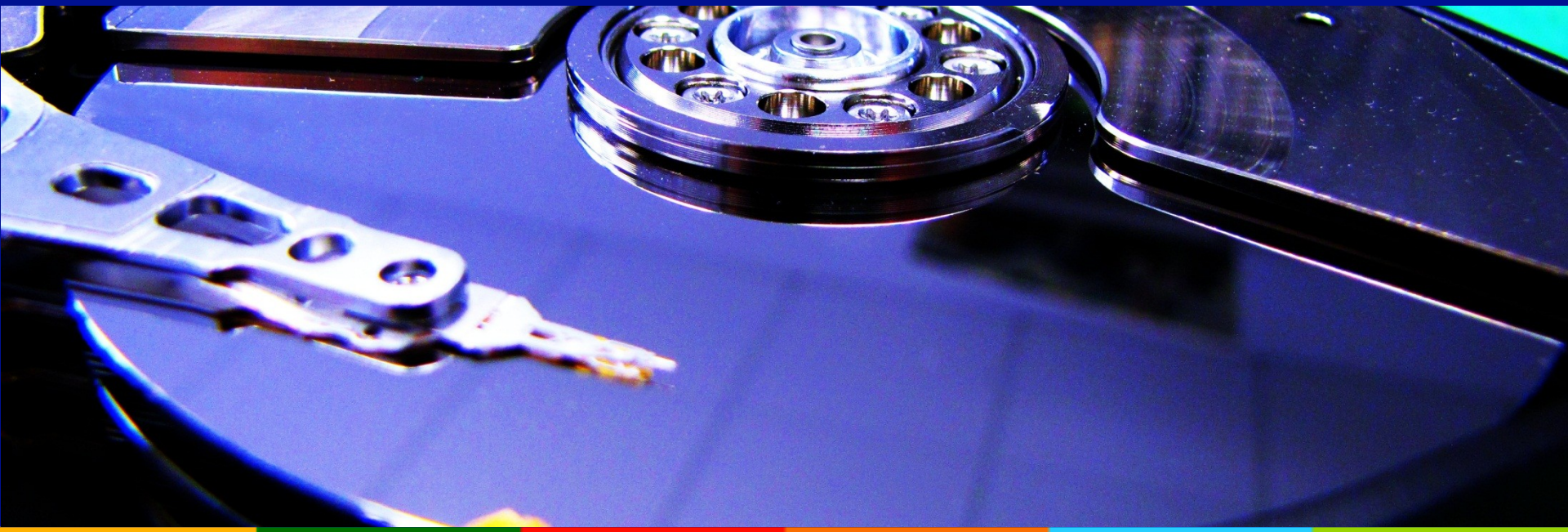
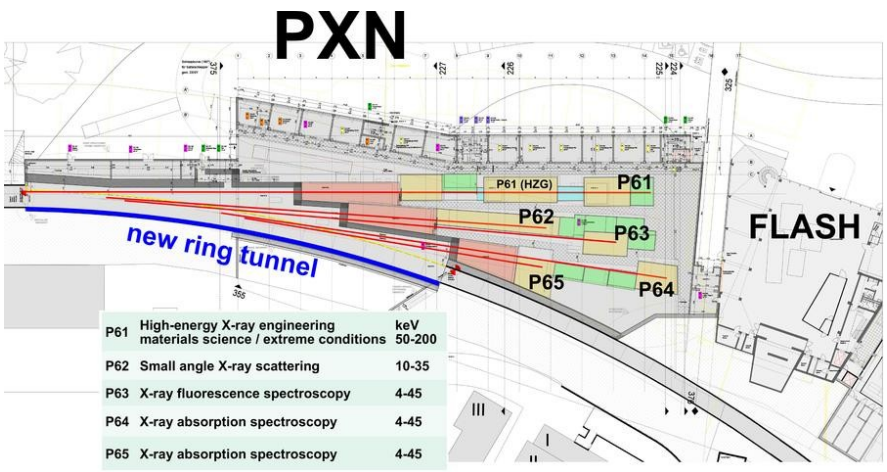
DLCL Matter

