

Introduction to Scientific Computing at DESY

Dealing with Scientific Data Challenges

Kilian Schwarz

DESY, August 16, 2024



Agenda

01 introduction to DESY

- International context
- Accelerator, photon science, particle physics

02 introduction to IT and SC

- IT in numbers, organisation
- Scientific Computing

03 Scientific Data Challenges

- Data ingest, storage, archiving
- Data access and processing

04 sustainable computing

- RF2.0

05 ML and AI

- Self adaptive dCache

06 3rd party projects

- PUNCH4NFDI, FIDIUM, HSC

07 student projects

- Summer student project
- bachelor/master theses, internships

08 Summary and outview

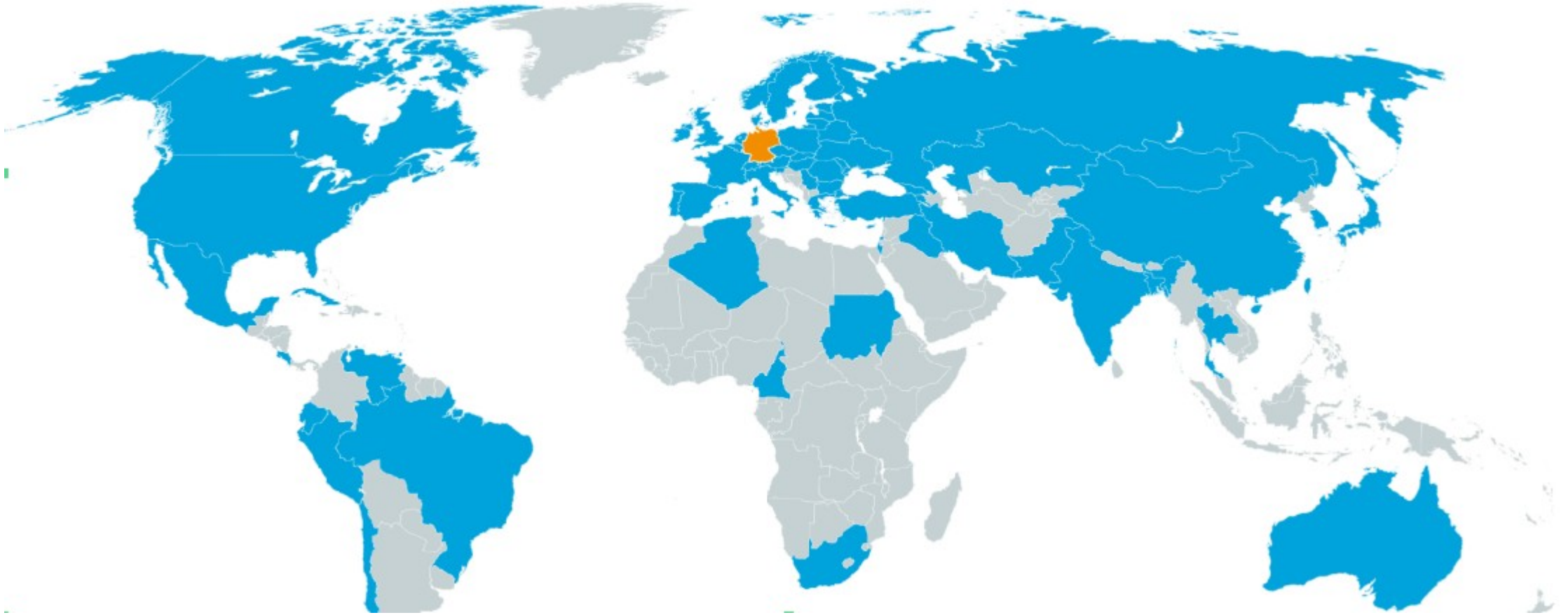
- User communities and motivation

Chapter I

Introduction to DESY

Guest Scientists at DESY

3000 scientists from over 40 countries visit DESY each year



What do we do at DESY ?

Accelerator physics, photon science, particle physics

Accelerator Physics

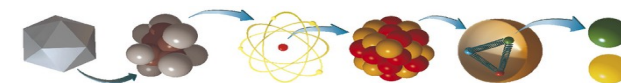
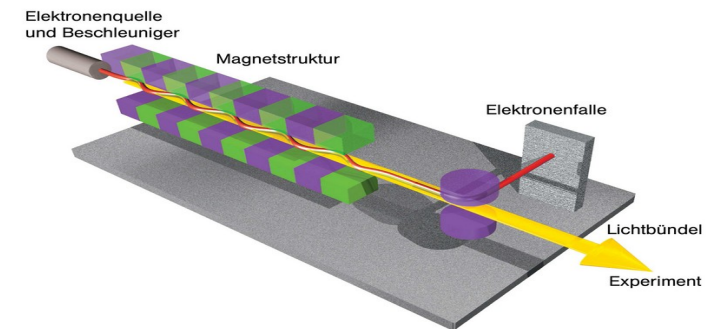
- Development, construction and operation of particle accelerators
- Modern light sources, plasma accelerators
- IT: highly parallel simulations in HPC environments

Photon Science

- Investigation of molecules and materials with special light from particle accelerators

Particle Physics

- What are the fundamental building blocks and forces in the universe ?
- How did the universe come to existence ?



What do we do at DESY ?

PETRA III

- Originally for particle physics, then pre accelerator for HERA
- Transition to most brilliant synchrotron source of the world 2007-09
- Experimental hall of 300 m with 14 beam lines for 27 experiments
- Nano technology and material research
- Currently extension for more beam lines
- IT: a set of pictures, like a movie which has to be analysed



What do we do at DESY ?

European XFEL (X-Ray Free Electron Laser, 2017)

