

COMPUTING FOR ALICE IN THE CZECH REPUBLIC in 2015/2016

D. Adamova,

with the help and support of :

J. Chudoba, J.Svec, V. Rikal, A. Mikula, M. Adam and J. Hampl

Outline

Status of the WLCG Tier-2 site in Prague

ALICE operations in 2015/2016

Report on the delivery of mandatory resources

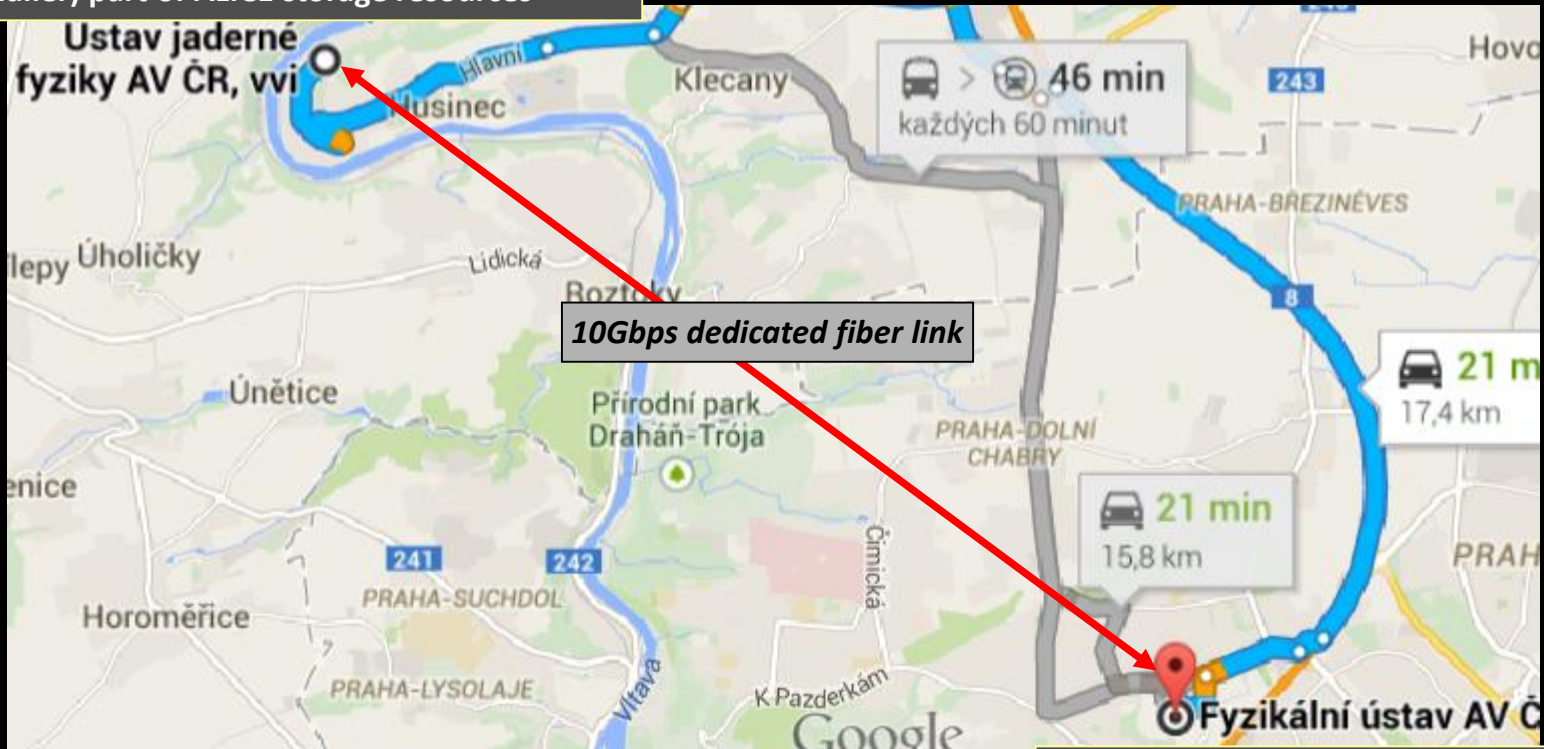
What is really mandatory?

A couple of various issues

Summary and Outlook

The geographical layout

Nuclear Physics Institute AS CR (NPI)
a large ALICE group, no ATLAS involvement
A (smaller) part of ALICE storage resources



10Gbps dedicated fiber link

Originally, the xrootd storage servers for ALICE were only at NPI. Subsequently, xrootd servers were also put in operation at FZU and now even the redirector is there.

