Scientific Computing and Data Management at KIT

Dr. Andreas Heiss, andreas.heiss@kit.edu

Steinbuch Centre for Computing

Steinbuch Centre for Computing (SCC)

Which demands do we satisfy?

- Computational Science & Engineering (CSE)
- Data-Intensive Science (DIS)
- For users in KIT, BaWü, Germany and international

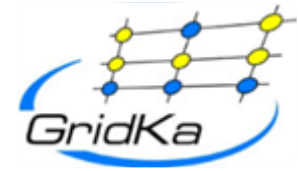**Who are we?**

Institute in KIT with service tasks

What do we do?

- Research, education and innovation in Supercomputing, Big Data and secure IT-federations
- Operation of large scale research facilities
- Operation of basic IT services

Steinbuch Centre for Computing

Data-Intensive Science (DIS) and HPC

- Operation of **Data Facilities**
  - GridKa as the German Tier-1 for LHC data
    - 15 PB disk netto, 18 PB tape used, 11 k cores
    - Starting point of data-intensive science at KIT (and even HGF)
  - Multi-disciplinary Large-Scale Data Facility
    - > 6 PB disk for climate, energy, systems biology, etc.
- Operation of **HPC systems**
  - ForHLR closely linked to data facilities
    - > 35 k cores, 1.4 PetaFlop/s peak, energy-efficient warmwater-cooling
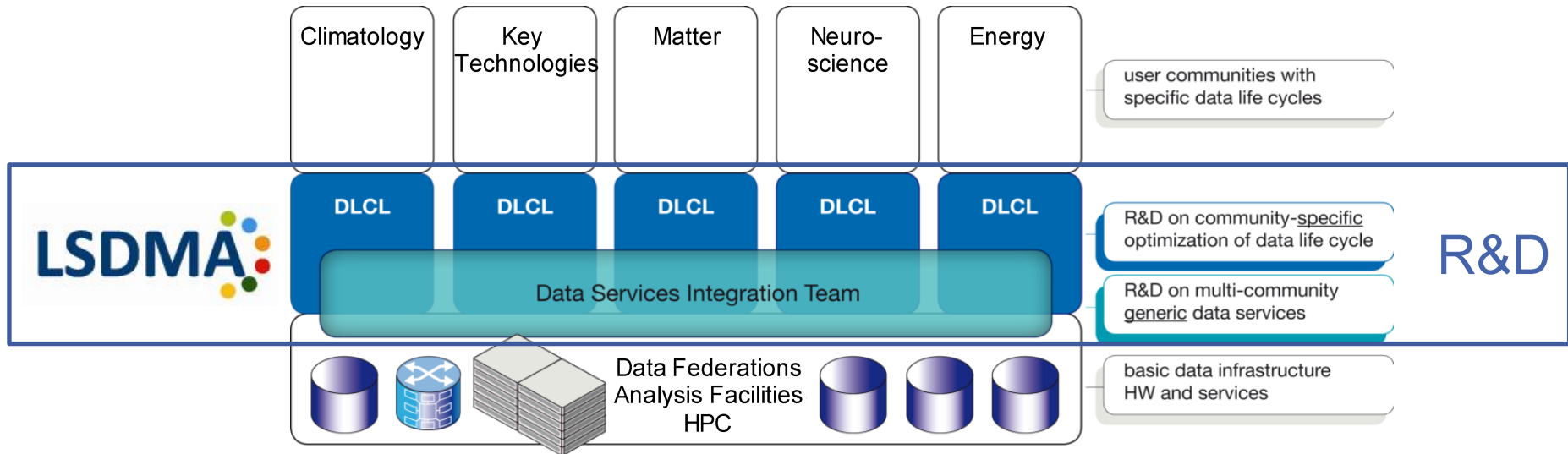- **Joint R&D** with scientific communities
  - Data Life Cycle Labs (DLCLs) and SimLabs
  - Generic data management and analysis tools
  - Generic computer science methods & algorithms
- **Innovation driver** for SMEs and big industry

Steinbuch Centre for Computing

Large Scale Data Management and Analysis
Helmholtz Portfolio project and cross-program activity, lead by KIT