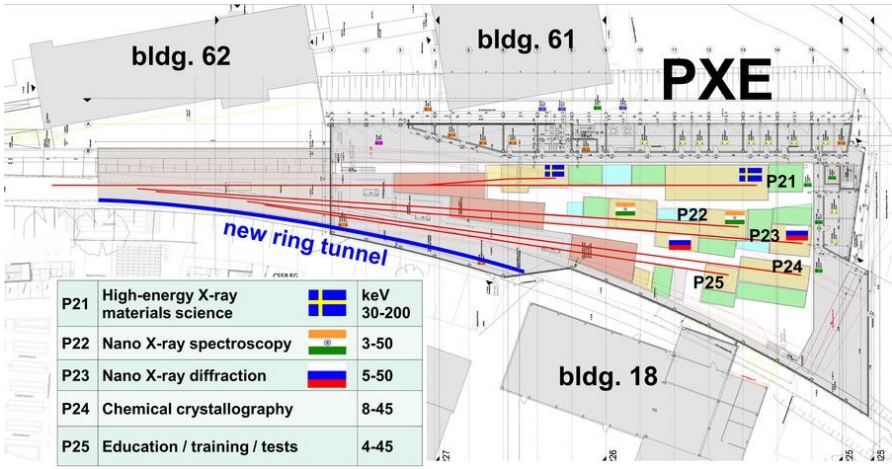
LSDMA All-Hands Meeting, October 2015



LSDMA

The Synchrotron Facilities At DESY: Petra III and PetraExt



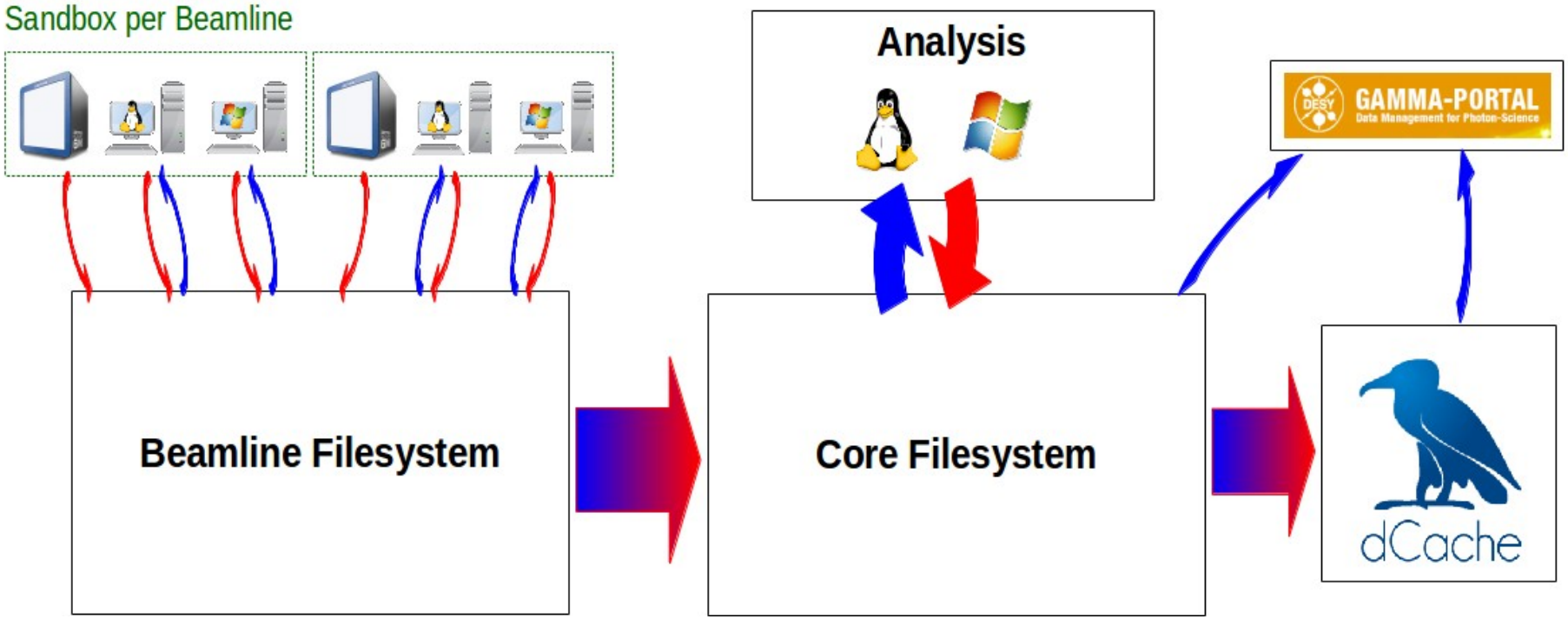


- Petra III 14 beamlines
  - 3 EMBL, 1.5 HZG
- 10 beamlines
  - Commissioning starts 2nd half 2015

Starting situation

- New challenges
  - New detectors achieve much higher data rates (15 times higher)
  - Old system hit limits
  - New storage system had to be installed during Petra III shutdown
- Requirements for the new storage system
  - High performance for single clients > 1GB/s
  - Handling of data peaks (burst buffer)
  - Privacy issues due to legal constraints of the data between beamlines and user groups
- Additional limitations
  - Distance to the data center ~ 1 km
  - Low space in the experiment hall and at beamlines
  - Only 10 Gigabit Ethernet available
  - Zoo of different operating systems and versions
  - Shared accounts for data-acquisition per beamline

