Introduction to Scientific Computing at DESY

Dealing with Scientific Data Challenges

Kilian Schwarz, Dwayne Spiteri, Jan Hartmann
DESY, July 24, 2025

HELMHOLTZ

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 student projects

- Summer student project
- bachelor/master theses, internships

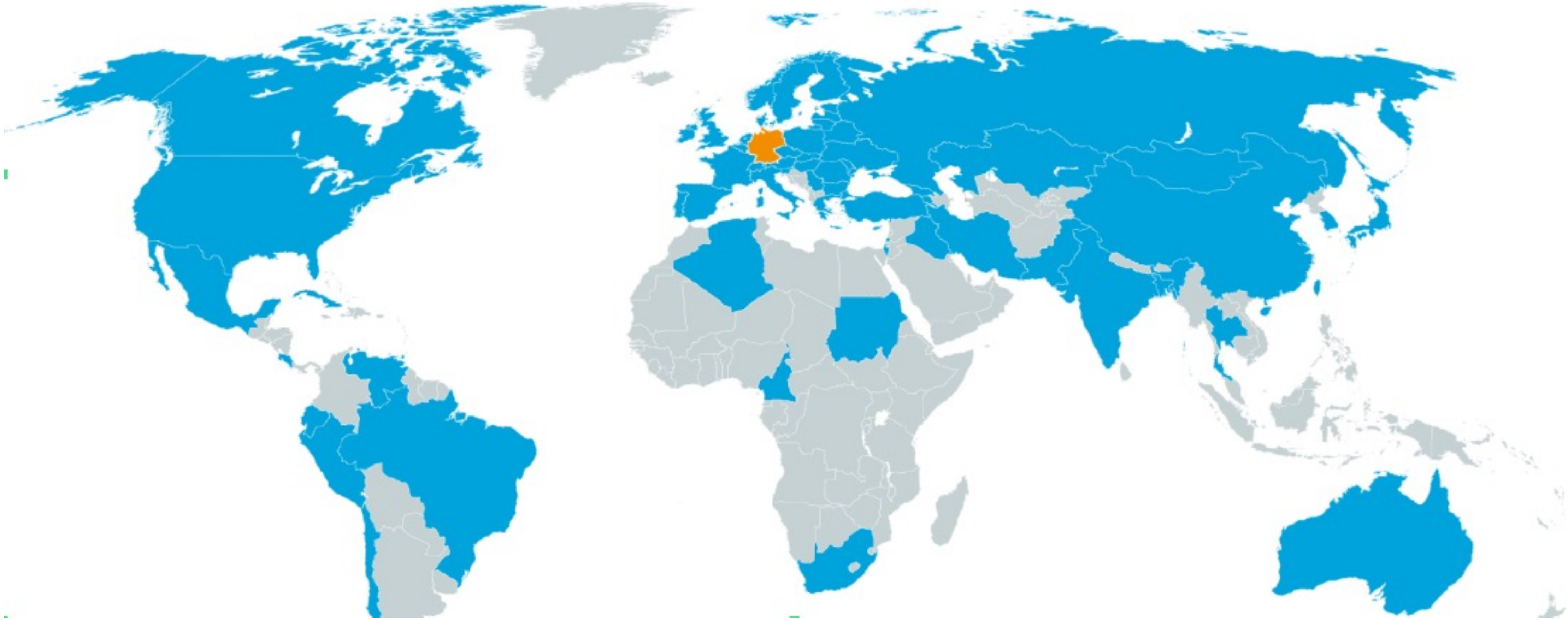
06 Summary and outview

- User communities and motivation

Chapter I
Introduction to DESY

Guest Scientists at DESY

3000 scientists from over 40 countries visit DESY each year

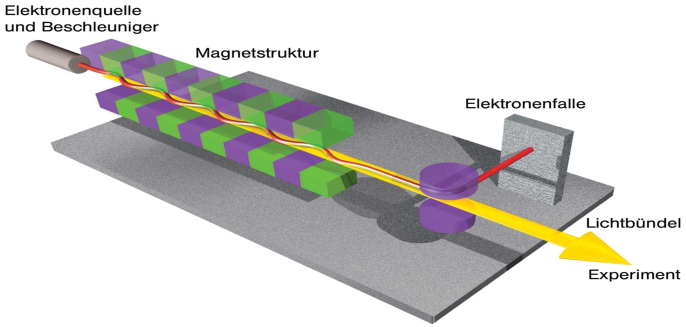


What do we do at DESY ?