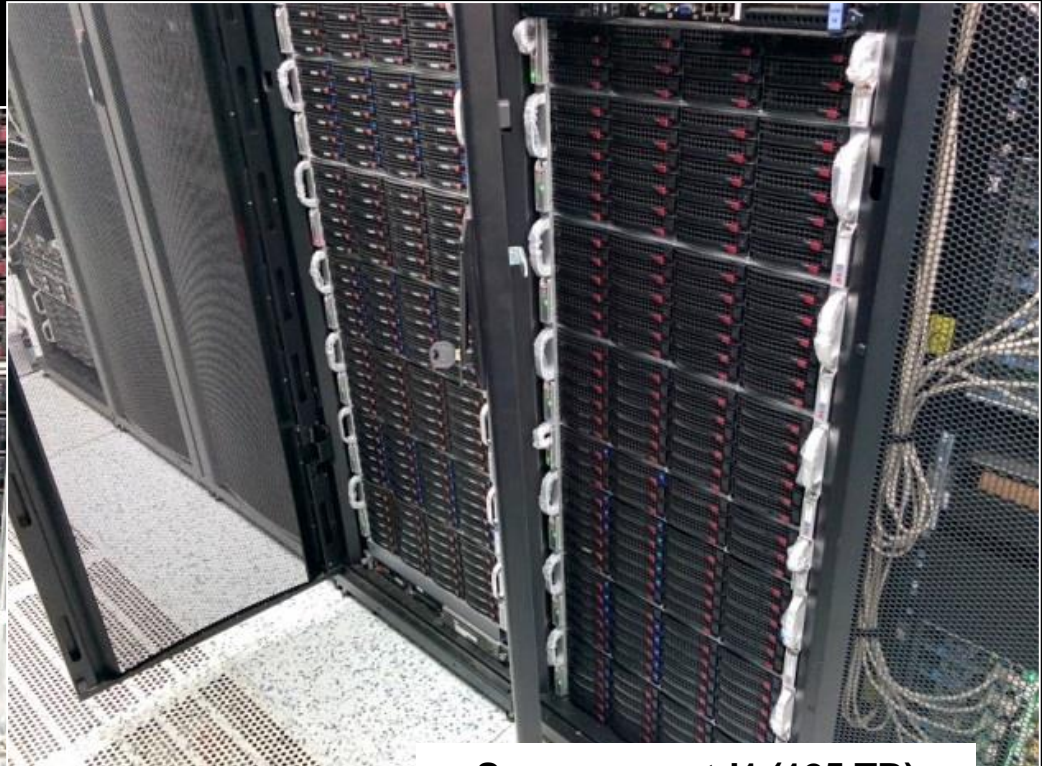
Institute of Physics AS CR (FZU)
Regional computing center
WLCG Tier-2 site praguelcg2
All the CPU resources for ALICE
A (larger) part of storage resources for ALICE
Quite small ALICE group, much larger ATLAS community

HEP Computing in Prague: site praguelcg2 (a.k.a. the farm GOLIAS)

- **A national computing center for processing data from various HEP experiments**
 - Located in the Institute of Physics (FZU) in Prague
 - Basic infrastructure already in 2002, but officially started in 2004
- **Certified as a Tier2 center of LHC Computing Grid (praguelcg2)**
 - Collaboration with several Grid projects.
- **April 2008, WLCG MoU signed by Czech Republic (ALICE+ATLAS).**
- **Very good network connectivity:** Multiple dedicated 1 – 10 Gb/s connections to collaborating institutions. 10 Gb/s connection to LHCONE.
- **Provides computing services** for ATLAS + ALICE, D0, Solid state physics, Auger, Nova, CTA ...

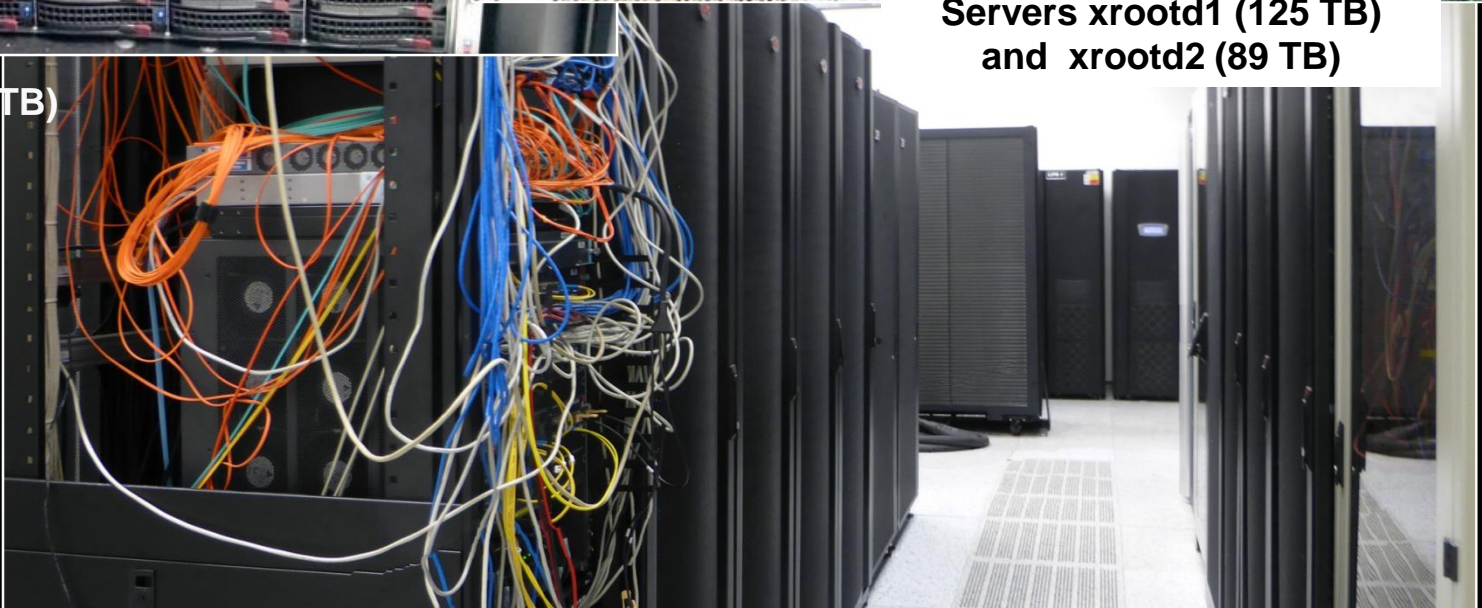


Some pictures



**Servers xrootd1 (125 TB)
and xrootd2 (89 TB)**

Server xrootd5 (422 TB)