Logical Dataflow – From the cradle to the grave

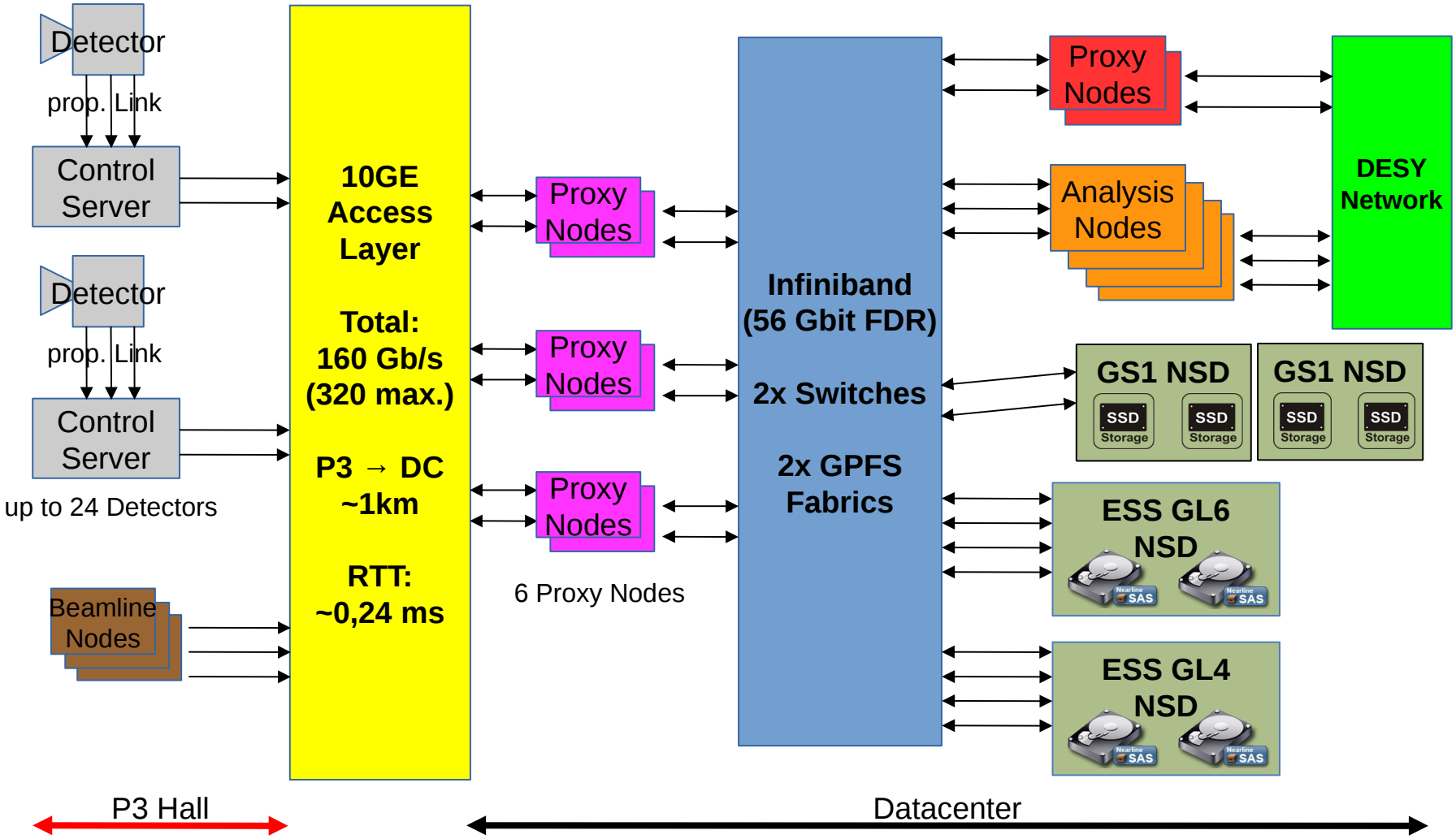


- Low latency
- Low capacity
- Host-based authentication

- 4 min latency
- High capacity
- Full user authentication

- ~20 min latency
- Very high capacity, tape
- Full user authentication

New Architecture

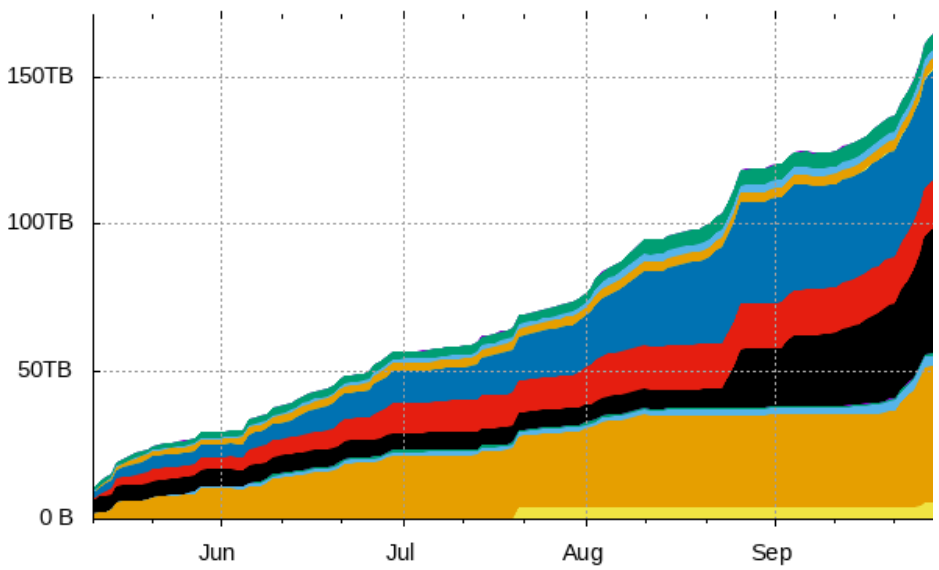


Current Status

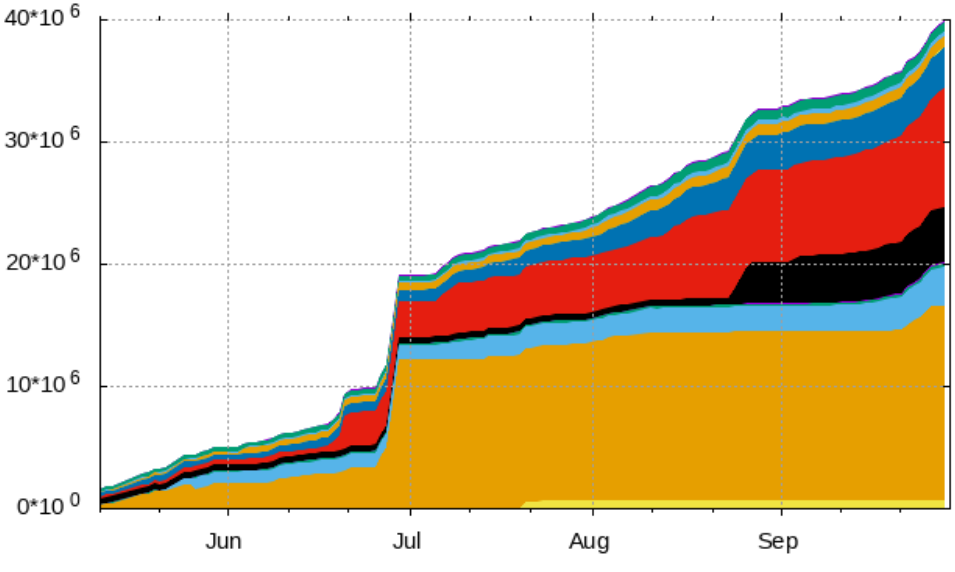
- System in production since April 2015
- No incidence since July
- Significant data rate improvement after optimization:
  - Windows (with SMB):
    - 20 MB files: ~ 19 Hz, 380 MB/s
    - 13 GB file: saturated 10 GE
  - Windows (with ZeroMQ):
    - Saturated the 10 GE from a file size of 5 MB and more
  - Linux: no improvement needed
- A (very) similar system is built for the European XFEL - however it needs much higher capacities and bandwidths (~50-60 GB/sec and several 10 PB)

Current Usage (Snapshot from 2015-09-29)

Storage consumption in size



Storage consumption in number of files

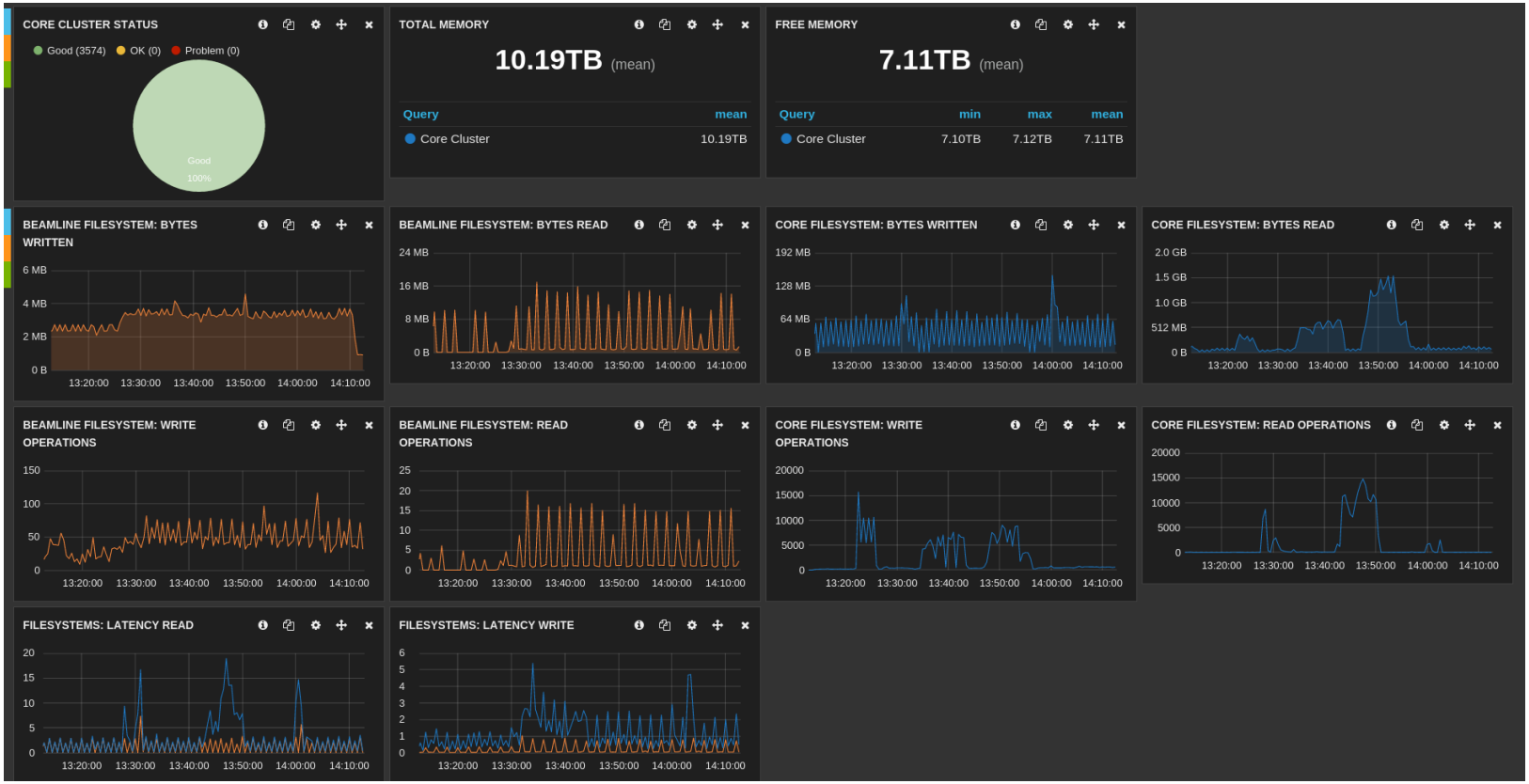


Different colors represent different beamlines

- Still in commissioning phase
- We expect significantly higher storage request in the near future (new detectors arrived in September)