Accelerator physics, photon science, particle physics

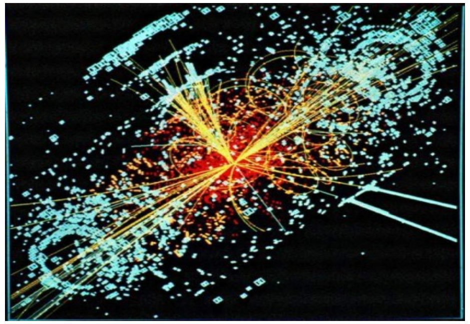
Accelerator Physics

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



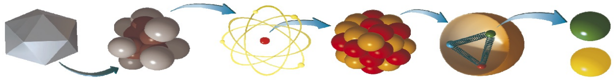
Photon Science

- 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

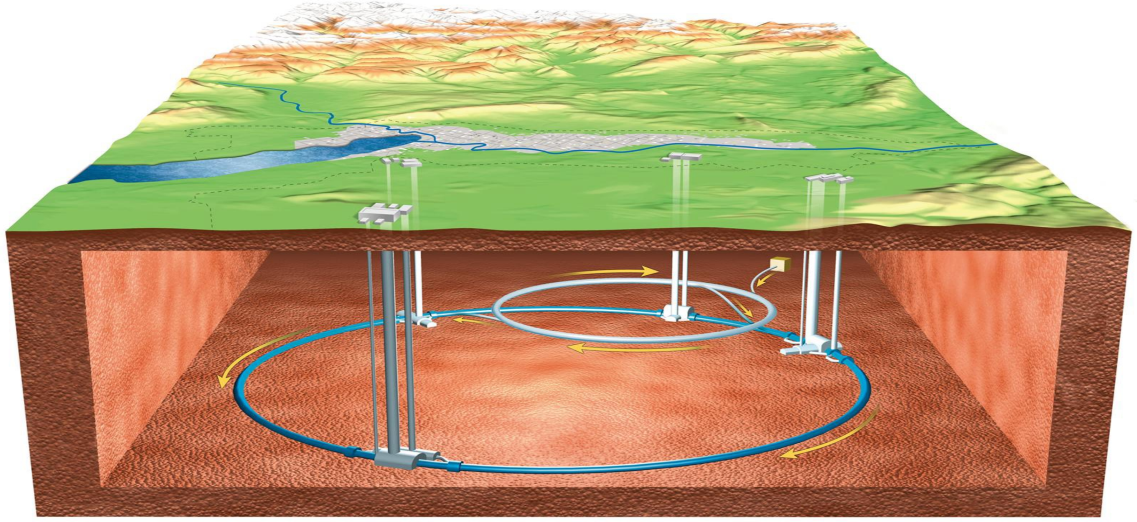
European XFEL (X-Ray Free Electron Laser, 2017)

- 17.5 GeV / 3.4 km / strongest X-ray laser of the world, first experiments 2017
- During peak times up to 5 PB of data on a single weekend

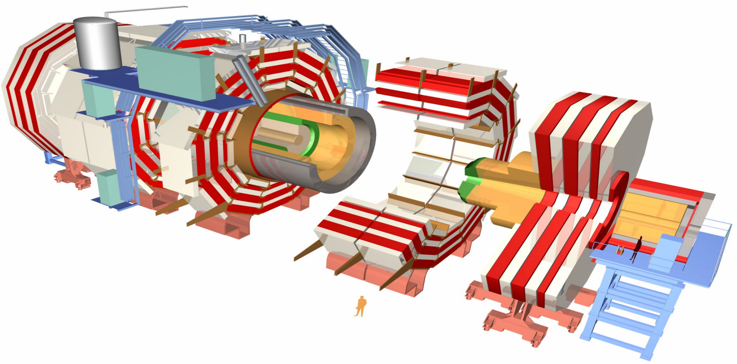
What do we do at DESY ?

What do we do at DESY ?

