DESY site-report for IBERGRID 2023

Tim Wetzel, Peter van der Reest, Yves Kemp, Patrick Fuhrmann
Benasque, Spain, 2023-09-29

HELMHOLTZ

DESY.

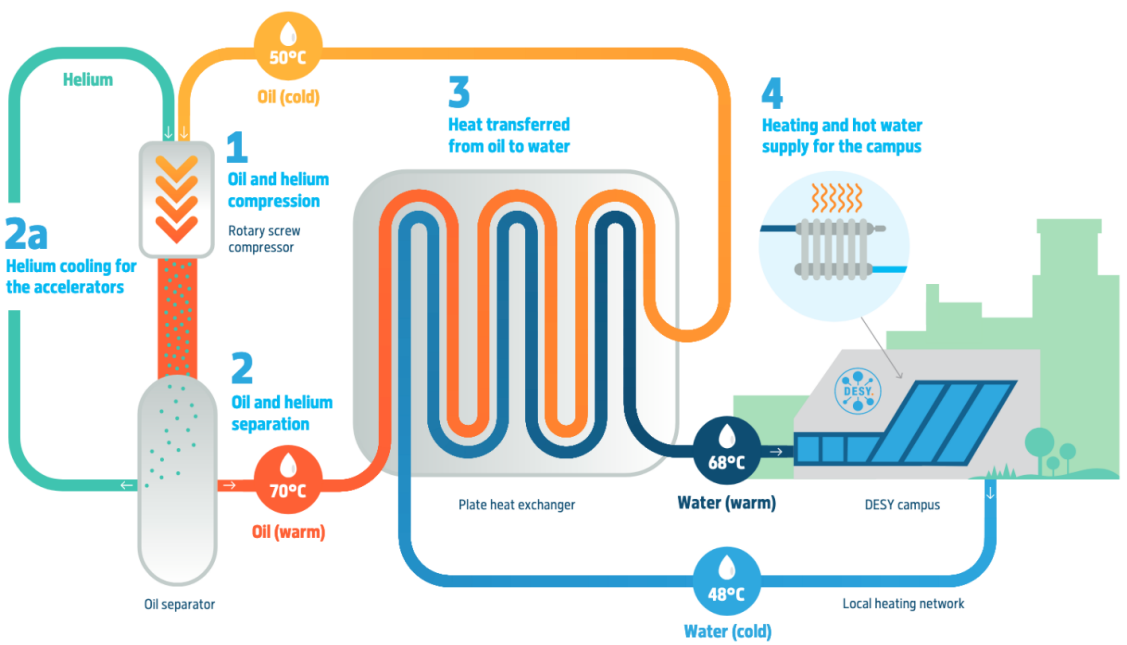
Activities for Ukrainian scientists, students and civilians

- **Fellowships for two Guest Scientists**
  - EU funded special fellowships to continue their work while the Russian invasion of their home country continues.
- **Winter School for students**
  - 22 students enrolled at Ukrainian universities have been working on DESY research projects for six weeks at the Hamburg and Zeuthen campuses
  - intensive course for students of any nationality who were enrolled at a Ukrainian university and who had at least completed two years of bachelor studies
- **Humanitarian aid both in Ukraine and Germany**
  - among others, families that have fled the war are helped with accommodation on campus, until more permanent housing can be found
- **... and further DESY and private initiatives**

IT computing center and sustainability projects

Funding for re-use of waste heat on campus

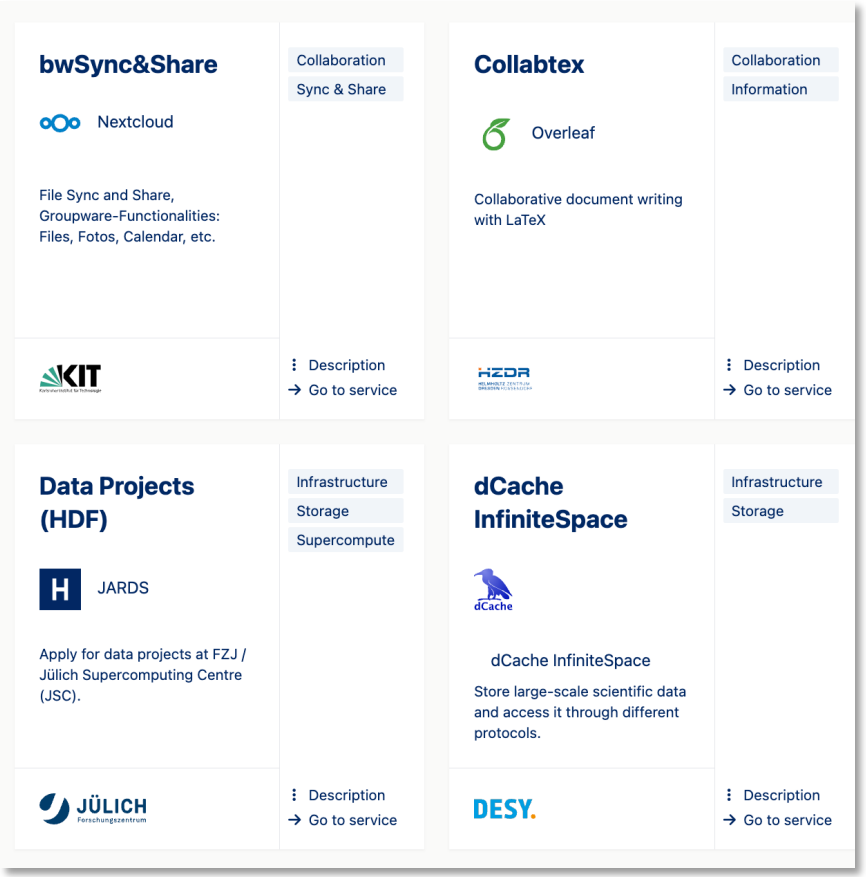
- DESY has long been engaged in the recycling of waste heat generated from the helium liquefaction plant.
- Additional funding has been procured to extend waste heat utilization from computing centers as well as other sections of the accelerator complex.
- The reclaimed heat will be integrated into a low-temperature campus heating network.
  - New buildings will be architecturally designed and constructed to leverage these temperature levels for space heating.