Current numbers

- 1 batch system (torque + maui)
- 2 main WLCG VOs: ALICE, ATLAS
 - FNAL's D0 (dzero) user group
 - Other VOs: Auger, Nova, CTA
- 258 WNs, 4756 CPU cores, published in the Grid
- 4.25 PB on disk storage (DPM, XRootD, NFS), a tape library
- Regular yearly upscale of resources on the basis of various financial supports, mainly the academic grants.
- The WLCG services include:
 - Apel publisher, Argus Authorization service, BDII, several UIs, Alice VOBOX, Cream CEs, Storage Elements, lcg-CA

Monitoring: Munin, Ganglia, Nagios w/multisite, MRTG, Netflow

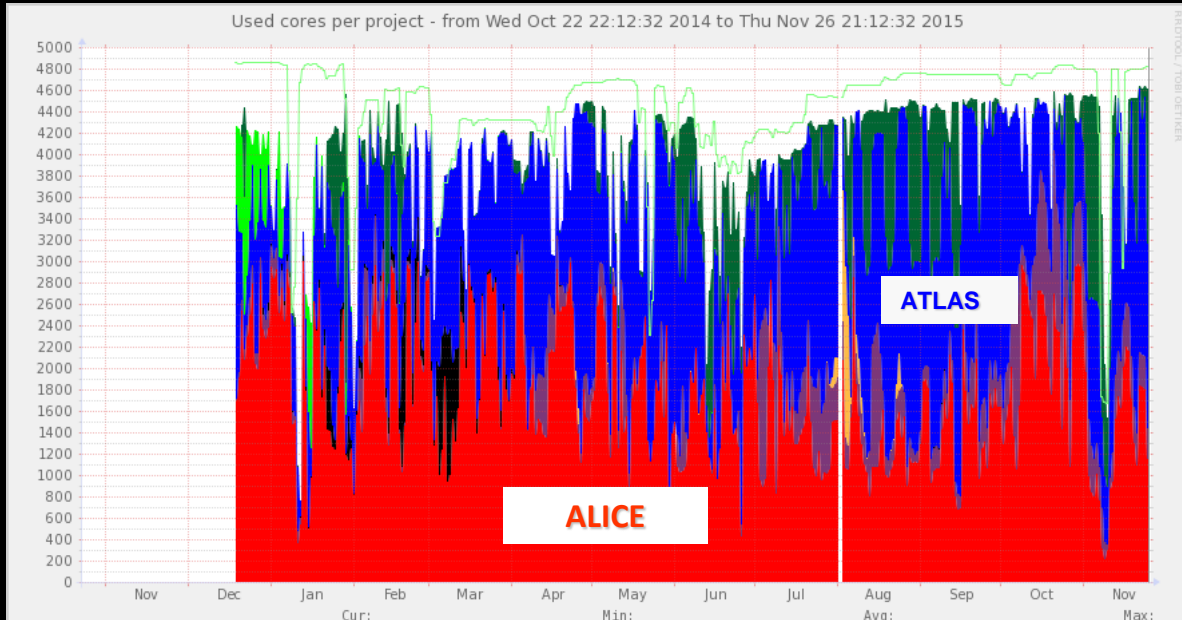
The use of virtualization at the site is quite extensive.

- ALICE disk XRootD Storage Element ALICE::Prague::SE
 - ~ 1.6 PB of disk space in total
 - Redirector+ 5 clients @ FZU, 4 clients @ NPI Rez
 - → a distributed storage cluster
 - 10 Gb/s dedicated connection from FZU to the NPI ALICE storage cluster