Climatology | Key Technologies | Matter | Neuro-science | Energy

user communities with specific data life cycles

LSDMA

DLCL · DLCL · DLCL · DLCL · DLCL

Data Services Integration Team

Data Federations
Analysis Facilities
HPC

R&D on community-_specific_ optimization of data life cycle

R&D on multi-community _generic_ data services

basic data infrastructure HW and services

R&D

Data Life Cycle Labs
Joint R&D with communities

- Optimizing the data life cycle
- Specific data analysis tools and services

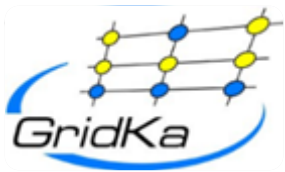## Data Services Integration Team
Generic, multi-community R&D

- Interface between federated data infra-structures and DLCLs resp. Communities
- Integration of data services in scientific working process
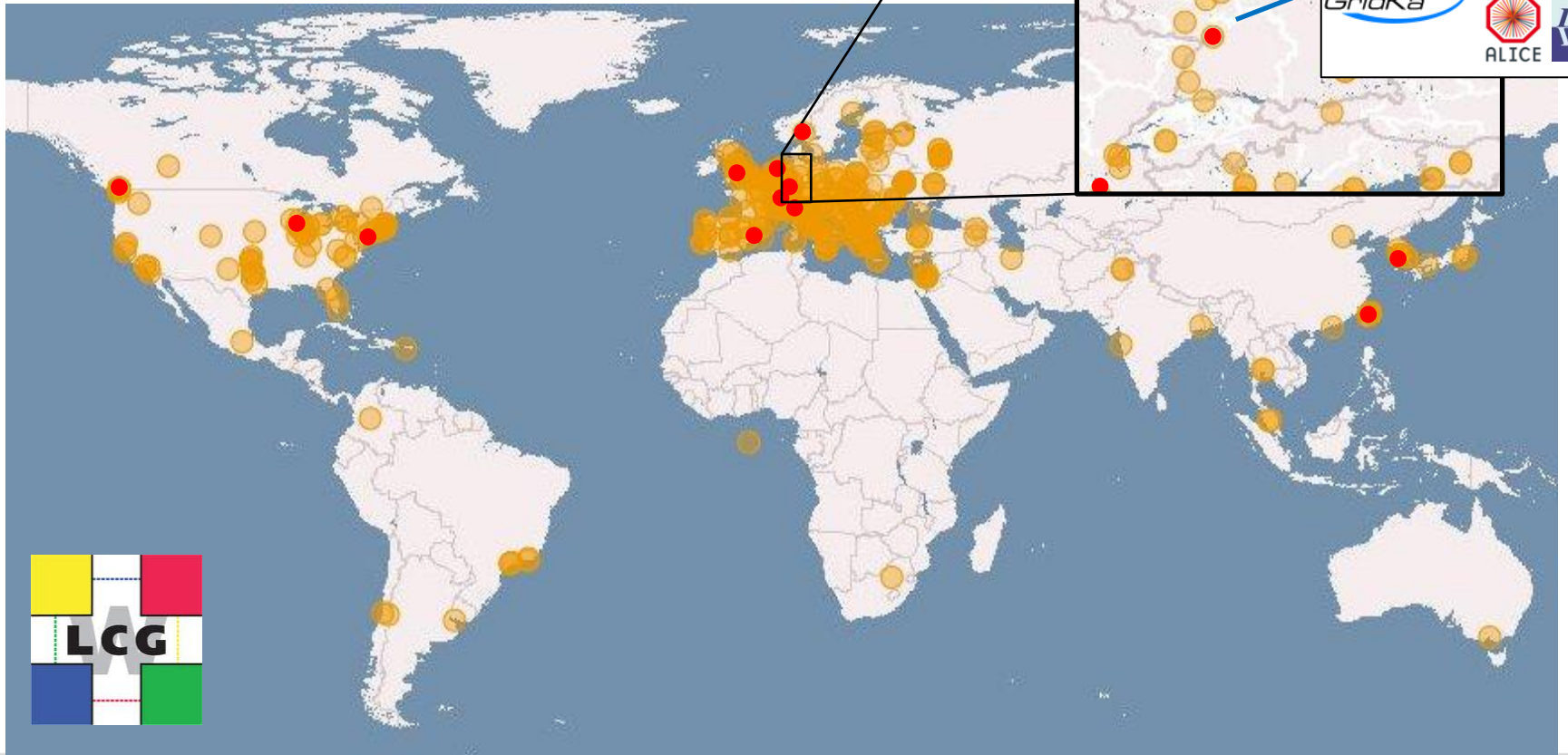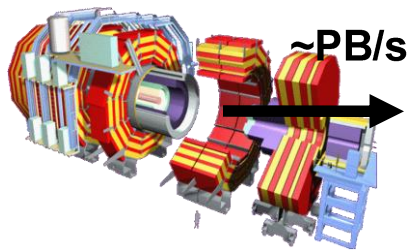
SCC Steinbuch Centre for Computing