Interlab cooperation and beyond

Helmholtz Federated IT Services (HIFIS) Program has been reviewed and very well received

- see https://hifis.net for further information
- Further activities between reseach organisations and universities for various scientific communities in Germany have been created and in operation
- National Research Data Infrastructure has launched discipline oriented projects to increase exchange and sharing of knowledge, infrastructure and services
- At DESY the Photon Science community is involved in DAta from PHoton and Neutron Experiments
  - DAPHNE - https://www.daphne4nfdi.de)
- The High Energy Physics community is involved in Particles, Universe NuClei and Hadrons
  - PUNCH - https://www.punch4nfdi.de/)



IT – Security

- **Cybersecurity**
  - Scans of accessible services have shown that assumptions of being secure by relying on Linux distros is false
  - not all flaws in products inside the distros are addressed by companies
  - not all releases get the same amount of love
  - we are contemplating how to get better control
- **Security incidents in academic institutions in Germany** in the last months have led to an increased focus on reviewing our own measures and procedures
  - real breaches have downed several (large) organisations
  - upper management is now very interested in the subject
  - we are reviewing our procedures and adapting where necessary
  - and accelerating some of the projects that we had on the board already

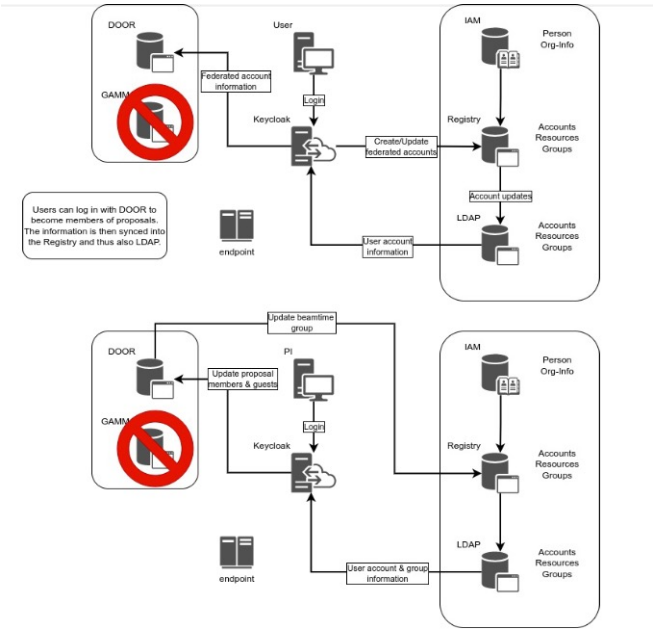
IT – Security, AAI and MFA

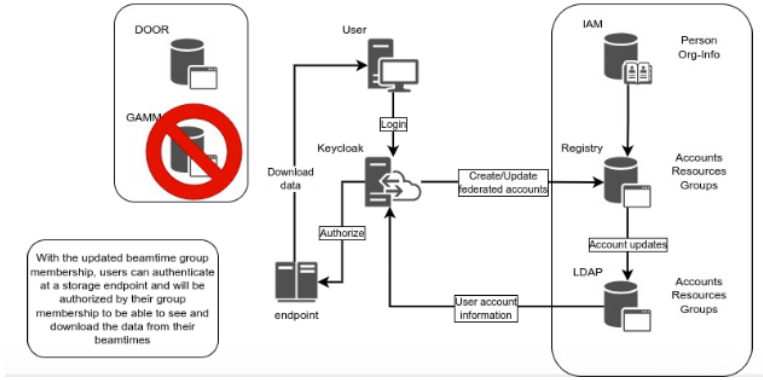
- Activities on multi factor authentication intensified/accelerated
- Plan is to have all interactive remote access handled first (VPNs, SSH)
- In parallel take care of email web interface
- and consecutively gather other services that allow for external access (web apps, secured web pages)
- using the technology that we already have deployed for handling federated identities (Keycloak)
- and adding a MFA service that communicates with Keycloak, so not all applications and services have to implement the MFA hook individually
- as a by-product we gain SSO for all services connected

Enhanced data management

inclusion of metadata catalogues for Photon Science

- Jointly with the Photon Science Division, IT is developing a service for metadata collection and search
- based on SciCat – https://scicatproject.github.io/
  - originated from Paul Scherrer Institute in Switzerland, now development in an international team at PSI, ESS, RFI, ALS
  - individual adaptations at sites reflecting local infrastructure and processes
- development at DESY includes:
  - automated setup of metadata collections from various sources of information in Photon Science and IT
  - automated ingest of metadata from experiments at Petra III and FLASH light sources
  - integration in the DESY Online Office for Research with Photons (DOOR) and the IT user & group management





Sync&Share services

- DESY operates two Sync & Share installations (for DESY and EuXFEL respectively)
  - based on Nextcloud and dCache
  - serving users from all over Helmholtz Association
- Functionality is steadily enhanced – as needed by user communities
  - Evaluation of useful plugins for Sync&Share (Nextcloud), also in HIFIS context
  - Connected to Helmholtz AAI (effectively including all of GÉANT federation), Umbrella and more as needed
- Development of infrastructure
  - Desaster Recovery in place based on dCache base and data keeping in orthogonal data storage based on IBM Spectrum Protect
  - allowing multi protocol access (NFSv4) by using event driven actions that notify Nextcloud databases
  - connection to Keycloak and local user & group management

