













Beyond HEP: Photon and accelerator science computing infrastructure at DESY

Christoph Beyer¹,  and Stefan Bujack¹, and Stefan Dietrich¹, and Thomas Finner¹, 
and Martin Flemming¹,  and Patrick Fuhrmann¹, and Martin Gasthuber¹,  and Andreas
Gellrich¹,  and Volker Guelzow¹, and Thomas Hartmann¹,  and Johannes Reppin¹,  and
Yves Kemp¹,  and Birgit Lewendel¹,  and Frank Schluenzen¹,  and Michael Schuh¹, 
and Sven Sternberger¹, and Christian Voss¹,  and Markus Wengert¹

¹DESY, Notkestraße 85, D-22607 Hamburg, Germany

Abstract. DESY is one of the largest accelerator laboratories in Europe. It develops and operates state of the art accelerators for fundamental science in the areas of high energy physics, photon science and accelerator development. While for decades high energy physics (HEP) has been the most prominent user of the DESY compute, storage and network infrastructure, various scientific areas as science with photons and accelerator development have caught up and are now dominating the demands on the DESY infrastructure resources, with significant consequences for the IT resource provisioning. In this contribution, we will present an overview of the computational, storage and network resources covering the various physics communities on site.

Ranging from high-throughput computing (HTC) batch-like offline processing in the Grid and the interactive user analyses resources in the National Analysis Factory (NAF) for the HEP community, to the computing needs of accelerator development or of photon sciences such as PETRA III or the European XFEL. Since DESY is involved in these experiments and their data taking, their requirements include fast low-latency online processing for data taking and calibration as well as offline processing, thus high-performance computing (HPC) workloads, that are run on the dedicated *Maxwell* HPC cluster.

As all communities face significant challenges due to changing environments and increasing data rates in the following years, we will discuss how this will reflect in necessary changes to the computing and storage infrastructures.

We will present DESY compute cloud and container orchestration plans as a basis for infrastructure and platform services. We will show examples of Jupyter notebooks for small scale interactive analysis, as well as its integration into large scale resources such as batch systems or Spark clusters.

To overcome the fragmentation of the various resources for all scientific communities at DESY, we explore how to integrate them into a seamless user experience in an *Interdisciplinary Data Analysis Facility*.

1 Introduction: DESY as an interdisciplinary laboratory

DESY is a large research laboratory in Germany, carrying out fundamental research. Historically, particle accelerators were the founding basis of DESY and until today are a key

tool for DESY scientists. Hence, DESY develops and operates accelerators and carries out science related to accelerators. This is mainly research with photons and high energy particle physics. As a complement, DESY also carries out astroparticle physics and other branches of science more loosely coupled to accelerator research and usage. We will briefly describe the different activities.

1.1 Developing and operating accelerators

DESY operates a multitude of accelerators of different sizes. The main large systems are the Petra III electron storage ring[8], acting as a synchrotron radiation source, and the European XFEL free-electron laser (EuXFEL)[7]. In addition DESY operates FLASH[9], the world's first XUV and soft X-ray free-electron laser and some smaller pre-accelerators and accelerators for research and development. DESY also does research and development for new accelerators (e.g. Petra-IV) and accelerating techniques (e.g. plasma wakefield).

These accelerator systems are mostly run by DESY staff, with some external, long-term collaborations from other institutes or universities.

The computational needs and the structure are diverse: IT resources for operating the accelerators are split between machine groups and the central DESY IT group. Resources for research and development are mostly run by central IT. The operating systems involve Windows and Linux. Since new accelerators are usually equipped with more and more sensors and actors, the needs for processing these telemetry data increase. Research and development of new accelerators and techniques relies to a large extent on simulation, carried out on HPC systems. Telemetry data, as well as results from simulations, might also be stored on long-term archive.

1.2 Research with Photons

DESY offers over 40 experiments stations at Petra-III and FLASH. Scientists from DESY and all over the world can access these after passing a proposal system. DESY scientists also take data at other light sources. EuXFEL scientists and guest also produce data at the EuXFEL facilities. The data of these experiments is collected and analyzed to large extents at DESY.

