Grid Operations in Germany

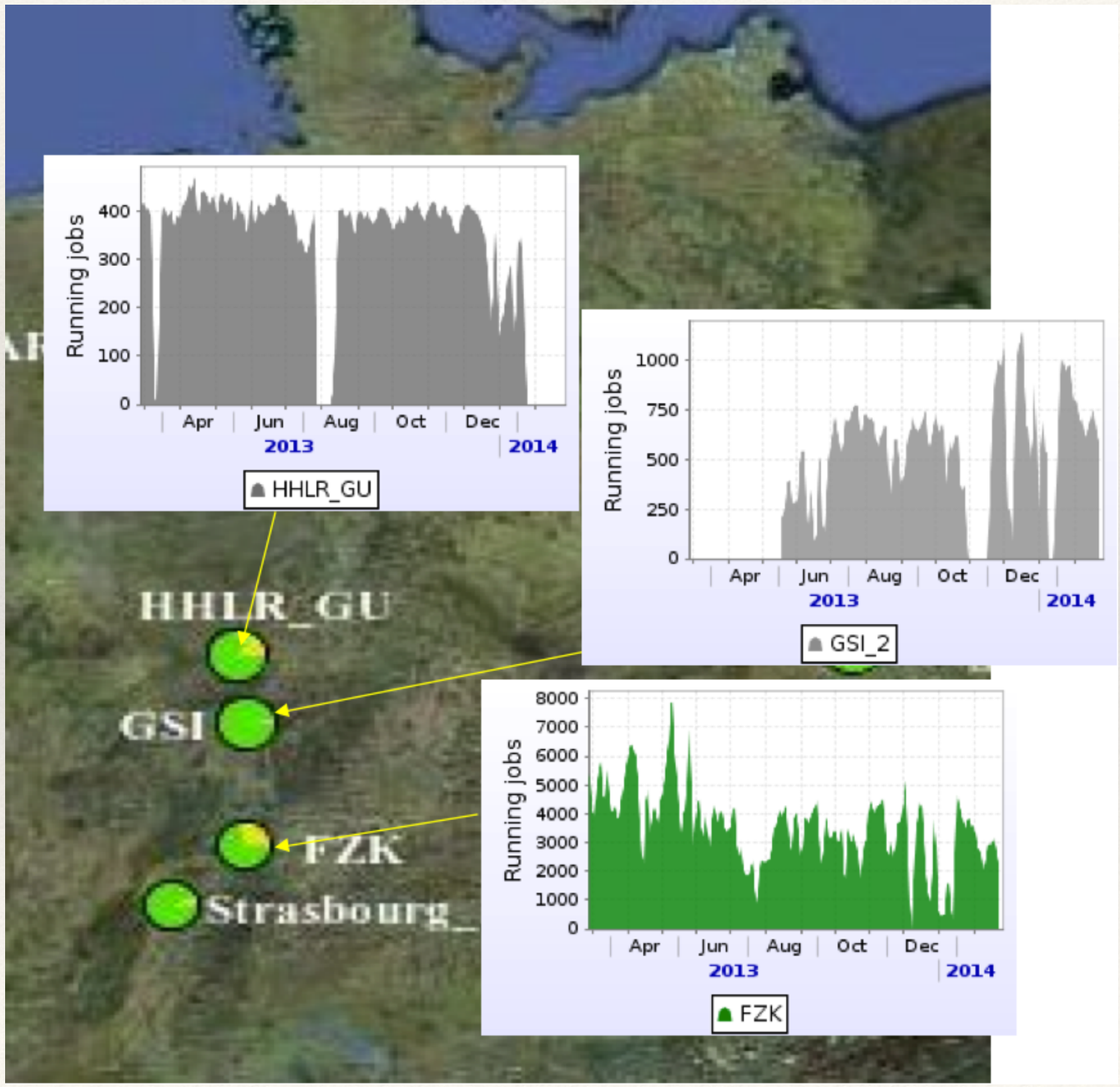
ALICE Tier-1/2 workshop
Tsukuba, Japan

Kilian Schwarz (GSI)
Christopher Jung (KIT)
*WooJin J. Park (KIT & GSI)**

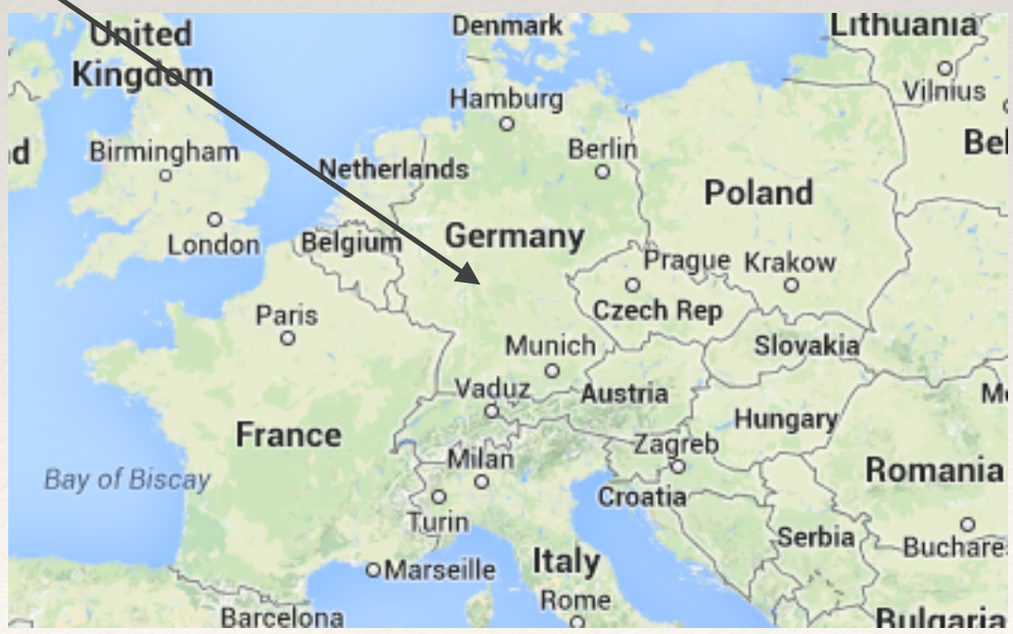
Table of Contents

Table of Contents

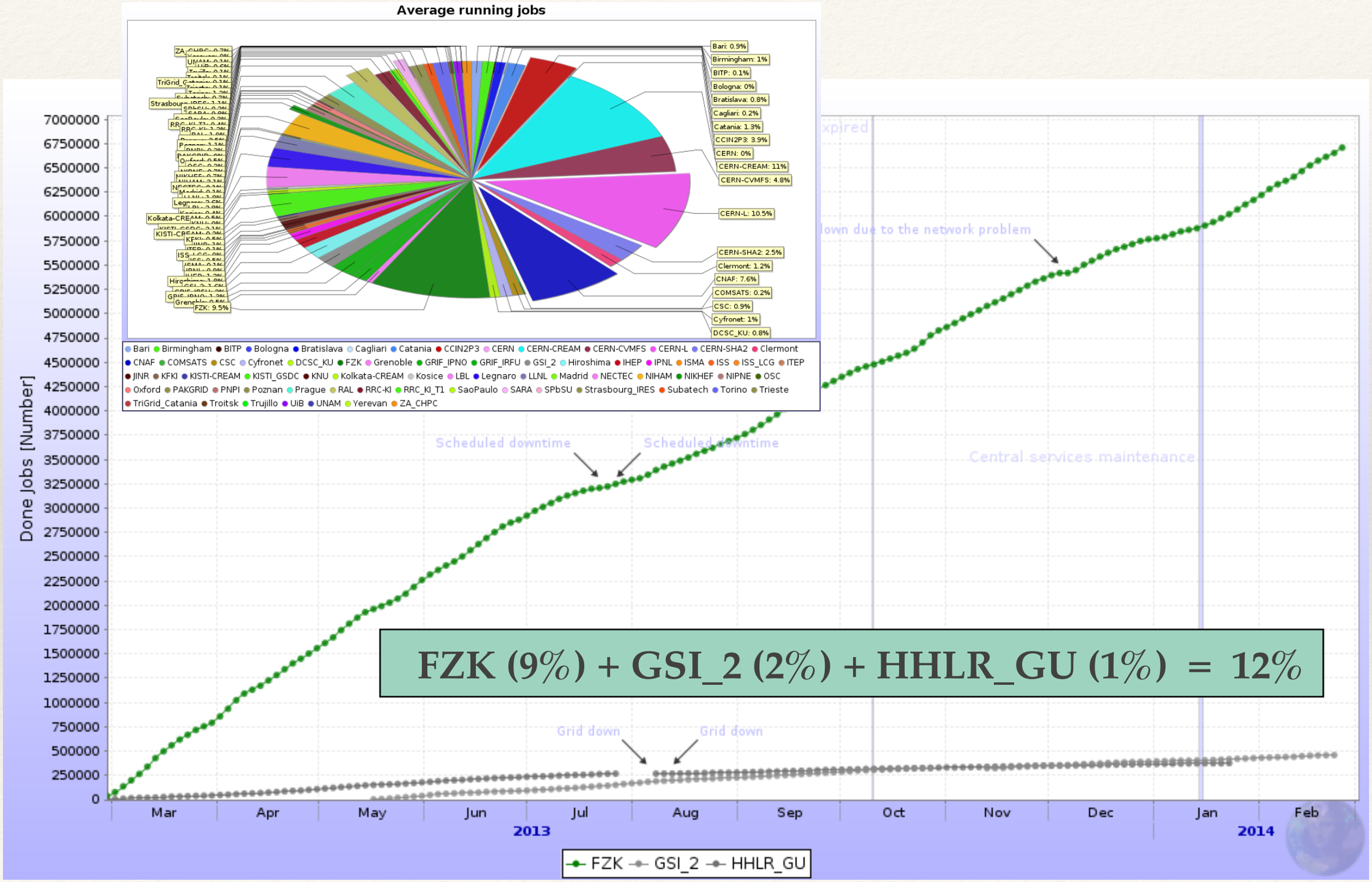
Map of German Grid Sites



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



Job Contribution (last year)