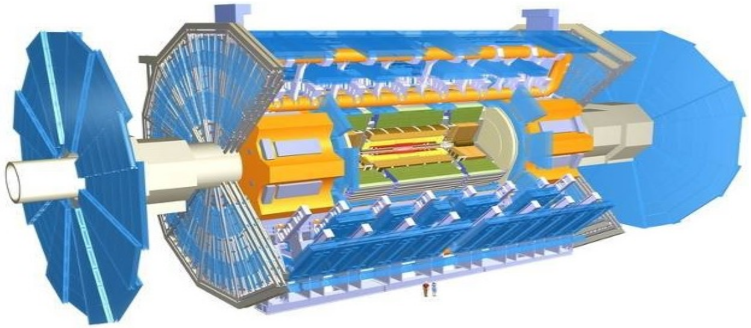
Large Hadron Collider (LHC) at CERN



- 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

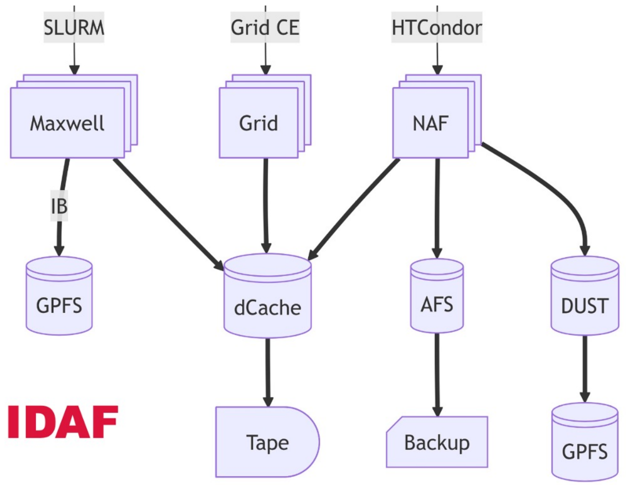


Chapter II
Introduction to IT and SC

Introduction to IT

DESY IT in numbers

- 1000 sqm space
- 1.2 MW power

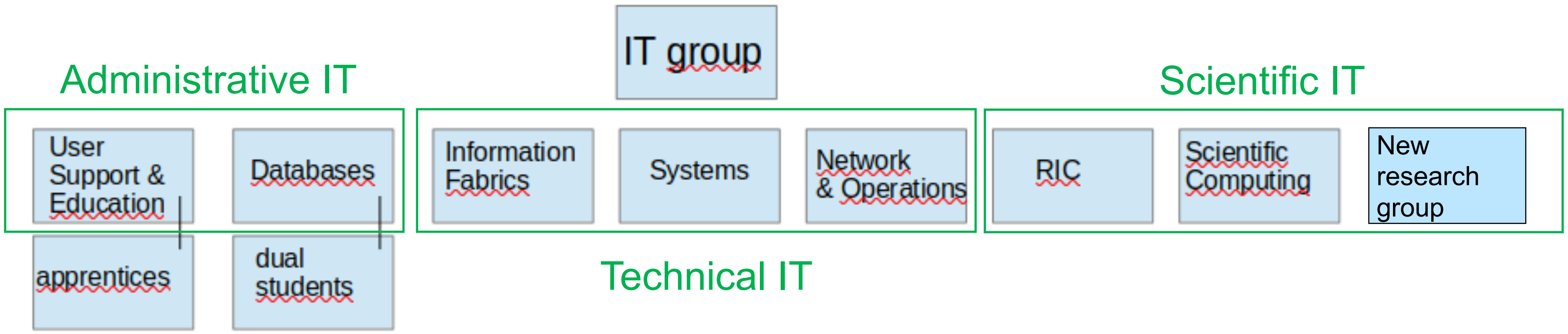


- 300 racks capacity, 2.500 servers (+1.600 virtual systems)
- IDAF: ca. 165.000 CPU cores (including virtual), 520 GPUs
- Storage: 160 PB dCache, 80 PB GPFS, 150 PB tape
- Connectivity: 200 Gb/s Infiniband, links with 2x50 Gb/s to global research networks
- ~ 50.000 IP addresses

- 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



Organisation of IT group:

- 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
- RIC: EU projects and Helmholtz-wide platforms
- Network & Operations: data network and telephone
- 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

ASAP::O

For the first mile after the detector

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

DOOR – DESY Online Office for Research with Photons ⇨ Proposal submission

GAMMA-PORTAL – Data Management for Photon-Science ⇨ Data management

HiDRA ⇨ Data transfer, source for live viewer

ASAP::O ⇨ Stream processing - data transfer, source for live viewer, online/offline data analysis

