handling datasets
- Split the list of files to transfer into work packages
- Start n processes to do the copy
- If configured, spread the processes across a list of hosts
- Copy from the infiniband (GPFS) to the ethernet (dCache) universe
- ACLs included

dCache use in Photon Science at DESY

Getting the data into dCache: The Archiver

- The same software is used for PETRA-III and XFEL
 - Different configurations (queues, limits, workflows, host lists, facilities, UTF-8 check, location of data and archive, ..)
- Core: *gdccp*, a DCAP application
 - Reads a list of paths to be transferred
 - Decode the URL encoded paths
 - Read GPFS ACLs using the native interface
 - Convert the ACLs into a NFSv4 ACL string
 - Send them (and the file contents ..) to dCache via DCAP
- Scripts to implement parallelism, queues, (some) triggers, policy runs, ..

dCache use in Photon Science at DESY

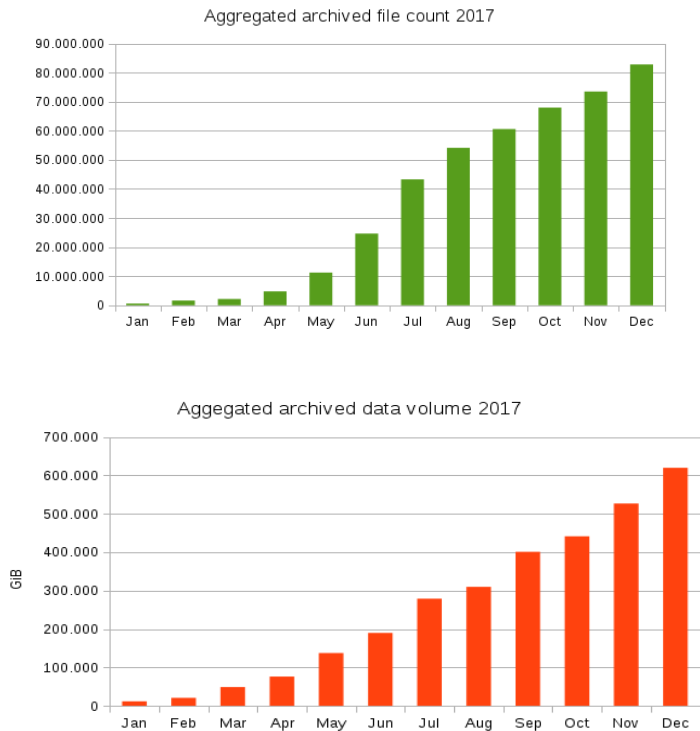
Getting the data into dCache: The Small File Plugin

- Problem: How to store millions of files efficiently on tape
- Solution: dCache transparently puts them into containers upon writing ..
- .. and transparently extracts them upon reading
- Done by the Small File Plugin developed at DESY
- Used in
 - The Photon dCache instance (PETRA-III, FLASH, others)
 - The DESY instance (e.g for mail backup)

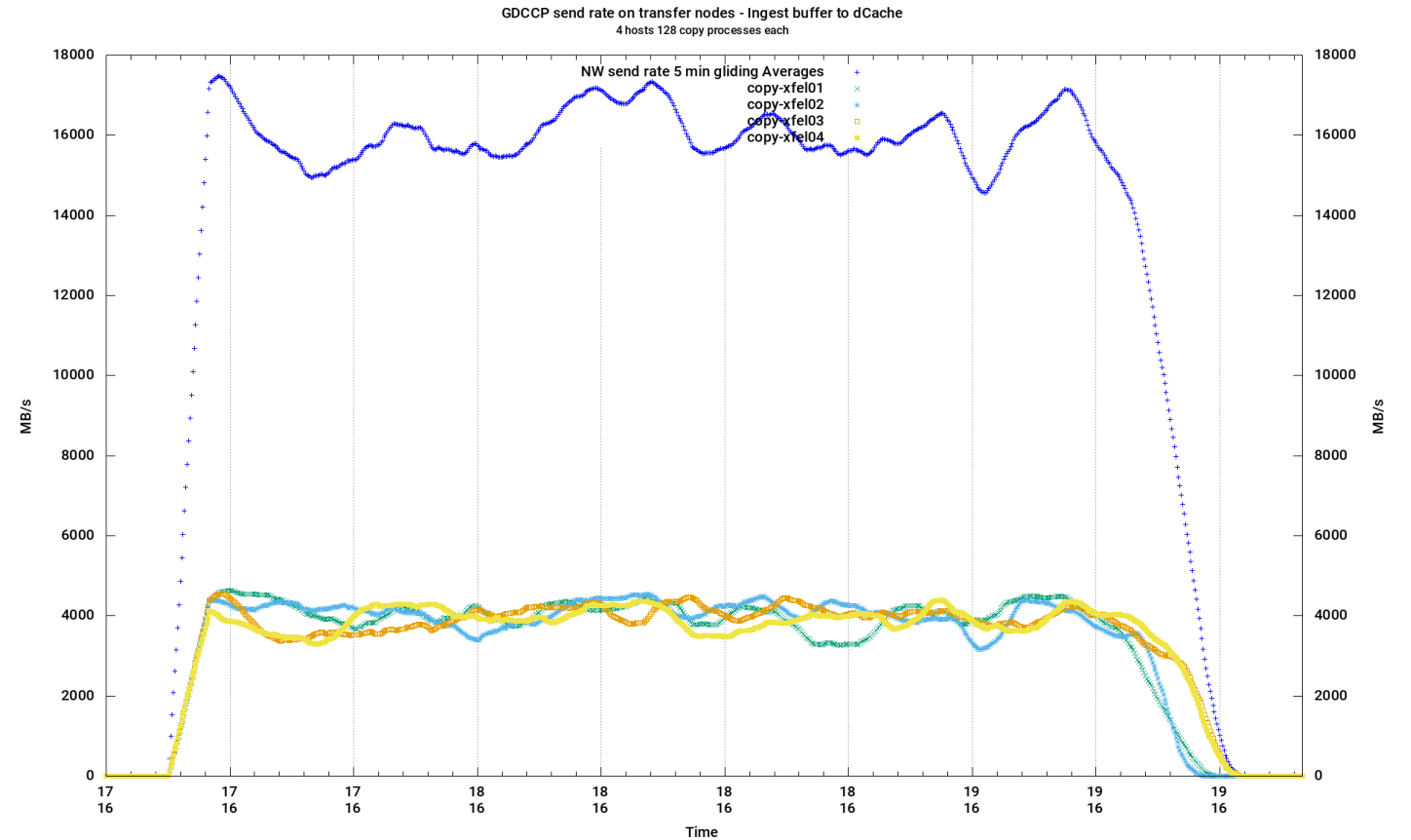
dCache use in Photon Science at DESY

Getting the data into dCache: Some Statistics

PETRA-III/FLASH – archived data



European XFEL – throughput during pre-production tests



Summary

- dCache at DESY for HEP, astroparticle, cloud, accelerator research, photon science.
- 7 instances with ~30 PB disk space .
- Kerberos authentication enabled.
- Dual stack IPv4 and IPv6 in use for Atlas and CMS SE.
- High Availability HA enabled for Photon, Cloud, Atlas.
- Archiving to tape for photon science data fully automated.