FZK (9%) + GSI_2 (2%) + HHLR_GU (1%) = 12%

Storage Contribution

AliEn name	Size	Used	Free	Usage	No. of files	Type	Size	Used	Free	Usage
ALICE::FZK::SE	1.694 PB	1.256 PB	449.2 TB	74.11%	26,751,467	FILE	14.43 PB	10.96 PB	3.466 PB	75.98%
ALICE::GSI::SE2	550 TB	236.6 TB	313.4 TB	43.02%	378,429	FILE	5.184 PB	4.108 PB	1.076 PB	79.24%
ALICE::FZK::TAPE	640 TB	2.858 PB	-	457.3%	1,801,824	FILE	640.4 TB	564.9 TB	75.51 TB	88.21%

Total size:

- GridKa: 2.7 PB Disk SE (0.7 PB tape buffer included)
  - xrootd infos not correct:
    - every redirector reports full disk space of all new servers
    - can probably be fixed in config
- GSI: xrootd shows the capacity of the complete Lustre cluster.
  - Can this be limited to the size of a directory or some quota space?
- 3.3 PB disk based SE (ALICE total: 23 PB)
  - 660 TB disk buffer with tape backend

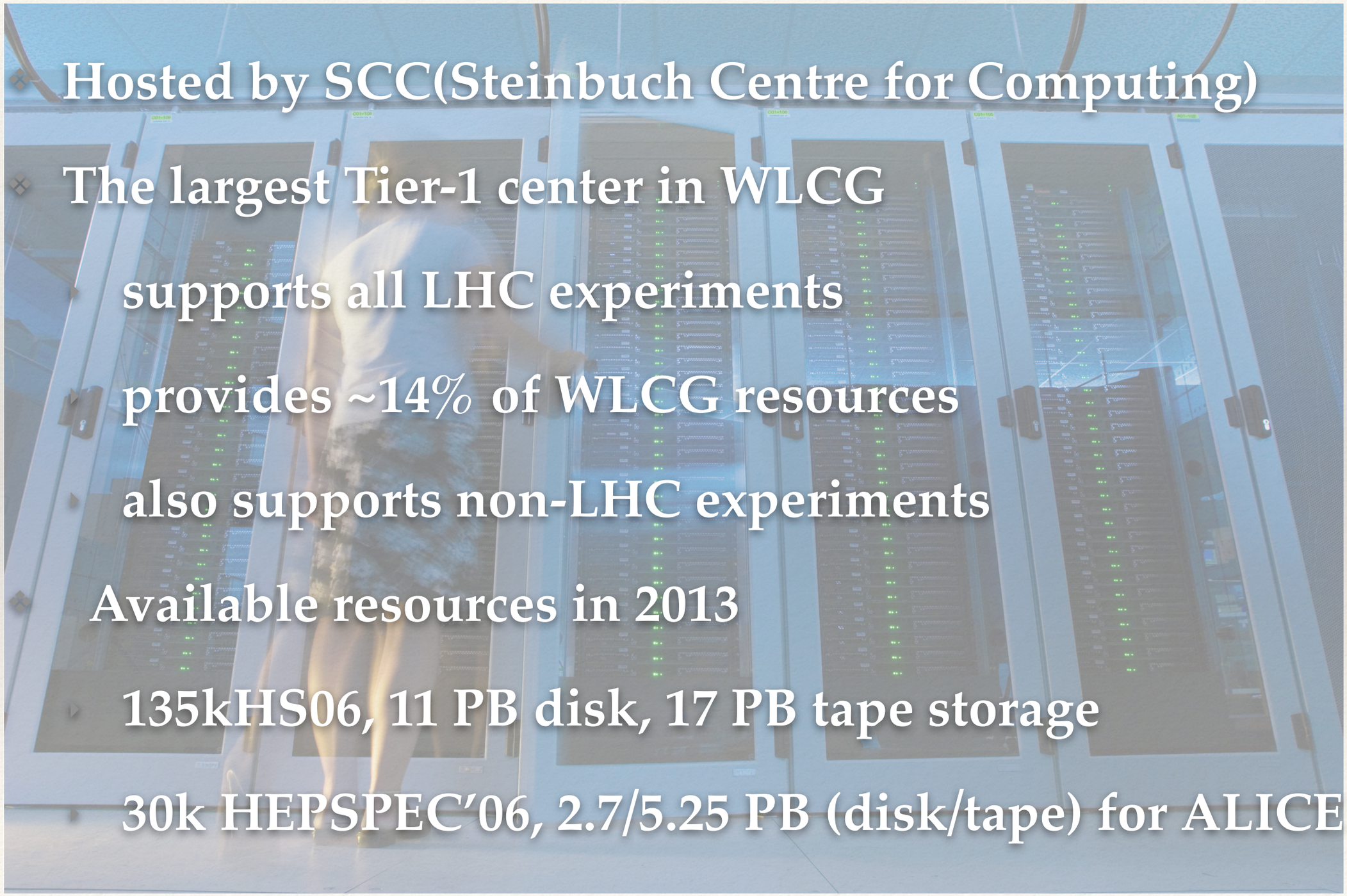
Table of Contents

Karlsruhe Institute for Technology (KIT)

Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

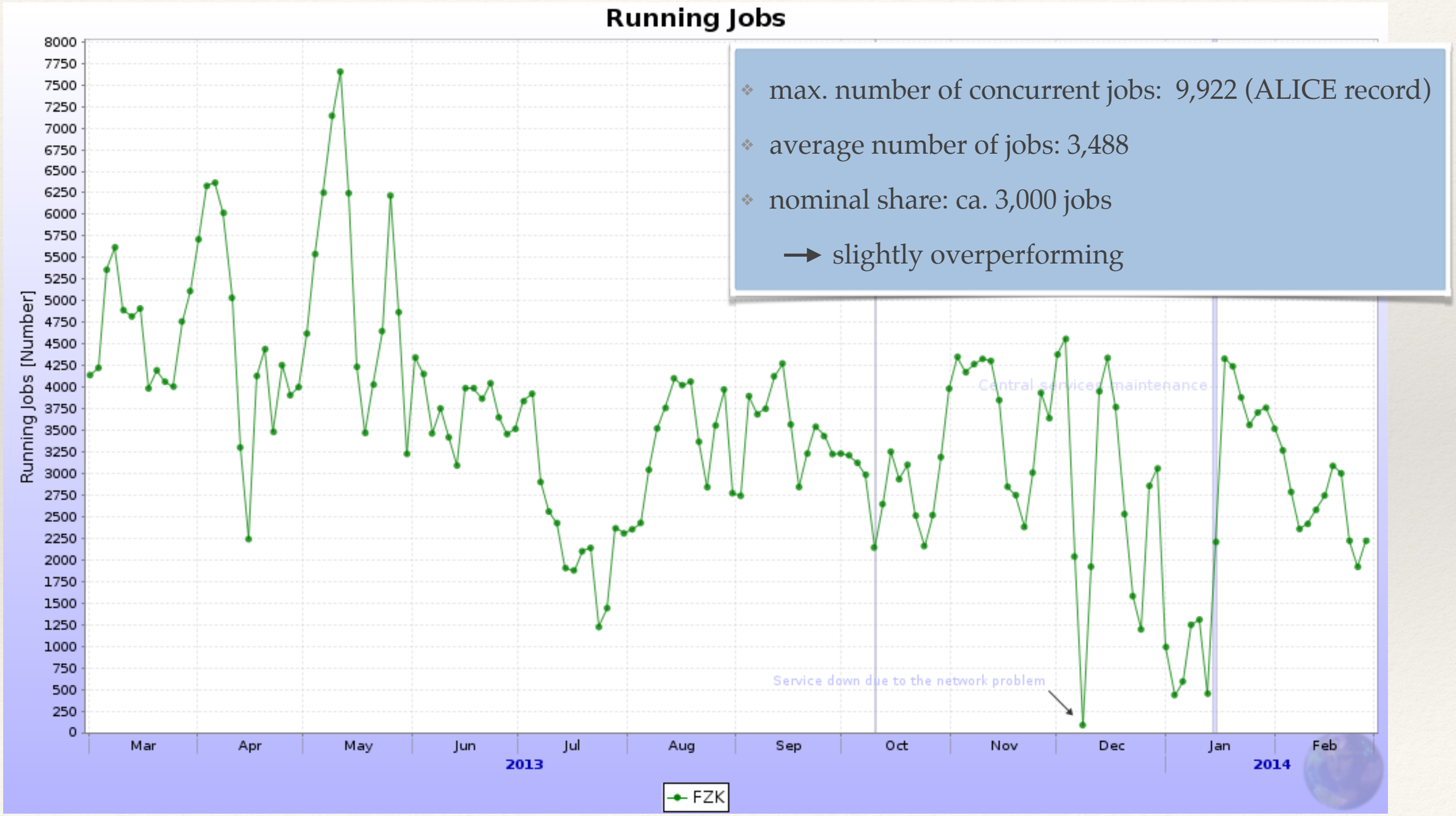
- Founded on October 01, 2009
  - merger of Forschungszentrum Karlsruhe (FZK) and Universität Karlsruhe
  - "FZK" is still being used as a site identifier in ALICE
- ~9,500 employees and ~24,000 students
- Budget (2013) : ~800 million Euros

