

# ***Grid Operations in Germany***

***T1-T2 workshop 2015  
Torino, Italy***

Kilian Schwarz

WooJin Park

Christopher Jung

# ***Table of contents***

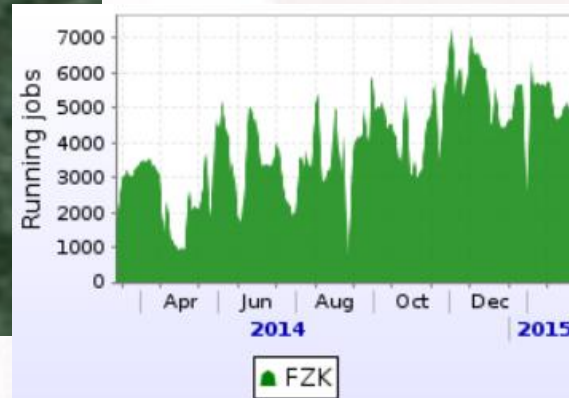
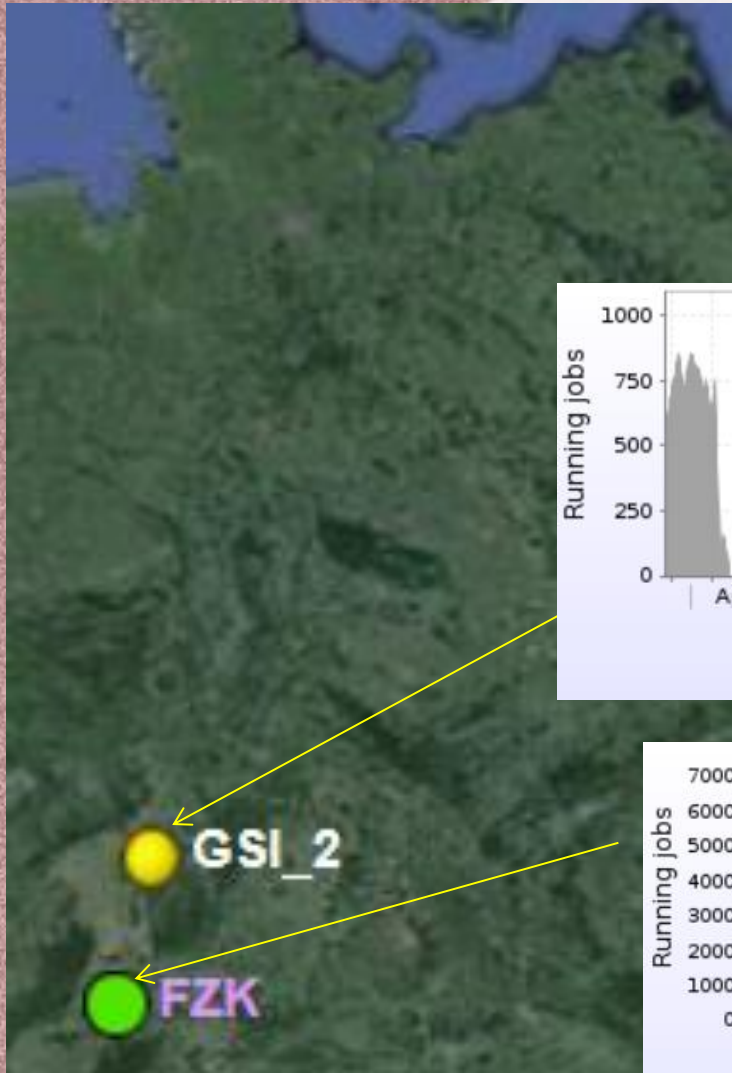
- Overview
- GridKa T1
- GSI T2
- Summary

# ***Table of contents***

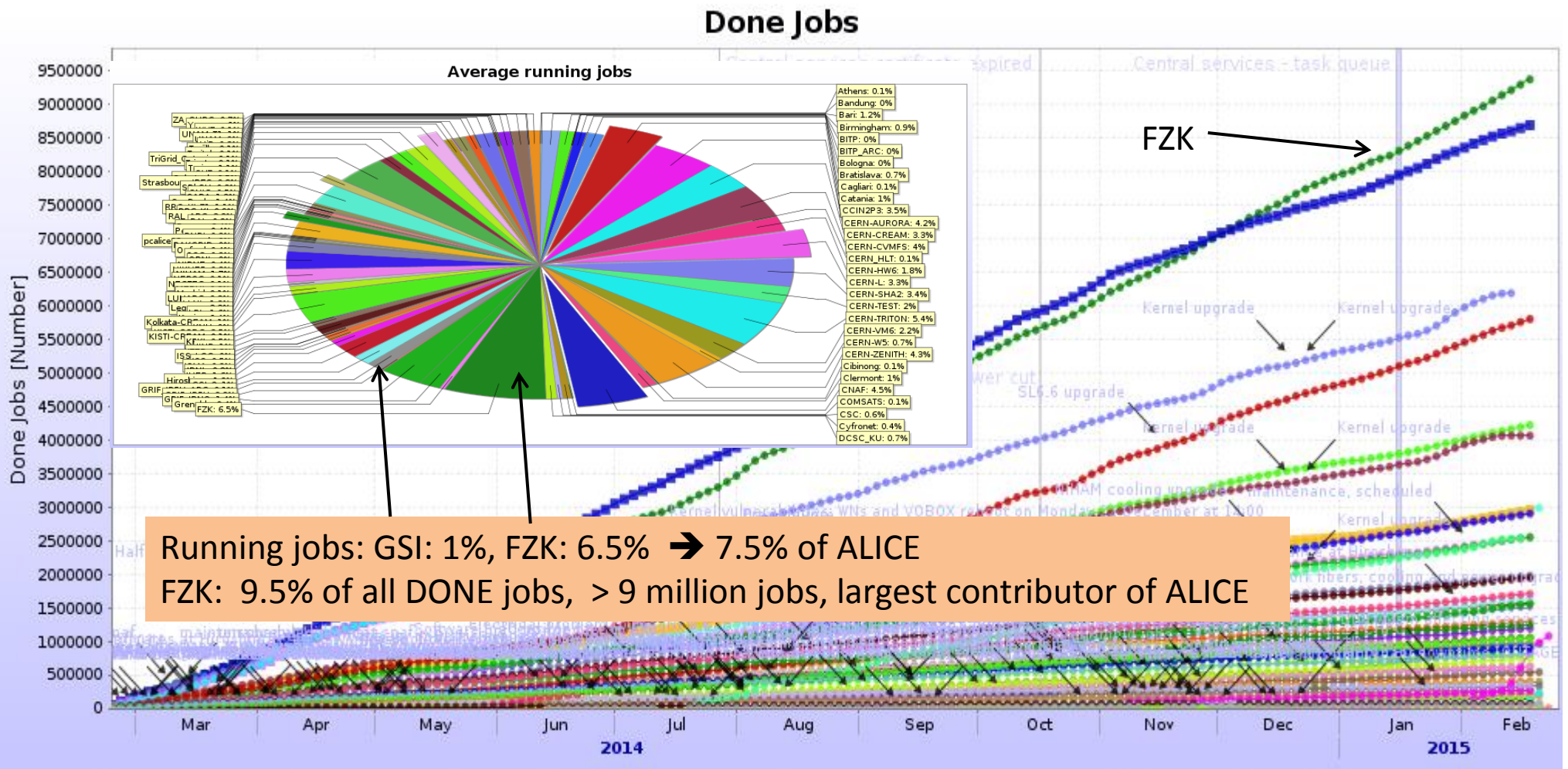
- **Overview**
- GridKa T1
- GSI T2
- Summary