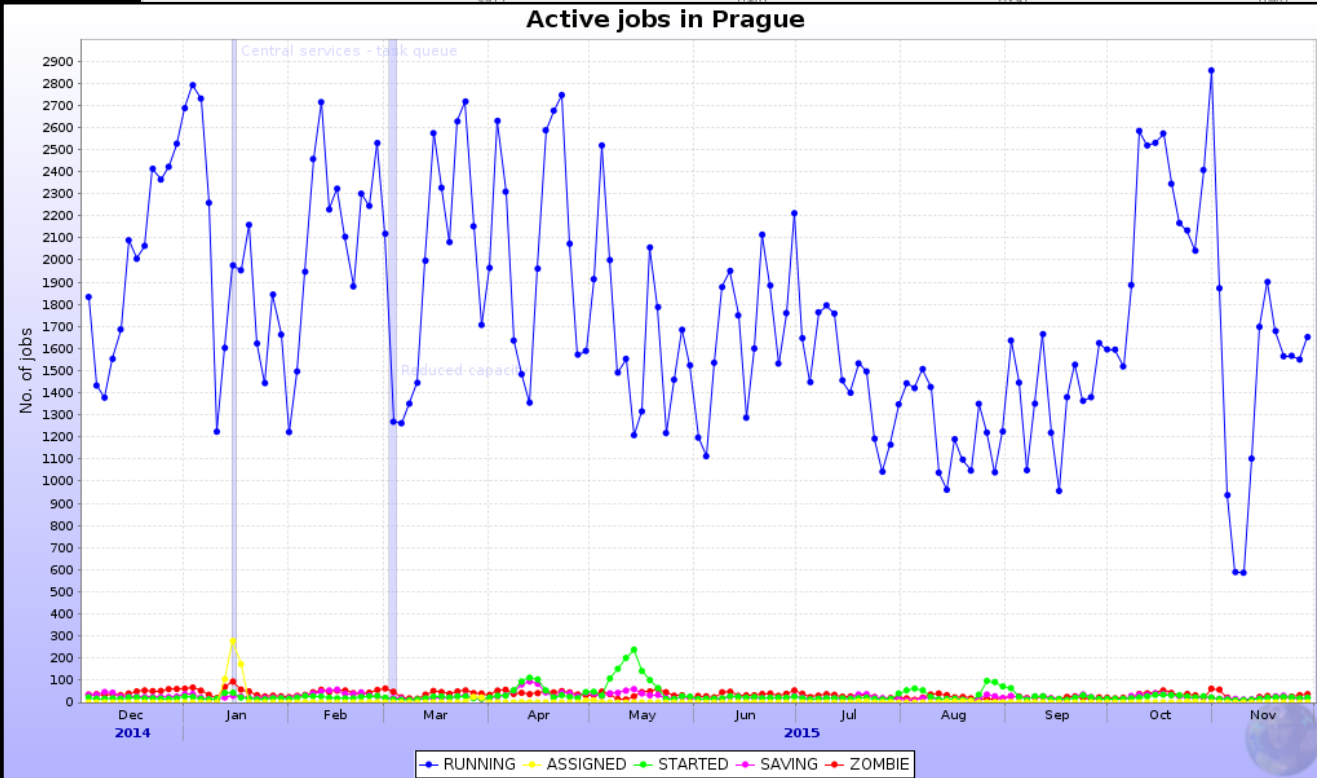


ALICE operations in 2015

The number of concurrently running ALICE jobs at FZU in 2015 was in average ~1370, in peaks over 3300. Due to a steady operation and active usage of idle resources, ALICE was the largest CPU consumer in 2015.

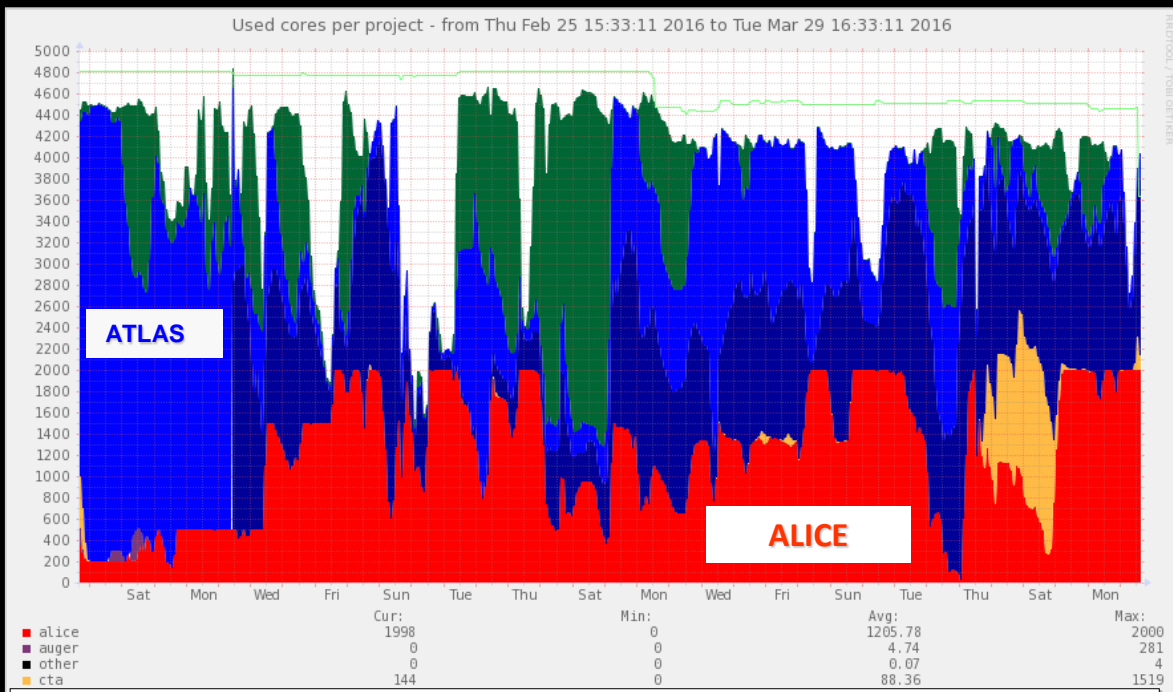


Running jobs profile as shown in ALICE MonALISA. Almost 5.5 million of jobs was processed in Prague during 2015.