Collaborations, Projects, Initiatives

Tier-1 centre of the Worldwide LHC Computing Grid (WLCG)

- 12 Tier-1 and ~140 Tier-2 sites
- ~ 10,000 users world-wide

Steinbuch Centre for Computing

~PB/s

Online System

200 - 500
MB/s
per Exp.

**Offline farm
CERN computer centre**

Tier-0

raw, reconstructed data
> 20 PB/year (GridKa)
~15% of total raw LHC data

Tier-1 (12 sites)
- **Archival of raw data**
- **Reconstruction and reprocessing**
- **(MC and analysis)**

Tier-1   **Tier-1**   **GridKa**   • • •

reconstructed data
MC

Tier-2   **Tier-2**   **Tier-2**   • • •

Tier-3

Institut   Institut   Institut   Institut
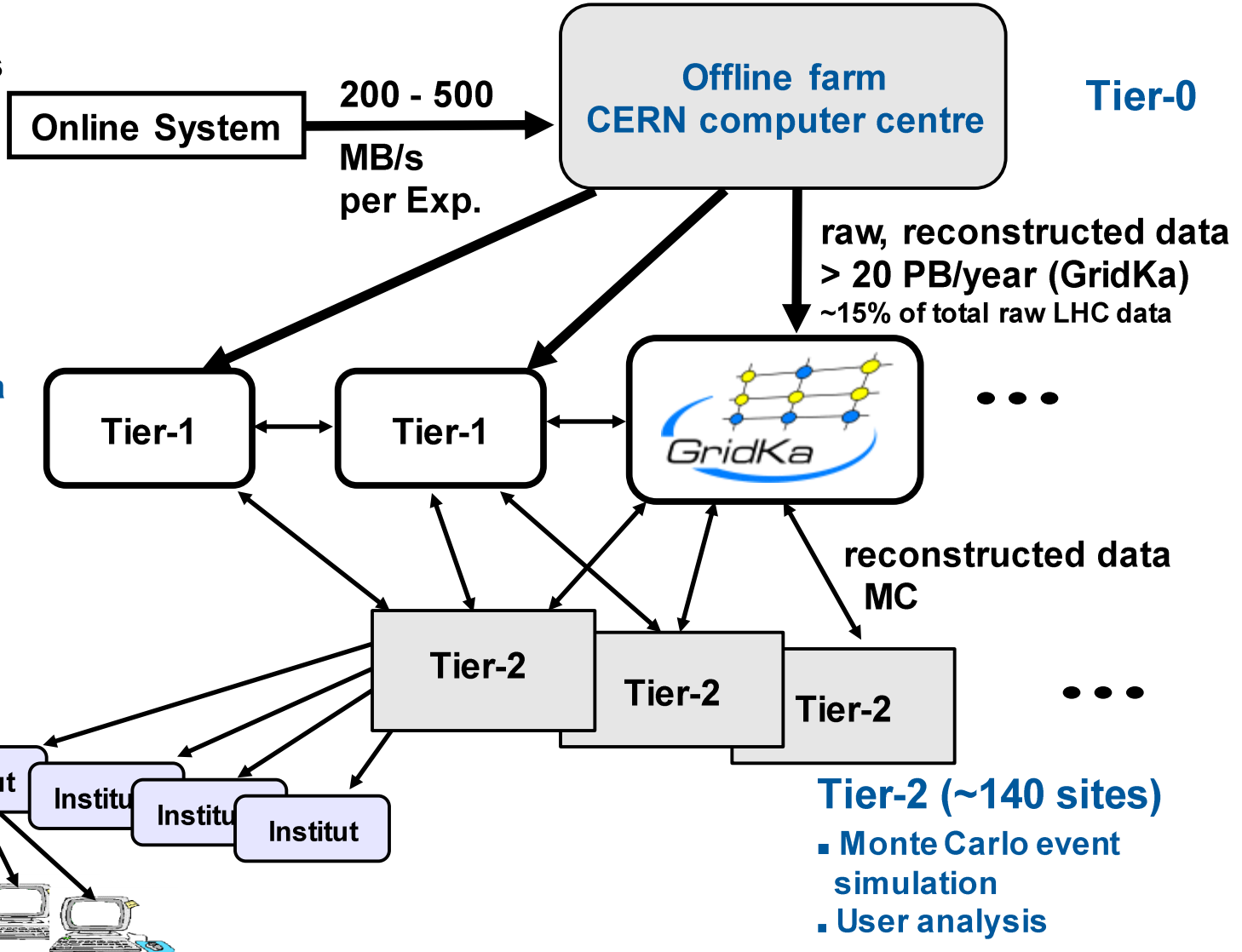
Tier-2 (~140 sites)
- **Monte Carlo event simulation**
- **User analysis**