# Map of German Grid sites

- T1: GridKa/FZK in Karlsruhe
- T2: GSI in Darmstadt



# Job contribution (last year)



- Athens — Bandung — Bari — Birmingham — BITP — BITP\_ARC — Bologna — Bratislava — Cagliari — Catania — CCIN2P3 — CERN — CERN-AURORA
- CERN-CREAM — CERN (Wigner) — CERN-HW6 — CERN-L — CERN (Meyrin) — CERN-TEST — CERN-TRITON — CERN-VM6 — CERN-W5 — CERN-ZENITH
- CERN\_HLT — Cibinong — Clermont — CNAF — COMSATS — CSC — Cyfronet — DCSC\_KU — FZK — Grenoble — GRIF\_IPNO — GRIF\_IRFU — GRIF\_IRFU\_ARC
- GSI\_2 — Hiroshima — IHEP — IPNL — ISMA — ISS — ISS\_LCG — ITEP — JINR — KFKI — KISTI-CREAM — KISTI\_GSDC — KNU — Kolkata-CREAM — Kosice
- LBL — Legnaro — LLNL — LUNARC — Madrid — MEPHI — NECTEC — NERSC — NIHAM — NIKHEF — NIPNE — ORNL — OSC — Oxford — PAKGRID
- pcalice92.cern.ch — PNPI — Poznar — Prague — RAL — RAL\_ARC — RRC-KI — RRC\_KI\_T1 — SaoPaulo — SARA — SNIC — SPbSU — Strasbourg\_IRES
- Subatech — SUT — Torino — Torino-Torrent — Trieste — TriGrid\_Catania — Troitsk — Trujillo — UIB — UNAM — UNAM\_T1 — WUT — Yerevan — ZA\_CHPC

# Storage contribution

AliEn name	Size	Used	Free	Usage	No. of files	Type	Size	Used	Free	Usage
ALICE::FZK::SE	1.694 PB	1.407 PB	294.3 TB	83.04%	28,418,238	FILE	3.803 PB	3.697 PB	107.7 TB	97.23%
ALICE::GSI::SE2	550 TB	226 TB	324 TB	41.1%	371,158	FILE	7.273 PB	4.557 PB	2.716 PB	62.66%
ALICE::FZK::TAPE	640 TB	2.843 PB	-	454.8%	1,795,011	FILE	640.4 TB	549.9 TB	90.5 TB	85.87%

## Total size:

- GridKa: 3 PB Disk SE including 0.7 PB tape buffer –
  - all 10 xrootd servers have been upgraded to v4.0.4
  - duplicated name space problem was solved with new xrootd version
- **GSI: xrootd shows the capacity of the complete Lustre cluster. Can this be limited to the size of a directory or some quota space ?**
  - an extra quota for the ALICE SE has been introduced, now
- 3.3 PB disk based SE
- 5.25 PB tape capacity
- 700 TB disk buffer with Tape backend