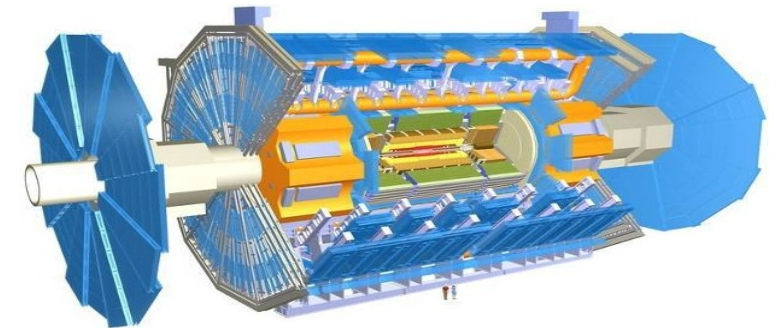
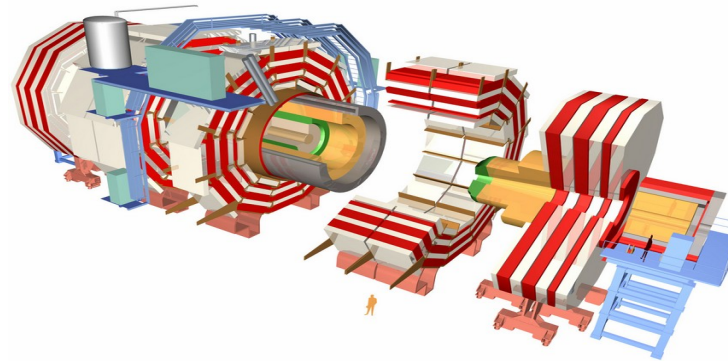
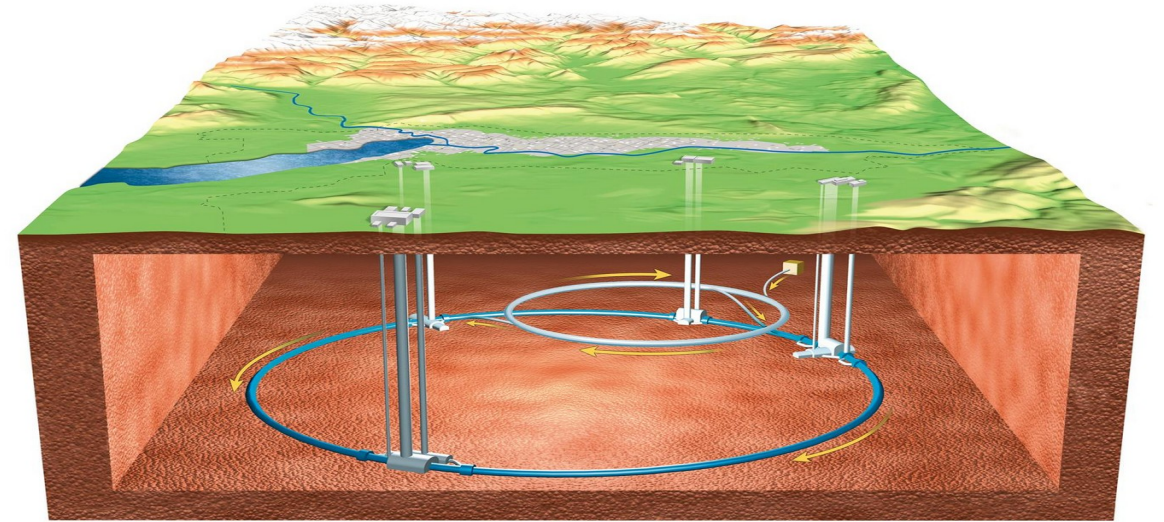
- 17.5 GeV / 3.4 km / strongest X-ray laser of the world, first experiments 2017



What do we do at DESY ?

Large Hadron Collider (LHC) at CERN

- Proton proton (ion ion) ring accelerator
 - Circumference: 27 km
 - Worldwide strongest particle accelerator
 - Measurements since 2009
- Targets
 - Higgs properties (discovered 2013)
 - New particles beyond standard model
- DESY involvement
 - Particle detectors CMS and ATLAS
 - Theory, Grid centre
 - IT: many million events in parallel, intrinsic parallelisation



Chapter II

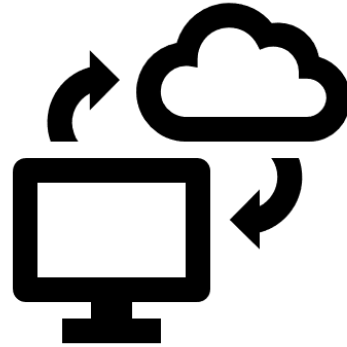
Introduction to IT and SC

Introduction to IT

DESY IT in numbers



- 1000 sqm space
- 1.5 MW power
- up to 4 MW cooling



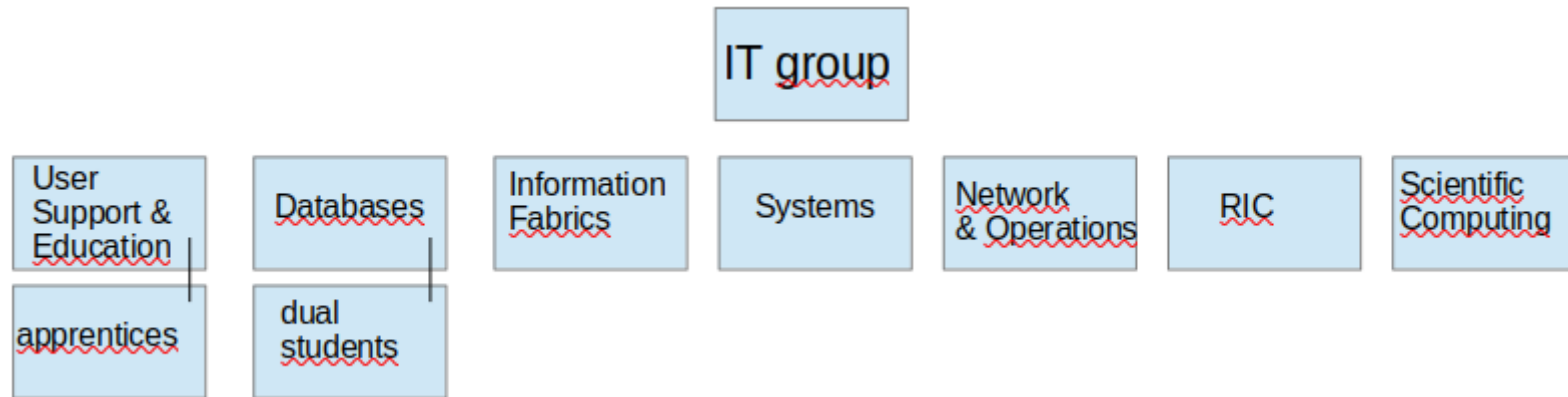
- ca. 300 racks, 5.000 servers (+1.350 virtual systems)
- IDAF: ca. 60.000 cores, 400 GPUs
- Storage: 150 PB dCache, 90 PB GPFS, 130 PB tape
- Connectivity: 100 Gb/s backbone, links with 2x50 Gb/s to global research networks
- ~ 40.000 IP addresses



- The team:
 - O(100) staff
 - O(10) apprentices
 - O(10) students (dual study)
- O(10 000) users (DESY, ntl, intl)

Introduction to IT

Current structure at DESY IT



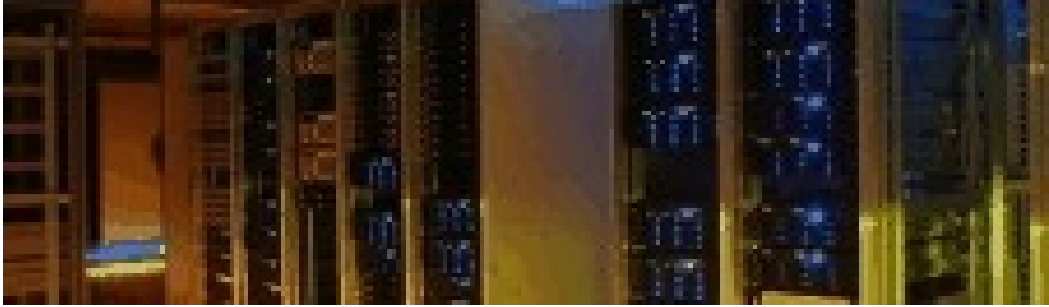
IT group:

about 112 employees without students and apprentices

- User support & education: User Consulting Office, Software
- Databases: databases and applications, training
- Information Fabrics: email, file service, registry, AAI, ...
- Systems: computing centres, operating systems, security together with other IT groups
- Network & Operations: data network and telephone
- RIC: EU projects and Helmholtz-wide platforms
- Scientific Computing: scientific experiment support, storage middleware

Introduction to IT and Scientific Computing

What do we do in Scientific Computing ?