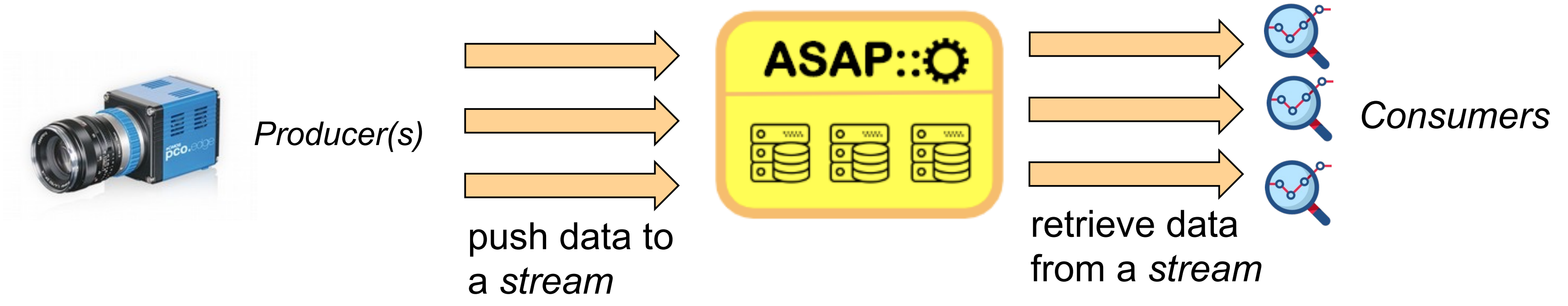
dCache ⇨ Data archiving

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

dCache.org

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

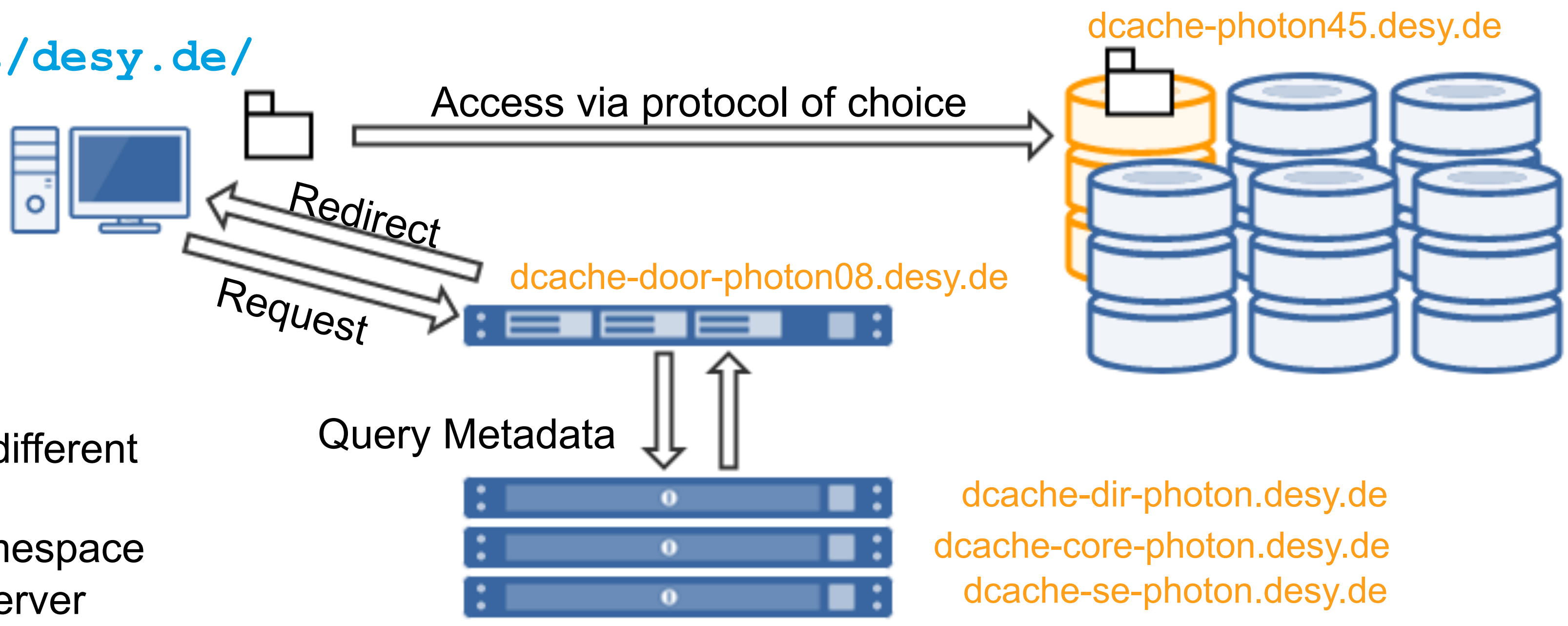
- FLASH, LINAC

- 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/`

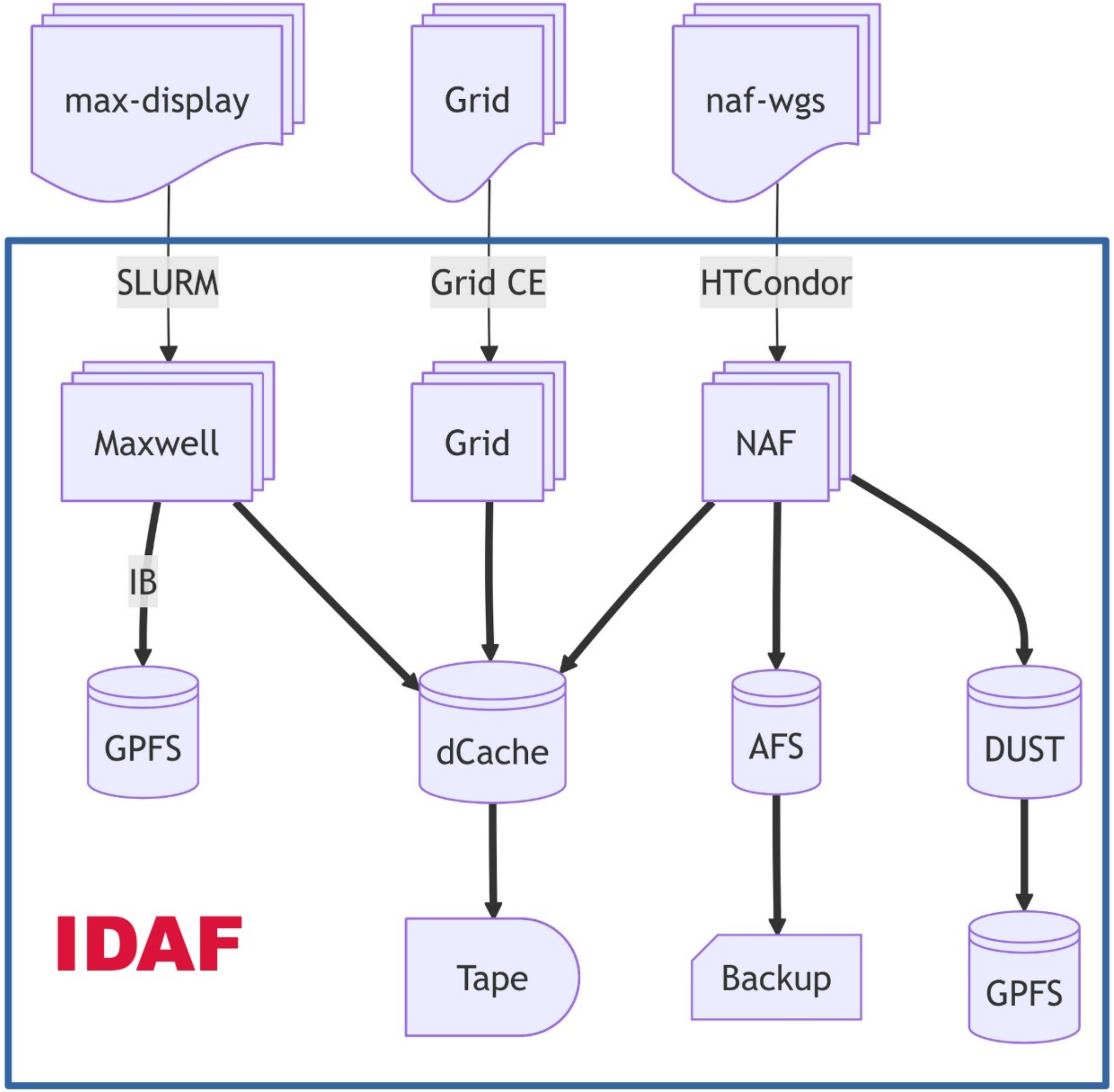


- Based on Micro-Services
  - `Doors` – supporting each a different protocol
  - `Heads` – pool selection, Namespace
  - `Pools` – data storage and server

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

Storage Systems at DESY

- GPFS
  - HPC filesystem, interconnected with low-latency network to Maxwell cluster.
- dCache
  - Large data store. Multi-protocol, multi-authentication support. Direct connection with tape library.
- AFS
  - Home directory on WGS and NAF nodes. Nightly incremental backups.
- DUST
  - Limited scratch space. Re-export of GPFS over NFS.



What I Should Use?

- GPFS
  - HPC workloads
    - Many processes accessing same or a small number of files for read or write.