System Monitoring on the Beamline

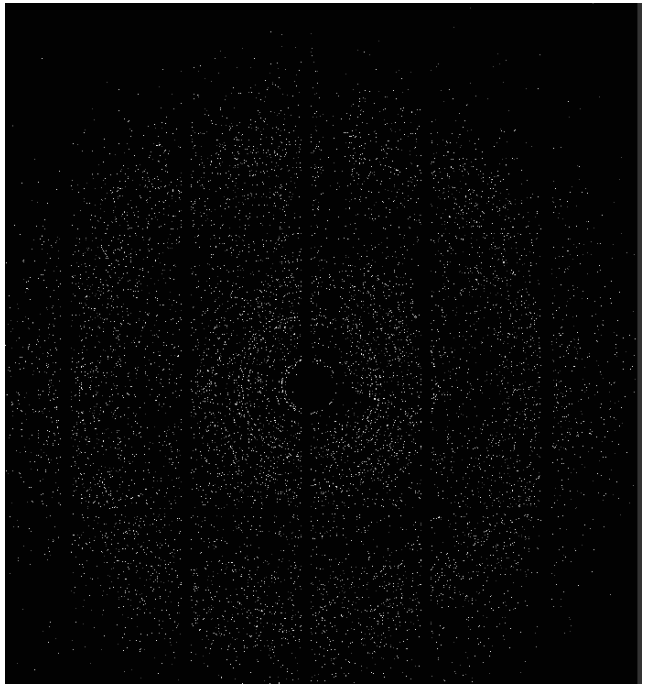
- Monitoring solution with Elasticsearch and Kibana
  - Collects syslogs, performance metrics, SMB statistics and beamline activities



Experiment-specific Adaptions: Live-Viewer and Fast Feedback Loop

- Short path data stream
- Based on ZeroMQ do distribute the data to parallel targets
- Independent from DESY's storage architecture (GPFS)
- Currently two applications:
  - To display images in real-time during the experiment
  - To analyze the data for reasonableness in near-real-time

    (25 Hz data ingest: 22-23 Hz analysis with ~ 1-2 seconds delay)



Further Interaction with DSIT

- From our experience with the data flow of the photon communities, we collected valuable use cases for modern scientific “Data Management Plans”, which we provided to LSDMA WP2.
- LSDMA WP2 has been recently integrated into INDIGO-DataCloud WP4 (DESY & KIT), which will provide a generic protoype for QoS and DLC solutions for IBM products (GPFS) and dCache, from which we hope to benefit in the future.

Publications

- Strutz, Gasthuber, Aplin, Dietrich, Kuhn, Ensslin, Smirnov, Lewendel, Guelzow: „ASAP3 – New Data Taking and Analysis Infrastructure for PETRA III“ - Chep 2015

Presentations

- Chep 2015
- Edge 2015
- HEPiX Spring 2015 Workshop
- IBM GPFS Workshop 2015
- PNI-HDRI Spring meeting 2015

Posters

- European XFEL Users' Meeting 2015
- solution.hamburg 2015

LSDMA
Autumn Meeting 2015
DLCL Struktur der Materie
GSI/FAIR

Status update

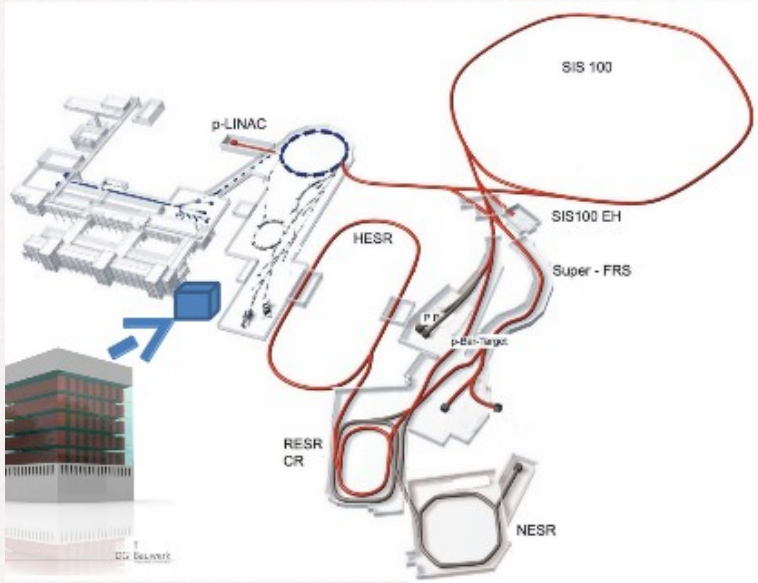
Kilian Schwarz, Sören Fleischer

GSI: a national Research Centre for heavy ion research
FAIR: Facility for Ion and Antiproton Research

GSI computing 2015

ALICE T2/NAF
HADES
LQCD (#1 in Nov‘ 14 Green 500)
~30000 cores
~ 25 PB lustre
~ 9 PB archive capacity

GreenCube
Computing Centre
- Construction finished
- Technical installation in final stage



FAIR computing at nominal operating conditions

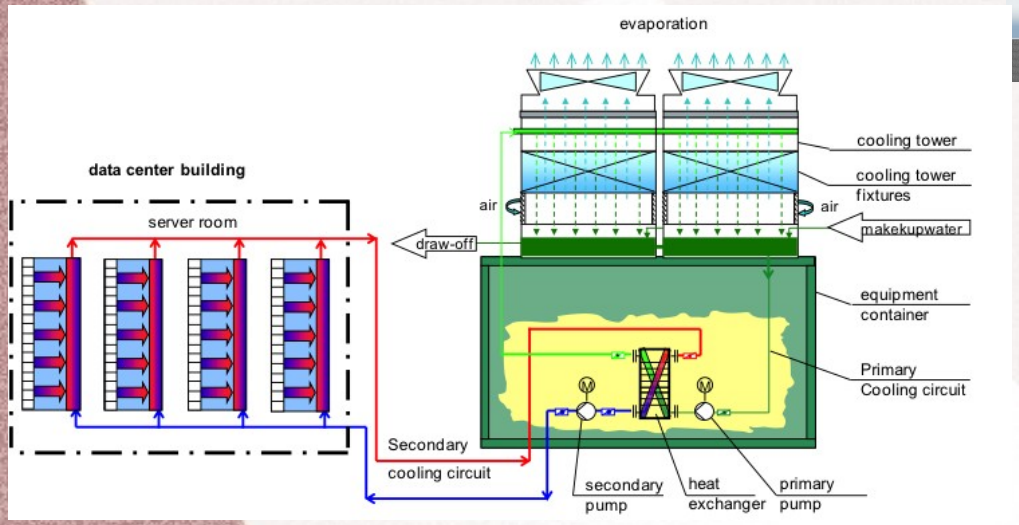
CBM
PANDA
NuSTAR
APPA
LQCD
300000 cores
40 PB disk/y
30 PB tape/y

View of construction site

- Open source and community software
- budget commodity hardware
- support different communities
- scarce manpower