- All computing aspects directly related to Physics research at DESY
 - Development and operations of mass storage systems including tape integration
 - Development of data ingestion and streaming platforms
 - Scientific experiment support on application level including software development
 - Scientific support of experiments in the areas of big data management and data processing, HPC and HTC computing, Grid services
 - Close collaboration with other SC groups at DESY
- 
- Help in planning and design of new research related computing projects
 - ML & AI
 - Bachelor and Master theses, internships
 - Summer Students

Chapter III

Scientific Data Challenges

Scientific Data Challenges

Ingest

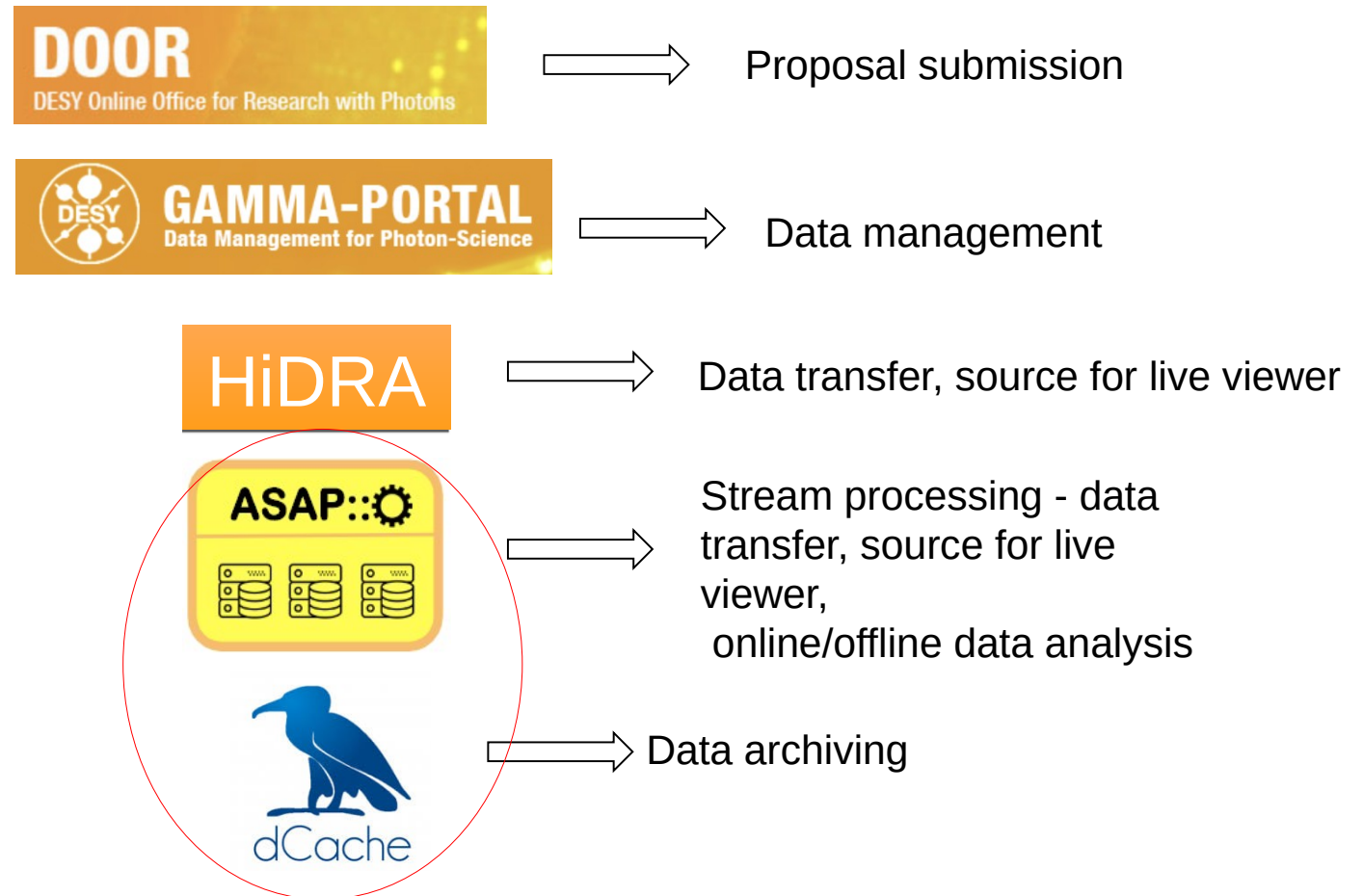
Ingest

- High data ingest rate
- Multiple parallel streams
- High durability
- Effective handling of large number of files

Scientific Data Challenges

ASAP – IT environment for Photon Science experiments at DESY

- Hardware
 - network
 - proxy nodes
 - compute nodes
 - storage
- Software
 - user portals
 - data transfer
 - online/offline data processing
 - archiving

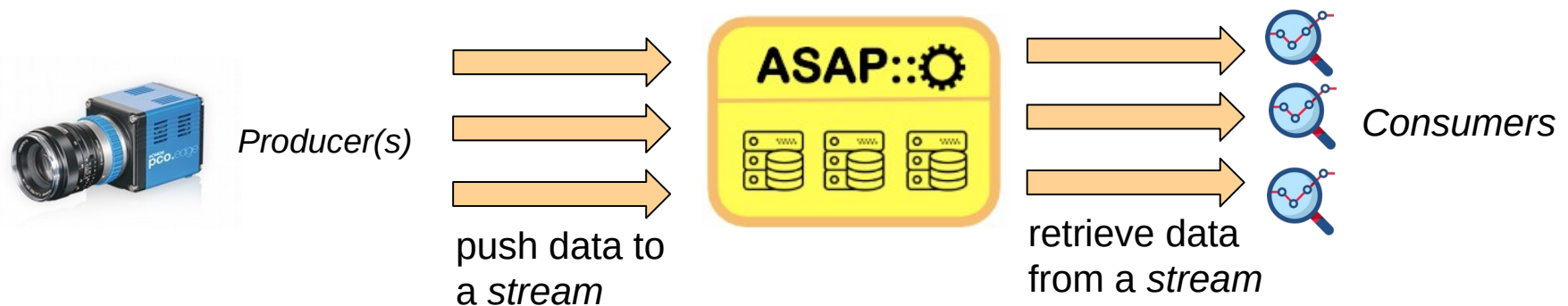


Scientific Data Challenges

ASAP::O

- > middleware for high-performance next-generation detector data analysis
 - Provides API to ingest data to the system - e.g. takes care of the “first mile” between the experimental hall and the compute center (high-performance data transfer)
 - Provides API to retrieve data from the system - e.g. for data analysis synchronous (online) and asynchronous (offline) to data taking
- > Basic characteristics
 - Scalable (N sources, K network links, L service nodes, M analysis nodes)
 - Highly available (services in Docker containers managed by Nomad/Consul or Kubernetes)
 - Efficient (C++, multi-threading, RDMA, ...)
 - Provides user friendly API interfaces (C/C++, Python, REST API)
 - Runs on Linux/Windows/...