DESY IT has worked together with Petra-III, FLASH and EuXFEL experts to set up the data taking, data storage and data analysis facilities. While usually experiment experts run the data acquisition systems, data transfer to the DESY IT computing facilities is tightly integrated in the data analysis chain and managed by DESY IT experts.

DESY and EuXFEL staff scientists prepare their work on DESY IT managed systems, e.g. perform simulations. DESY and EuXFEL staff scientists, as well as DESY and EuXFEL guests use DESY IT managed systems for data management and analysis. The increasing needs for online processing of data is accounted for in the setup of the analysis systems. In a more batch-like approach, data is reprocessed. DESY IT also saves raw data to archive.

1.3 High energy physics

DESY is part of the LHC experiments ATLAS and CMS, as well as the Belle II experiment at KEK. DESY is also involved in smaller and upcoming experiments such as ALPS, or the ILC. Some legacy analyses are still being carried out on HERA data.

DESY IT has, since long, provided resources, support and consultancy for the DESY HEP community. DESY IT not only offers services for the local DESY staff, but also to national and global computing efforts of the supported experiments, for all parts of their analysis

workflows. For some experiments, DESY saves raw and simulation data on tape, and offers additional services besides just CPU and storage.

1.4 Other activities

While the accelerator based science is mostly concentrated at the DESY Hamburg site, the DESY Zeuthen site is mostly involved in astroparticle physics. Neutrino astronomy is carried out with the IceCube and IceTop, Amanda and Baikal experiments. DESY participates in the gamma astronomy experiments H.E.S.S., Magic, Veritas, CTA and Fermi. Some of these experiments have high demands in computing and storage, and DESY is actively supporting these communities.

The DESY theory groups are an important pillar for innovation and research for all DESY science and beyond. The DESY IT groups in Hamburg and Zeuthen actively support their computing efforts in simulation and computational physics.

2 Central computing facilities: Past and present setup

DESY IT provides several systems for data management and analysis. We will briefly present the past and present setup of the most important ones.

2.1 Grid Services and Compute

In 2003, the first Grid site at DESY was installed. The DESY Grid infrastructure provides a large portfolio of Grid services such as Virtual Organization Management Services (VOMS) or file catalogs[1]. The actual workhorses are a large common batch system based on HTCondor, with about 20.000 cores, and high-capacity dCache storage elements. These resources serve different projects, most of them with a high energy particle physics background, as multipurpose infrastructures. Most notably, the DESY contributions to WLCG for ATLAS, CMS and LHCb as well as Belle II are provided using the shared DESY Grid infrastructure.

The batch system of the Grid is a standard high-throughput computing cluster: Jobs are placed to a scheduling system, which handles them to a pool of commodity server nodes. The network is a standard peer-to-peer Ethernet network. Inter-node-communication is not foreseen. Scheduling is based on group fairshare, and carried out on a per-core basis with dynamic core count request possible per job on a node.

2.2 Mass storage and archival: dCache

dCache is a storage system for extremely large amounts of data[3]. It was originally developed in the context of scientific data management and has been shown to scale far into the Petabyte range. dCache unifies the storage space from many, potentially quite heterogeneous, storage servers into a single virtual file system. It takes care of data hot spots and failing hardware and can ensure that at least a minimum number of copies of each dataset reside within the system for resilience or performance optimization. Furthermore, dCache supports a diverse set of access protocols and authentication and authorization methods. dCache is Free Software, available under a permissive AGPL license without licensing costs, and has been under active development for almost two decades. Its developers have close ties to the research community and understand the needs of scientific data management firsthand.

The development started at DESY, which today is still the lead site, in collaboration with NDGF and Fermilab.

DESY operates seven instances of dCache, managing a total of ~ 60 PB of disk and interfacing ~ 30 PB of archive storage (the tape archive itself is managed by another system) for all scientific communities.

2.3 The National Analysis Facility (NAF)

In addition to the Grid infrastructure, the National Analysis Facility (NAF)[2] has been in operation at DESY since 2007. It complements the DESY and German Grid resources and serves similar projects. As the NAF supports direct interactive access, it allows for fast-response workflows necessary for development, debugging, testing and small-scale private productions - important complements to the Grid infrastructure, which provides computing resources for a continuous massive production albeit with higher latencies. The NAF provides around 8.000 CPU cores and a small number of GPUs. The HTCondor batch system of the NAF is set up in a similar high-throughput computing scheme as the Grid batch system, with the addition of interactive Jupyter[14] jobs and whole-node scheduled GPU jobs, and a preference for jobs with small requested resources to reflect the fast-response requirements of the NAF.