  - Latency sensible analysis

- AFS
  - Point access
    - Startup scripts
    - Jobs configurations

- dCache
  - Large volumes of immutable data.
  - HTC workloads
  - Data import/export with other sites
  - Multiple access protocols
  - Tape integration

- DUST
  - Small reproducible data sets
  - Job outputs
  - Concurrent writers
  - Local container images

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

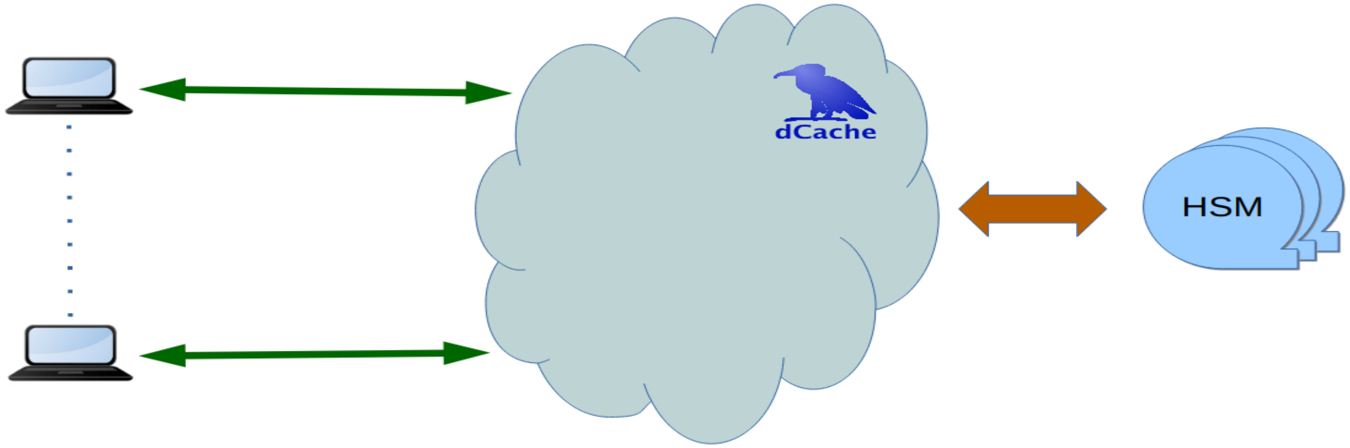
- High Reliability
- Self-healing
- Automatic technology migration
- Persistent identifier

- Scientific data need to preserved for a long time, at least 10 years
- Many scientific data are still analysed for publications long time after the experiment already stopped
- FAIR and Open Data allow discipline overarching data analysis and cross checking
- FAIR data is a requirement from EU and German government
- FAIR: Findable, Accessible, Interoperable, Reusable
- These days data can be published alongside with the publication using DOIs …