Steinbuch Centre for Computing

Facts and figures

- Started in 2003.
    - >10 years of experience in large scale data management.
- Helmholtz LK-II facility
- Experiments (collaborations) supported:
    - One of four Tier-1 centres supporting all 4 LHC experiments
    - Belle-II and other non-LHC HEP experiments
    - Pierre Auger experiment
- Custodial responsibility for LHC raw data

- On-site experiment representatives for LHC experiments
- 24/7 service (on-call service at night and on weekends)
- Reliability remotely monitored

- Technical Advisory Board and Overview Board with representatives of user communities and funding bodies (OB)

Facts and figures

- Optimized for high-throughput computing
  - Excellent wide area network connectivity (2x10 Gb/s to CERN, 100 Gb/s to DFN)

  - 610 compute nodes (11000 cores)
    - 20 Mio. Compute jobs per year (av. 54500/day), 106 Mio. CPU hours

  - ~ 250 servers
    - 160 file servers
    - Login-nodes, Grid-services, databases, monitoring, …. > 80 servers

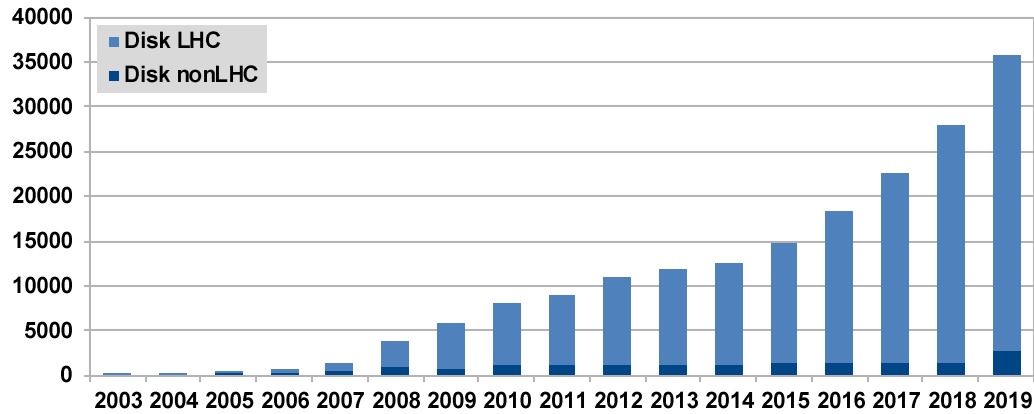 - 4 dCache and 2 xrootd data management systems