GridKa Tier-1 Center



- Hosted by SCC(Steinbuch Centre for Computing)
- The largest Tier-1 center in WLCG
  - supports all LHC experiments
  - provides ~14% of WLCG resources
  - also supports non-LHC experiments
- Available resources in 2013
  - 135kHS06, 11 PB disk, 17 PB tape storage
  - 30k HEPSPEC'06, 2.7/5.25 PB (disk/tape) for ALICE

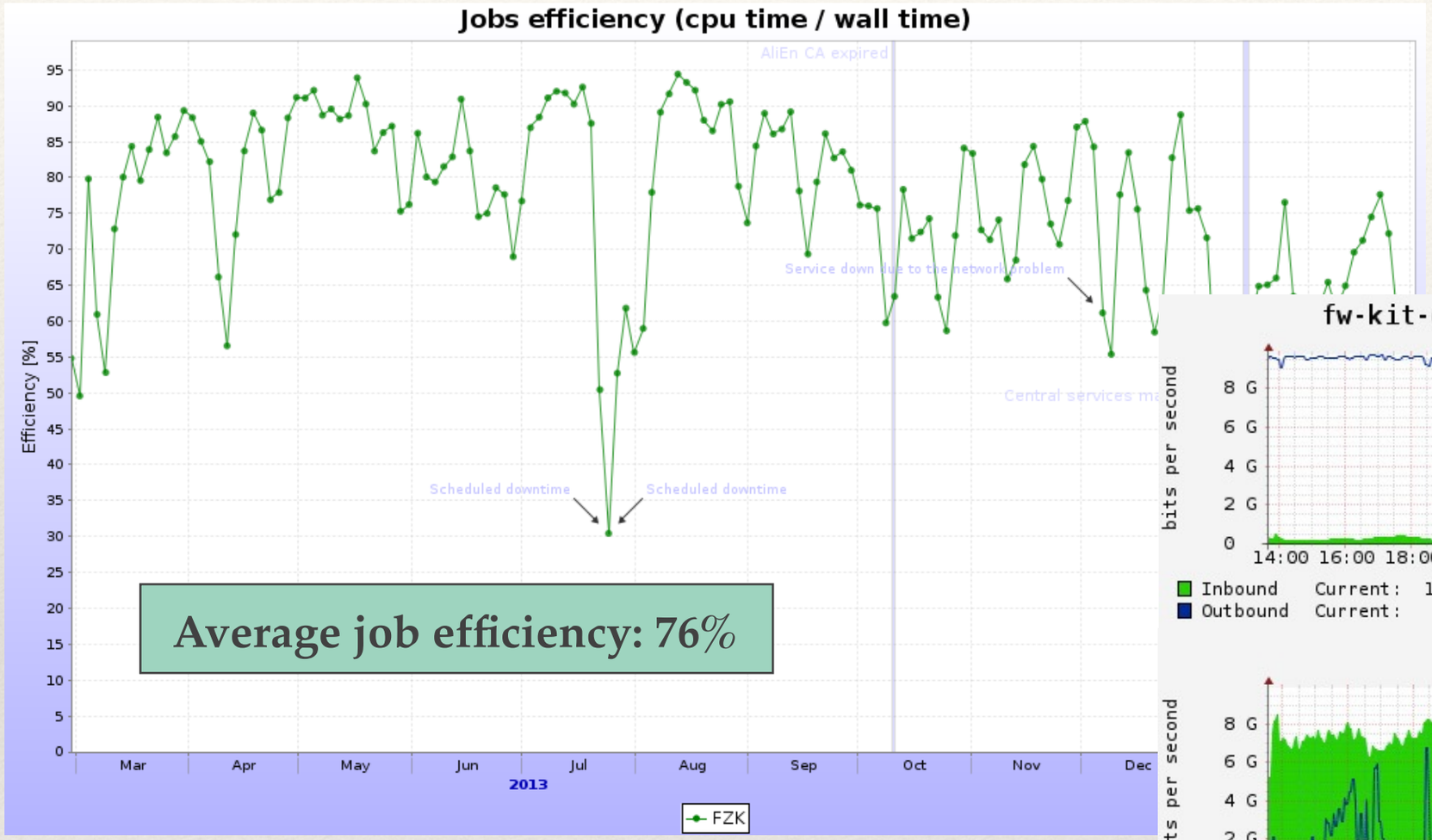
Jobs at GridKa within last year

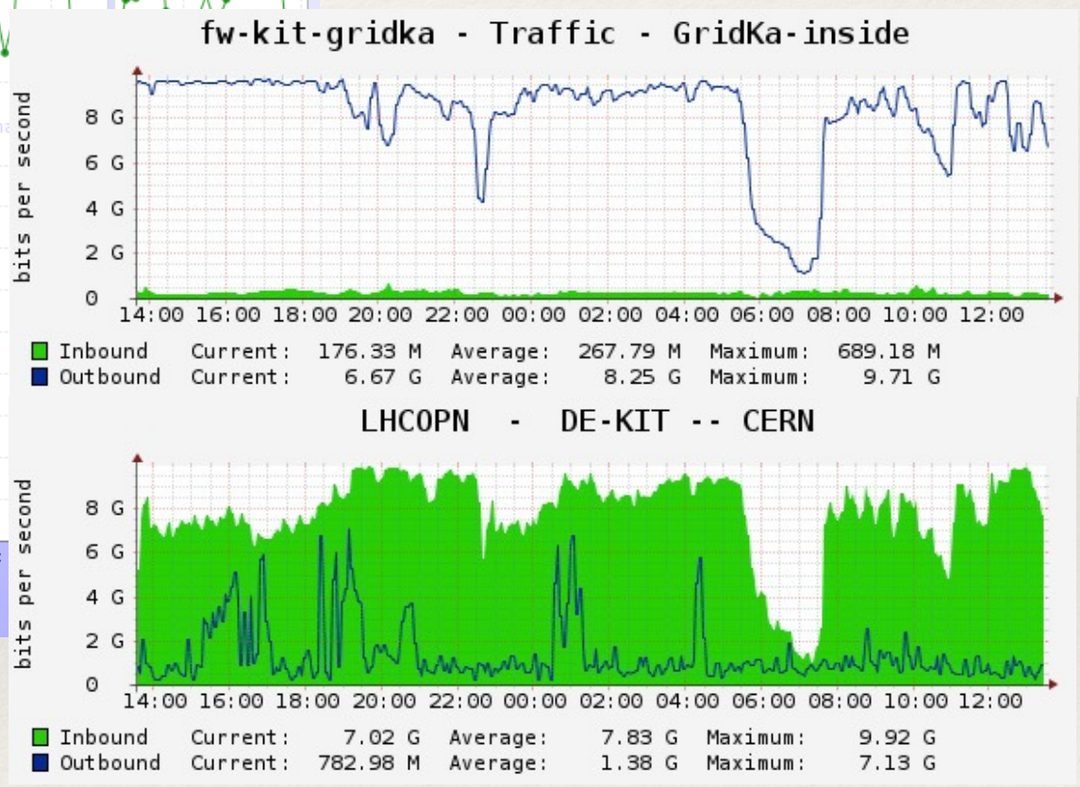


- max. number of concurrent jobs: 9,922 (ALICE record)
- average number of jobs: 3,488
- nominal share: ca. 3,000 jobs