2.4 Maxwell HPC Services and Compute

Launched in 2011, the Maxwell HPC cluster[5] has quickly grown into a massive HPC platform. The Maxwell cluster is a truly collaborative and cross-disciplinary compute infrastructure also open for different applications and tenants via a buy-in model. The applications range from data analysis (PETRA III, EuXFEL, etc.) over machine/deep learning (high energy physics) to massively parallel computational tasks (simulations for laser wakefield accelerators and PETRA IV design studies, molecular dynamics and photon-matter interactions, etc.). Maxwell provides around 30.000 CPU cores and 100 GPU. SLURM[11] is used as a batch system, with an integration of interactive Jupyter notebooks. Maxwell scheduling relies on a per-node scheduling and supports inter-node-communication over a dedicated InfiniBand network, also used for fast storage access. The Maxwell cluster is defined by IT such that it is homogeneous in OS and software, and complies with some minimal hardware standards. Groups can buy their own hardware and place it into the shared cluster. Groups have prioritized access to their own resources, all users can place jobs also opportunistically on resources of other groups, which might however be preempted.

2.5 Storage and Data Taking: ASAP3 and EuXFEL GPFS

Petra III, FLASH and XFEL data is stored on the GPFS cluster file system[12]. DESY IT, together with experts from the experiments, has put infrastructure in place to manage beamtimes, user access, data transfer and lifecycle[6].

Currently, DESY IT stores over 20 PB of data on 7 GPFS clusters. These systems are primarily accessible from the Maxwell cluster, since they rely on InfiniBand network. Export services for NFS and SMB exist for limited purpose and bandwidth. A dedicated long range InfiniBand line connects EuXFEL online and offline systems.

3 Interdisciplinary Data Analysis Facility (IDAF)

As one can see, DESY IT provides several compute and storage infrastructures. Based on the experiences DESY IT has gathered from running these infrastructures, we have started

to prepare for the future by laying the ground works for the IDAF: Interdisciplinary Data Analysis Facility.

The basic idea behind the IDAF is to bring all compute and storage systems together, for better utilization, efficient management, better support and services.

In this section, we will discuss which boundary conditions by users, use cases and technology have changed, and how these changes reflect in an opportunity for a new, consolidated setup.

3.1 HPC vs. HTC, batch vs. interactive

Computing in HEP was, for a very long time, strictly "embarrassingly parallel", and relied on the one-job-one-core paradigm. This is no longer the case: Multi-core-jobs have come up as a way to more effectively use resources. While we do not yet see inter-node-communication as enabled by traditional HPC systems, the use of HEP of such HPC systems has grown. HEP has learned how to use these often very special, large HPC systems.

Conversely accelerator research and development is a classical example of HPC use case. However, data analysis of accelerator data is most likely a task which does not require traditional HPC systems. A classical batch system might be sufficient, maybe extended by advanced and interactive access methods and workflow tools.

Scientists working in photon science at DESY have only become familiar with batch queuing systems in the recent past. Their use cases before just did not need such large-scale computing facilities. Their job profiles today (and most likely in future) directly benefit from an HPC like system by the fast and low-latency InfiniBand network to access their data, while this network is only rarely used for parallel jobs and inter-node communication. Photon science use cases often need a large amount of RAM for their computations. The Maxwell HPC systems offer a relatively large amount of RAM per CPU, when compared to the other batch systems at DESY as well as compared to most of the systems on the HPC TOP500 list[15]. Besides of the large memory footprint, workflows are rather well suited for serial batch processing, thus would fit an HTC system. Photon science experiments also need more interactive resources during data taking for online computing than in the past. The exact needs may vary for each individual experiment, hence, pooling resources and using a resource management system can increase the utilization.

Thus, communities tend to have been using *their* resources more out of habit rather than from a point of optimal resource utilization. Moving forward towards the IDAF will need technological efforts, but at a same time, support efforts: We see the need to expand and adapt our support for the different user communities with their various computing experiences. Else we see our resources being not efficiently utilized as well as the groups inefficiently spending their work and time.