FAIR Data Center

A common data center
for FAIR (Green Cube)



6 Floors, 4.645 sqm
Space for 768 19” racks (2,2m)
4 MW cooling (baseline)
Max cooling power 12 MW
Fully redundant (N+1)
PUE <1.07

Green Cube

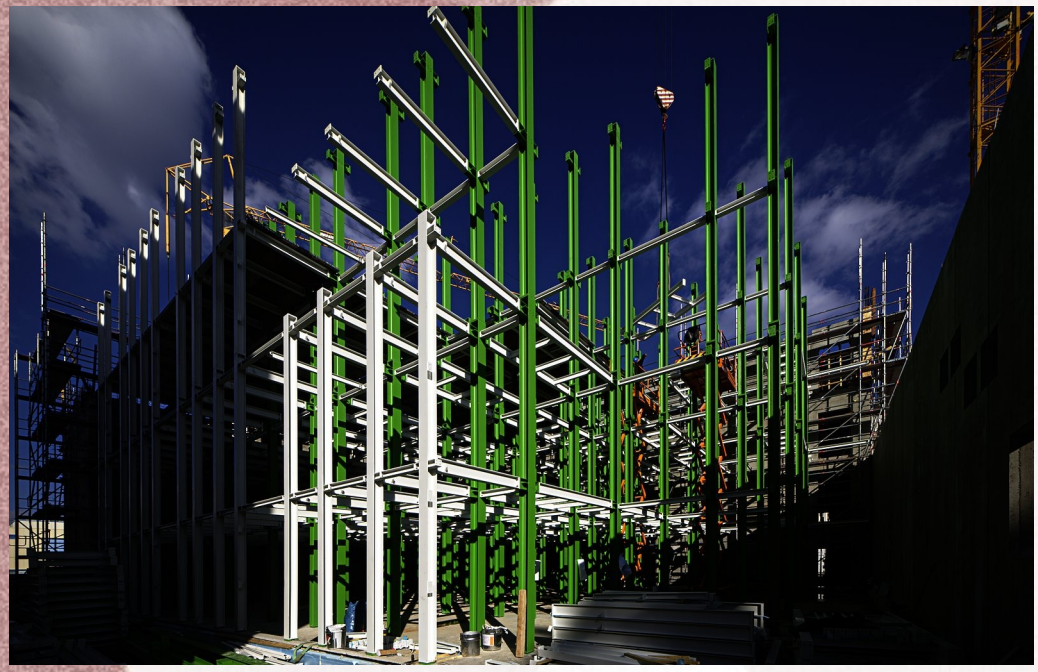
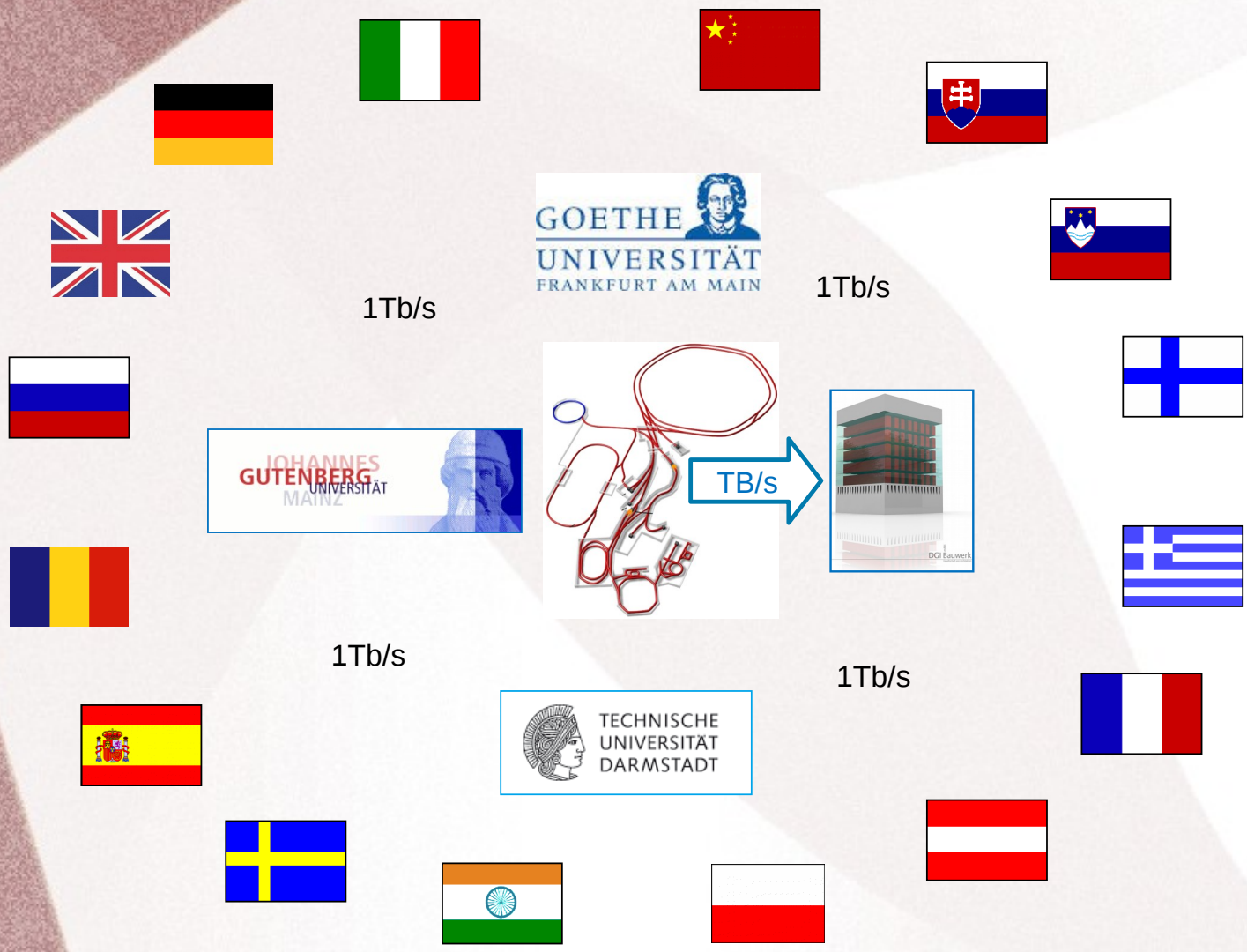


Foto: C. Wolbert

- Construction finished (started in December 2014)
- Installation of technical infrastructure in final stage
- On schedule for start of operation in Q4/2015

FAIR Computing: T0/T1 MAN (Metropolitan Area Network) & Grid/Cloud



Status Update LSDMA DLCL FAIR Structure of Matter

GSI: manpower

- 2 positions have been advertised
- 1 person has been hired for the LSDMA FAIR DLCL (Structure of Matter)
  - S. Fleischer (April 1, 2015)
  - still looking for a suitable person for the second position

General objectives
(according to document „Ausführliche Bedarfsanalyse“)

- Global Federations
  - prototype solutions based on xrootd exist
- Parallel and distributed computing
  - Triggerless „online“ system
    - FairRoot: integrating GPUs, parallelisation via ZeroMQ
  - Metropolitan Area Network
    - high speed links to surrounding universities
      - prototypes exists, e.g. 120 Gb to Frankfurt
    - federated identity management: prototype exists and is being tested, planned to be integrated in FAIR AAI, integration in DFN still outstanding