→ slightly overperforming

ALICE Job Efficiency at GridKa

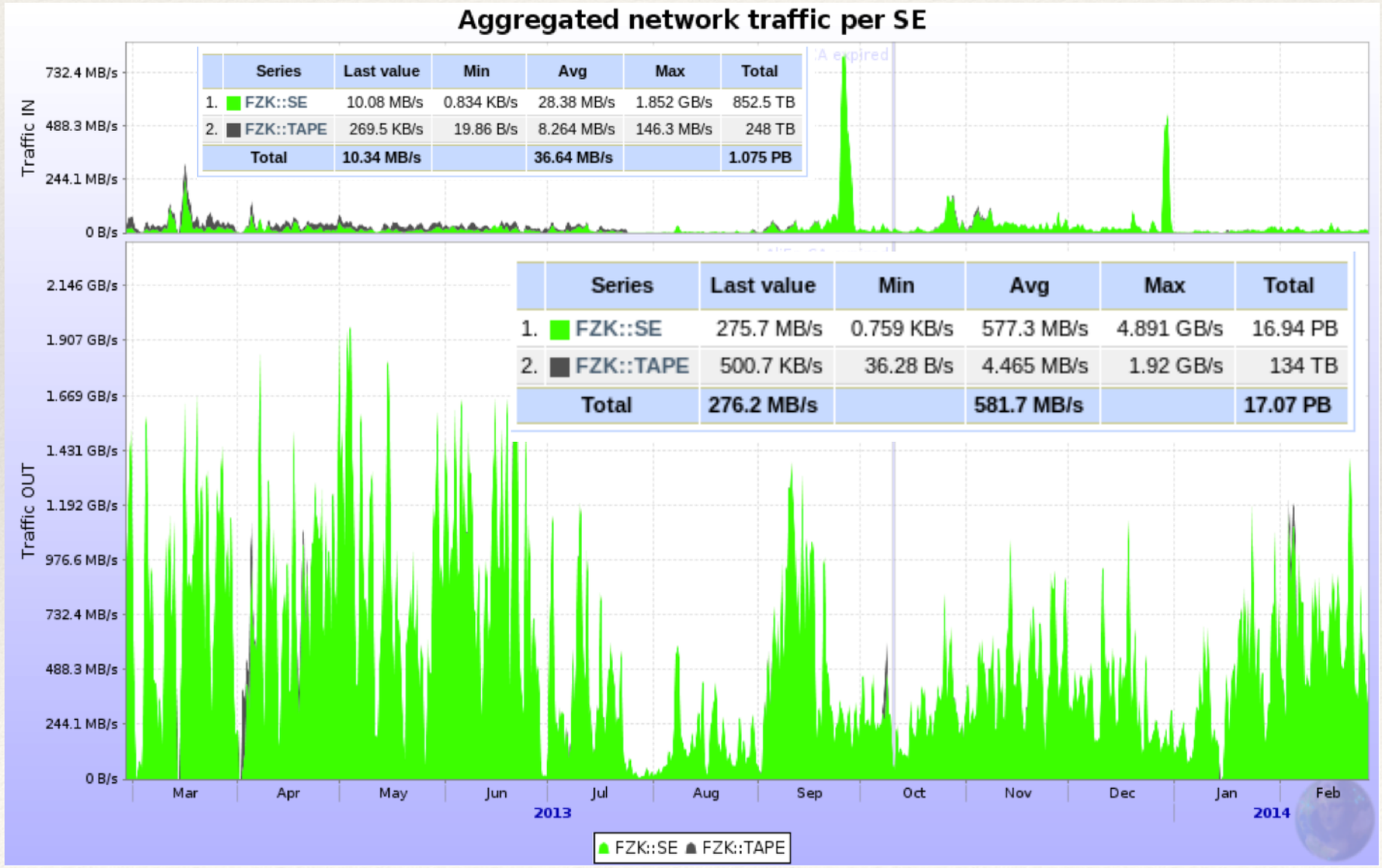




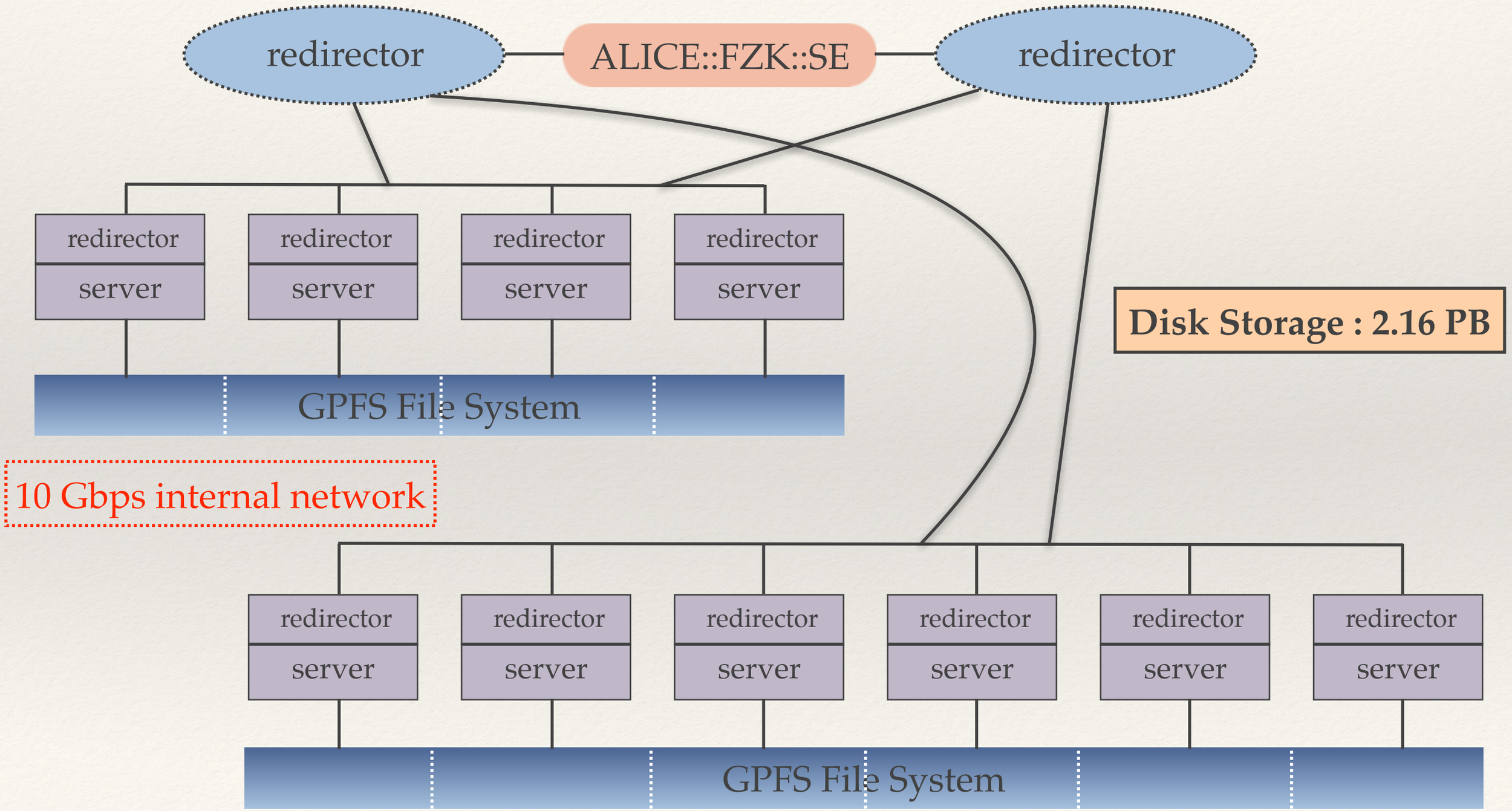
- Firewall upgrade has been done (on 24th of July, 2013)
- New FW is able to cope with 10 Gb/s throughput
- Before that, ALICE jobs reading from remote SEs were affecting other experiments due to limited FW capacity.
- Remote data storage will not affect site performance anymore
- ALICE jobs still read frequently from remote SEs, though, which may bring down the job efficiency

GridKa Storage

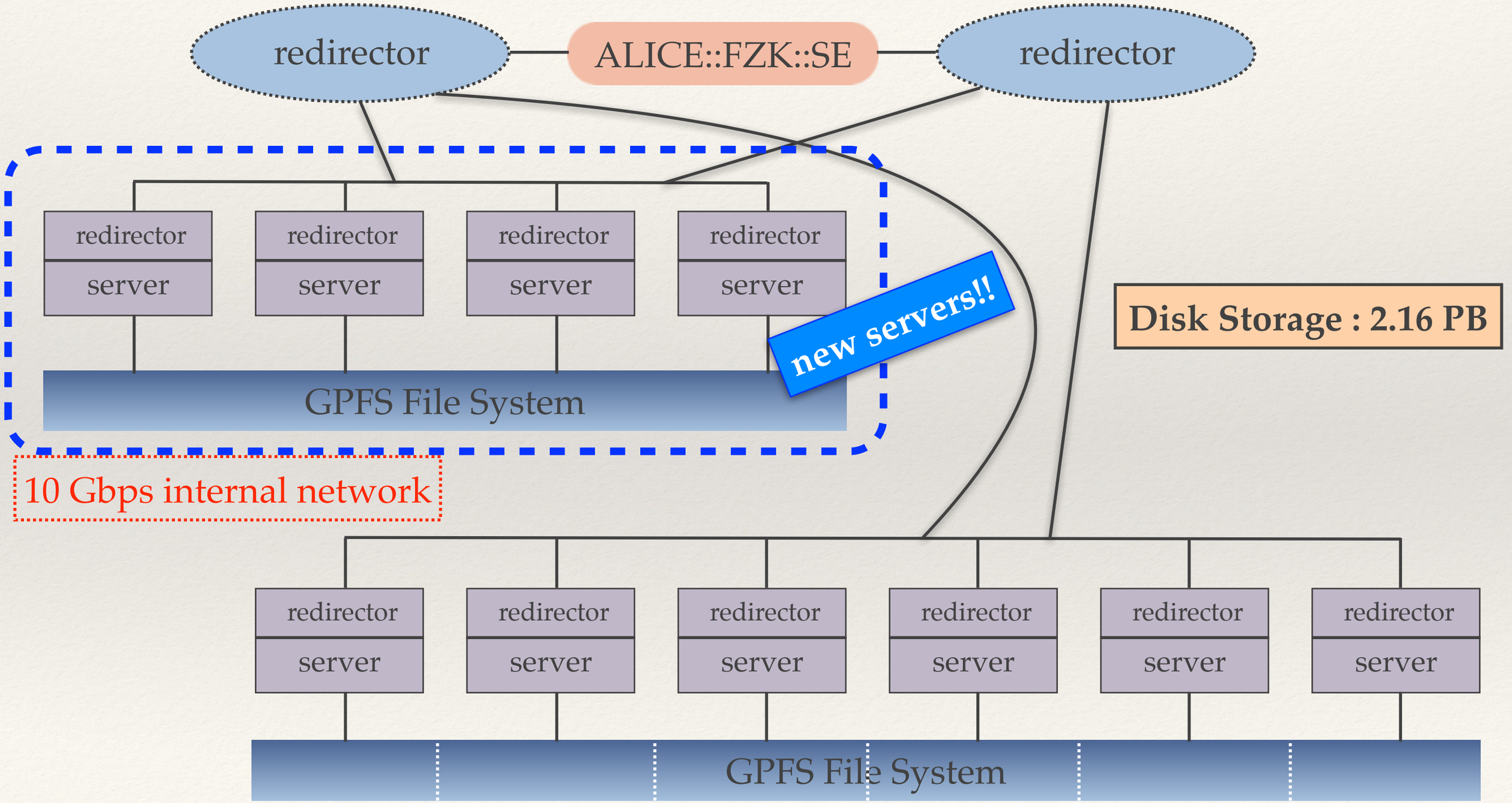
- XRootD based SE works well and is heavily used
- 17 PB of data have been read from FZK::SE last year
- FZK::TAPE is almost not used anymore