3.2 New computing models and needs

Machine learning has become increasingly important, and hence the usage of GPUs. However, we see different approaches: Grid workflows usually bring a fully trained network with them, and only inference is done on DESY resources. Therefore, the usage of normal CPU's is sufficient. In the NAF, some training is done, but so far, the usage is small, but increasing. The accelerator research and development groups do currently not use GPUs a lot. Machine learning is more a topic for accelerator operations, and will get stronger in the future. Photon science use cases at DESY benefit most from GPU accelerators when compared to other DESY scientist use cases, which can be seen by the relatively large number of GPUs present in the Maxwell cluster. These GPUs are used for training of neural networks, but also for

normal computations with specialized algorithms optimized for GPUs. For efficiency reasons, most GPU systems are available through batch systems. So far, we do not make use of multi-user GPU technologies, since most users require all GPU RAM. A small number of GPUs are, however, available in interactive, multi-user machines. These are mostly provided for development and small-scale testing purposes. No further job control is applied. GPUs also help serving FastX graphical logins in interactive display nodes. Most users rely on CUDA, thus limiting the choice of GPU manufacturers. Most GPUs have good double precision capacities, since this is mostly required by users, or users simply ignore the details of their application. We purchase GPU systems only in small quantities at a time, and often customized to the needs of one group. We thus have a mixture of systems with different GPU generations, GPU capabilities and number of GPUs per system. All of our batch provided GPU systems tend to be well utilized. Investigation of new scheduling concepts is especially important for further usage of GPU systems in the IDAF.

LHC experiments are moving away from a hierarchical organization of homogeneous compute and storage sites to a data-lake model with a mixture of dedicated, opportunistic and commercial resource providers[16]. It is likely, that other HEP experiments will follow this path.

Resource provisioning for Photon science will become more streamlined and organized. Computing requirements might even become parts of the proposal submission and review process, and computing consultancy before, during, and after data-taking part of the offer of DESY experiments.

New computing paradigms, based on compute cloud methods and workflows, or based on container creation and deployment, will change the technology utilized for resource provisioning. They might also change the way users interact with IT, and redefine users and IT roles.

4 Implementation

It is clear, that the change from the current setup to the IDAF facility is a process rather than a one-shot migration.

4.1 Combining clusters, and making cluster usage transparent

As a very first step towards implementing the IDAF, we analyze whether clusters can be combined. It has turned out, that the NAF and the Grid compute clusters are very similar in systems setup. They differ in scheduling policies, driven by different resource access expectations. The systems setup on an operation system, application level and storage access has been made identical. This already now eases administration. In the last years, a common batch and scheduling system HTCondor[10] has been deployed, replacing the systems in place before. This further makes administration more efficient. We are investigating on how best to merge the two clusters, such that jobs from the Grid and the NAF share the same hardware, and make overall usage even more efficient[13].

A further step has been to open Maxwell to dedicated HEP users. First, users strong in machine learning profit from the more numerous GPU systems in Maxwell. Second, dedicated HEP production users can launch simulation jobs opportunistically with very low priority on free Maxwell resources that would otherwise be idle.

It has shown that mere compute jobs with only little I/O footprint can run transparently independent of the cluster. In order to enable data analysis jobs, local storage access needs to be made available cross clusters. Currently, local storage access is bound for performance,

stability and security reasons to the realm of a cluster. We are investigating on how to offer local storage to all users, independent of the cluster in a site wide consistent namespace and authentication scheme. As long as the clusters are centrally managed by IT, this turns out to be a performance, stability and costs challenge. Expanding the access of POSIX network file systems to compute clouds or container orchestration backends brings in a severe additional security challenge.

4.2 Further combining clusters: The case of HPC and HTC

As we have shown, the Maxwell HPC clusters has a dedicated InfiniBand network, and is usually equipped with more RAM than the other HTC clusters. In an ideal world, all systems of a combined IDAF compute cluster would have an identical setup, that would be with InfiniBand and high RAM configuration. While this would ease utilization and administration, it would be economically difficult to realize, and would not give a benefit to many applications.