Goodbye ATLASSIAN

Continuing our exit path from ATLASSIAN tools

- Users have been and are migrating from Bitbucket and Bamboo to the DESY gitlab service
  - we also have ways of migrating Jira software issues to gitlab
  - goal is to have completed these migrations by end of summer 2023
  - IT provides consulting for groups on how to migrate. Groups do work themselves
- PoC with Xwiki.com to replace Confluence ongoing
  - verification of on-premises infrastructure setup, tuning of instance, HA measures
  - structure of Wikis jungle
  - pilot migration of representative spaces until start of summer break
- Jira functionality for Project Management and esp. Resource Management is still an issue
  - we are interested in experience from other sites to the latter
  - project management will be included in larger DESY effort to unify tool set

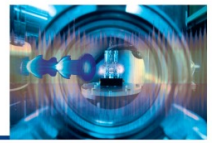
OS and energy saving

- Deployment of AlmaLinux 9 has started
  - server installations first
  - waiting for official EGI/WLCG middleware
- Deployment of Windows 11 with pilot user groups
- Large deployment later in the year
- Activities around energy conservation in computing centers
  - measures in place to automatically shut down worker nodes when energy supply becomes criticial
  - and to turn machines back on automatically, when the situation allows
- Zeuthen site was hit harder by rise of energy cost
  - decision was taken shut down all Compute Systems (Grid+HPC+Batch) older than 5 years
  - resulting in ~40% less cores in operation on Zeuthen site
- More info in dedicated presentation later this week

Start with Review of PoF IV Proposal

Goal of the IDAF

The Evolution of the LK II Tier-2 Facility



- **Recommendation**: Tier-2 LK II Facility should support additional user communities
- Observation throughout all Programs in Matter
  - Growing data deluge ➞ Important to **access** and **analyse** large amounts of data

Necessity for a facility to store and analyse data with access for all scientists within Matter.

M T IDAF

↳ From LK II Tier-2 ➞ Interdisciplinary Data and Analysis Facility

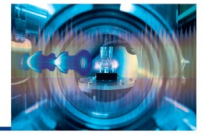
- Association with M T
  - MT is interdisciplinary ➞ IDAF is moved from M U to M T
  - Current setup planned at DESY (very broad matter community, experience with Tier-2)

Start with Review of PoF IV Proposal

Plans for the IDAF

Building the IDAF:

Merging High Performance Computing (HPC) and Tier-2 Clusters



HPC Cluster

Tier-2 Centre

- Users from Program M M L & M U
- Users from Program M T

- Users from Program M U
- Users from Program M T

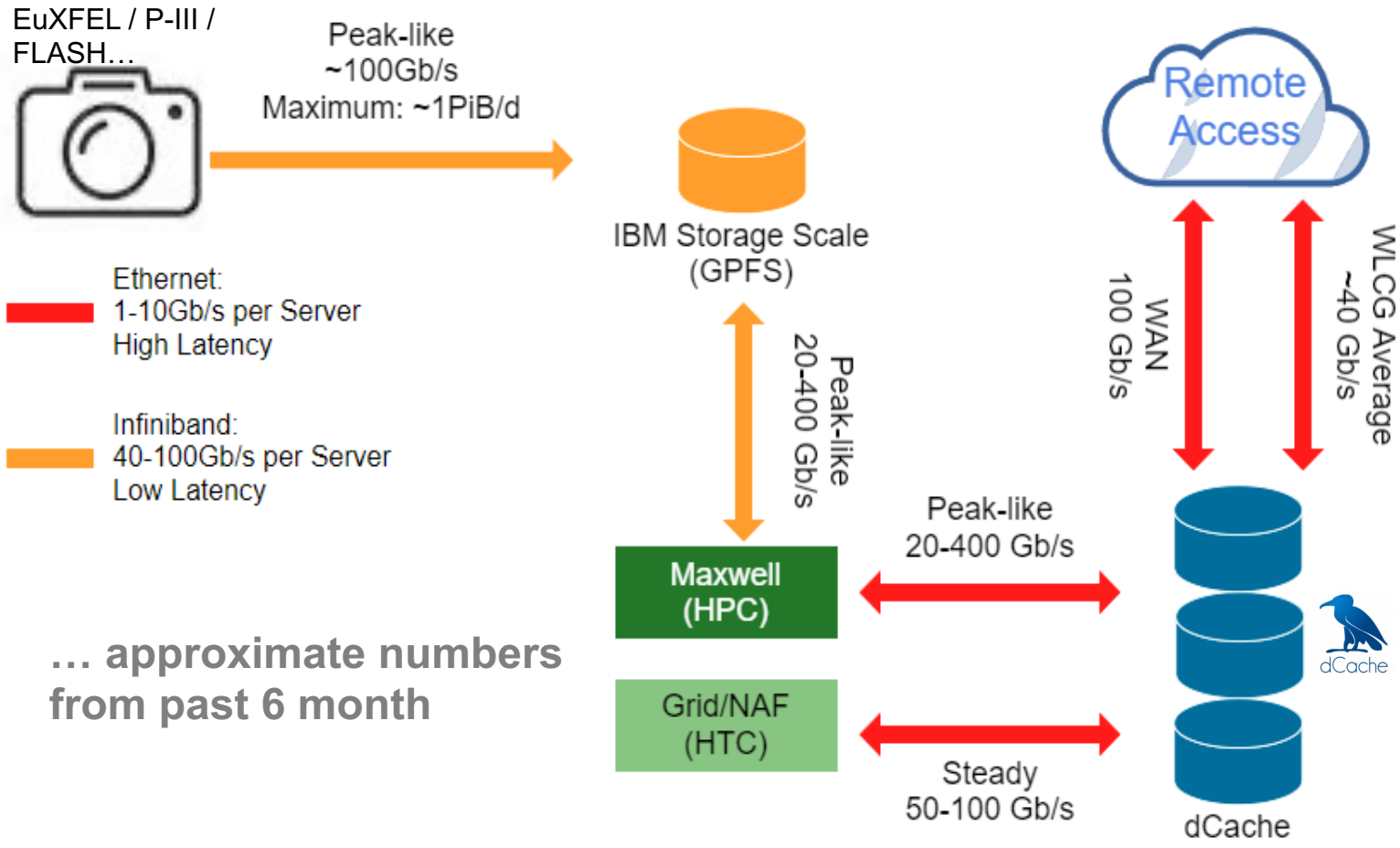
M T IDAF

Single infrastructure open for all scientists in Matter

Paradigm: Scientific Analyses are Data Driven

Strategy: Keep the Paradigm that Made the Tier-2 Successful

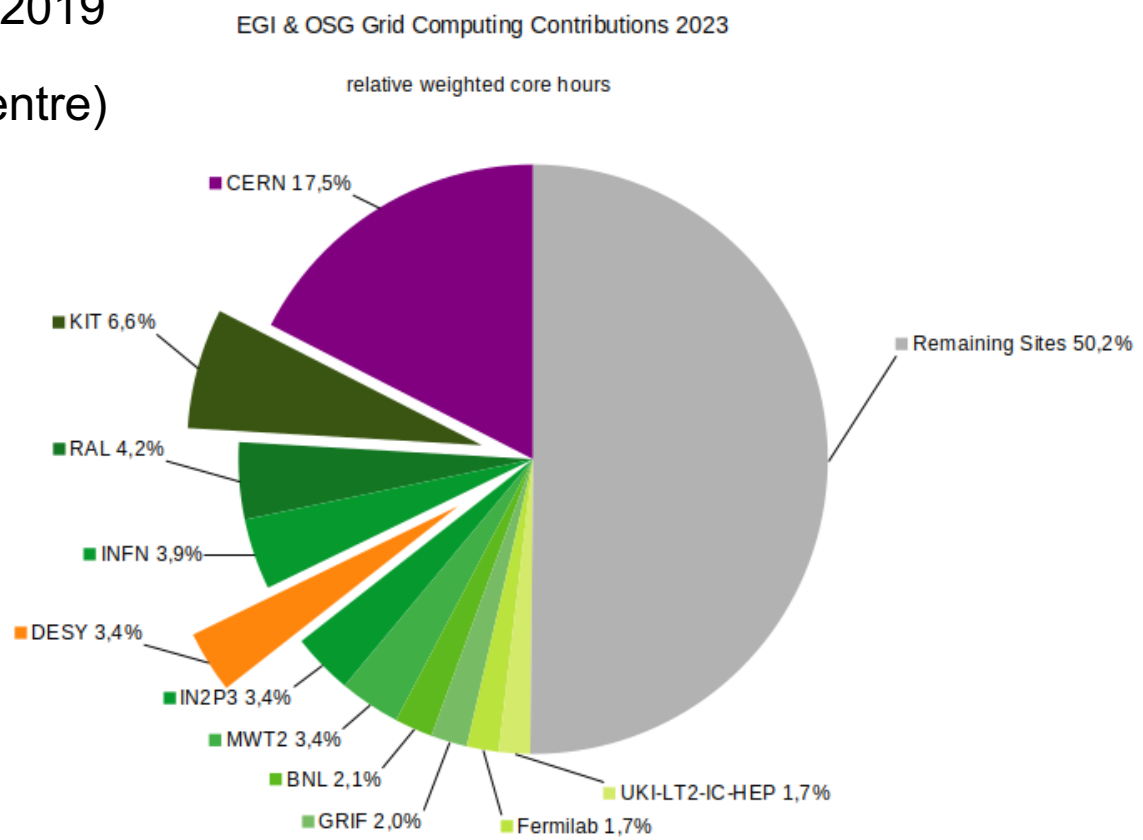
- Example: Traffic pattern in IDAF, approximate numbers from 2023H1



Continued International & new National Commitments

IDAF Inherited Previous MU Commitments for the (Astro-)Particle Physics Experiments from Tier-2

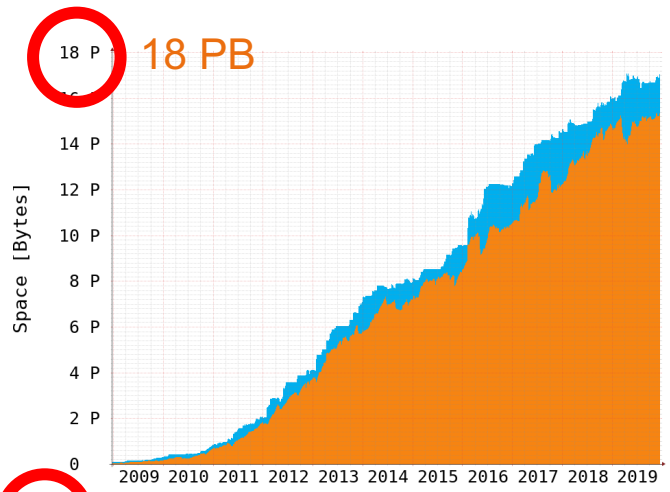
- IDAF contributed around 4% to (Astro-)Particle Physics in 2019
- 2022: share ~3.4% (IDAF still largest contributing Tier-2 centre)
- Expanded responsibilities
  - Raw Data Centre (Tier-1 equivalent) for Belle II
  - **Offer tape storage to LHC experiments (compensate affected Russian Tier-1 sites)**
- **Take over storage share from German universities**
  - KET: University Tier-2 centres to be discontinued
  - CPU shares to be taken over by some NHR sites
  - Storage to be split among Helmholtz Sites (KIT/DESY)
  - Investment in part covered by the BMBF (Verbundantrag)
  - Some additional investment expected in kind by DESY past 2025
  - New workflows expected. Will need research, and support. Close eye on network, might need expansion