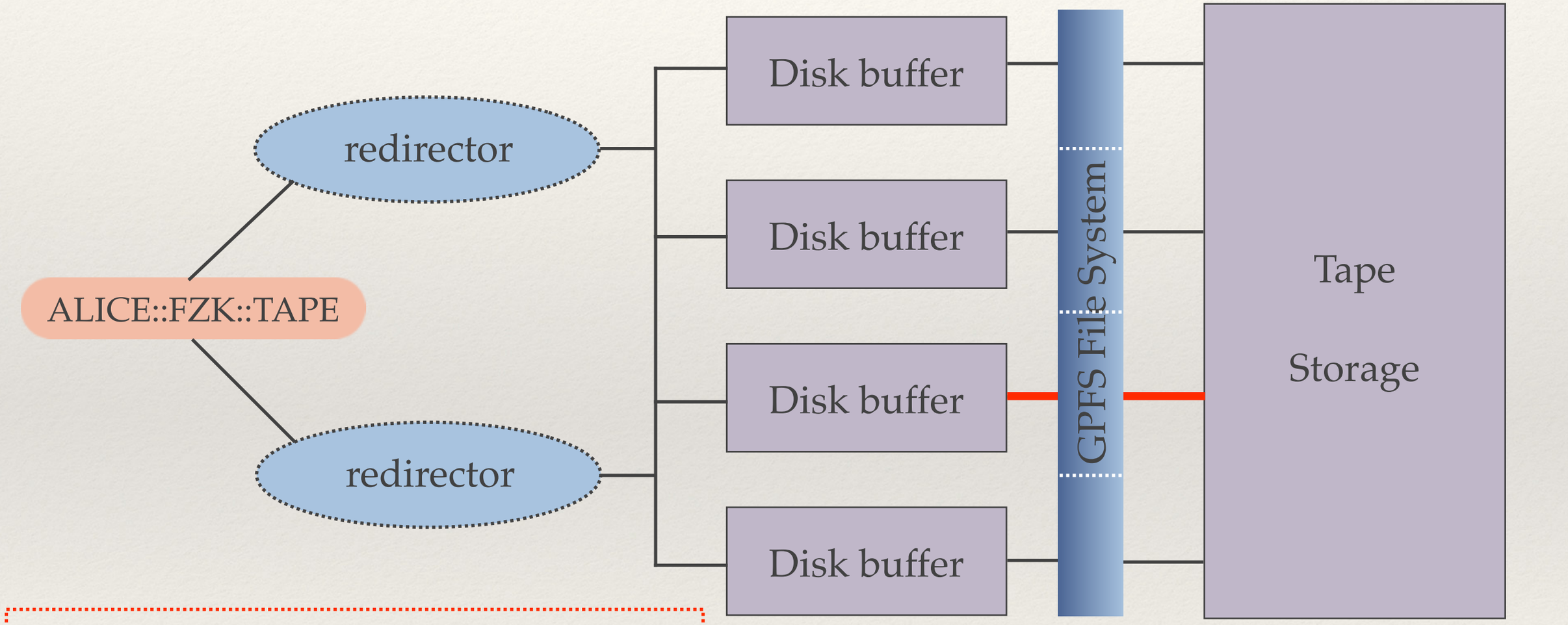
XRootD Architecture at GridKa



XRootD Architecture at GridKa



XRootD Architecture at GridKa



Disk buffer : 660 TB

Tape Storage : 5.25 PB

Various Points of Interest

- IPv6
  - GridKa internal test ongoing (dCache and FTS3)
  - willing to provide a testbed for ALICE services
- ALICE requested trade of tape to disk:
  - No major hardware change planned within the next 2 years
  - This is not feasible at the moment due to limited budget

ALICE Requirements at GridKa

※ 25% of the T1 requirements!

	CPU	DISK	TAPE
	HEPSPEC06	PB	PB
2014	30k	2.7	5.3
2015	30k	3.9	2.3
2016	40k	4.7	3.9
2017	52.5k	5.5	4.9

※ numbers are not yet final

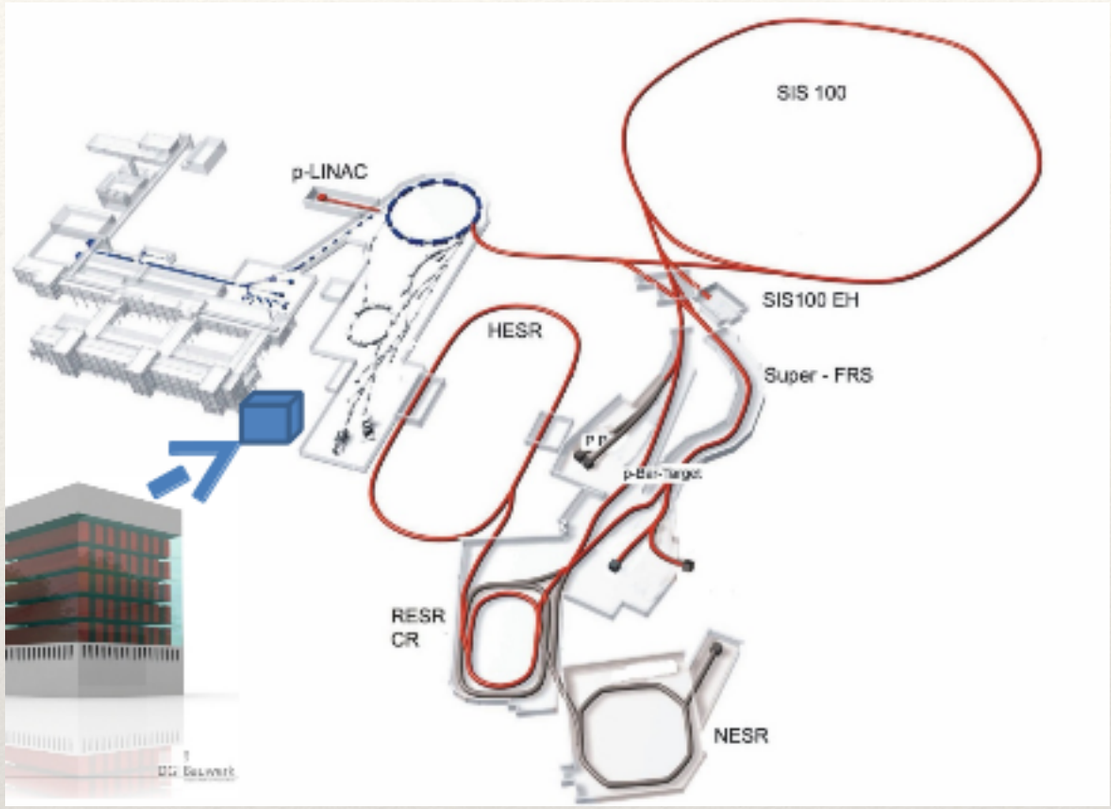
Table of Contents

Gesellschaft für Schwerionenforschung mbH (GSI)

- Dedicated to heavy-ion research
- Located in Darmstadt
- Employs about 1,000 people

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

- **GSI computing today**
  - ALICE T2/NAF/HADES
    - ~10,500 cores
    - ~ 6.7 PB lustre
    - ~ 9 PB archive capacity