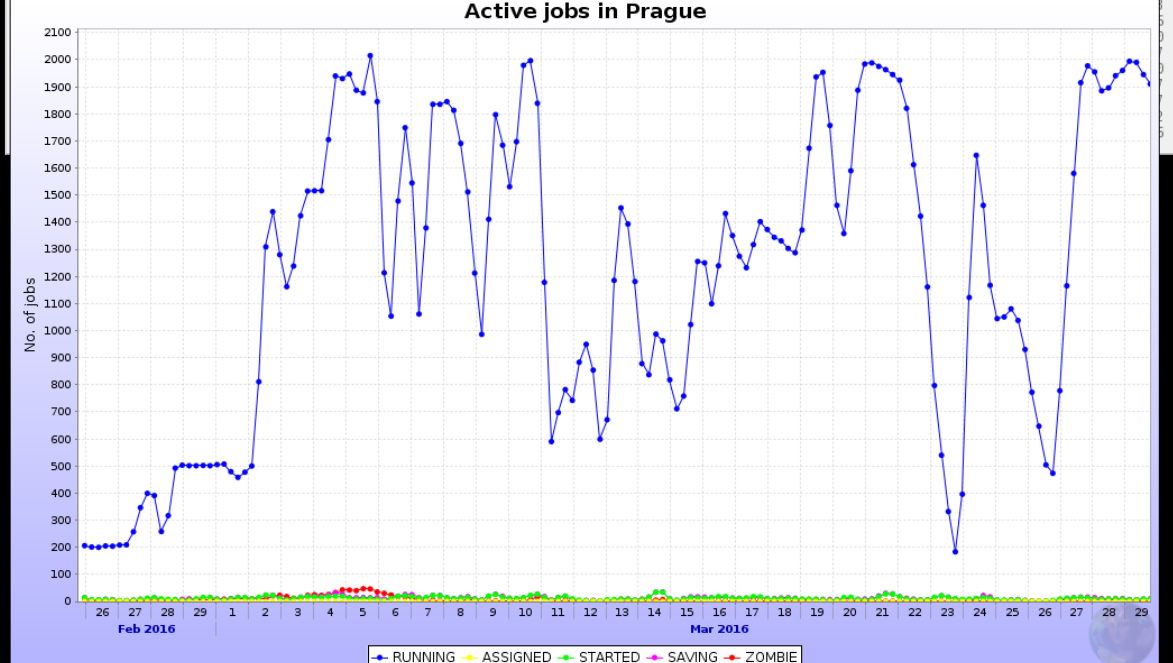


ALICE operations in 2016

Running jobs profile at the Prague Tier-2 in March 2016. The ALICE share is lower than in 2015.



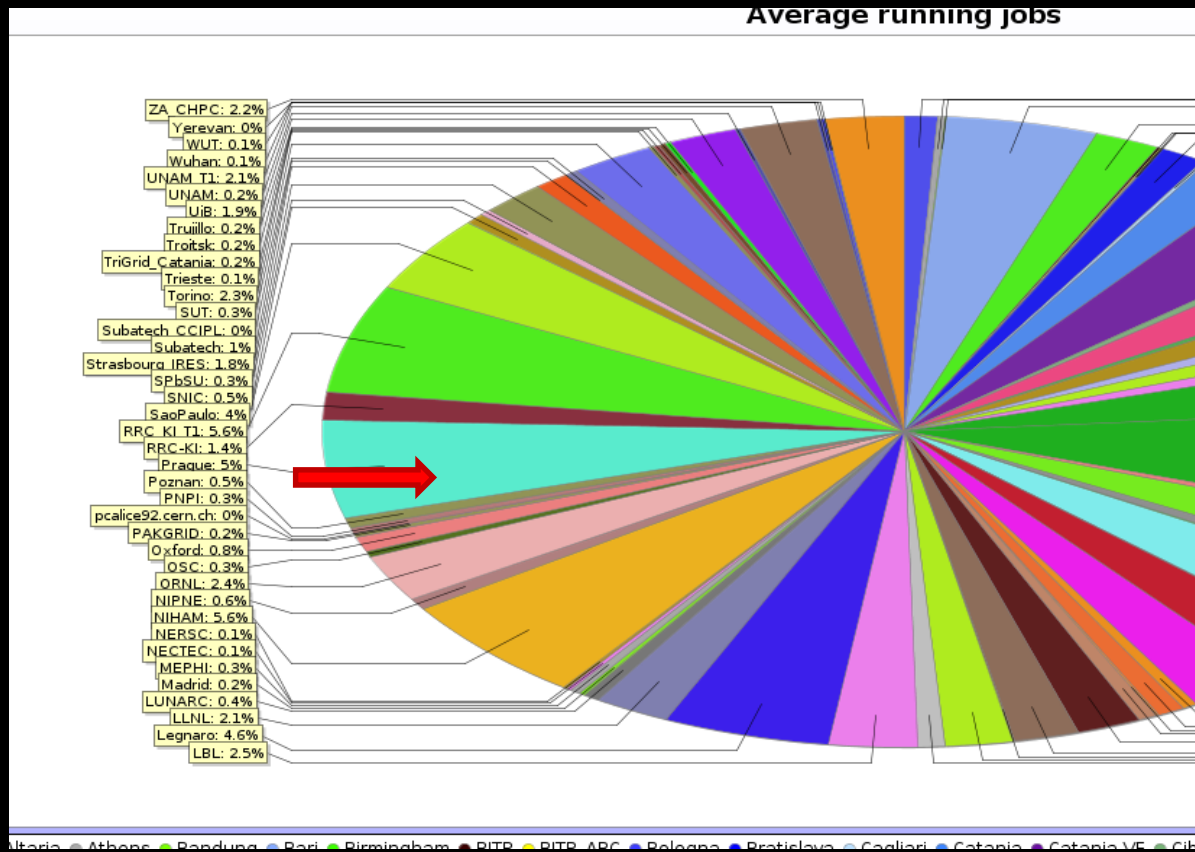
Running jobs profile as shown in ALICE MonALISA. The number of concurrently running ALICE jobs at FZU in March 2016 was in average ~1200, in peaks ~2100. This will very likely drop down later in 2016 due to problems discussed in next slides.