It is very likely, that there will be two clusters, or two parts of a common cluster, that would have different hardware setups in terms of InfiniBand and RAM configuration. A common cluster with otherwise identical system setup could, however, give the possibility for current users of the Maxwell HPC system to transparently use smaller systems when their applications does not need high RAM or fast InfiniBand network. This would reduce the overall cost of systems, and also reduce the complexity of the InfiniBand network.

4.3 Scheduling, Meta-Scheduling, and no visible scheduling

DESY uses two batch scheduling products: HTCondor and SLURM. At the time of decision, HTCondor was well suited for large scale high-throughput computing, very extendable, and integrated well in HEP workflows. SLURM featured excellent capabilities for HPC parallel computing, which HTCondor was lacking. We currently see no major change in the positioning of these products, so it is unlikely, that we can consolidate to one of these two in the foreseeable future. However, some form of coexistence is possible, either on a meta-scheduling layer, or even on a scheduling level. This is a direction we will further investigate.

It is important that our clusters can integrate into existing meta-schedulers outside of DESY, e.g. pilot factories of the LHC experiments. Other collaborative systems are being set up, e.g. in the HIFIS or EOSC context, which often base on Cloud computing or container orchestration workflows. DESY resources also need to integrate into those systems.

On a more general note, we observe that some users tend to stick to a compute infrastructure because they have learned how to use the respective scheduler, even if the infrastructure itself might not be most efficient. Such cases could profit from a side wide meta-scheduling system, which would chose the best matching infrastructure for job execution.

Batch systems, and their inherent waiting times and additional submission overhead, are often only accepted by users because no interactive, easy scaling alternative is available. DESY has started to offer Jupyter notebooks on Maxwell and the NAF, and integrate those interactive tools with the HTCondor and SLURM batch queuing systems. Launching such a notebook happens with a very small latency. The integration of such interactive workloads into batch systems was both a technical and conceptional challenge. Currently, these systems are well received by users, and we will report in more details about our work and experience at a later stage.

In addition, the usage of Apache Spark as an easy scaling, interactive analysis tool has been set up for selected pilot users.

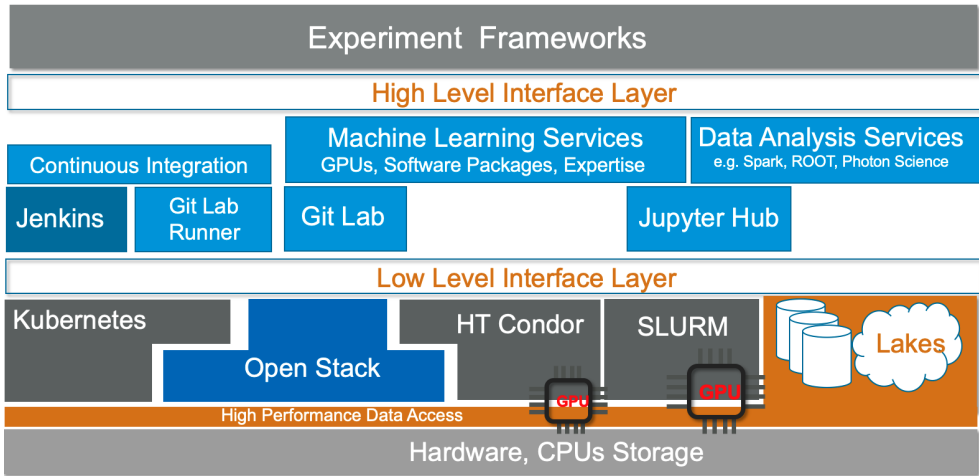


Figure 1. One possible vision of the setup of the IDAF: How the different layers and technologies interact, and interface to the users.

4.4 More on Cloud, Virtualization, and Containers

With the availability of commercial cloud resources, not only the expectation of immediate availability has changed. Also a more customizable and portable environment, and more standardized access methods are requested.