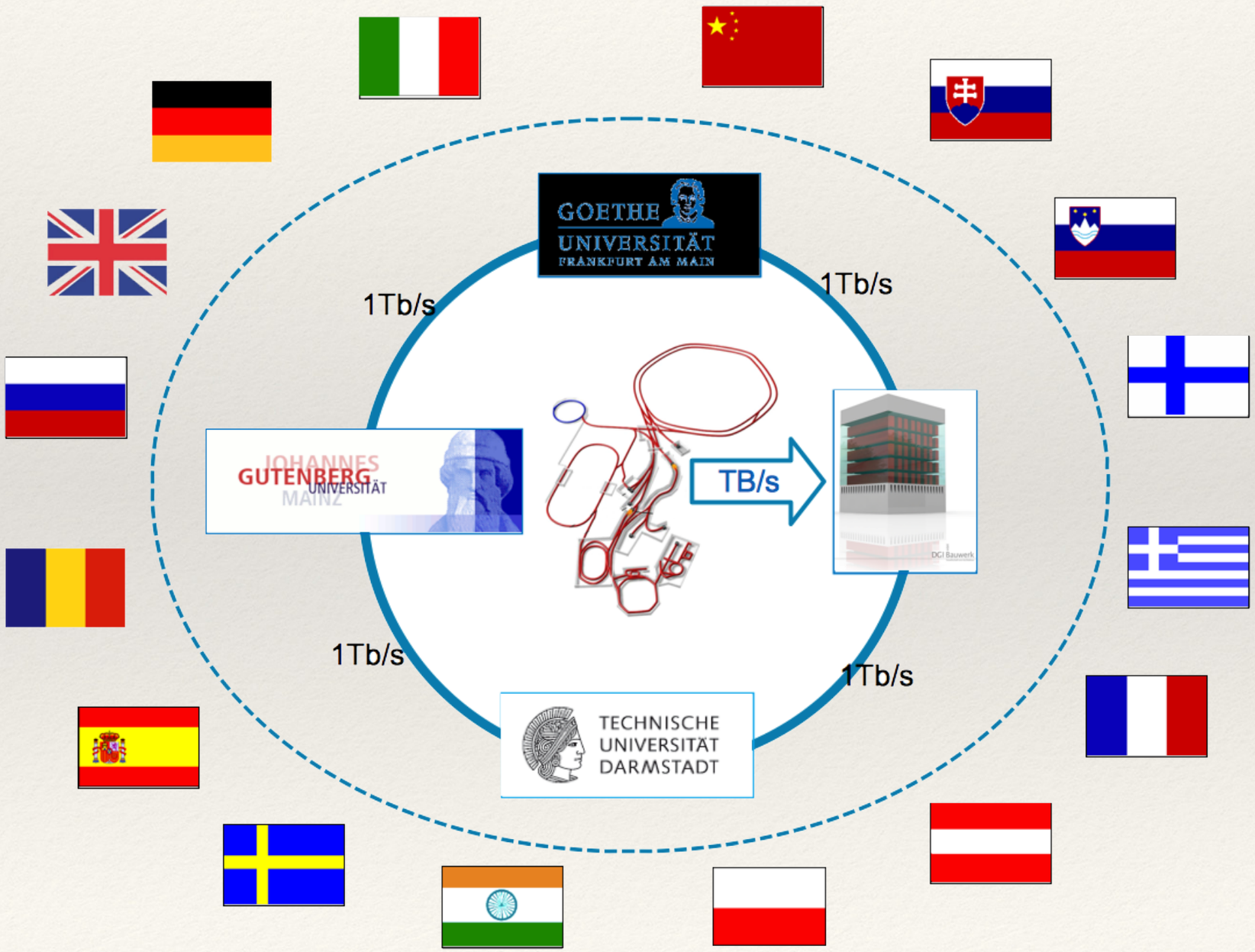


(view of construction site)

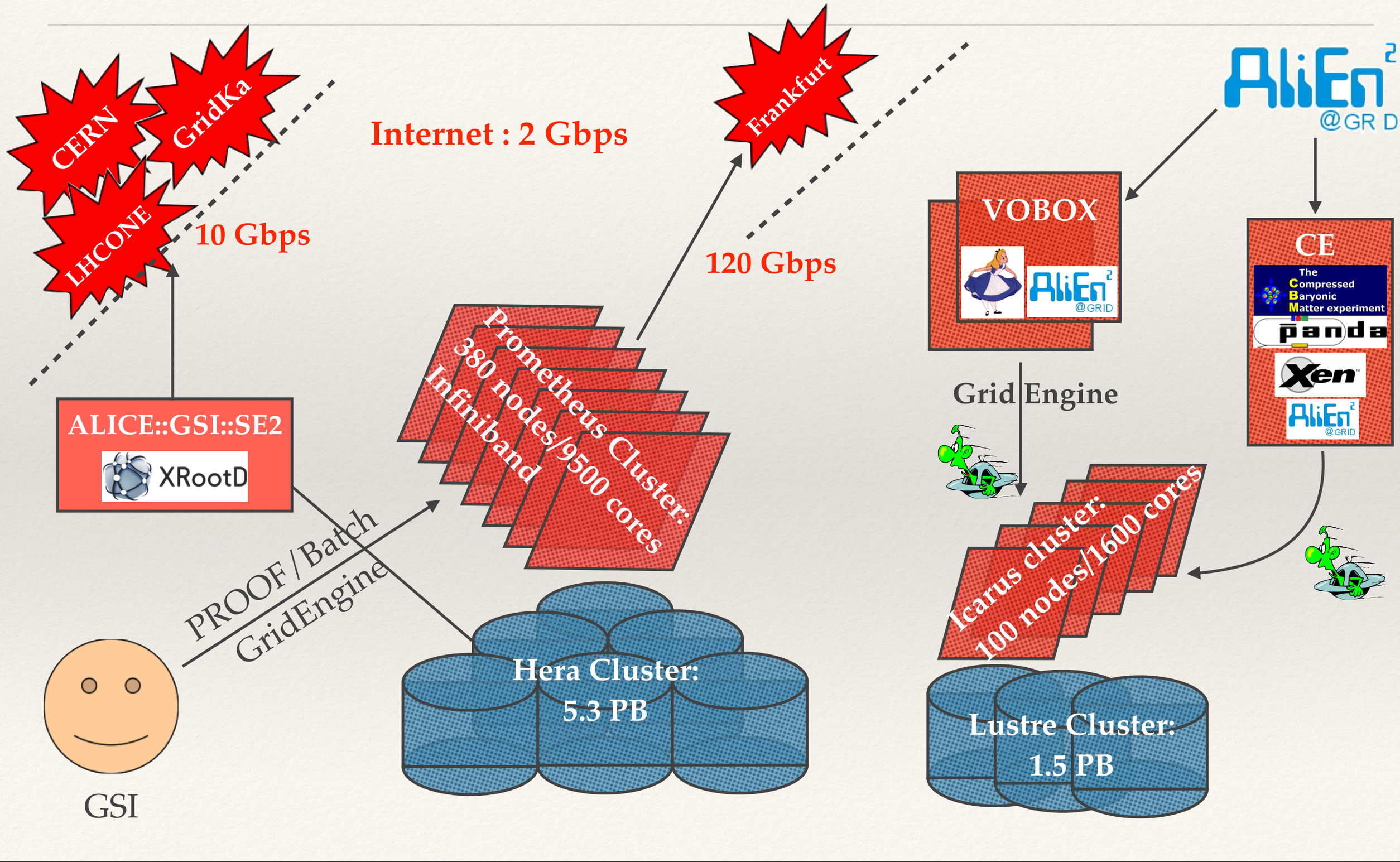
- **FAIR computing 2018**
  - CBM/PANDA/NuSTAR/APPA/LQCD
    - 300,000 cores
    - 40 PB disk
    - 40 PB archive