Scientific Data Challenges

Long term archive

dCache+HSM Tandem



- **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

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

Scientific Data Challenges

In collaboration with CERN

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

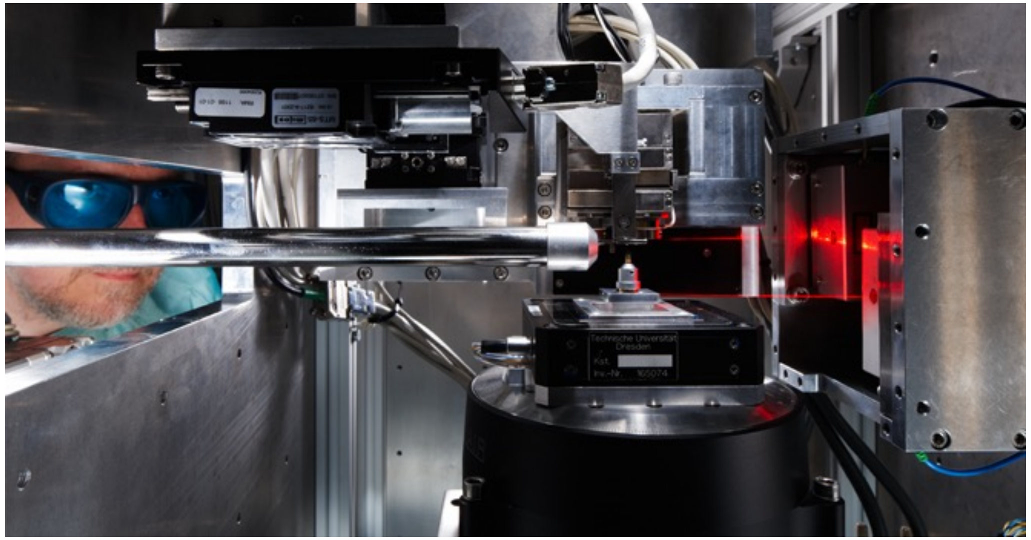
Analysis

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

Scientific Data Challenges

Data processing – user community

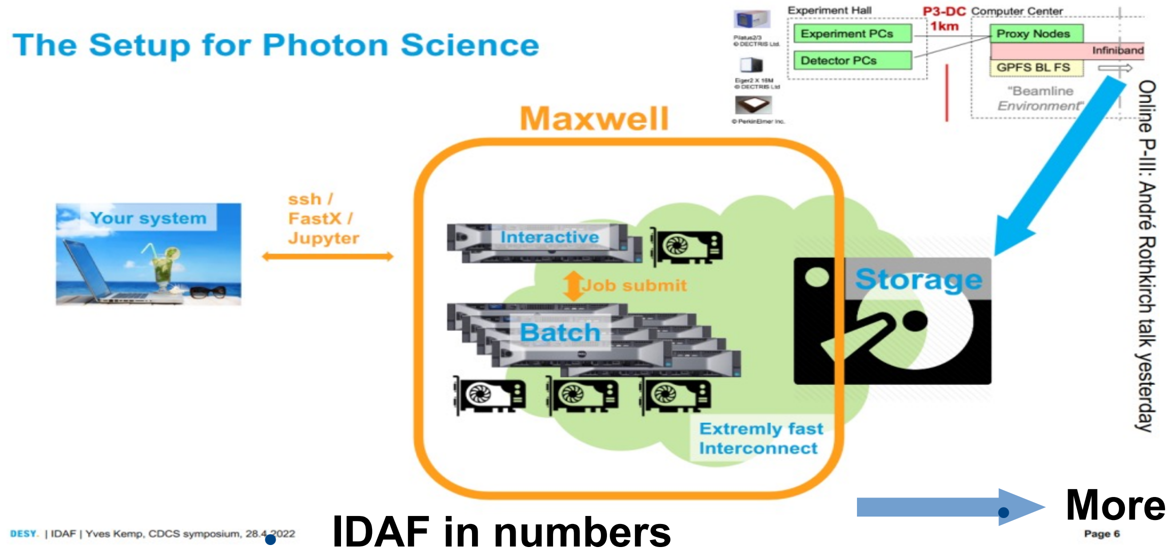
- **Particle physics**
  - ATLAS (CERN)
  - Belle II (KEK)
  - CMS (CERN)
  - ILC
  - LHCb (CERN)
  - ALPs II, BabyIAXO, MADMAX, … (DESY)
- **Photon Science**
  - XFEL (DESY)
  - FLASH (DESY)
  - PETRA III (DESY)

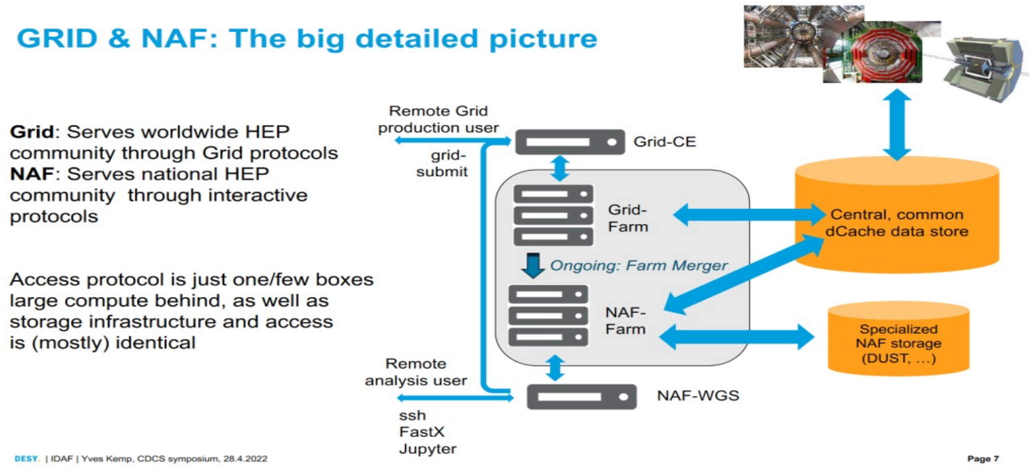


Scientific Data Challenges

Cluster infrastructure
provided by IT/Systems

Data processing – IDAF (Interdisciplinary Data Analysis Facility)





→ **More about Maxwell HPC cluster in next slides**

- **IDAF in numbers**
  - Compute: Maxwell + Grid + NAF
  - dCache + GPFS
  - 165.000 CPU cores + 520 GPUs
  - 240 PB data on disk*
  - 2.500 servers (compute, storage, management)
  - 1.2 Megawatt power consumption

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

*: 1 PB corresponds to a 25 Mb/s video stream running for 10 years !!!