Steinbuch Centre for Computing

Facts and figures
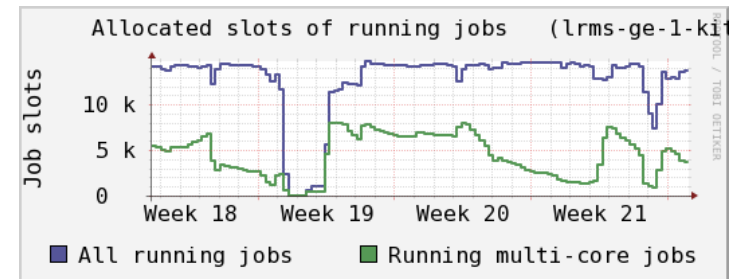
- 2016 total compute power: 179000 HepSpec'06
  - 130000 for LHC
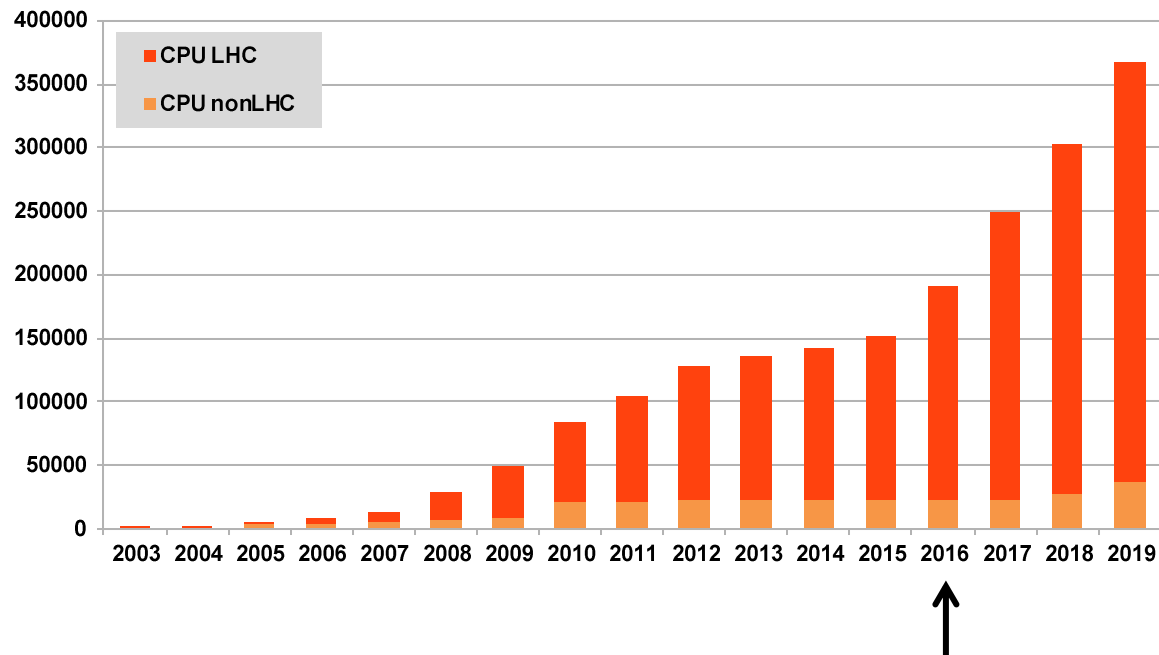
(expected) increase of compute power