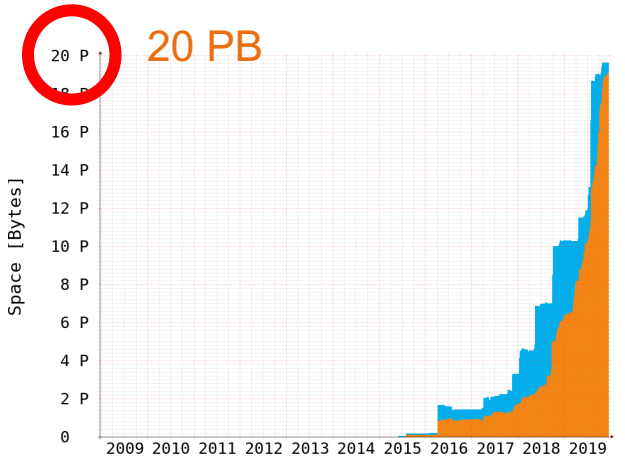


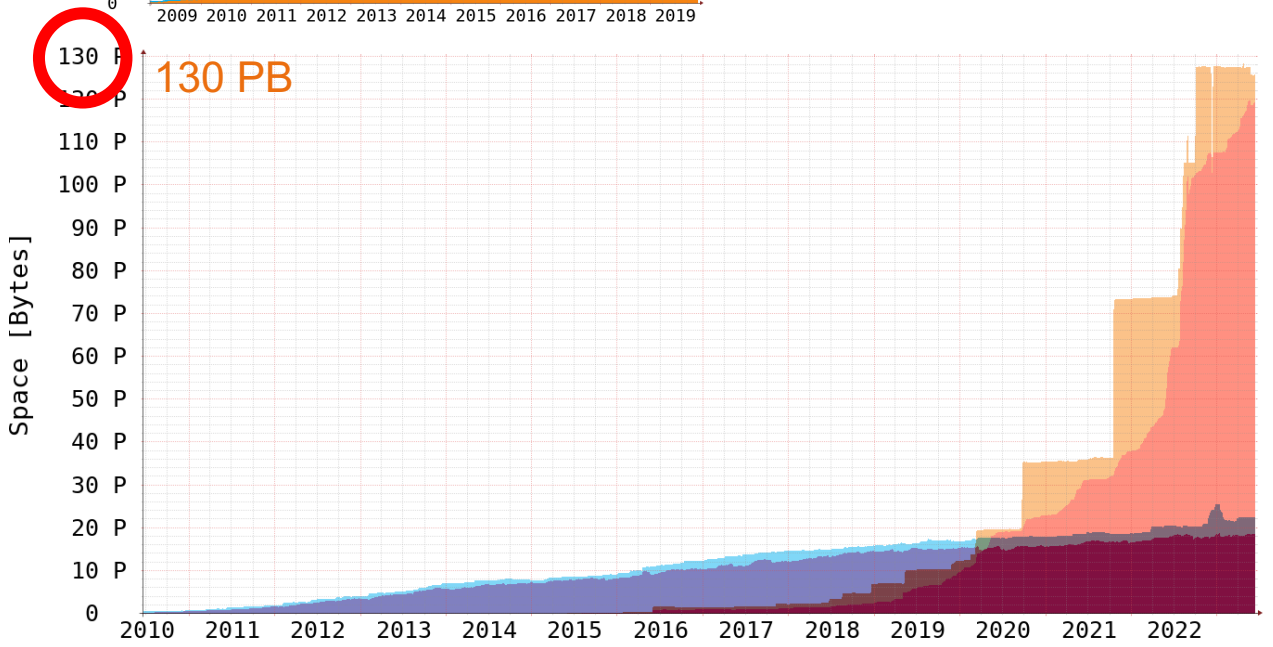
Challenges: Data Deluge in Photon Science

Photon Science and Especially European XFEL Continued to Grow Exponentially

- Data stored since beginning of PoF IV more than doubled
- **Accelerator division starts to contribute**
- HPC cluster storage similarly increased
- Capacity growth slow down/halt during end of 2022 due to funding situation
- Alternative usage of existing capacity
- **More heavy involvement of tape storage** (as done by ATLAS in the WLCG)
- European XFEL still expects to collect 50PiB in 2024
- **Data reduction** on the horizon?
- **Observe scaling issues for the IDAF**