Example:



Scientific Data Challenges

Storage

Ingest

- High data ingest rate
- Multiple parallel streams
- High durability
- Effective handling of large number of files

Sharing & Exchange

- 3rd party copy
- Effective WAN Access
- In-flight data protection
- Identity federation
- Access control

The dCache Storage System

Distributed Scalable Mass Storage System

- Central element in overall storage strategy
- Collaborative development under open source licence by
 - DESY (leading laboratory)
 - Fermilab
 - Nordic E-Infrastructure Collaboration (ex. NDGF)

Particle Physics

- 75% of all remote LHC data stored on dCache

Astronomy & Radio-Astronomy

- LOFAR Long Term Archive (~40 PB) & CTA

Photon Science

- European XFEL and others for archival

Accelerator and Detectors

- FLASH, LINAC

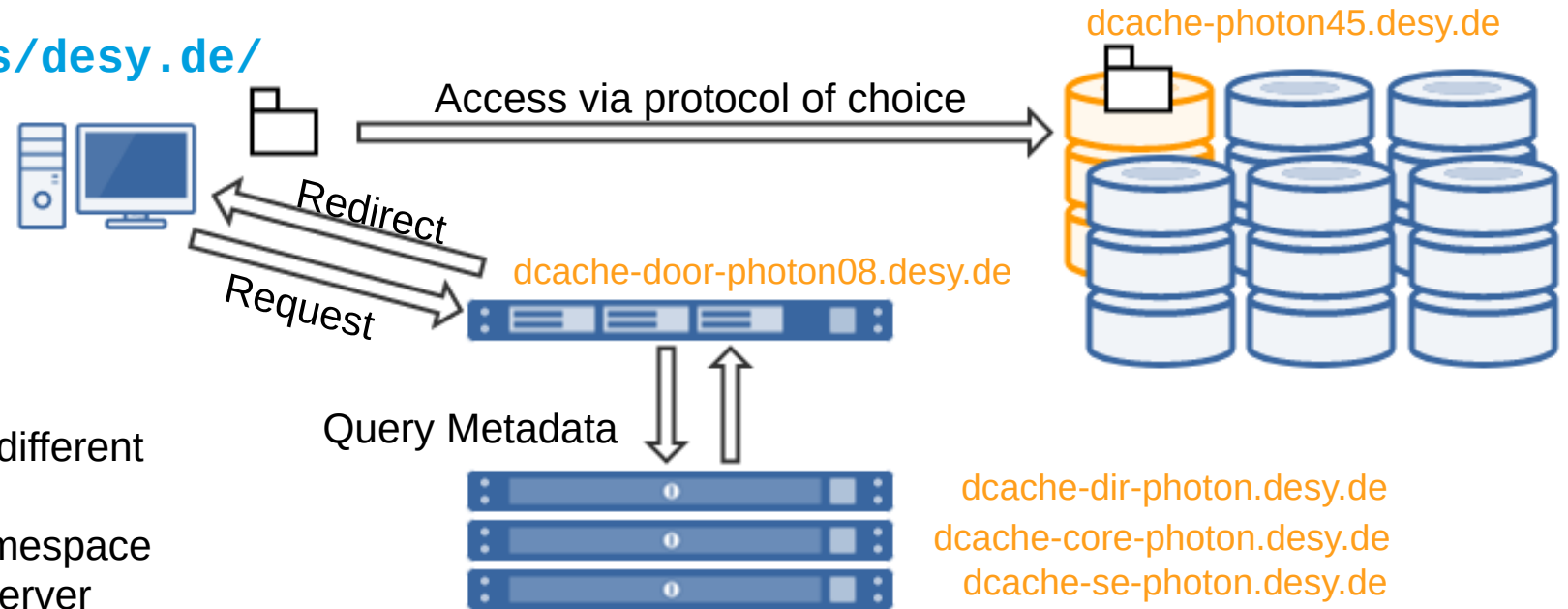
- Distributed Storage System with single name space
- Big Data application
- Micro service architecture
- Open Source



dCache Architecture

User access to dCache

- Use dCache: Access to [/pnfs/desy.de/](https://pnfs.desy.de/)



- Based on Micro-Services

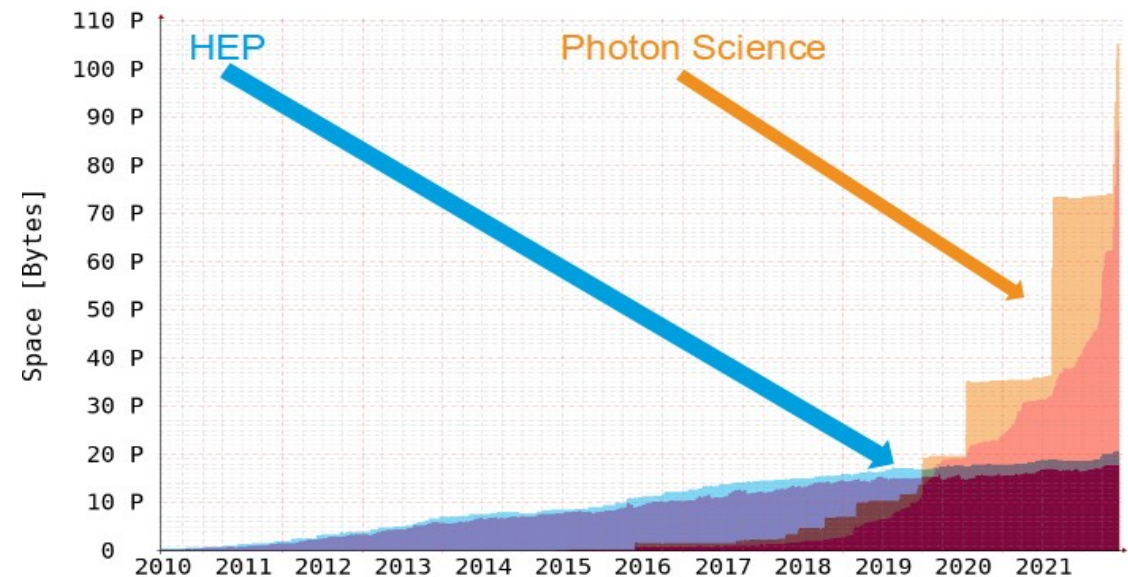
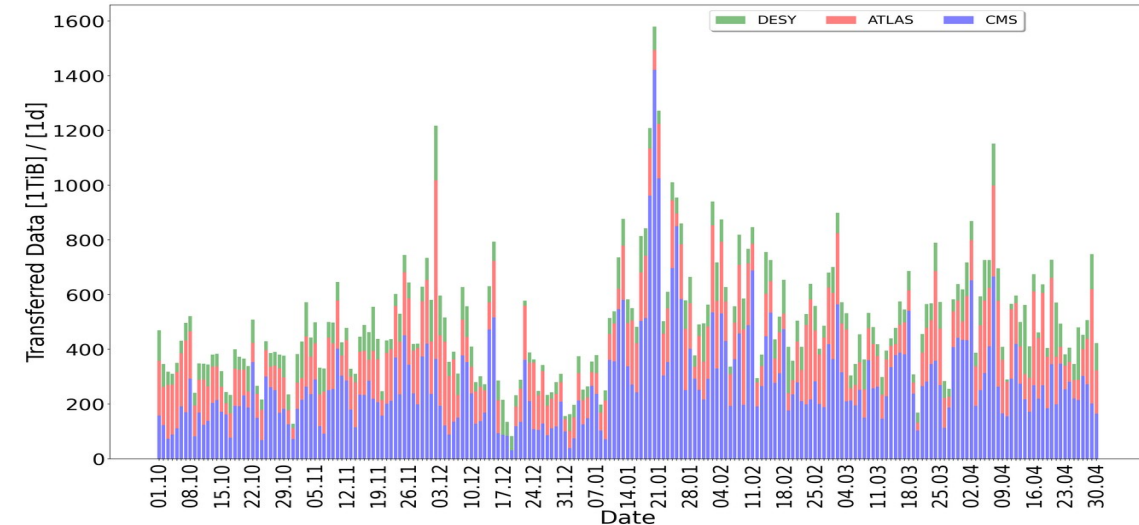
- **Doors** – supporting each a different protocol
- **Heads** – pool selection, Namespace
- **Pools** – data storage and server

- dCache instances for Photon Science/Machine, European XFEL, ATLAS, CMS, Belle/ILC/DPHEP, Sync&Share