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.

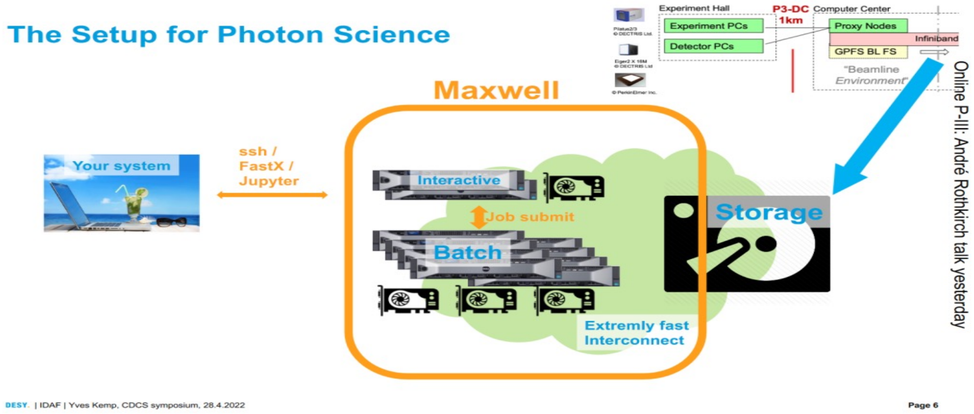
- DESY is a site in the WLCG and in Belle II Grid

Grid computing is a form of distributed computing that uses geographically dispersed, networked computers to tackle large-scale computational problems. It essentially creates a virtual supercomputer by linking together idle resources from multiple machines to process complex tasks that would be difficult or impossible for a single computer to handle. Grid is an HTC infrastructure

Scientific Data Challenges

Data processing – Maxwell HPC Cluster

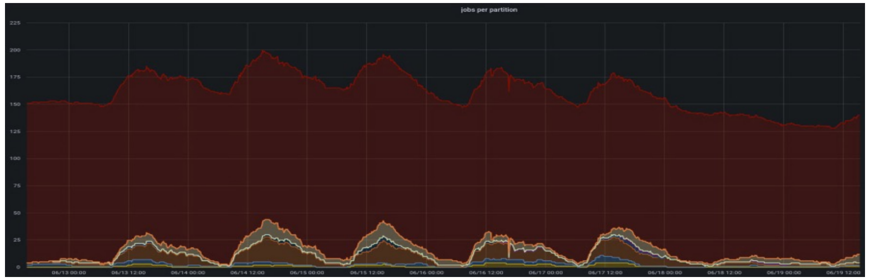
- HPC characteristics:
  - low latency network (IB)
  - fast cluster storage (GPFS)
  - mass storage (dCache)
  - substantial GPU resources
  - no afs, krb support
  - targeted at massively parallel and GPU computations