FAIR Computing: T0/T1 MAN & Grid/Cloud

※ **MAN** : **M**etropolitan **A**rea **N**etwork

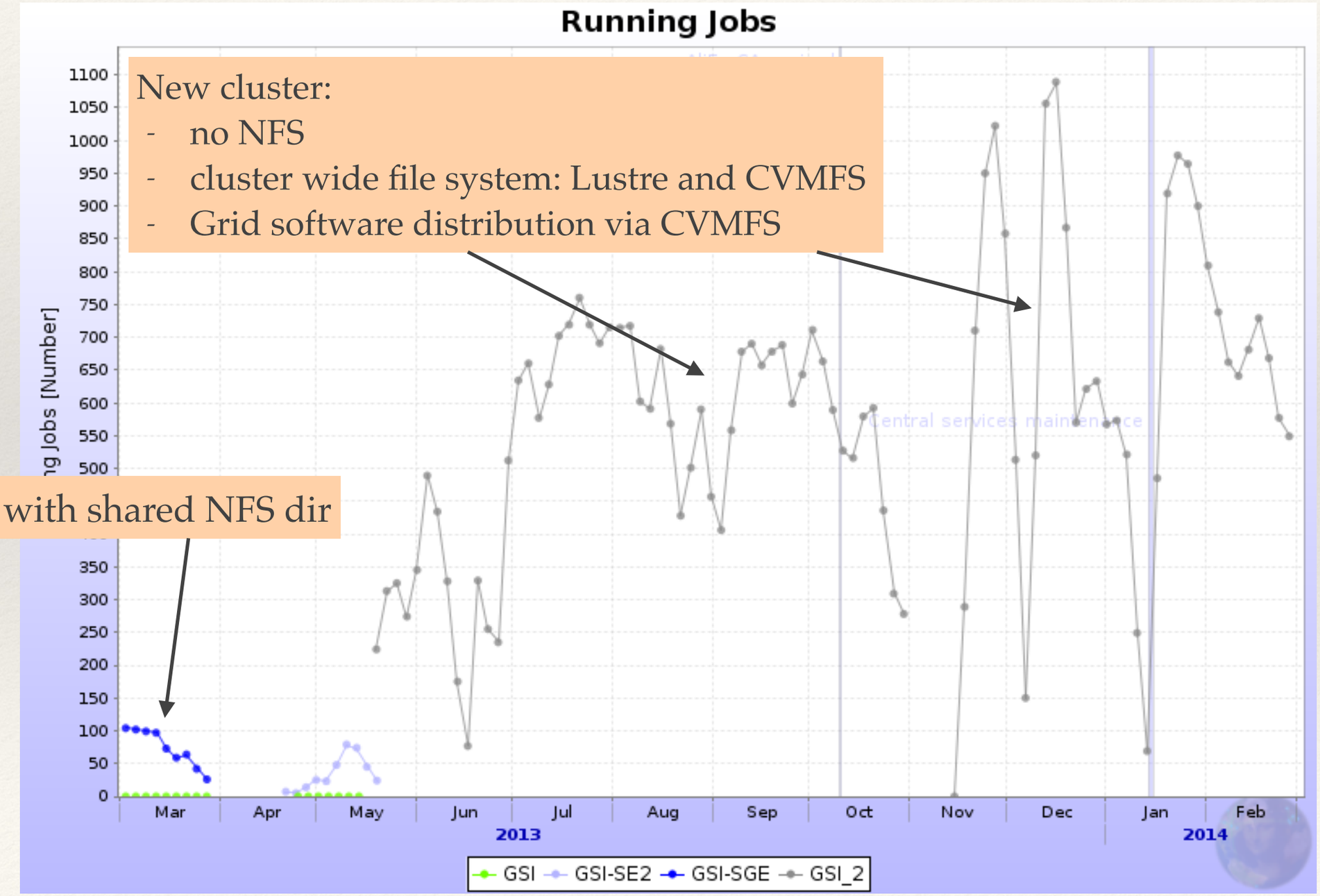


GSI Grid Cluster – Present Status



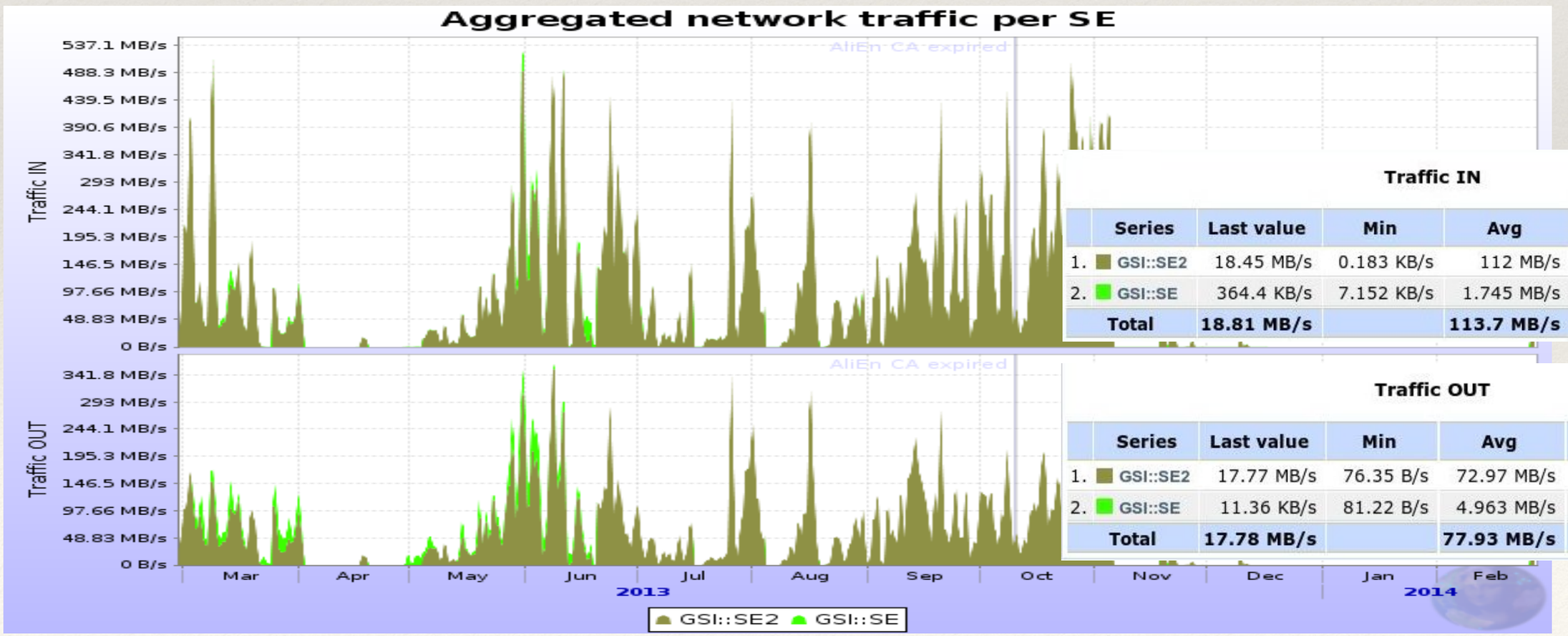
Jobs at GSI within last year

※ software distribution has been switched from BitTorrent to CVMFS on Nov. 2013



GSI Storage Elements

- ALICE::GSI::SE has been phased out in Dec. 2013
- ALICE::GSI::SE2 is now the production SE at GSI
  - 2 × 10 Gb interface connected: one to LHCONE, the other to LNET routers (Lustre cluster mounted)
  - LFN names are also stored as symlinks: usable in a transparent way for local users
- Remaining issue
  - "successful" 0-byte transfer from remote sites (USA, Russia and Asia) to lxalise1: under investigation



Traffic IN

	Series	Last value	Min	Avg	Max	Total
1.	GSI::SE2	18.45 MB/s	0.183 KB/s	112 MB/s	8.703 GB/s	3.284 PB
2.	GSI::SE	364.4 KB/s	7.152 KB/s	1.745 MB/s	215.9 MB/s	42.06 TB
	Total	**18.81 MB/s**		**113.7 MB/s**		**3.325 PB**

Traffic OUT

	Series	Last value	Min	Avg	Max	Total
1.	GSI::SE2	17.77 MB/s	76.35 B/s	72.97 MB/s	6.681 GB/s	2.14 PB
2.	GSI::SE	11.36 KB/s	81.22 B/s	4.963 MB/s	495.2 MB/s	119.6 TB
	Total	**17.78 MB/s**		**77.93 MB/s**		**2.257 PB**

GSI: Next Activities

- Fix pending issues
- Include the big GSI compute farm (Prometheus) in ALICE Grid
  - so far, grid jobs run only on the small cluster with open network connection
  - Prometheus is based on Infiniband and is principally closed to/from the outside world
- Plans for IPv6
  - new firewall will be brought in operation (probably 6 months from now)
  - after that, introduction of IPv6 is planned
    - including all needed tests and planning, at least another 6 months

    ➡ IPv6 not before Spring 2015

ALICE Requirements at GSI

- 20% of the T2 requirements (including NAF)

※ 1 core ~ 12 HEPSPEC06

	CPU		DISK
	HEPSPEC06	# of cores	PB
2014	**7k*** /40k	**583*** /3,333	**0.55*** /2.8
2015	**7k*** /40k	**583*** /3,333	**0.55*** /4.5
2016	48k	4,000	5.2
2017	54k	4,500	6.14

※ numbers are not yet final

***: pledged resources for GSI T2**

Table of Contents

HHLR_GU

(Hessisches Hochleistungsrechenzentrum Goethe Universität)

Center for Scientific Computing Frankfurt

Excellence in High Performance Computing

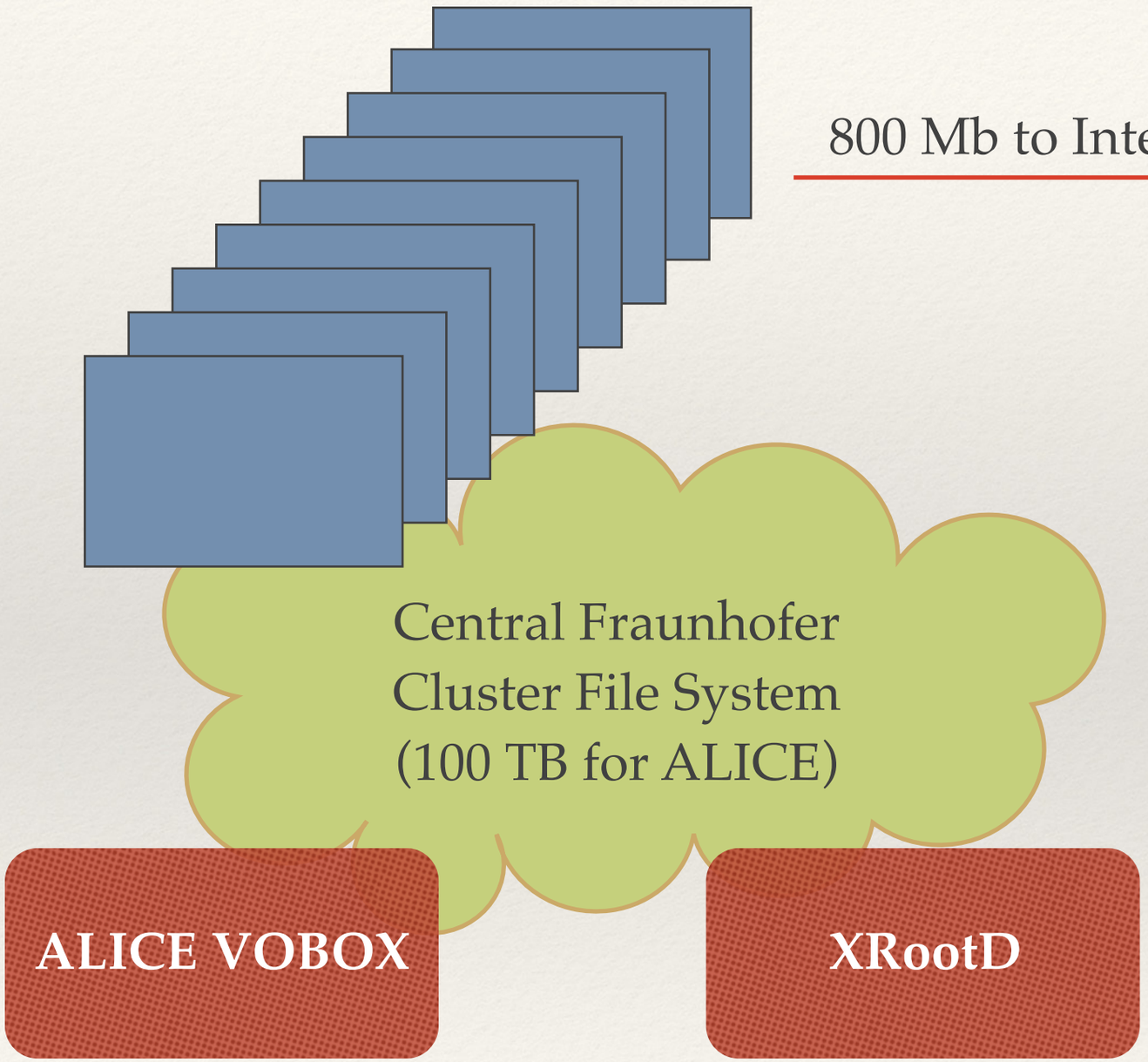
GOETHE UNIVERSITÄT FRANKFURT AM MAIN

- CSC Home
- CSC Clusters
  - LOEWE-CSC
    - Quickstart
  - FUCHS
  - SCOUT
  - Ancient Clusters
  - Access
- Master Program
- Research Groups
- People @ CSC

CPU/GPU cluster "LOEWE-CSC"

- Cluster Performance:
  - CPUs performance (dp): 176 TFlop/s (peak)
  - GPUs performance (sp): 2.1 PFlop/s (peak)
  - GPUs performance (dp): 599 TFlop/s (peak)
  - **Cluster performance HPL: 299.3 TFlop/s**
  - **Energy efficiency Green500: 740.78 MFlop/s/Watt**
- Hardware:
  - 832 nodes in 34 water-cooled racks,
  - 20,928 CPU cores plus 778 GPGPU hardware accelerators,
  - 56 TB RAM and over 2 PB aggregated disk capacity,
  - QDR InfiniBand interconnects,
  - parallel scratch filesystem with a capacity of 764 TB and an aggregated bandwidth of 10 GB/s.
- Installed in late 2010 on Industriepark Höchst.