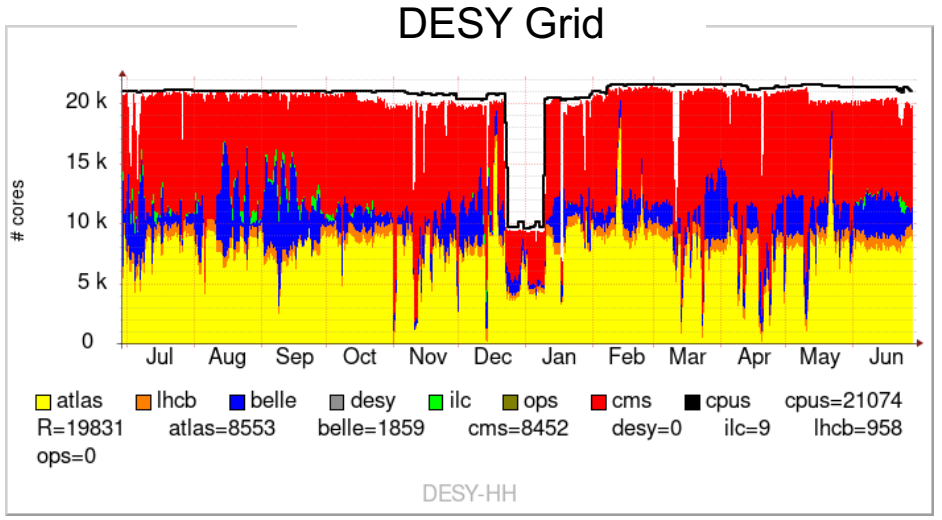


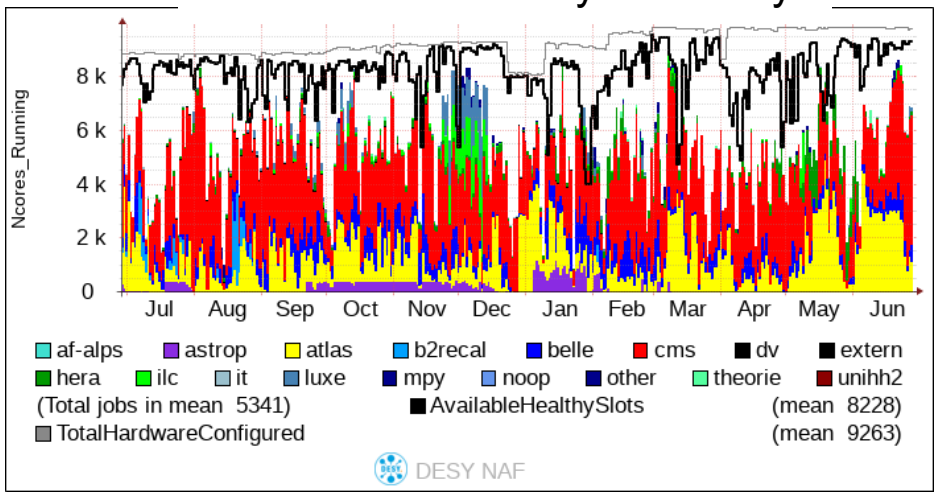


Resource and usage status IDAF

- **High Performance Cluster: Maxwell**
  - ~900 nodes (inkl. ~250 GPU), ~50k Cores. 2700 users (~1000 active in past 3 month)
  - Storage: GPFS, dCache, (BeeGFS). InfiniBand, SLURM scheduler
- **High Throughput, Production: Grid**
  - 400 nodes, 20.000 cores
  - Storage: dCache, CVMFS. Ethernet, HTCondor Scheduler – Integration in WLCG/Experiment frameworks.
- **High Throughput, Interactive: NAF**
  - 350 nodes, 8.000 cores.
  - Storage: dCache, DUST (GPFS/NFS), CVMFS, AFS. Ethernet, HTCondor Scheduler.

→ **Planning for consolidation, unification**





Challenges: Accessing Data

Users Prefer to Use POSIX — IDAF Needs to Adapt to that Fact

- Continued trend to access data ‘directly’

```
def read_frame_from_file(frame_id: int, data_file: str):
    start_time = time.time()
    with h5py.File(data_file, 'r') as h5in:
        tmp_arr = h5in['/PATH:xtdf/image/data'][frame_id]
    read_time = time.time() - start_time
    return read_time
```

- Usually only option for MML and MT
- Trend includes MU despite remote read capabilities
- **Poses the challenge of having uniform name-space across the IDAF**

HPC

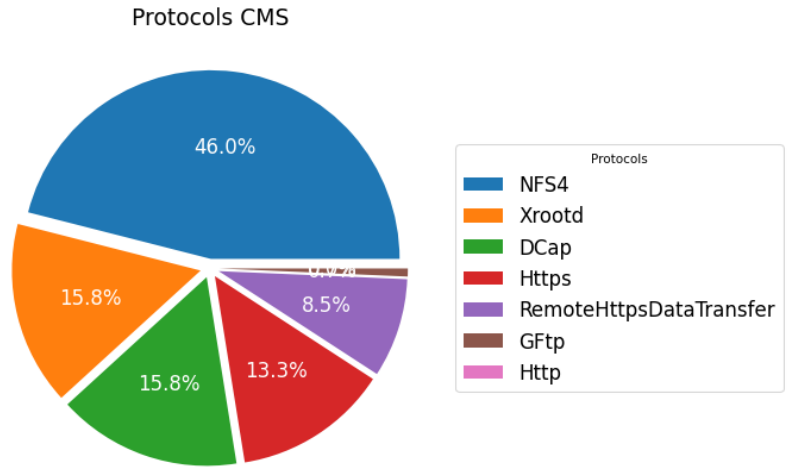
```
[vossc@max-display008] ~ $ md5sum /gpfs/dust/belle2/user/vossc/stage-rest-api.out
0108f37dbbb38103bba6d836f356d7b7  /gpfs/dust/belle2/user/vossc/stage-rest-api.out
```