General objectives
(according to document „Ausführliche Bedarfsanalyse“)

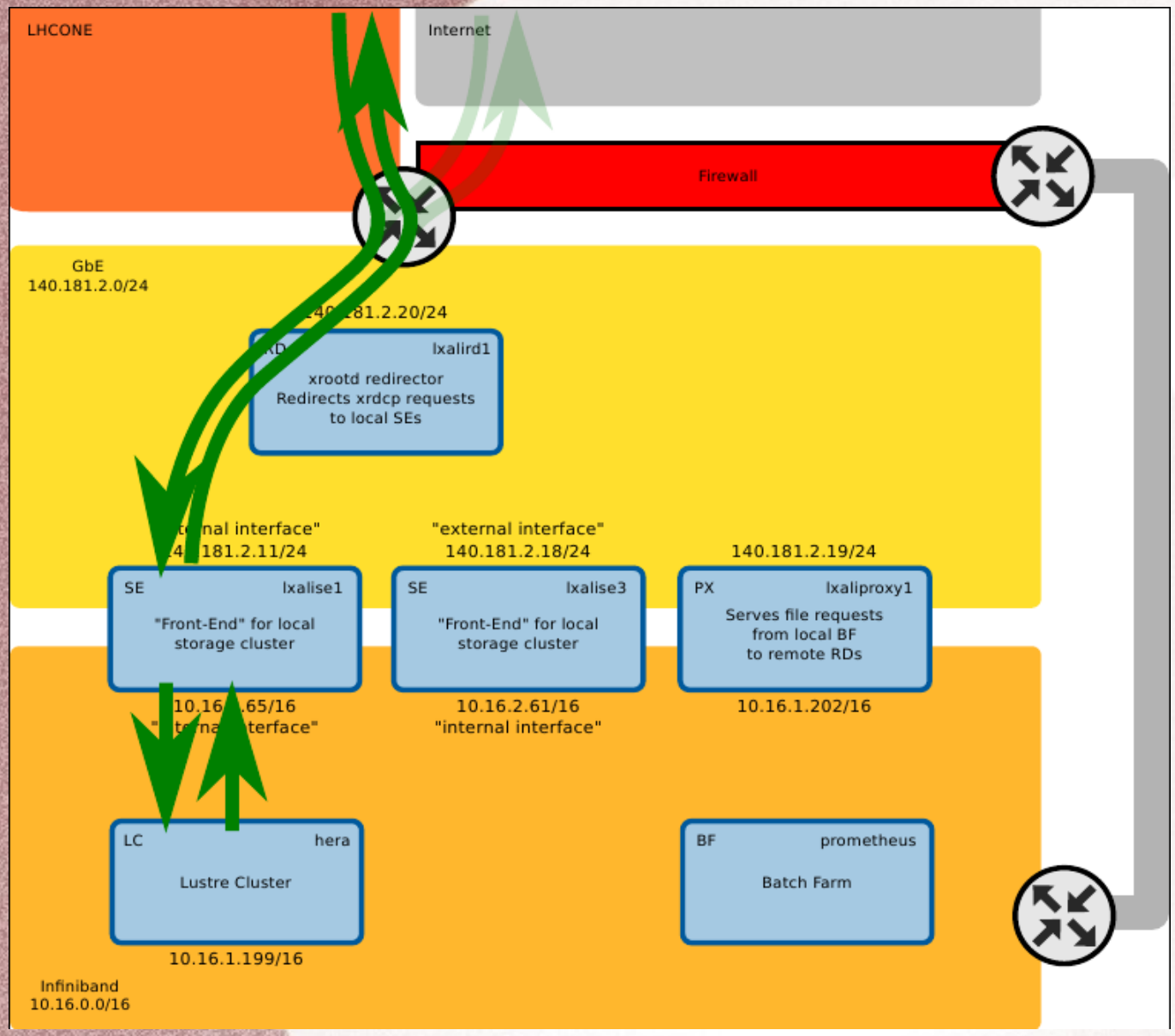
- Parallel and distributed computing
  - Grid/Cloud infrastructure
    - Prototype for PANDA, PandaGrid is up and working
    - Integration of FAIR Metropolitan Area System into global Grid/Cloud infrastructure
      - interface between Grid and HPC systems with storage being a key component (a working prototype based on xrootd has been developed: for details see next slides)
        - main components:
          - xrootd forward proxy (in production already)
          - xrootd client plugin for directly accessing Lustre (in testing stage)

interface between HPC and Grid storage

a) ALICE SE and the xrootd Forward Proxy

- SE: xrootd running on top of Lustre file system
- an xrootd redirector with split directive redirects external clients to the external interfaces of the local SE data servers, internal clients to the internal interfaces
  - meanwhile two xrd data servers are connected
  - both servers have an Infiniband interface connected to the HPC system (compute farm and storage) and a 10 Gb Ethernet interface pointing to LHCOne
- jobs running inside the HPC cluster