HHLR_GU: ALICE Setup (up to now…)

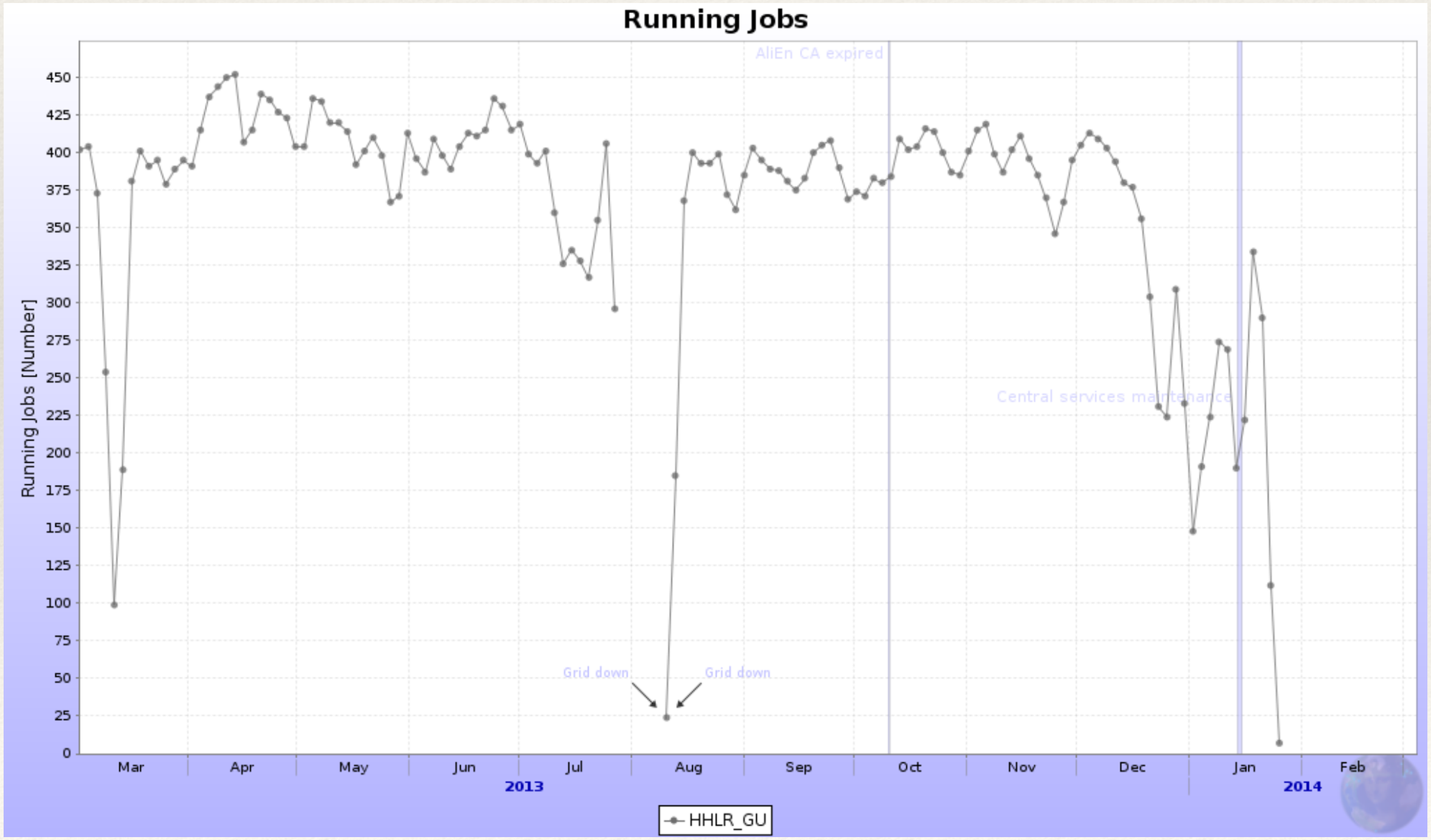


- 1,200 cores (now 600) exclusively for ALICE
- Job Submission System: Slurm
- for this the native AliEn Slurm interface has been reactivated (A. Montiel Gonzalez)

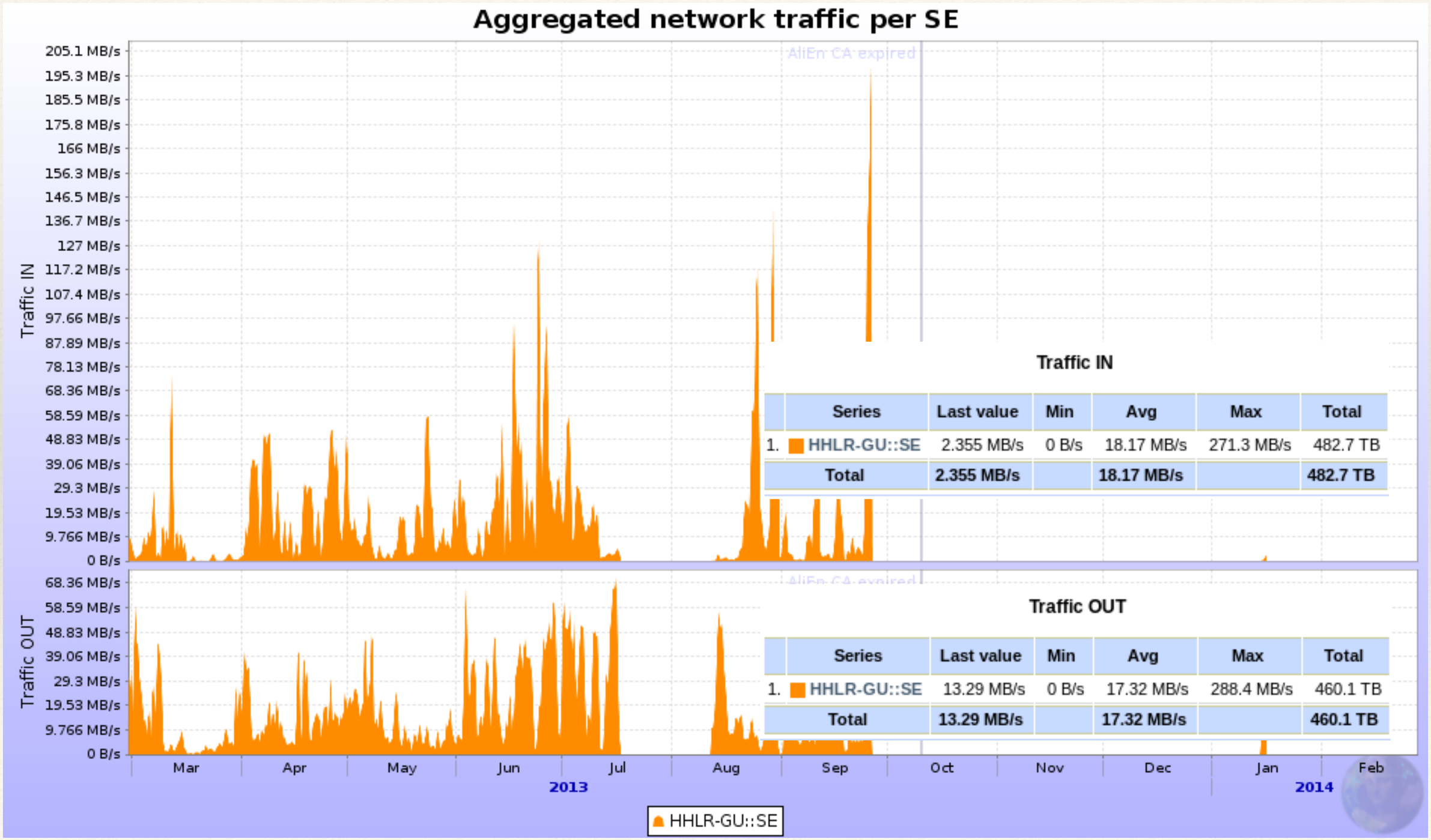
Currently in Maintenance!!

Jobs at Loewe CSC

- Stable operation until January 2014
- Computing centre is being restructured
- HHLR-Loewe CSC will be in maintenance up to May 2014 !!!



Storage at Loewe CSC



HHLR-GU: Status and Plans

- Currently HHLR-GU is undergoing major restructuring
  - new nodes will be added and storage will be restructured
  - maintenance up to May 2014
- Up to then plan will be made
  - if and how the centre will be re-integrated into ALICE Grid
- Contribution of HHLR-GU is voluntary
  - goes on top of resources pledged by GSI
- No plans for IPv6 yet

Table of Contents

Summary

- German sites provide a valuable contribution to ALICE Grid
- Operation stable and continuous.
  - thanks to the centres and to the local teams !!
- New developments are on the way
- FAIR will play an increasing role (funding, network architecture, software development and more ...)