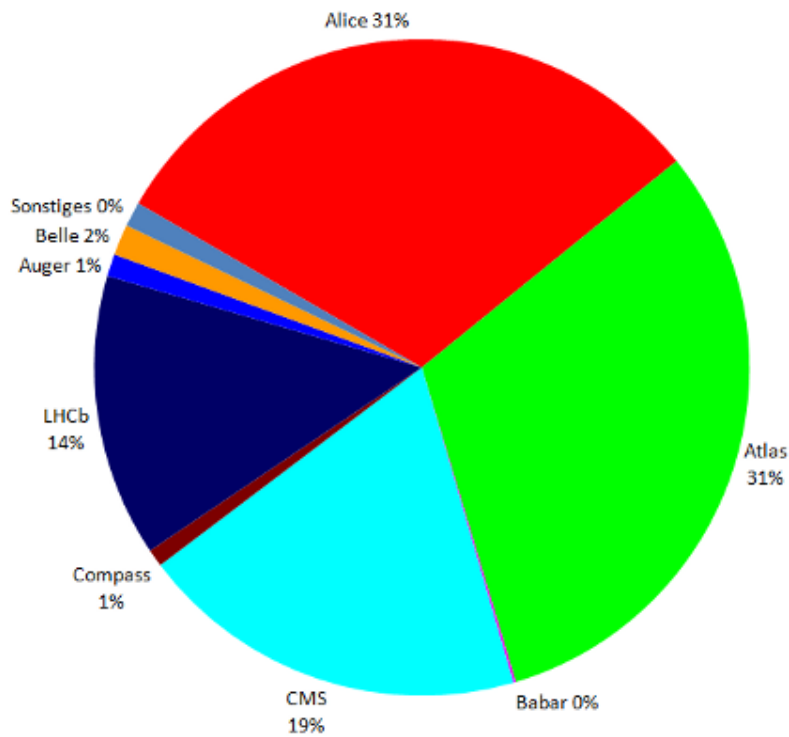
# ***Table of contents***

- Overview
- **GridKa T1**
- GSI T2
- Summary

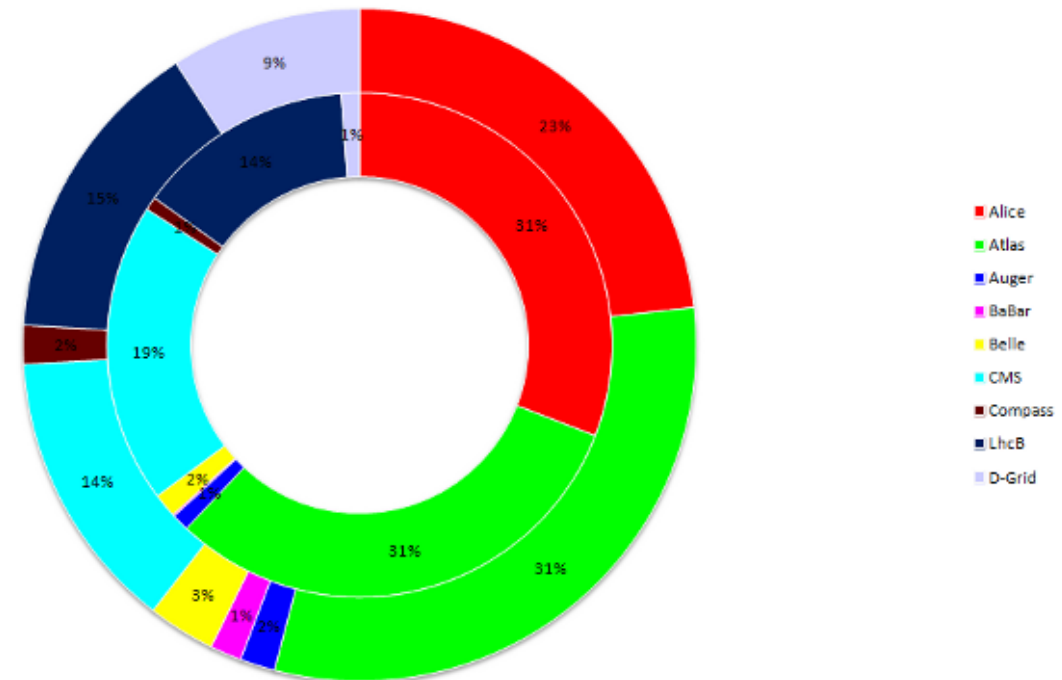
# GridKa Usage

(annual statistics 2014)

- The two largest users are ATLAS und ALICE
- The 4 LHC experiments together use > 90%



- ALICE and CMS use more than the nominal share
- ATLAS and LHCb use about their nominal share