Steinbuch Centre for Computing

Facts and figures

- 2015: 15 Petabytes online storage , 18 PB on tape

(expected) increase of storage capacity

Steinbuch Centre for Computing

Facts and figures
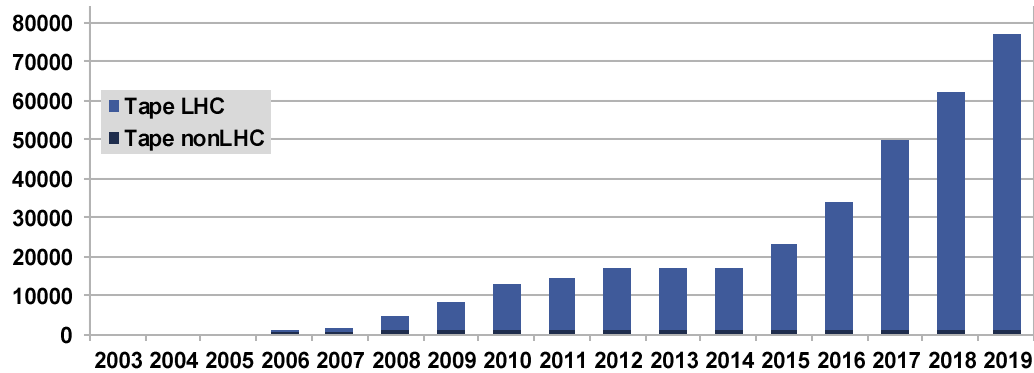
■ Data rate

Transfer Volume
2015-01-01 00:00 to 2016-02-01 00:00 UTC
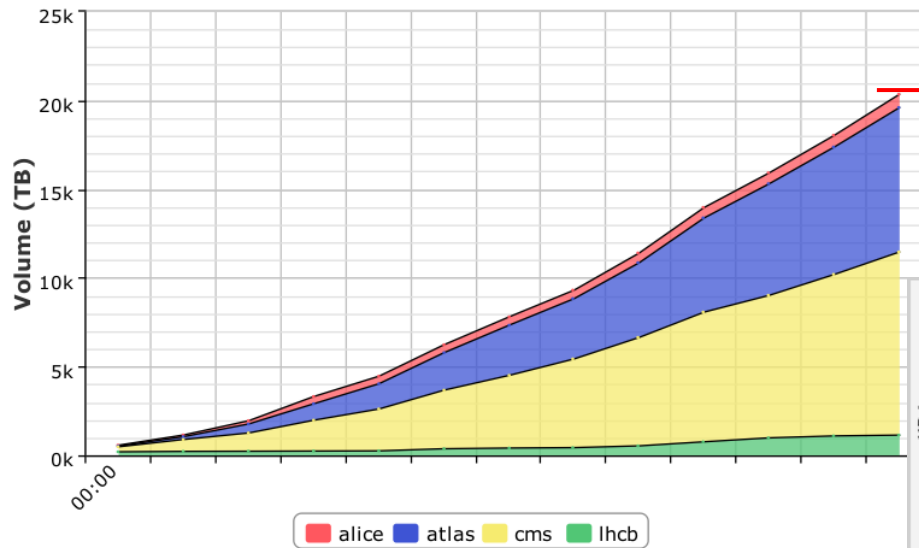
20 PB in 2015
(av. of 55 TB/day)

write transfers for all VOs last week

read transfers for all VOs last week

Max. of ~10 GB/s daily av.
(~ 860 TB/day)
- Transfer to other Grid centres or processed in GridKa compute farm.

Steinbuch Centre for Computing
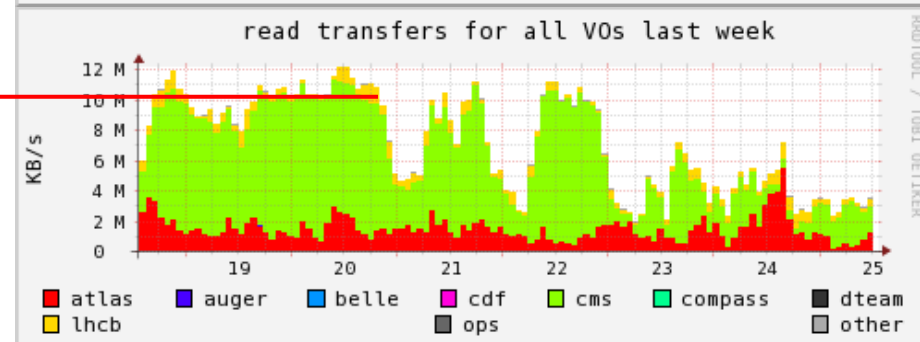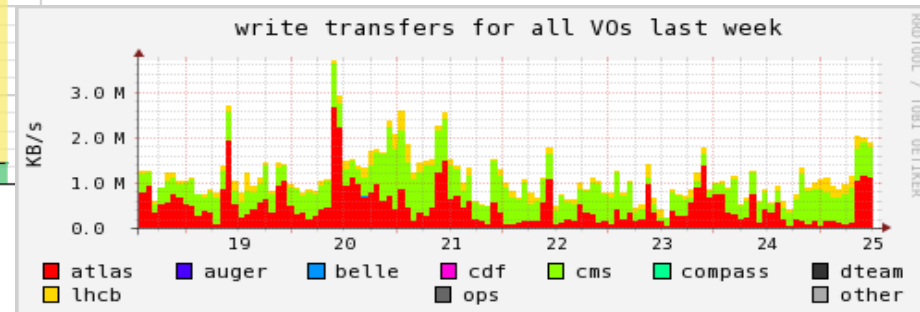
KIT (SCC) has many years of experience in large scale scientific data management and data intensive computing.

It was the starting point of data-intensive science in the Helmholtz-Association

Thank you!

Steinbuch Centre for Computing