CPU resources share & site availability in 2015

During 2015, there was a continuous uninterrupted ALICE data processing at the computing center at FZU.

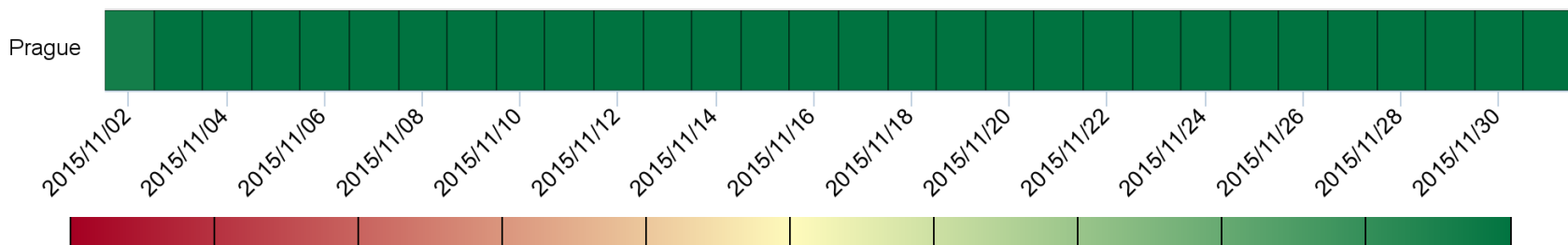
The Czech Republic share in the Tier-2 CPU resources delivery in 2015 was ~5%.