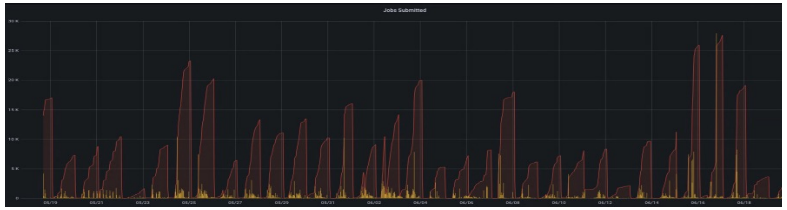


Scientific Data Challenges

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





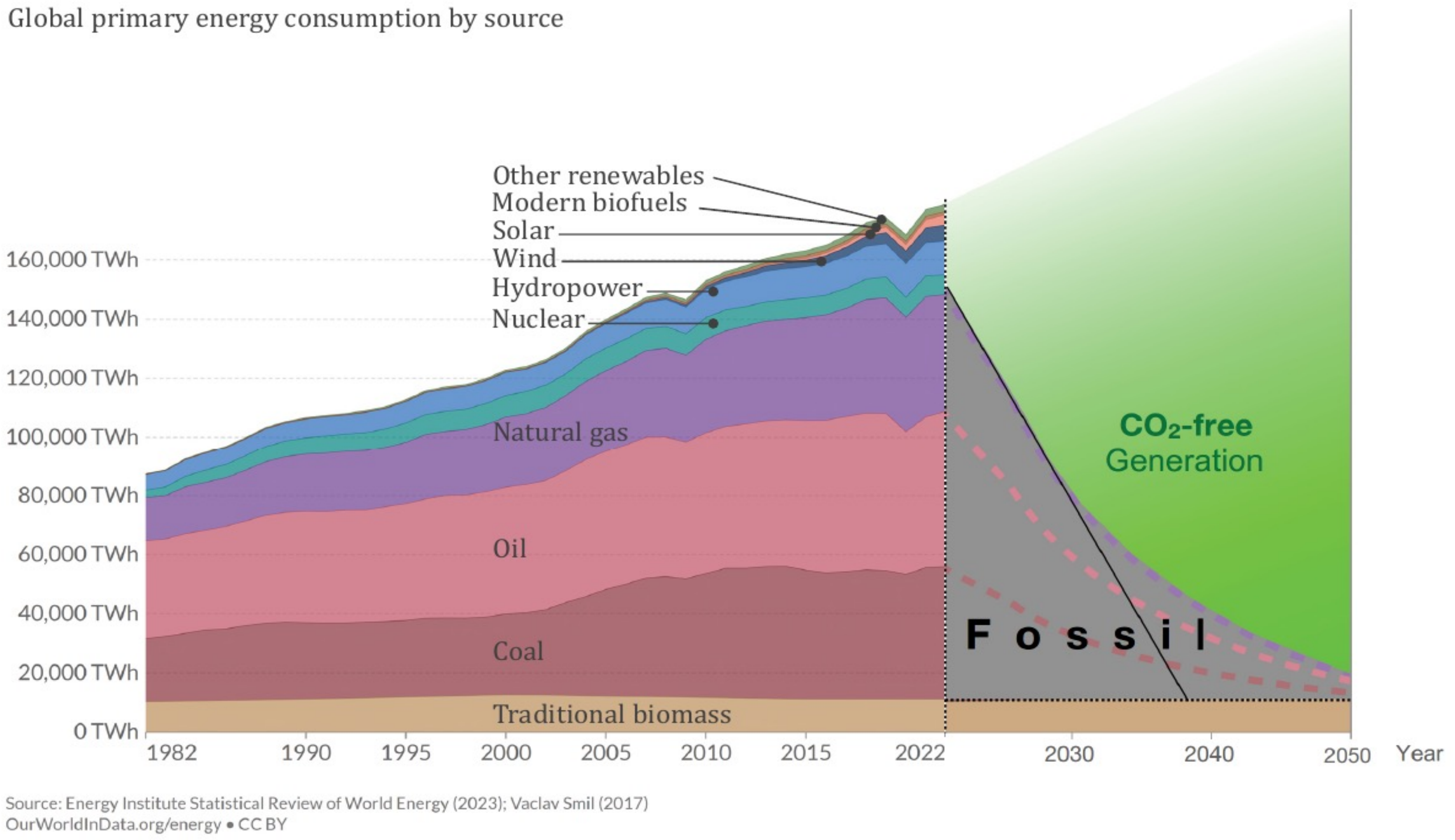
Run as buy in model:
- IT provides infrastructure and services
- storage to a large extent group financed, shared GPFS HOME

Chapter IV
Sustainable Computing

Sustainable Computing

The Challenge

- Paris agreement goal: greenhouse gas emissions must be reduced by **50% in the next 5 years!**
- Energy consumption for processing and storing data has a significant impact on computing CO2e footprint
- Transition to renewable energy operation but additional savings are also required



- Data centres should be able to dynamically ramp up/down resources
- Avoid unnecessary computations
- Increase efficiency of calculations

→ See presentation of D. Spiteri, J. Hartmann tomorrow

Chapter VII
student projects

Student projects

Educational programs

And work at DESY IT

- **Exciting projects available at DESY IT/Scientific Computing**
- Internships
- dual study programms
- bachelor/master theses
- Clearly defined projects in the context of software development, research data management, scientific experiment support, and more
- Summer student programm
- **With an IT education you are welcome to work at DESY IT !!!**

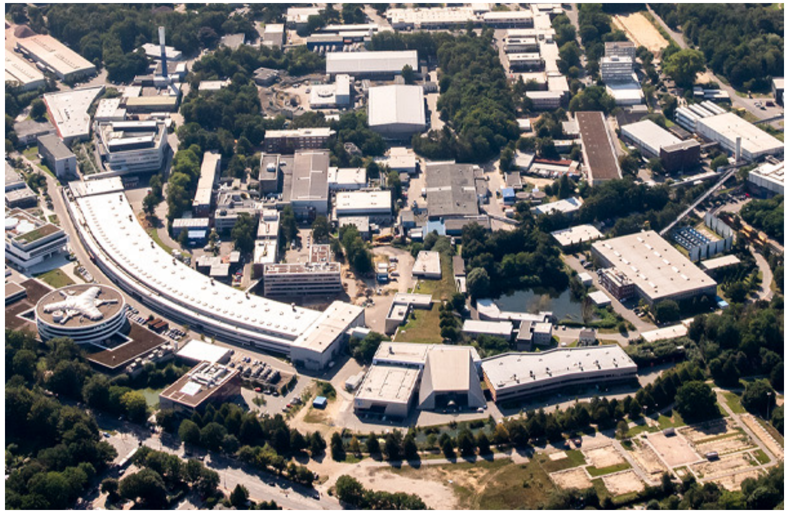
Chapter VIII
Summary and Outview

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

scientific computing infrastructure is grouped into storage, compute, networking and services

Thank you

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

Kilian Schwarz

IT/Scientific Computing

kilian.schwarz@desy.de

+49 40 8998 2596