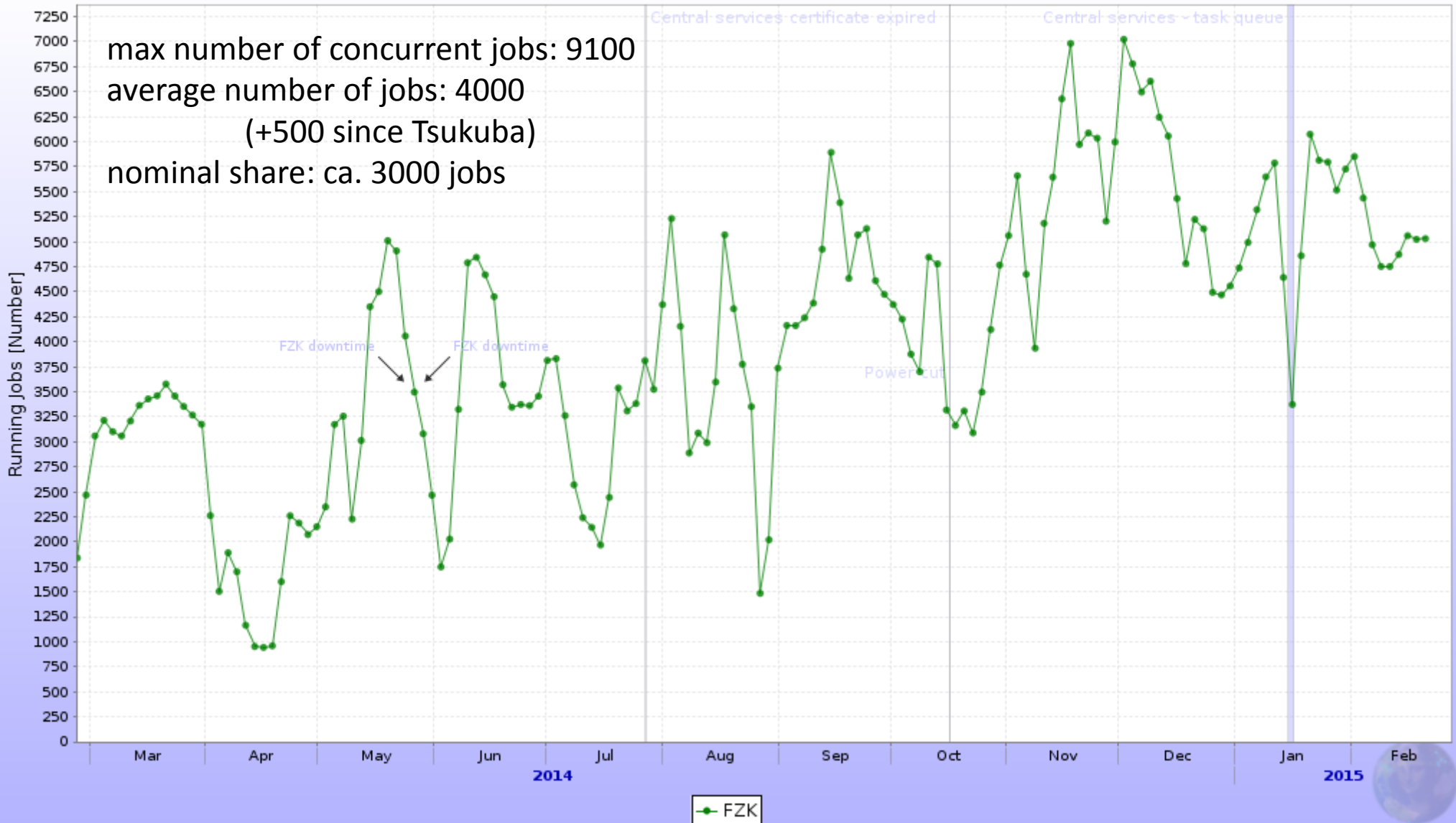
8

total used wall time: 107 372 920 hours



# Jobs at GridKa within last year

## Running Jobs



# ALICE Job Efficiency@GridKa



- average job efficiency: 77%
- 3 major efficiency drops:
- before April 2014: bug in the SE selecting algorithm  
(fixed by ALICE Offline team April 2014, created heavy load on GridKa firewall)
- December 2014 and February 2015: raw data processing  
(quite a number of mismatched/missing files between Alien catalog and GridKa tape storage leads to many jobs reading data from remote)

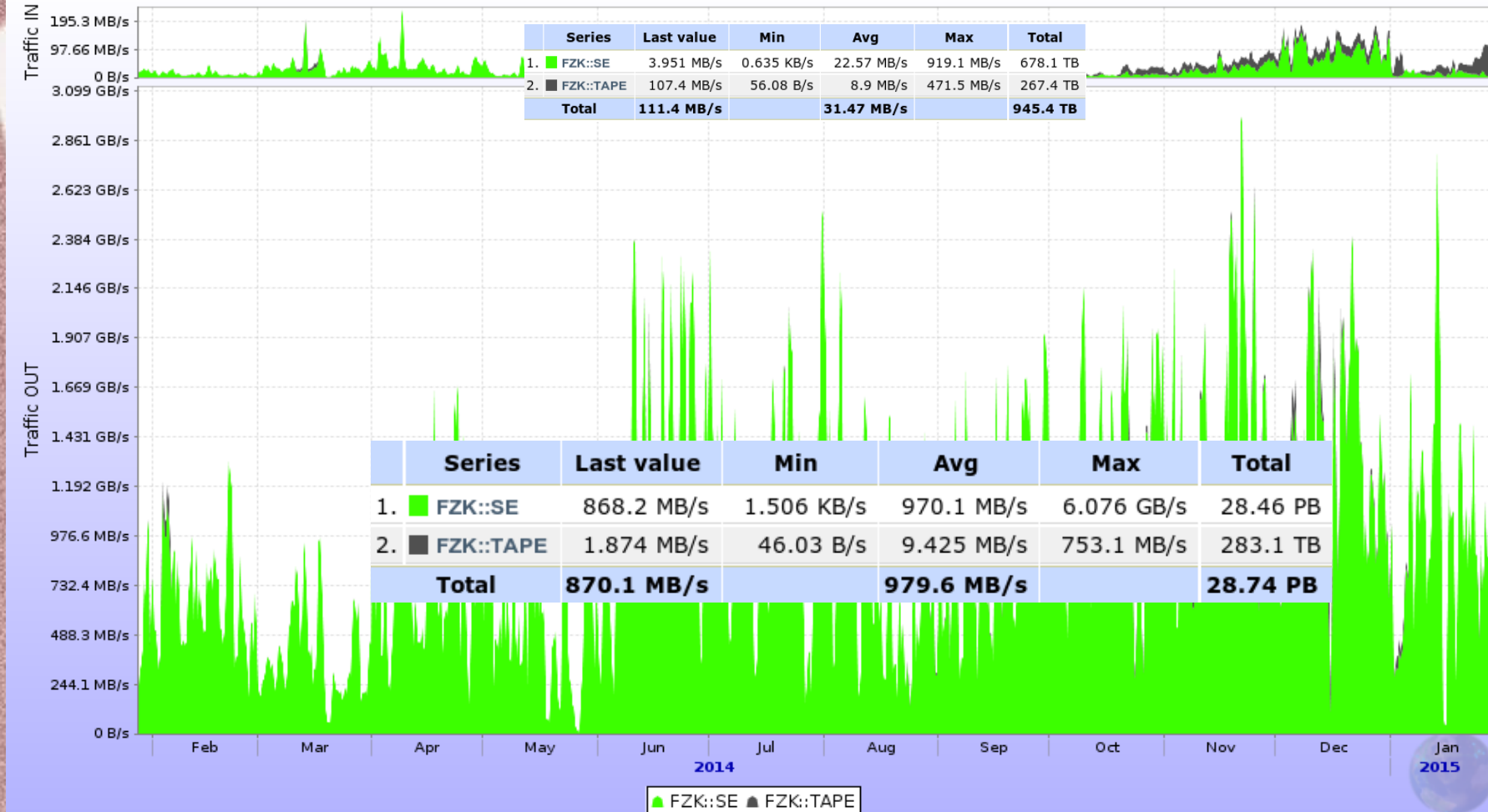


# xrootd based SEs work fine and are used intensively.

- Disk-based xrootd-SE has a size of 2.16 PB (+ 660 TB tape buffer)
- 28 PB have been read from ALICE::FZK::SE in 2014 (average reading rate: ~1 GB/s)
  - (almost twice as much as reported in Tsukuba)
- Tape usage increased again