The dCache Storage System

DCache at DESY

- **ATLAS** → 6,0 PiB disk capacity, 496 pools, ~26.700.000 files, 4.4 PiB overall stored data
- **CMS** → 10,1 PiB disk capacity, 788 pools, ~51.700.000 files, 7.8 PiB overall stored data
- **DESY** (Belle-II, ILC detector development, central IT-Services, ...) → 4,5 PiB disk capacity, 294 pools, ~30.400.000 files, 7.3 PiB overall stored data
- **XFEL** → 91,4 PiB disk capacity, 3844 pools, ~147.400.000 files, 77.2 PiB overall stored data



Scientific Data Challenges

Long term archive

Ingest

- High data ingest rate
- Multiple parallel streams
- High durability
- Effective handling of large number of files

Sharing & Exchange

- 3rd party copy
- Effective WAN Access
- In-flight data protection
- Identity federation
- Access control

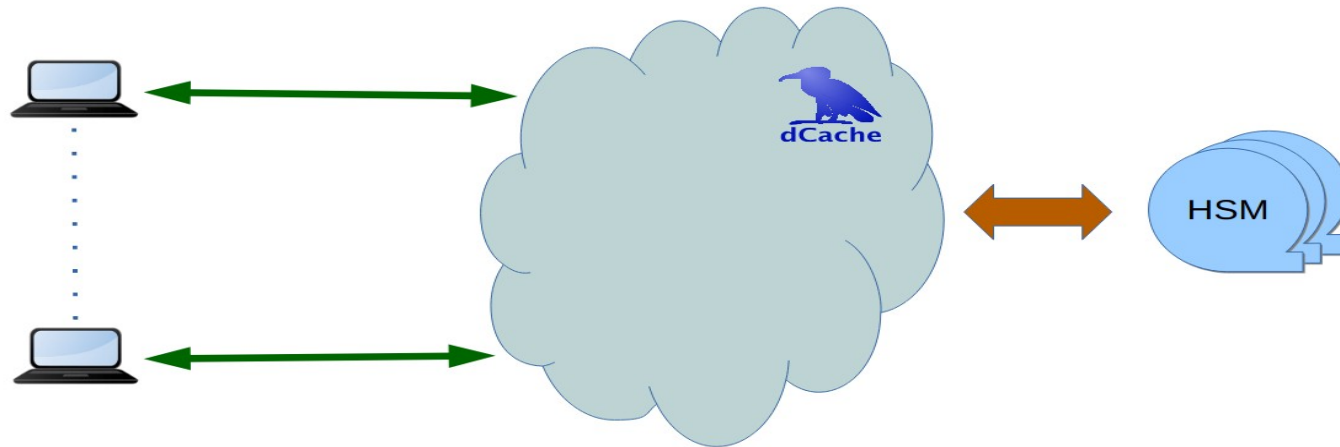
Long Term Preservation

- High Reliability
- Self-healing
- Automatic technology migration
- Persistent identifier

Scientific Data Challenges

Long term archive

dCache+HSM Tandem



All access to scientific data on tape goes exclusively through dCache!

- **HSM (Hierarchical Storage Management)** enables an outsourcing of file to cheaper storage media, as e.g. tape.
- For the user the files are still visible in the online file system
- When data are accessed HSM automatically triggers staging
- Criteria for outsourcing can be number of access, disk filling state, age or size of files

Scientific Data Challenges

Long term archive / CERN Tape Archive



- **CTA has been integrated into dCache**
- Pros: CERN product, GPL3, well defined software development process
- All DESY experiments moved to new system already

Scientific Data Challenges

Data processing

Ingest

- High data ingest rate
- Multiple parallel streams
- High durability
- Effective handling of large number of files

Sharing & Exchange

- 3rd party copy
- Effective WAN Access
- In-flight data protection
- Identity federation
- Access control

Long Term Preservation

- High Reliability
- Self-healing
- Automatic technology migration
- Persistent identifier

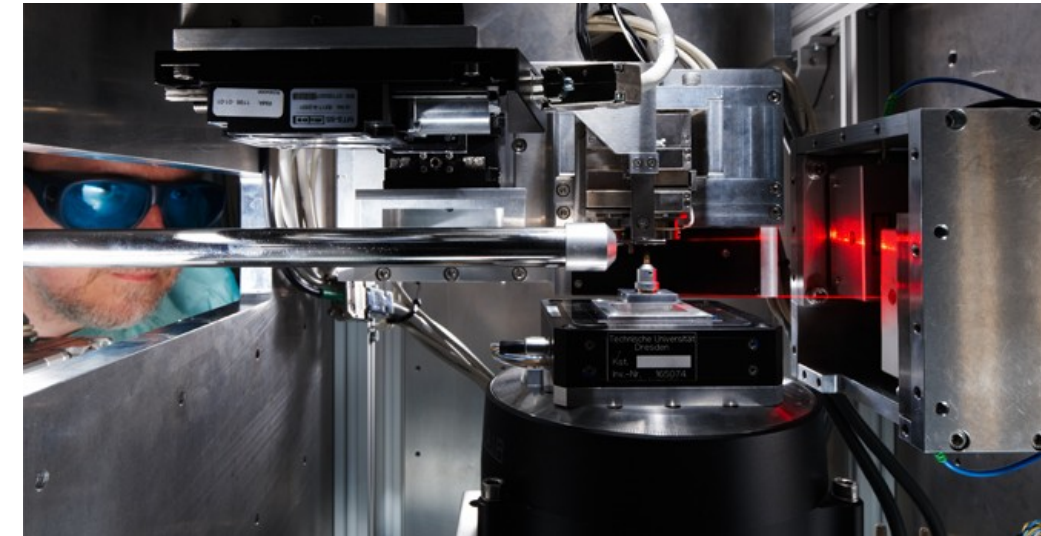
Analysis

- High CPU efficiency
- Unstructured access patterns
- Standard access protocols
- Access control
- Local user management

Scientific Data Challenges

Data processing – user community

- **Particle physics**
 - ATLAS (CERN)
 - Belle II (KEK)
 - CMS (CERN)
 - ILC
 - LHCb (CERN)
 - ALPs II, BabyIAXO, MADMAX, ... (DESY)
- **Photon Science**
 - XFEL (DESY)
 - FLASH (DESY)
 - PETRA III (DESY)

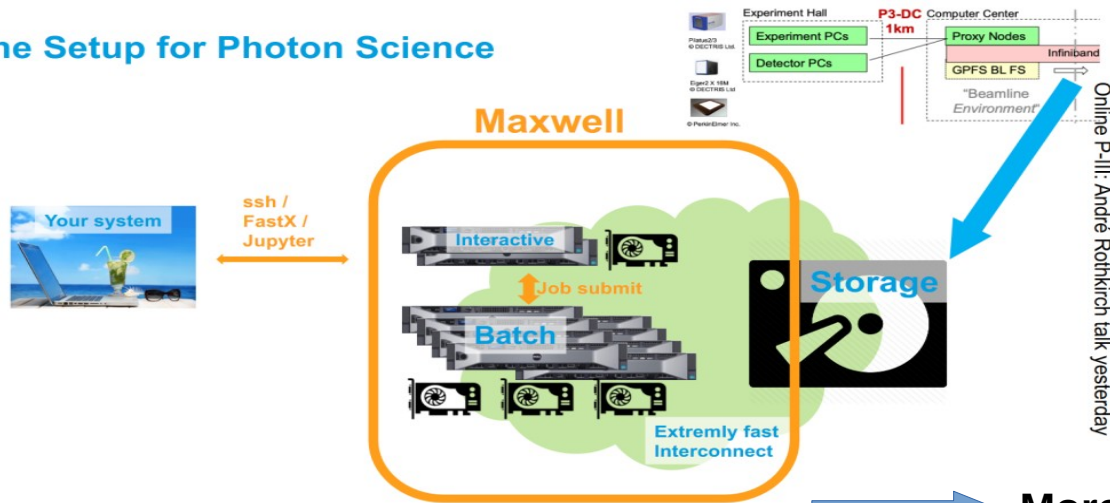


Scientific Data Challenges

Cluster infrastructure provided by IT/Systems

Data processing – IDAF (Interdisciplinary Data Analysis Facility)