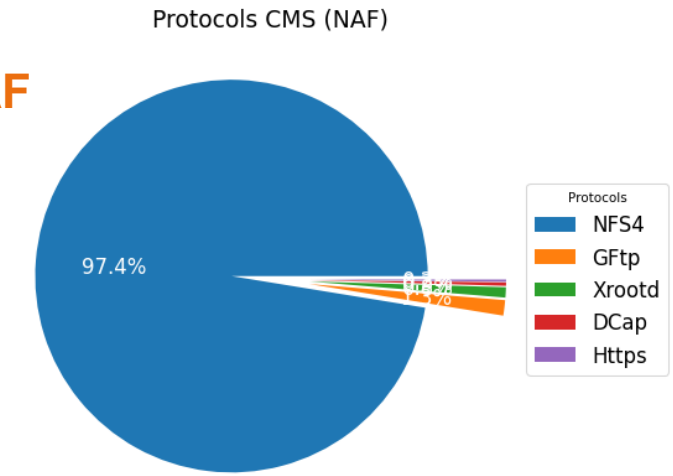
HTC

```
[vossc@naf-belle12] ~ $ md5sum /nfs/dust/belle2/user/vossc/stage-rest-api.out
0108f37dbbb38103bba6d836f356d7b7  /nfs/dust/belle2/user/vossc/stage-rest-api.out
```

- I would need to change my analysis depending on the cluster I’m on

Data Access CMS May 2023

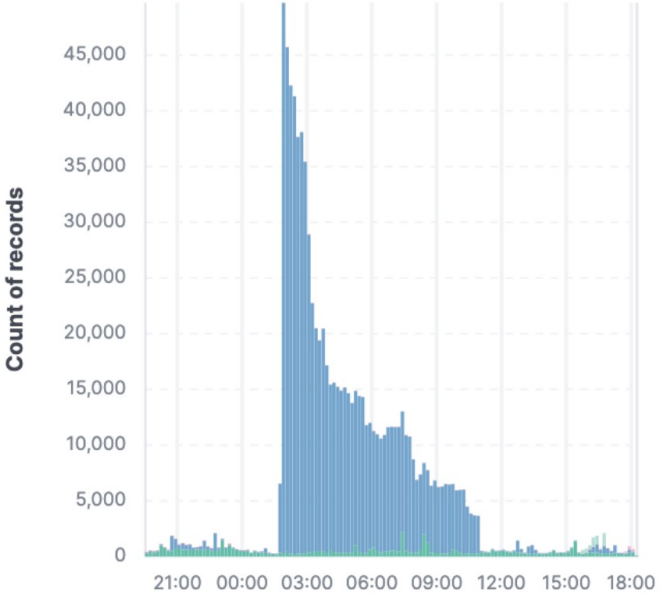


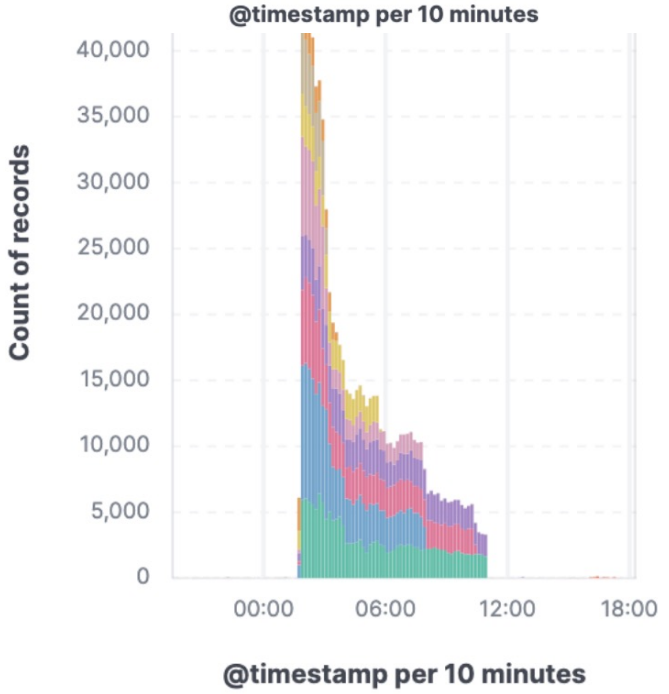


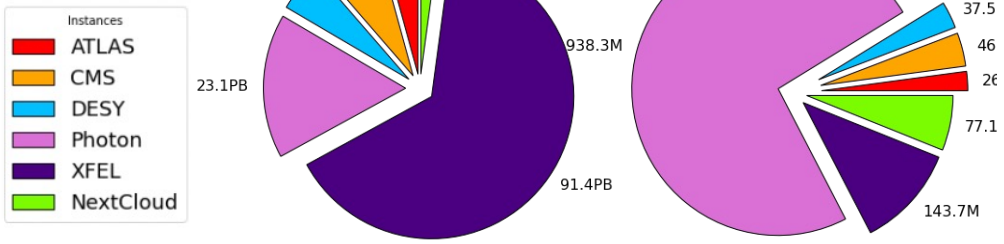
Challenge: Improved Monitoring and Analytics

Managing and Understanding the Change User Access Patterns

- Increasing capacity found to be manageable
  → read/write patterns found to be more challenging
- Departure from classic C/C++ or FORTRAN driven batch analysis
- Ease-of-Use of Python leads to higher memory footprint and excessive, repetitive data access (open files to read <1MiB)
- Increased WAN/Tape access will escalate this further
- Profit from research in :
  - Self adapting systems (e.g. Smart file replication) MTDMA
  - Improved I/O pattern, e.g. through portals (Coffea-Casa)
- Profit from research in MTDTS / MTDMA
  - Reasonable file sizes/numbers
  - Streaming/Online Analysis







Challenges: Sustainability

How to Make the Infrastructure more Sustainable

Constant improvement on PUE in DESY CC and infrastructure on DESY Campus … ongoing for years

- Hardware life cycle under close watch

Compute: Adapt hardware availability to power availability and/or user needs

Storage: Unused data on tape → Tape?