SE Name	AliEn SE		Statistics					Xrootd info			
	AliEn name	Size	Used	Free	Usage	No. of files	Type	Size	Used	Free	Usage
14. FZK - SE	ALICE::FZK::SE	1.694 PB	1.407 PB	294.3 TB	83.04%	28,418,238	FILE	3.803 PB	3.604 PB	203.5 TB	94.77%

### Aggregated network traffic per SE



5. FZK - TAPE	ALICE::FZK::TAPE	640 TB	2.843 PB	-	454.8%	1,795,011	FILE	640.4 TB	543.5 TB	96.
---------------	------------------	--------	----------	---	--------	-----------	------	----------	----------	-----

# XRootD Architecture at GridKa

- 08/14: all 10 xrd servers of disk SE upgraded to v4.0.4
  - flawless operation
  - duplicated name space problem solved
- ALICE::FZK::Tape
  - upgrade to v4.0.4 planned
  - plan to use FRM (File Residency Manager) feature of xrootd
    - local scripts need to be modified/tested
      - staging, reading, writing DONE
      - migrating, purging tests still ongoing
  - file consistency check alien FC ⇔ FZK::Tape done
    - **number of inconsistencies found**
    - list handed over to ALICE::Offline
- Internal monitoring : individual server monitoring page has been setup:
  - <http://web-kit.gridka.de/monitoring/xrootd.php>
  - number of active files, daemon status, internal reading/writing test and network throughputs, internal/external data access are displayed.
  - will be connected to the icinga alarm system

# *Various points of interest*

- Resources at GridKa:
  - ALICE has requested (pledged number in WLCG REBUS) 4.01 PB of disk space in 2015
    - **GridKa will fulfill the pledges** later this year.
    - The deadline (1st of April) will be missed.
    - WLCG has been informed.
- IPv6
  - GridKa internal tests ongoing (currently with ATLAS services)
  - can provide a testbed for ALICE services (but due to lack of manpower not with high priority)