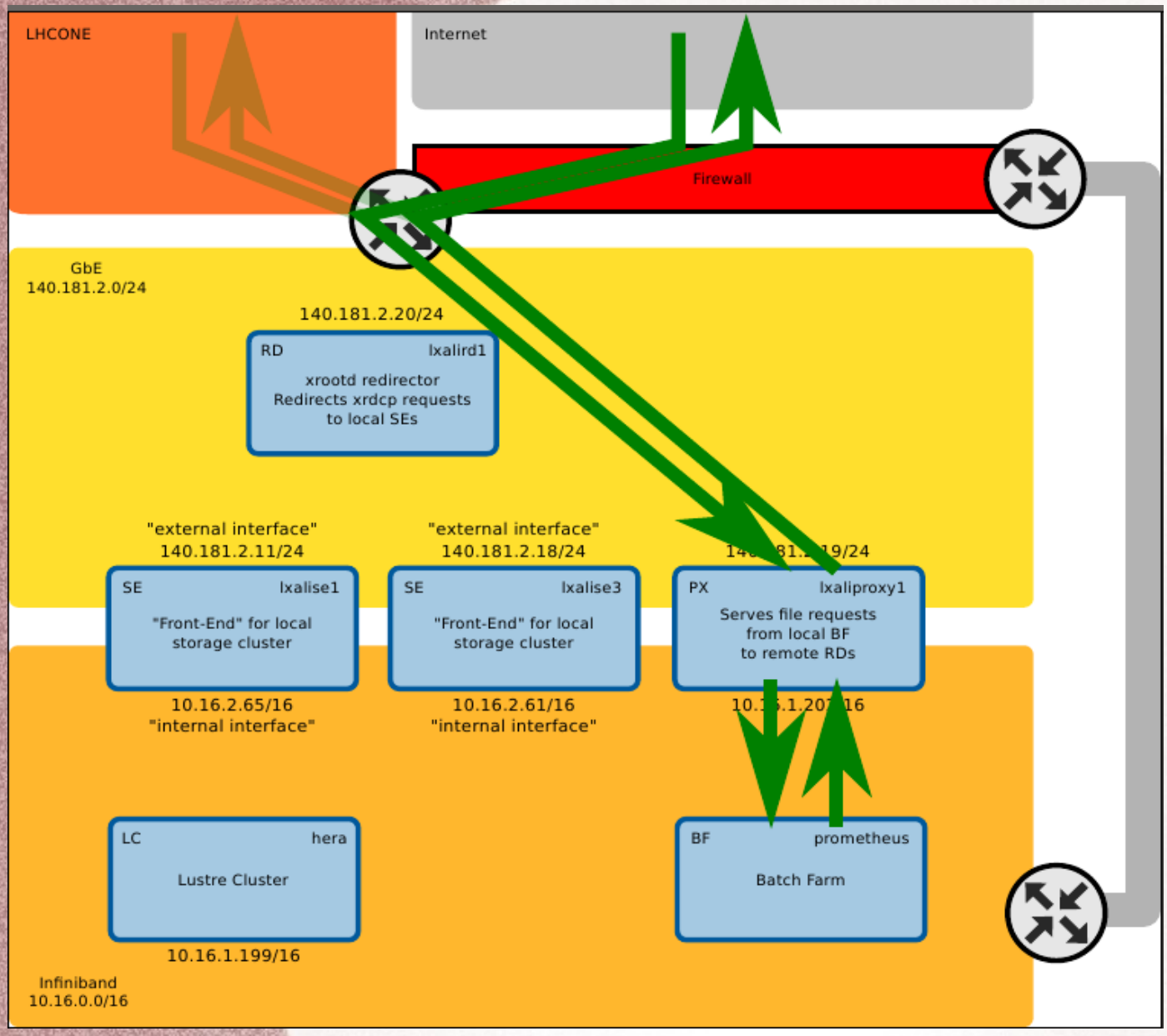
ALICE T2 – new storage setup



xrootd redirector

- points external clients to external SE interfaces
- points internal clients to internal SE interfaces

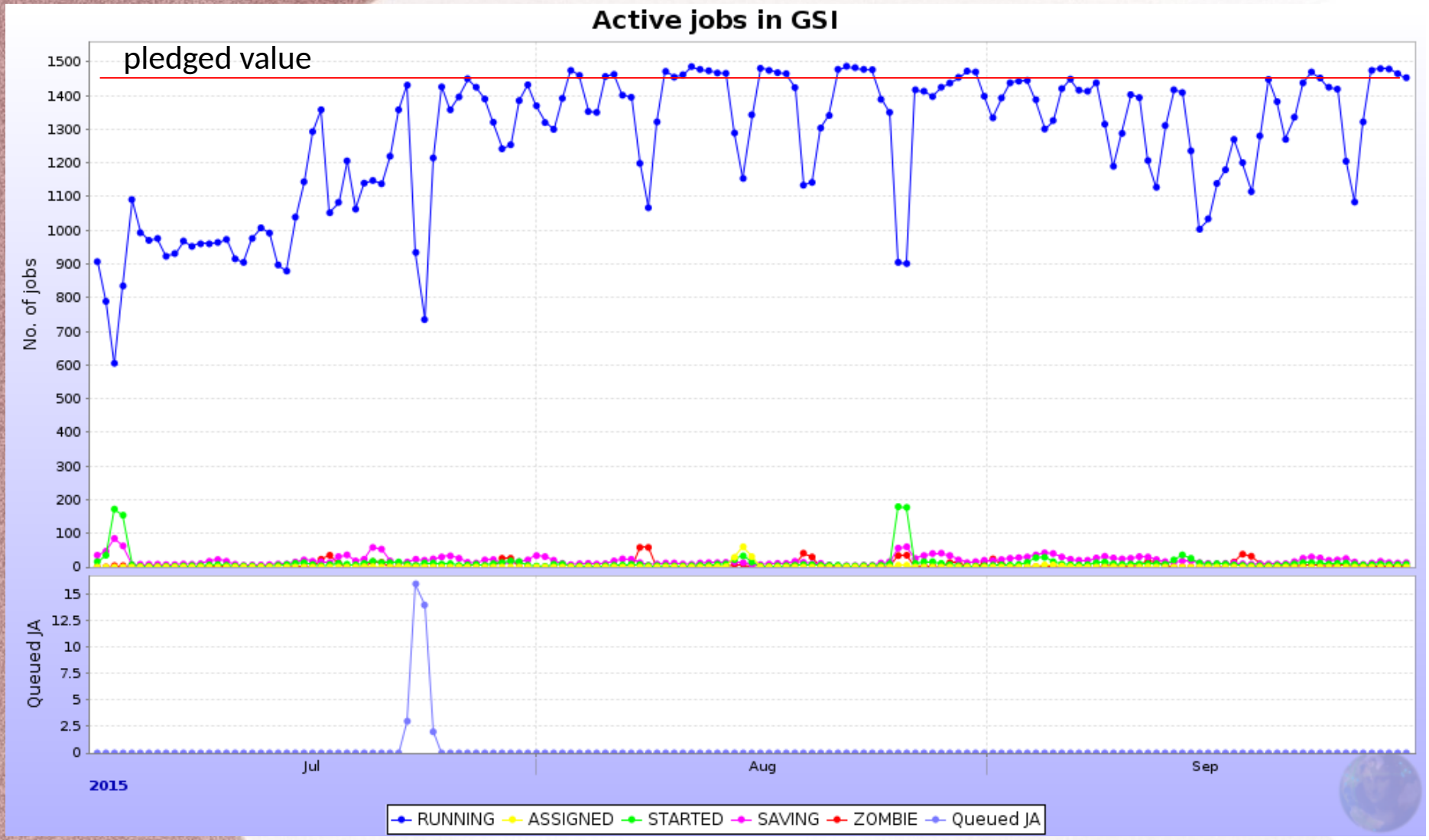
ALICE T2 – new storage setup



xrootd proxy tunnels traffic from/to clients within the HPC environment to/from Storage Elements outside GSI