The Setup for Photon Science



DESY | IDAF | Yves Kemp, CDCS symposium, 28.4.2022

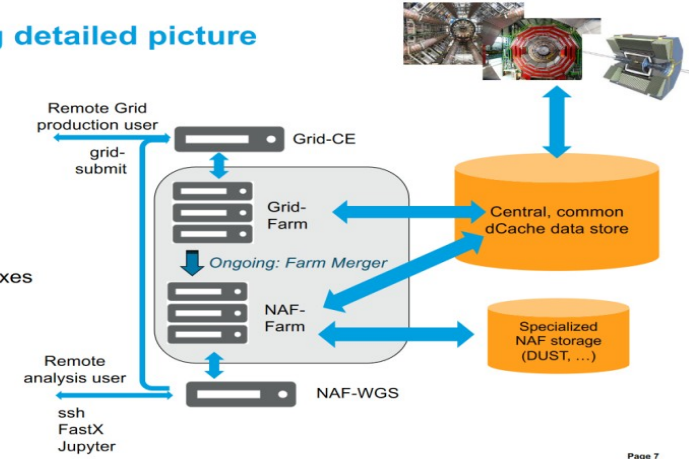
IDAF in numbers

- Compute: Maxwell + Grid + NAF
- Dcache + GPFS + BeeGFS
- 60.000 CPU cores + 320 GPUs
- 150 PB data on disk
- 2.700 server (compute, storage, management)
- 0.5 Megawatt power consumption

GRID & NAF: The big detailed picture

Grid: Serves worldwide HEP community through Grid protocols
NAF: Serves national HEP community through interactive protocols

Access protocol is just one/few boxes large compute behind, as well as storage infrastructure and access is (mostly) identical



DESY | IDAF | Yves Kemp, CDCS symposium, 28.4.2022

Page 7

More about Maxwell HPC cluster in next slides

Page 6

• HPC vs. HTC

- HPC: large amounts of compute resources in short time
- HTC: maximum number of job throughput in given (longer) time period

Scientific Data Challenges

Data processing – WLCG (Worldwide LHC Computing Grid)

- The Worldwide LHC Computing Grid (WLCG) project is a global collaboration of around 170 computing centres in more than 40 countries, linking up national and international grid infrastructures. The mission of the WLCG project is to provide global computing resources to store, distribute and analyse the ~200 Petabytes of data expected every year of operations from the Large Hadron Collider (LHC) at CERN on the Franco-Swiss border.

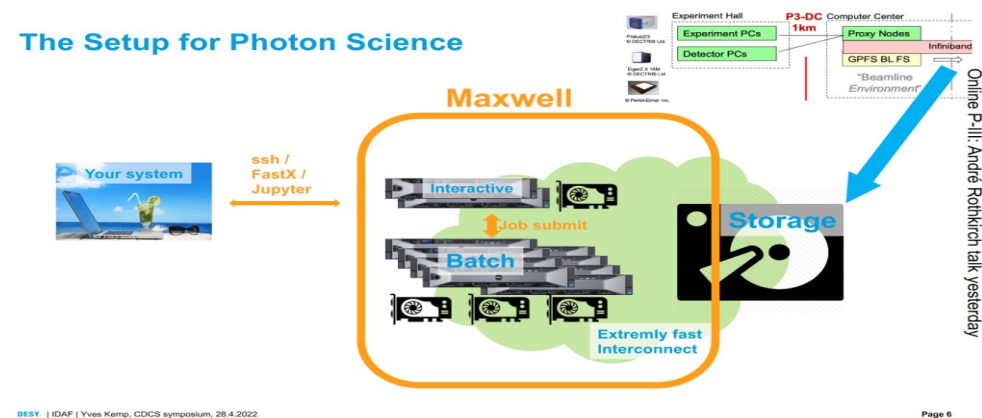


- DESY is a site in the WLCG and in Belle II Grid

Scientific Data Challenges

Data processing – Maxwell HPC Cluster

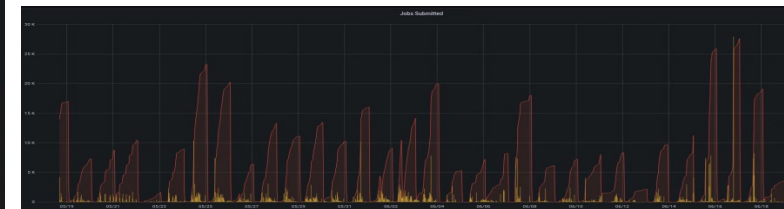
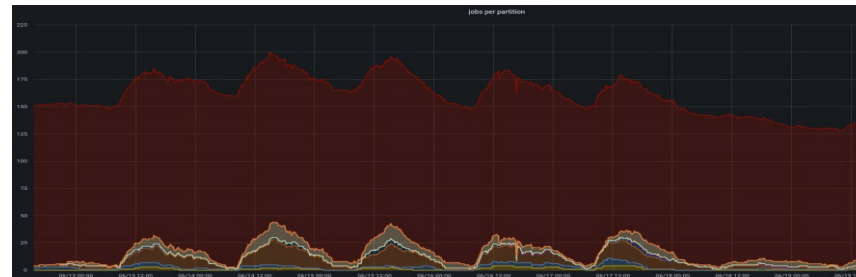
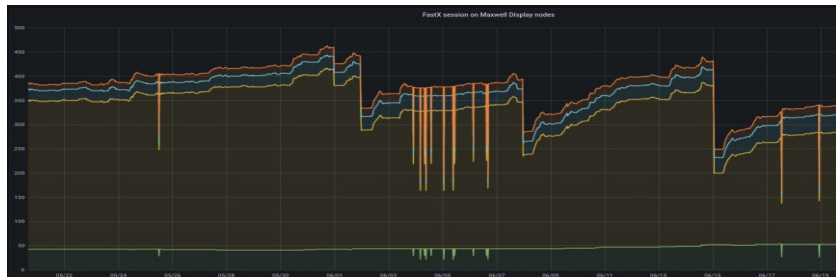
- HPC characteristics:
 - low latency network (IB)
 - fast cluster storage (GPFS, BeeGFS)
 - mass storage (dCache)
 - substantial GPU resources
 - no afs, krb support
 - targeted at massively parallel and GPU computations



Scientific Data Challenges

Data processing – Maxwell HPC Cluster

Users	2300
Concurrent graphical login	up to 500
Concurrent JupyterHub users	up to 200
Jobs	~11.500.000
Concurrent Jobs in Queue	up to 30.000
Scientific Publications	~50/yr



Run as buy in model:

- IT provides infrastructure and services
- storage to a large extent group financed, shared GPFS HOME