100% availability in November 2015

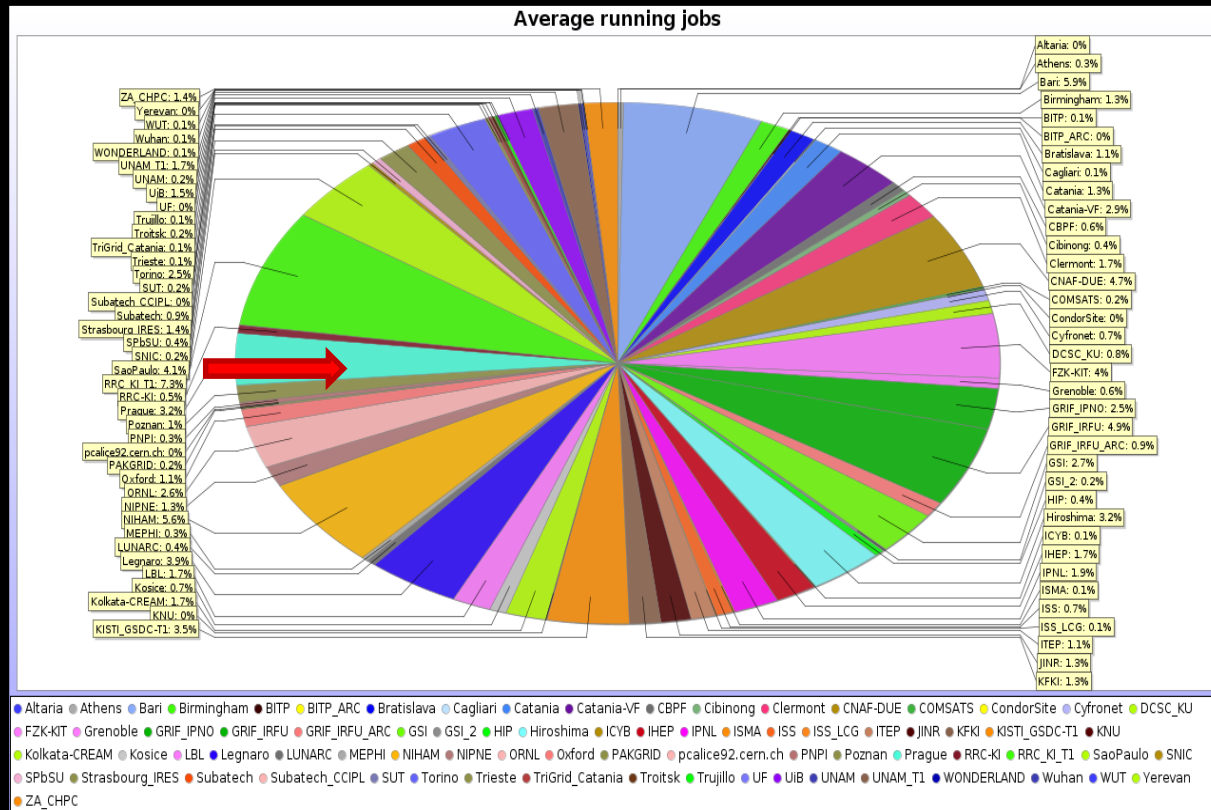
Site Availability using ALICE_CRITICAL

720 hours from 2015/11/02 to 2015/12/01

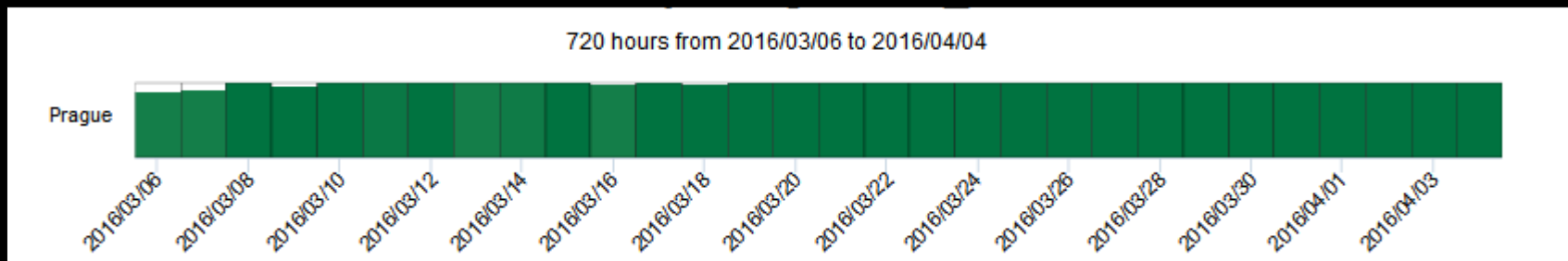


CPU resources share & site availability in 2016

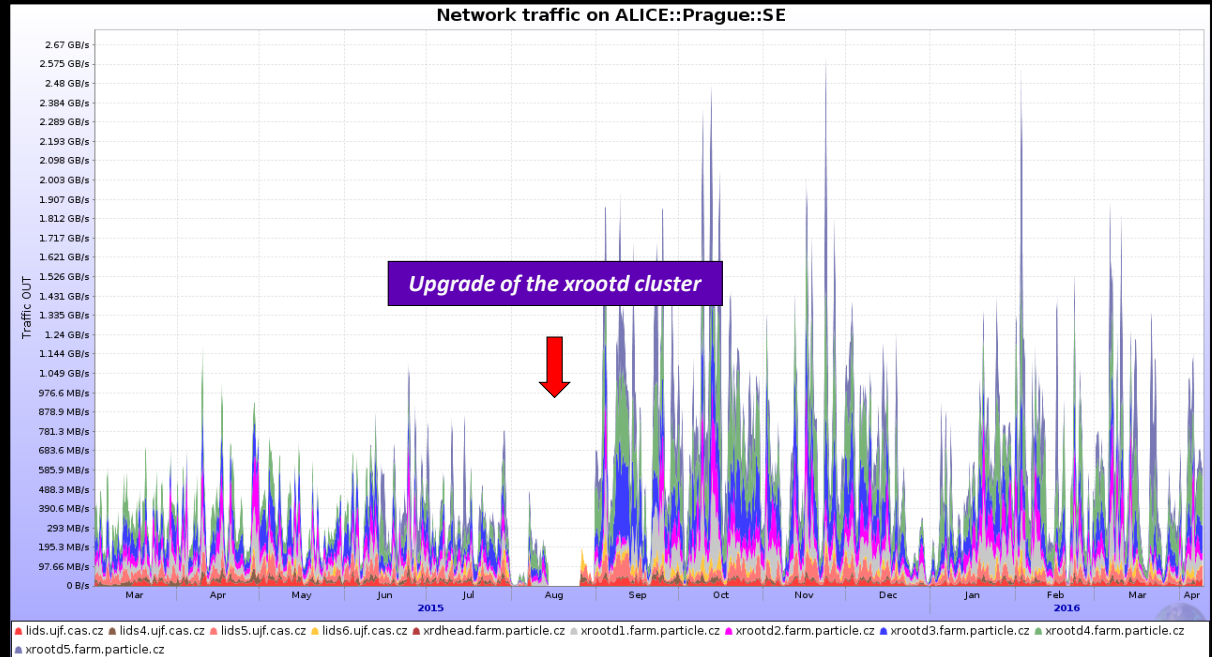
During the first months of 2016, the Czech Republic share in the Tier-2 CPU resources delivery dropped to ~3.2%.