# ***Table of contents***

- Overview
- GridKa T1
- **GSI T2**
- Summary

**GSI: a national Research Centre for heavy ion research**  
**FAIR: Facility for Ion and Antiproton Research ~2018**

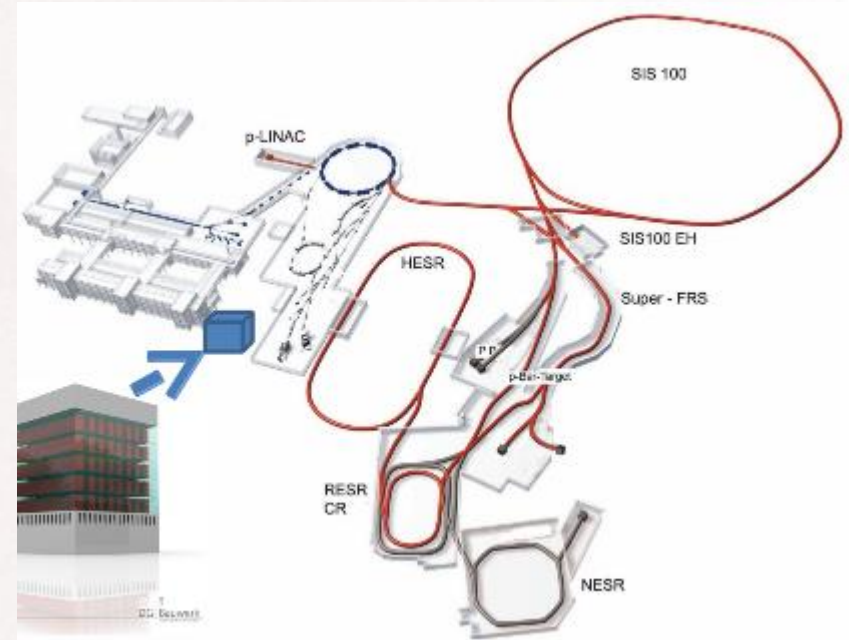
**GSI computing today**

ALICE T2/NAF  
HADES  
~10500 cores  
~ 7.2 PB lustre  
~ 9 PB archive capacity

**FAIR computing 2018**

CBM  
PANDA  
NuSTAR  
APPA  
LQCD  
300000 cores  
40 PB disk  
40 PB archive

GreenCube  
Computing Centre  
Currently being constructed

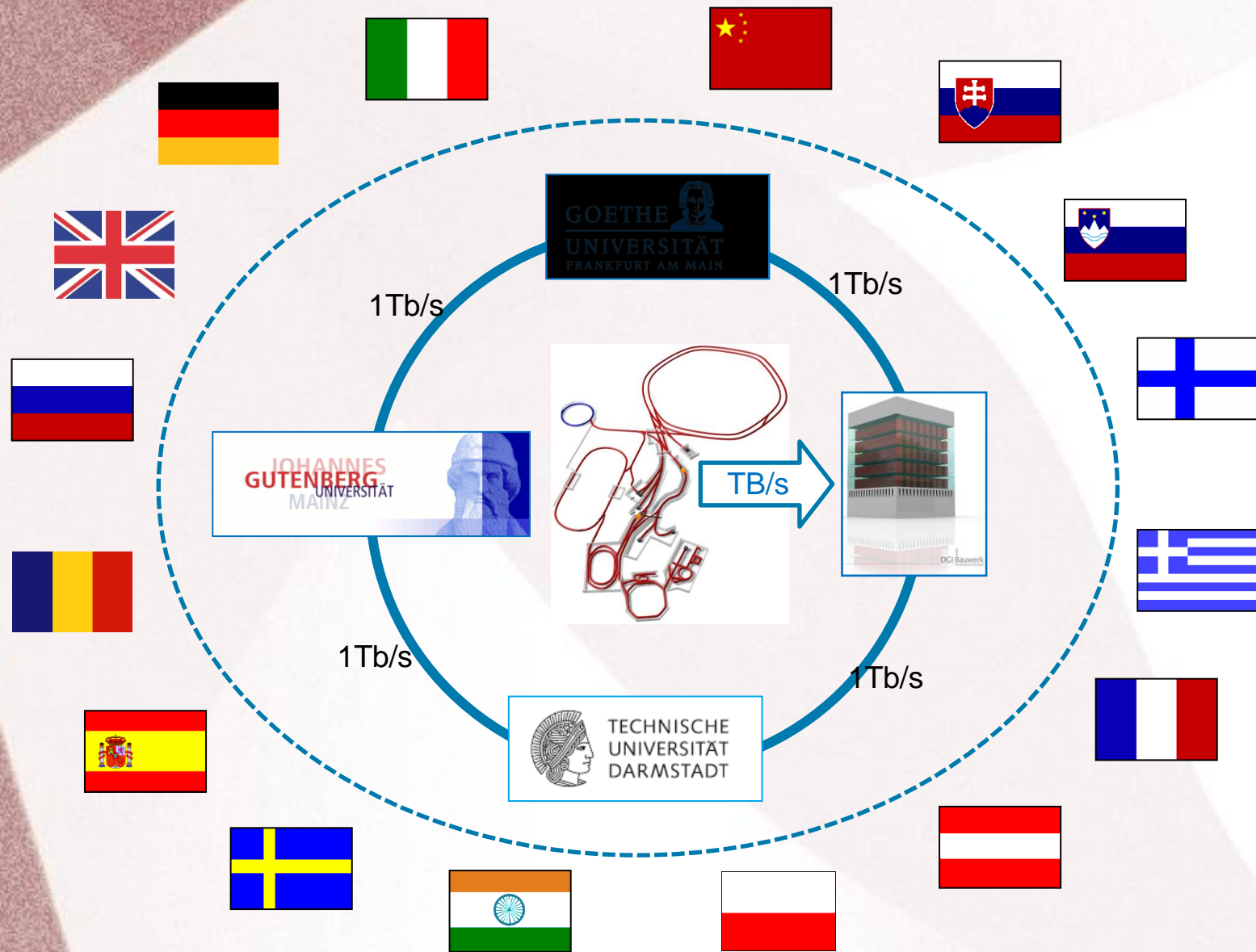


Open source and community software  
budget commodity hardware  
support different communities  
scarce manpower