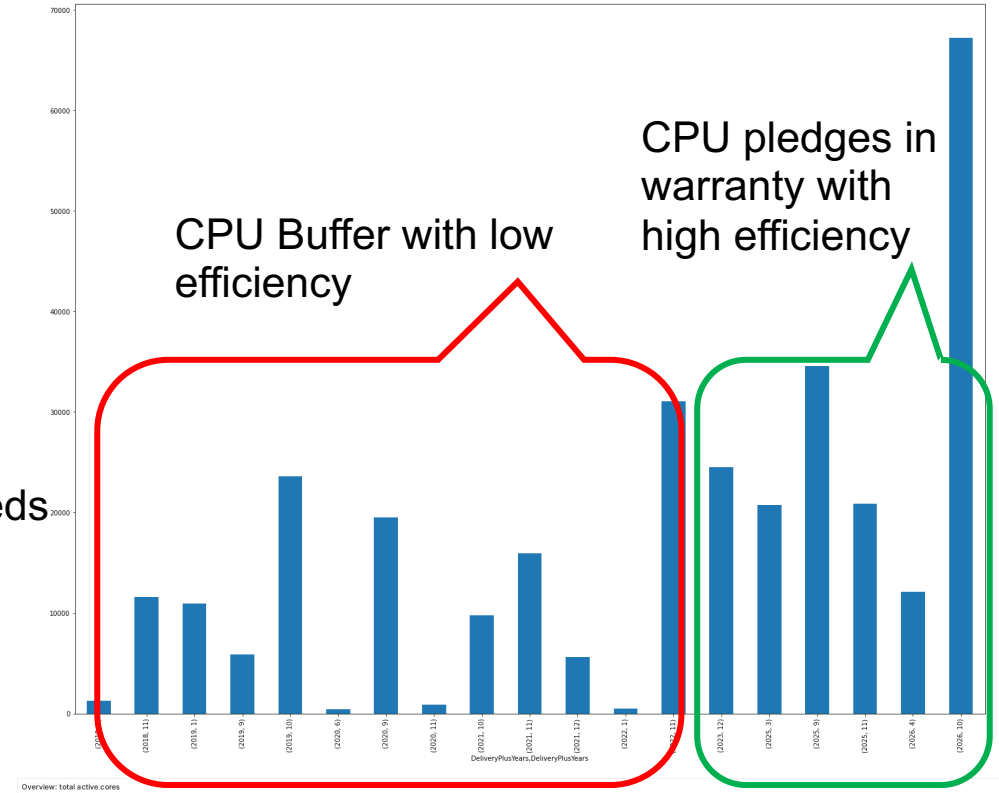
Raising **awareness** of users

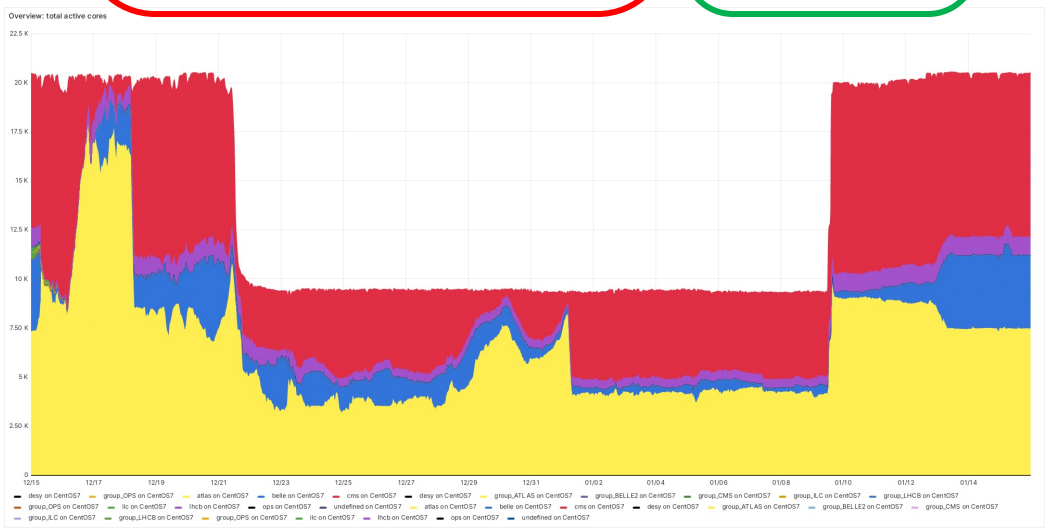
Train users on most efficient use of IDAF

Train users on tooling and optimal algorithms

Interactivity and fast reaction come with inefficiencies:

- Re-evaluate how much is needed
- Eventually tax users
- Work on scheduling and availability





Provided by T. Hartmann

Challenges: Hardware evolution and Person Power

Difficulty Acquiring Hardware and Filling Open Positions

Hardware evolution

- Short-term: Supply chains have still not returned to full capacity after end of pandemic
- Short/mid-term: GPU: NVIDIA dominance is not healthy, need combined effort to overcome
  - many interesting architectures / accelerator products out there, we should be more open and flexible
- Mid/long-term: Cloud providers driving technology … and making it private
  - Started to offer tape for ‘ultra-cold storage’ ➜ profound effect on design of tape libraries not well suited to the IDAF
  - Some architectures already now only available in commercial clouds
- Mid/long-term: First quantum computer commercially available. QC might become an additional IDAF platform

Person Power

- More and more difficult to fill open positions
- ML/AI can be filled eventually
- Regular IT positions often cannot be filled and get cancelled

Thank you

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

Tim Wetzel
Desy FH / IT / RIC
tim.wetzel@desy.de