Both the NAF and Maxwell support launching jobs in containers. The batch systems and the launching methods were adapted such that transparent access to POSIX mounted file systems is enabled from within the container, guaranteeing security and privacy of data. Containers, "guided" in such a way, are an option for implementing workflow systems for users of beamlines. While groups and their probes change often, the beamlines and their detectors persists. A workflow engine, hiding the complexity of the batch systems, could take the user data set and the payload as containers, and transfer to large scale computing power. Ideally, such a system would also allow for container development, creating, testing and deployment - a CI/CD system, where also users can build software and analysis pipelines with the batch systems as computing backends.

The DESY compute cloud, running on OpenStack, is currently in a pre-production phase, serving selected projects. A well performing, secure and easy access to the large data repositories dCache and GPFS at DESY is under investigation, and is one of the prerequisites for this infrastructure to be fully production ready for analysis. Once fully available, the compute cloud will be integrated into the IDAF as well, to complement the existing resources. One possible vision is depicted in figure 1.

A long-term vision is the offering of "virtual data centers" per experiment: Users define their needs in term of hardware (CPU, GPU, storage, network), access, applications, and support. Out of the flexible setup of the IDAF, these resources are allocated for the duration of data taking and later analysis, such that users can work in their environment. With such a setup, users can also be isolated against other users and their usage of the system.

4.5 Storage systems and access

Most DESY science is data centric. Taking data, performing a fast online analysis, safely storing data, having fast, easy and secure access to data for analysis and later long-term archive are of utmost importance.

In the last decade, DESY has spent a lot of efforts making storage systems easy to access by users, and scalable and fast. This is a matter of choosing the right product, the right access protocols, and the right environment.

The main storage systems, and their access protocols, for the IDAF are:

- **dCache:** used as the main storage for HEP data: WAN import/export is done using several protocols (GridFTP/HTTPS/XrootD). Local access to dCache can be done via WAN protocols as well as dCap or via a NFS v4.1 mount[4]. NFS access has been pushed in the past years, since it offers a very easy and well know interface for users. For the NFS mount to work properly, a well managed client environment is necessary, in order to get ACLs based on UID and GID right. dCache is also used as secondary mass storage for Photon science data, as well as data for accelerator research, development and operation. Long-term archival is also done through dCache.
- **IBM SpectrumScale (GPFS):** is used as main storage system for online and offline storage for Photon science experiments. The core of the system is a cluster file system, and uses the NSD protocol in the Maxwell InfiniBand network. All systems within the Maxwell InfiniBand network are centrally managed by DESY IT, UID and GID are enforced and hence ACLs based on them can be safely used. Import and export of data from other systems is done via special cluster export services, using NFS in controlled environments or authenticated SMB.
- **IBM SpectrumScale as backend for the NAF user space:** While the core filesystem again uses the NSD protocol in a dedicated InfiniBand fabric, all access by users is done via NFS cluster export services. In order to use NFS ACLs, UID and GID (and client IP addresses) need to be enforced.

One can see, that mounted file systems play a key role in the past and current setup of clusters. They rely on UID and GID (and client IP addresses) being controlled in order to realize data privacy. In a well managed cluster under central control, this is secure: Client IP addresses cannot be spoofed, and UID and GID cannot be faked (and in the event of IP spoofing of UID/GID faking in a central cluster, stronger security models, e.g. relying on Kerberos, will not help, since they would most probably also be compromised). For systems outside of the well managed cluster, this is different: Authentication and authorization other than just IP address and UID/GID must be used. It shows that the SMB protocol is much easier to be handled by the users on their desktops and laptops than e.g. an NFS mount, e.g. it is well integrated into Ubuntu Nautilus or macOS Finder - plus it brings the Kerberos security with it for free. If the storage system providers only use NFSv4 ACLs on their systems, both NFS and SMB can understand these, and serve in an identical manner to their respective clients. dCache and GPFS both use NFSv4 ACLs, which facilitates user access from both systems.

So far, mounting storage was restricted to a certain cluster, e.g. Maxwell GPFS data was not available on the NAF, and vice-versa. Sometimes this was bound to technology (missing InfiniBand links) or just policy decisions. Usually, there was no need for cross-cluster availability of data, since user communities are distinct.