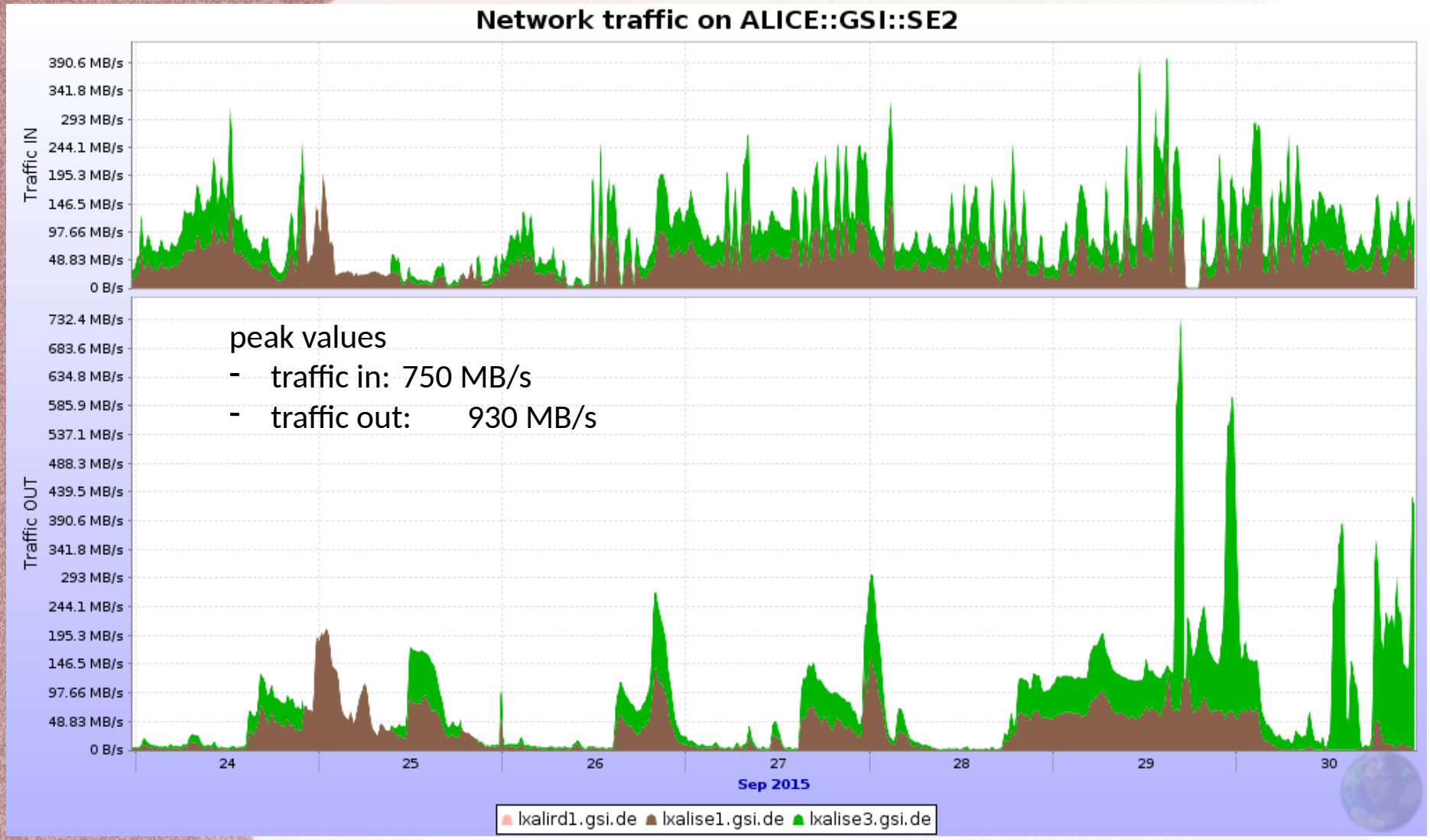
ALICE T2 – new storage setup

- in production mode since several months
- increasing job numbers



ALICE T2 – new storage setup

- in production mode since several months
- reliable storage infrastructure



ALICE T2: new storage infrastructure

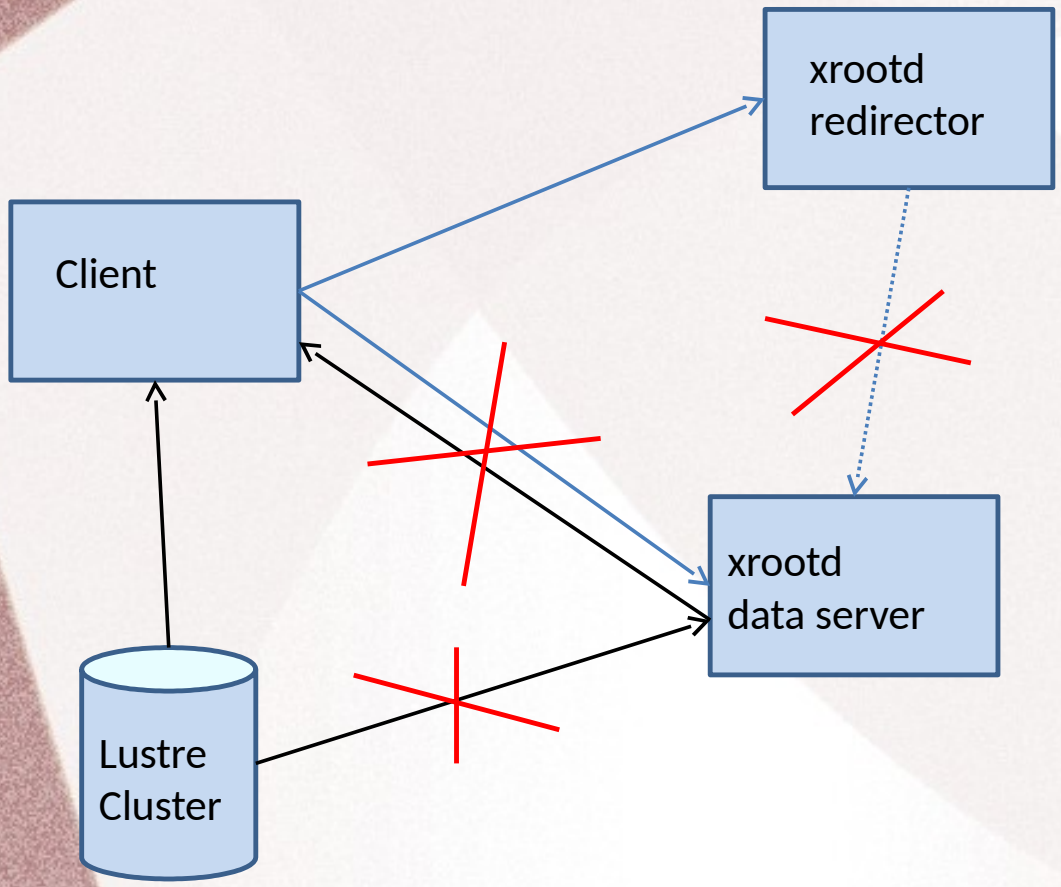
- all storage services have been implemented with
  - auto configuration (Chef cookbooks to deploy new machines and to keep existing ones consistent)
  - monitoring and auto restart as well as information service (Monit)
  - [illegible]

```
# Make a symlink from /usr/lib/x86_64-linux-gnu/libcrypto.so.1.0.0 (actual file) to /usr/lib64/libcrypto.so (link)
link "/usr/lib64/libcrypto.so" do
  to "/usr/lib/x86_64-linux-gnu/libcrypto.so.1.0.0"
  link_type :symbolic
end

# xrootd manual package "installation"
remote_directory "xrootd" do
  path node["xrootd"]["installdir"]  # Path on target machine
  mode "0755"
  owner node["users"]["storage-user"]["user"]
  group node["users"]["storage-user"]["group"]
  files_owner node["users"]["storage-user"]["user"]
  files_group node["users"]["storage-user"]["group"]
  files_mode "0755"
  action :create
end
```

ALICE T2: new storage infrastructure client plugin for Lustre URL

(still to be tested in production environment)



outview

- the GSI LSDMA FAIR DLCL is happy to continue working closely together with LSDMA DSIT and will continue to integrate DSIT solutions into the FAIR DLCL environment whereever it seems feasible
- - examples so far:
  - DSIT Lustre developments
  - DSIT AAI

LSDMA Spring Meeting 2016

- will take place at GSI
- March 7-13, 2016
- details will be announced by A. Streit