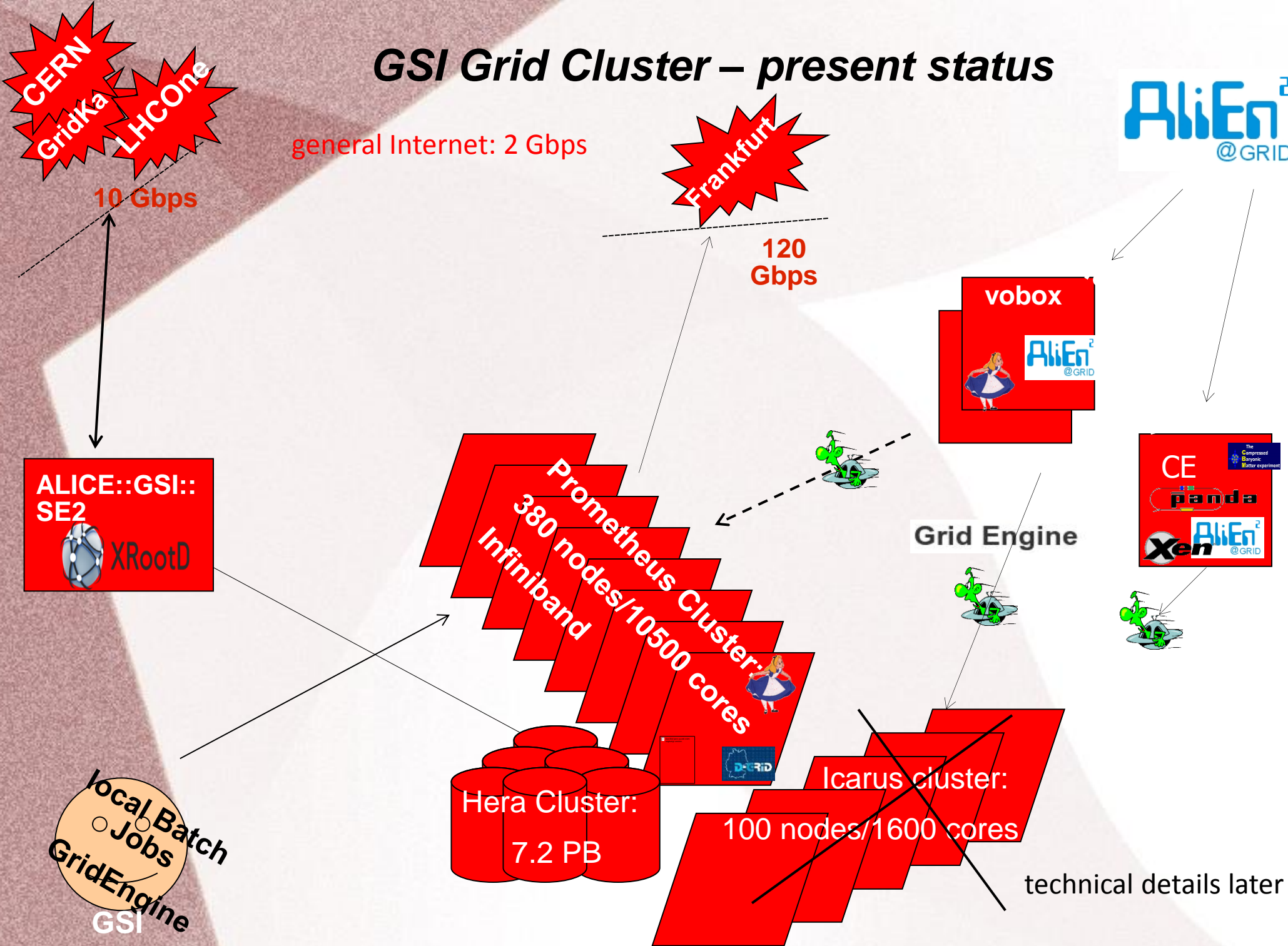
View of construction site

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

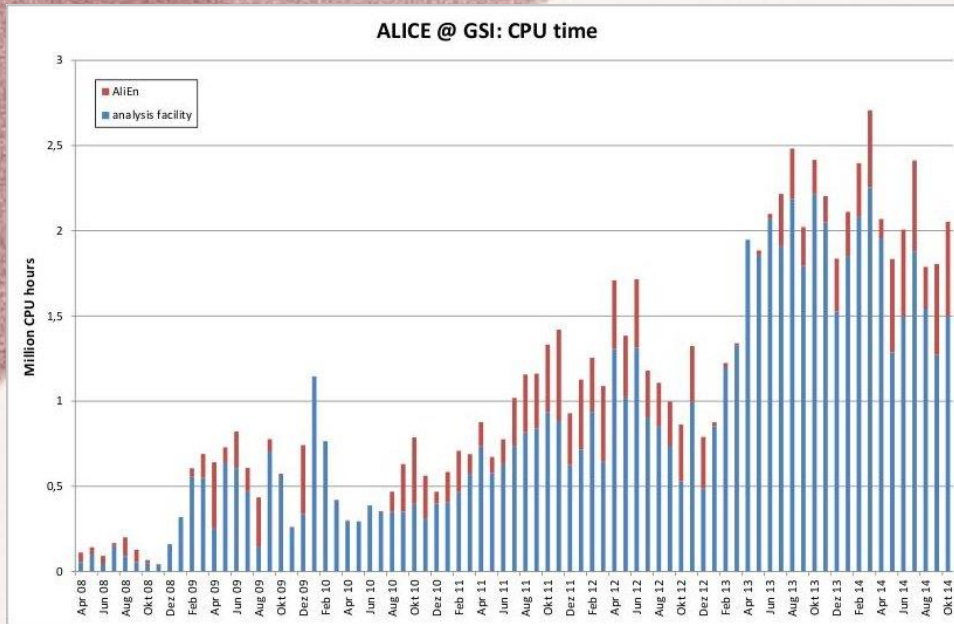




# GSI Grid Cluster – present status



# GSI Darmstadt: 1/3 Grid, 2/3 NAF



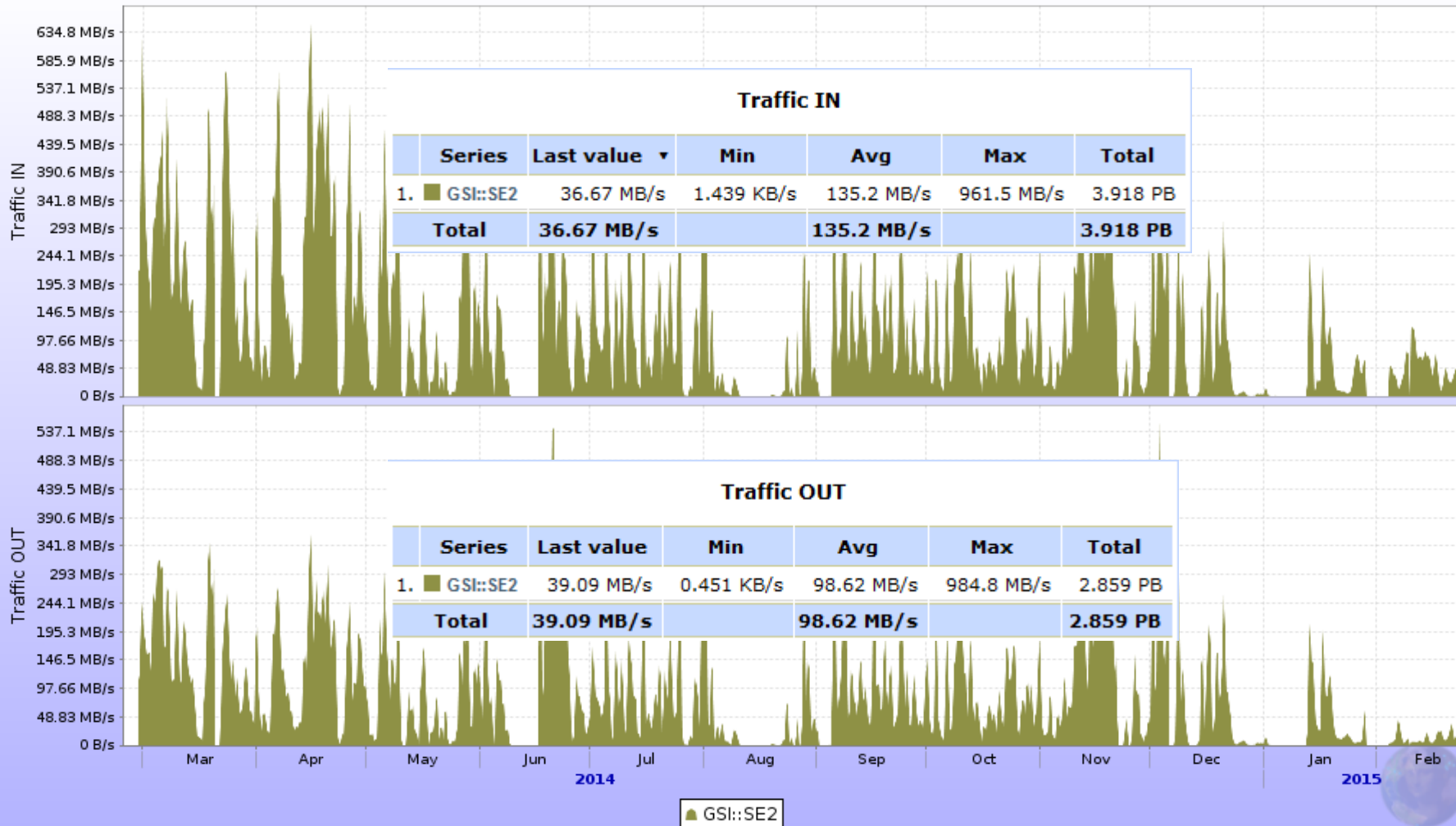
- GSI Darmstadt :
  - Batch System: Grid Engine
  - average: 800 T2 Jobs
    - max (1500 jobs)
  - about 1/3 Grid (red), 2/3 NAF (blue)
  - Intensive usage of the local GSI cluster through ALICE NAF (ca. 30% of the cluster capacity)