In an IDAF, where clusters might be combined, and hence user communities share the same resources, the clear separation of file system availability cannot be maintained. This leads to a more general discussion about file systems, file system availability, data provisioning other than file systems, provisioning to new compute entities:

Is a global namespace what we need or even want? If several distinct communities are served that do not share data, there is no need for a global namespace. Establishing it nevertheless might be a technical burden, and mix failure domains, thus lowering the overall user experience and availability. On the other hand, only global namespaces allow for transparent migration of workloads in the whole IDAF.

Is an object store a viable alternative? Mounted file systems do have advantages, e.g. in usability. However, due to the restrictions imposed by POSIX, and the tight integration into the operating system, mounted file systems also have disadvantages. Object storage might be an alternative.

How to make all storage systems universally accessible? Not only needs data be available on all DESY cluster systems, but also world wide access: Users work in international collaborations, and have affiliations to institutes all over the world. Data will be exchanged between data sources and institutes. Already now, data is taken outside DESY (CERN, KEK, IceCube, different photon sources, ...), but analyzed at DESY. Authentication and authorization, combined with easy-to-use, secure and scalable data transfers - fitting each individual community - are topics to be addressed. While some communities are very advanced in this subject, other communities will need adapted solutions.

How to translate the ACLs into the Container? Or user-managed VMs in a compute cloud? Workloads will be put into containers or VMs in a compute cloud - on-site and potentially also off-site. Currently, enforcing ACLs relies on well managed IP addresses and UID/GID, something that is not trivial if possible at all in user created containers or VMs. How to enable ACLs in this scenario will be another dimension of the storage discussions ahead for IDAF.

Answering these questions will be one of the major challenges, and will have a huge impact on the direction the IDAF implementation will take.

5 Summary and outlook

DESY IT serves the computing needs of different communities, gathered around developing, operating and using accelerators for science. An increasing need for computing and storage capacity, as well as changing user workflows, have brought up the necessity to bring the DESY compute landscape to a new level. The Interdisciplinary Data Analysis Facility (IDAF) is the vehicle, under which these changes will take place.

References

- [1] The DESY Grid Centre, Haupt A., Gellrich A., Kemp Y., Leffhalm K, Ozerov D. and Wegner P., Journal of Physics: Conference Series vol **396** p 042026 (2012)
- [2] Evolution of Interactive Analysis Facilities: from NAF to NAF 2.0, Andreas Haupt et al J. Phys.: Conf. Ser. **513** 032072 (2014)
- [3] The dCache project website <http://www.dcache.org/>
- [4] Elmsheuser J., Fuhrmann P., Kemp Y., Mkrtchyan T., Ozerov D. and Stadie H., Journal of Physics: Conference Series vol **331** (IOP Publishing) p 052010 (2011)
- [5] DESY Maxwell HPC Cluster website <https://confluence.desy.de/display/IS/Maxwell>
- [6] ASAP3 - New Data Taking and Analysis Infrastructure for PETRA III, M. Gasthuber et al., Journal of Physics Conference Series **664(4)**:042053 (2015)
- [7] The European XFEL, <http://www.xfel.eu/>
- [8] The Petra-III storage ring, <https://petra3.desy.de>
- [9] The Free-electron Laser, <https://flash.desy.de>

-
- [10] Distributed Computing in Practice: The Condor Experience. Thain D., Tannenbaum T. and Livny M., *Concurrency and Computation: Practice and Experience*, Vol. 17, No. 2-4, pages 323-356, February-April, 2005.
 - [11] SLURM: Simple Linux Utility for Resource Management. Jette M. and Grondona M., United States: N. p., 2002. Web.
 - [12] GPFS: A Shared-Disk File System for Large Computing Clusters. Schmuck F. and Haskin R. (January 2002). *Proceedings of the FAST'02 Conference on File and Storage Technologies*.
 - [13] Consolidating the interactive analysis and Grid infrastructure at DESY. Gellrich A. et al., *CHEP 2019*, to appear in these proceedings
 - [14] Jupyter Notebooks - a publishing format for reproducible computational workflows. Kluyver et al., doi:10.3233/978-1-61499-649-1-87 (2016)
 - [15] TOP500 list of HPC systems (Nov 2019 edition) <https://www.top500.org/lists/2019/11/>
 - [16] Architecture and prototype of a WLCG data lake for HL-LHC, Bird I. et al., *The European Physical Journal Conferences* **214(5)**:04024 (2019)