Scientific Data Challenges

Data processing – Maxwell HPC Cluster

Total Number of compute nodes (CPU + GPU)	776 (~50% Eu.XFEL)
Total number of cores with hyperthreading	59456
Total number of physical cores	29856 (~50% IDAF)
Number of CPU nodes	600
Theoretical CPU peak performance	1021 TFlops
Total RAM	418 TB
Number of GPU nodes	176
Total number of GPUs	331
Theoretical GPU peak performance	2104 TFlops
Total peak performance	3125 TFlops

Cooperative model: IT owns only a small fraction of compute resources (gen.purpose)

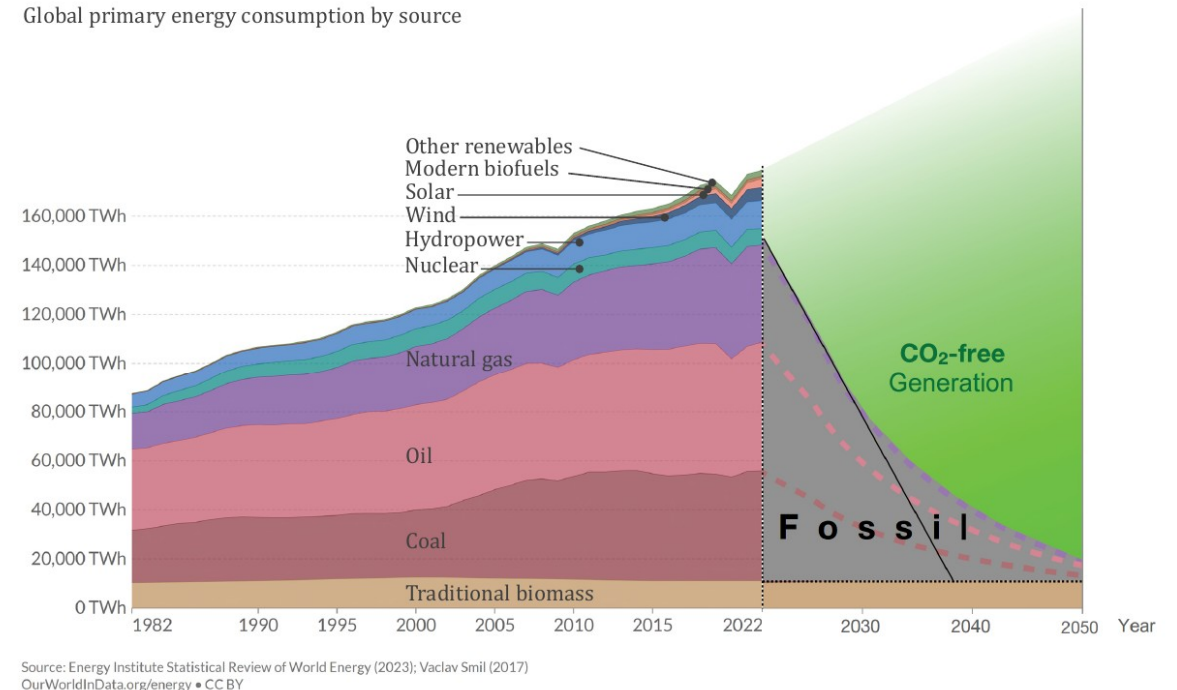
Chapter IV

Sustainable Computing

Sustainable Computing

The Challenge

- To meet the goal of Paris agreement: greenhouse gas emissions must be reduced by 50% in 6 years
- Energy consumption for processing and storing data has a significant impact on CO₂e footprint of computing
- Next to transition to regenerative energies additional savings are required
- Avoid unnecessary computations
- Increase efficiency of calculations
- Key target is to operate data centres with green energy
- Depending on availability data centres must be able to dynamically ramp up/down resources



==> see presentation of Yves Kemp about „Sustainable Computing“ today 11 am

Chapter V

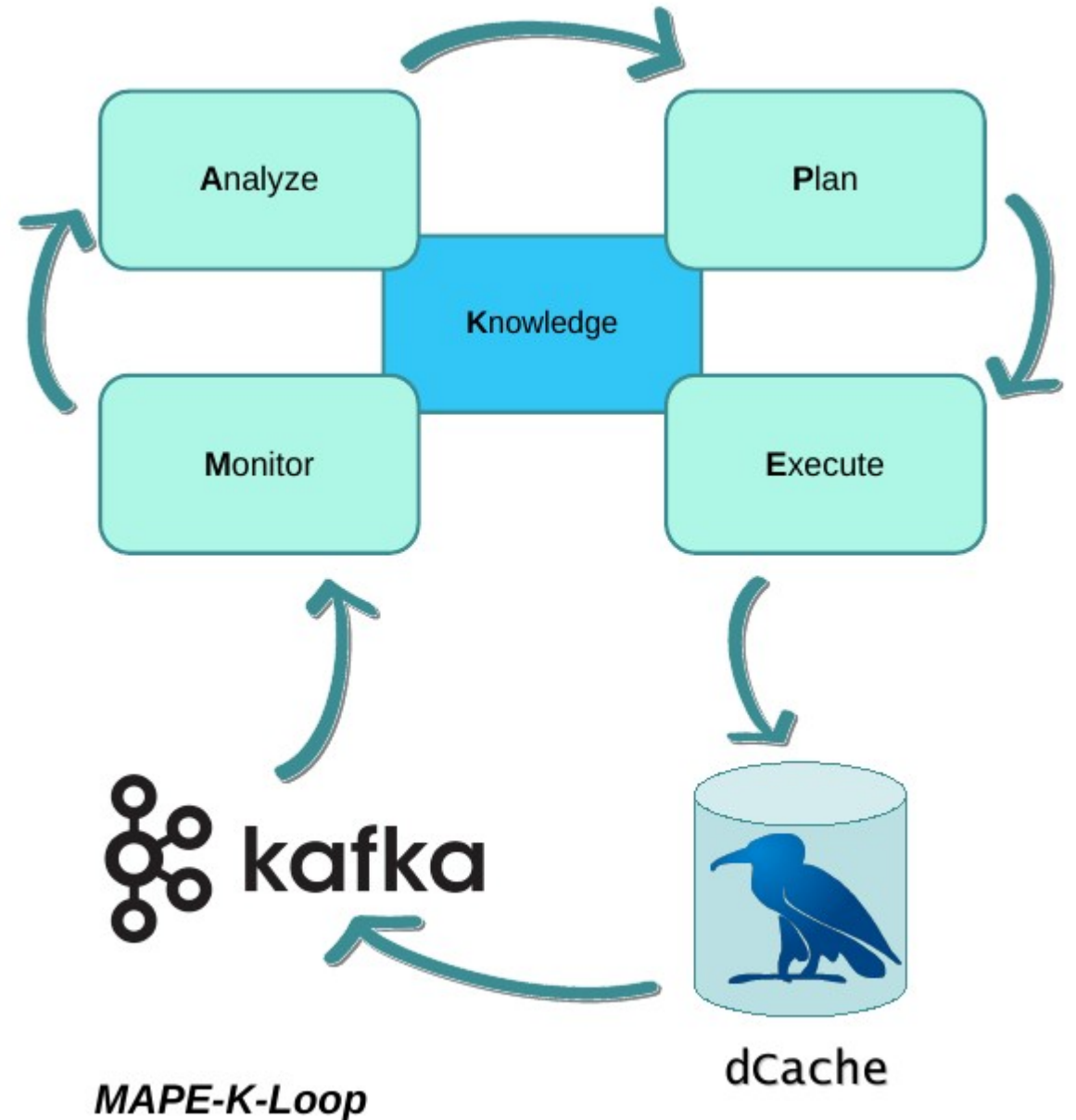
ML and AI

ML and AI

Self Adaptive dCache system

Together with HAW in context of KAI project:

- A prototype of a MAPE loop has been implemented
- This enables the dCache storage system to monitor and analyse anomalies in the system
- Correcting measurements can be planned
- And finally via REST API be executed
- examples are the replication of a data set or the restart of specific services



Chapter VI

3rd party projects

3rd party projects

A few selected examples

- DESY is financed by the Federal Government of Germany and the State of Hamburg
- In addition to the basic budget 3rd party projects can help in financing dedicated projects
- Money can come from various sources
 - EU, DFG, BMBF, Helmholtz, ...