# GSI Storage Element

- *ALICE::GSI::SE2 works well and is intensively used*
- *in order to fulfill new requirements the setup of the SE has been changed*
  - *technical details follow*
- *Lustre quota for SE has been introduced (separation of local and Grid usage)*
- *in 2014 about 4 PB have been written, 3 PB read*
  - *(both about 0.5 PB more than in Tsukuba report)*
- *remaining issue: ALICE SE tests fail once in a while (under investigation)*

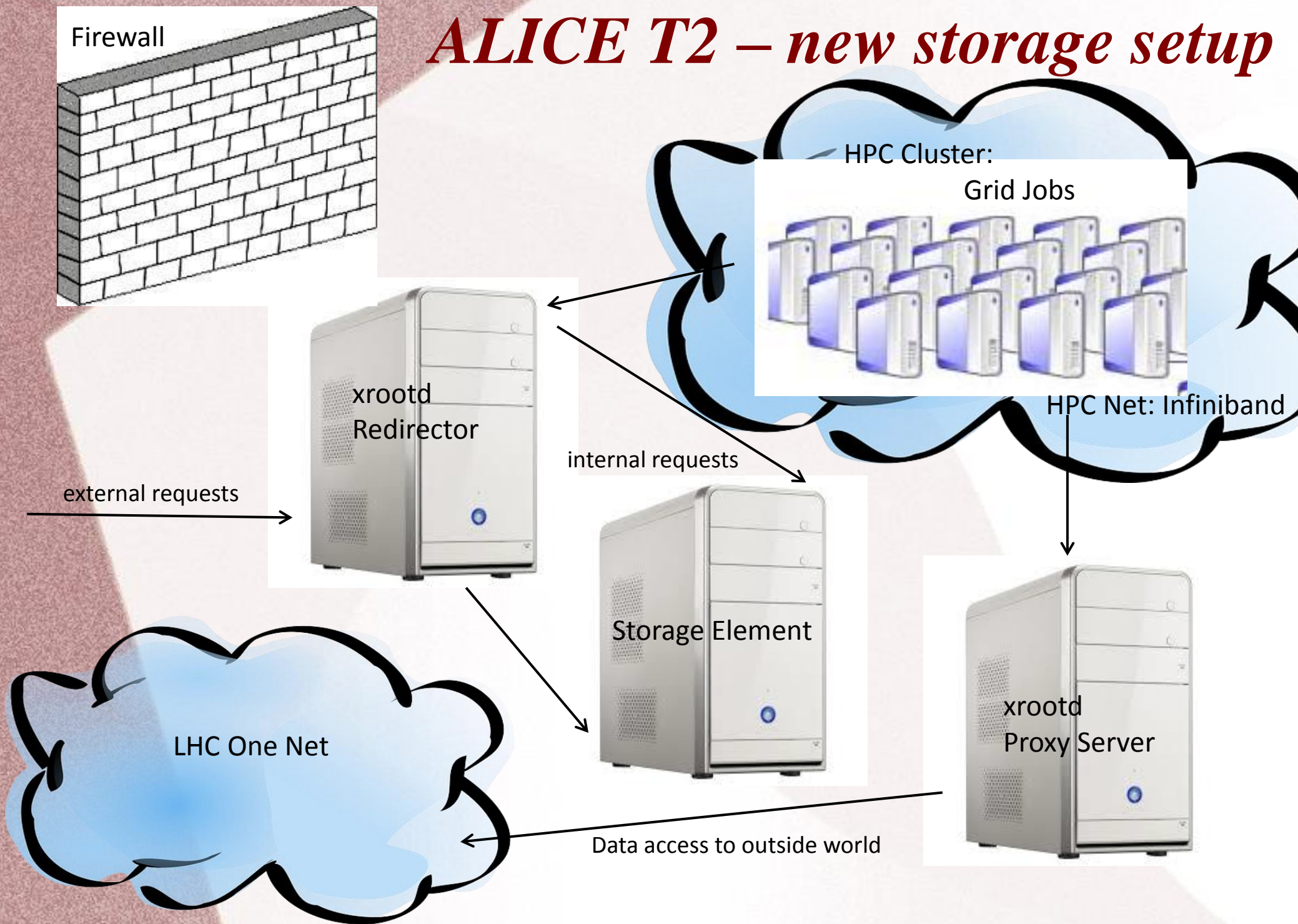
Aggregated network traffic per SE



# *new setup of GSI ALICE T2*

- old test cluster (Icarus) has been decommissioned
  - cluster was a Grid only cluster with almost no Firewall protection
- ALICE T2 needs to run in GSI HPC environment (completely firewall protected Infiniband world)
  - needs to fulfill HPC requirements
    - AliEn services will be tunnelled through
    - data access needs to go through xrootd proxy
    - redirector needs to distinguish between inside/outside access
  - advantage: parasitical use of local resources

# ALICE T2 – new storage setup



# *GSI: next activities*

- Fix pending issues with new ALICE T2 setup
  - a new set of AliEn code has been provided in order to be able to create xrootd Proxy URLs when using remote SEs
    - *work in progress*
- new manpower will be hired
  - also ALICE T2 will profit from this
- plans for IPv6:
  - first test environment end of 2015

# ***Table of contents***

- Overview
- GridKa T1
- GSI T2
- Summary**

# *Summary*

- German sites provide a valuable contribution to ALICE Grid
  - Thanks to the centres and to the local teams
- new developments are on the way
- Pledges for 2015 can be fulfilled to a large extent
- FAIR will play an increasing role (funding, network architecture, software development and more ...)