3rd party projects

A few selected examples

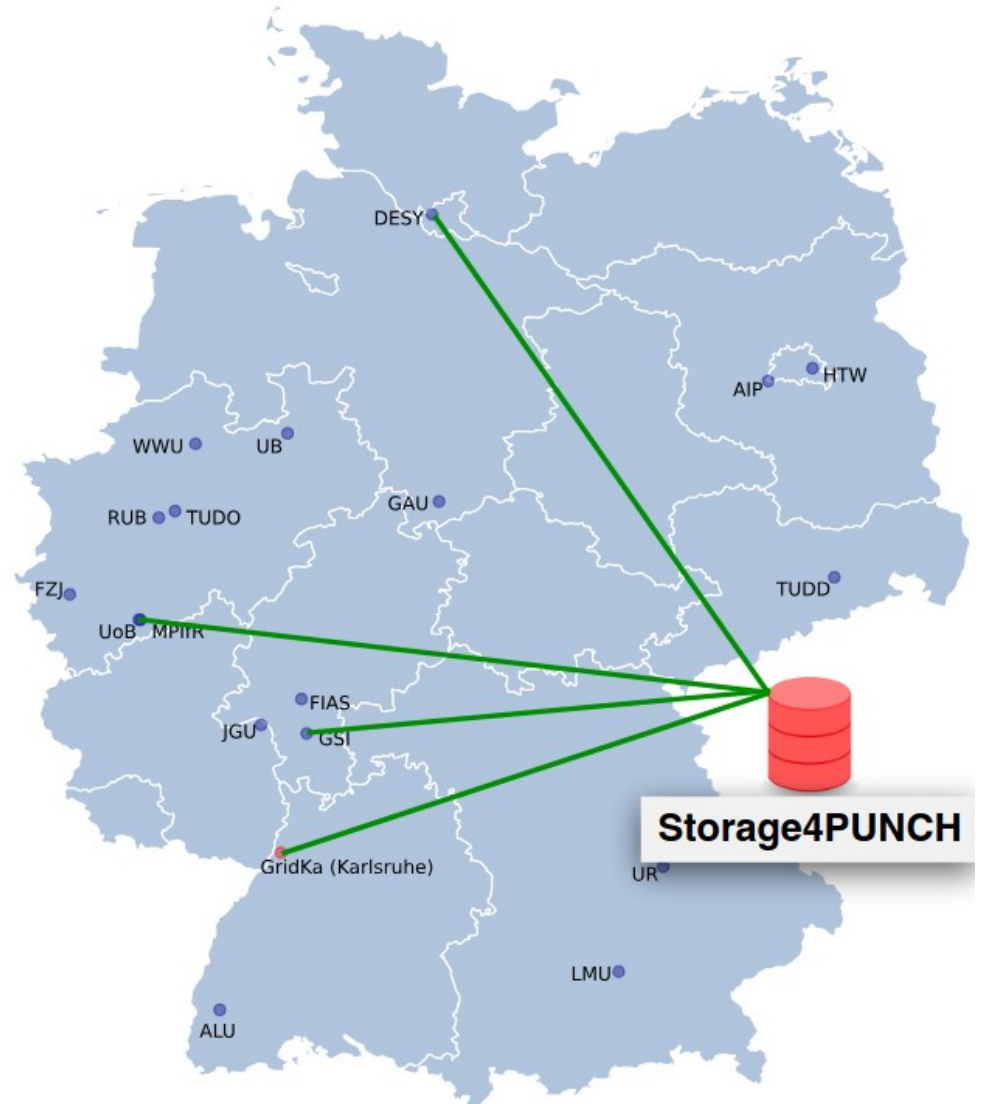
- PUNCH4NFDI, DFG, 5+5 years, 20 funded partner institutes in Germany, research data management for particle, nuclear, astroparticle physics, and astronomy, <https://www.punch4nfdi.de>
- FIDIUM, BMBF, 3+1 years, 10 funded partner institutes in Germany, distributed computing and storage for particle physics, <https://fidium.erumdatahub.de/>
- Hanseatic Science Cloud, EU, 3 years, 20 partner institutes in Germany, Southwest Scandinavia, collaboration of hospitals and life science research with large scale facilities as DESY, XFEL, ... <https://halric.eu/>
- PATOF, Helmholtz Metadata Collaboration, 2 years, 2 partner institutes in Germany, metadata for nuclear Physics and DESY experiments



3rd party projects

PUNCH4NFDI / Storage4PUNCH

- Federated Storage Infrastructure
- Distributed over Germany
- Based upon different technologies
 - dCache (DESY & KIT)
 - XrootD (GSI & University of Bonn)
- Token based authentication using PUNCH AAI



Chapter VII

student projects

Student projects

Educational programs

Deutsches Elektronen-Synchrotron DESY
A Research Centre of the Helmholtz Association

[DESY HOME](#) | [RESEARCH](#) | [NEWS](#) | [ABOUT DESY](#) | [CAREER](#) | [CONTACT](#)



SUMMER STUDENTS | Program 2024

Summer Students Home /

The 2024 program will take place from 16 July - 5 September 2024.

Exciting projects available at DESY IT/Scientific Computing

- Internships
- dual study programm
- bachelor/master theses
- Clearly defined projects in the context of software development, research data management, scientific experiment support, and more
- Summer student programm

Chapter VIII

Summary and Outview

Summary and Outview

Particle Physics

- Grid and NAF, large amounts of data from the experiments and simulations stored in dCache

Accelerator development

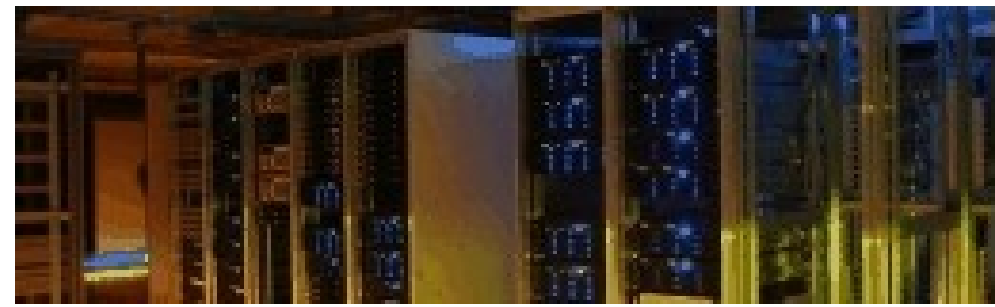
- HPC and local, moderate amounts of data from sensor data and simulations stored local, in dCache, and also on tape

Photon Science

- HPC, IDAF: huge amounts of data from experiments in GPFS, dCache, and on tape

Motivation

- The data need to be recorded, stored, archived, and analysed
- This is one of the main motivations why we do Scientific Computing at DESY IT



Thank you

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

Kilian Schwarz
IT/Scientific Computing
kilian.schwarz@desy.de
+49 40 8998 2596