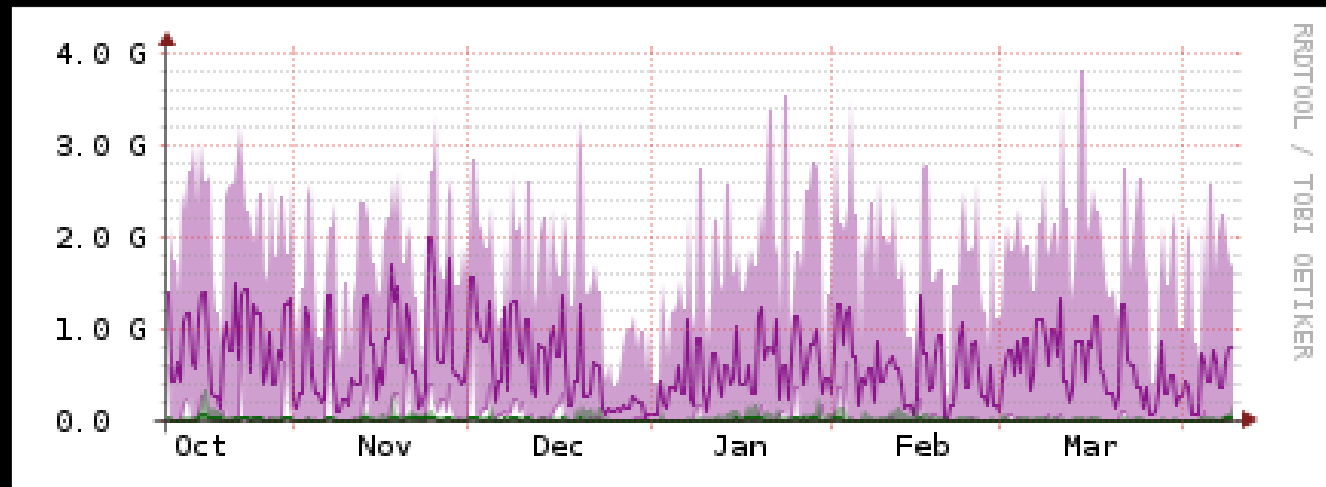
Almost 100% availability in March 2016



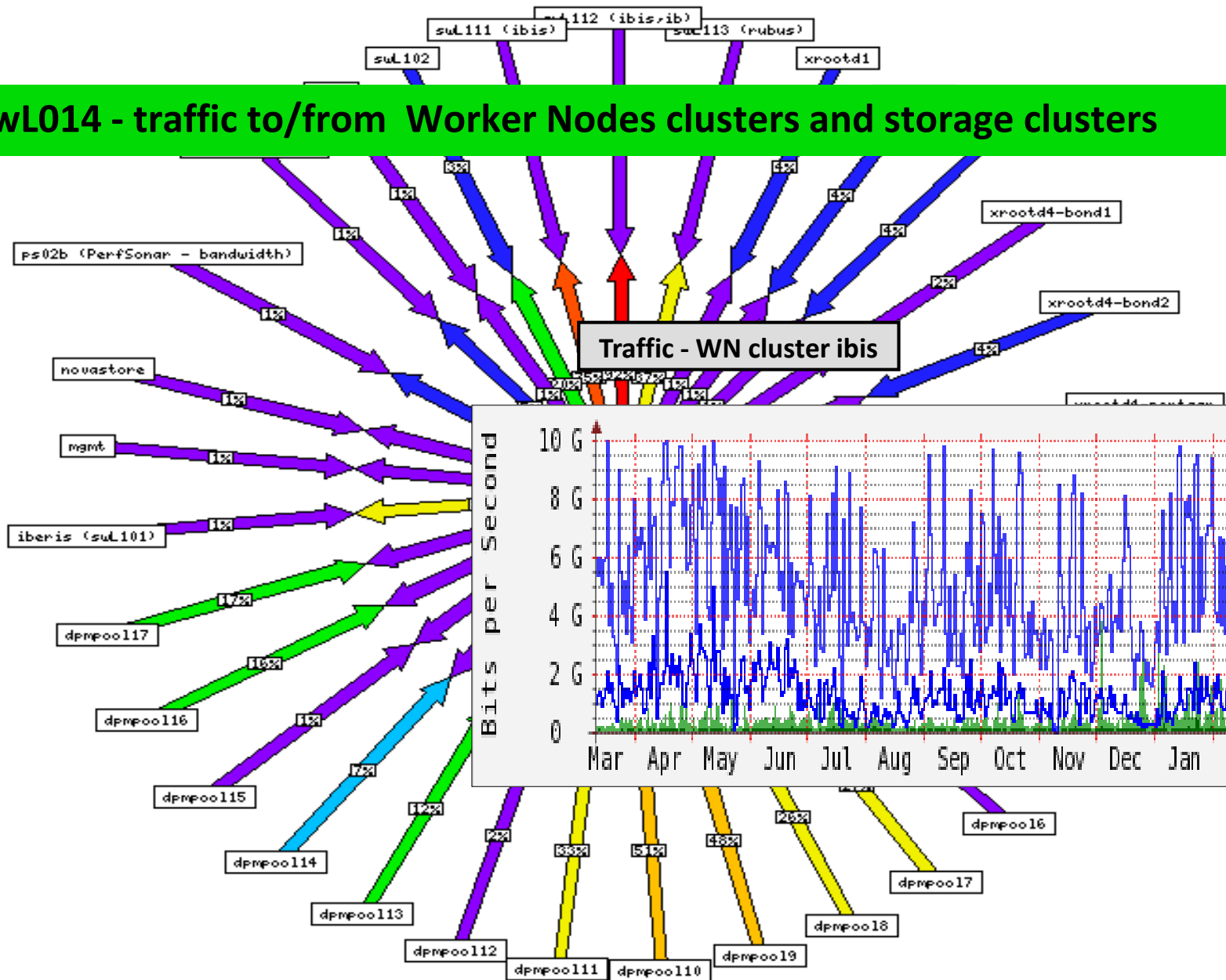
March 2015 – April 2016:
 outgoing traffic
 at the Prague/Rez
 ALICE storage cluster:
 > 3 GBytes/s in peaks
 almost 19 PBytes of data
 read out



Traffic on the dedicated line
 NPI <-> FZU. Maximum rate close
 to 4 Gbits/s, Which corresponds
 to 4 servers with 1 Gb/s Ethernet
 cards. This year we plan to upgrade
 the performance of the NPI cluster:
 the servers will be equipped with
 the 10 Gb/s Eth cards and the
 corresponding switch will be
 replaced with a more efficient one
 (Cisco Catalyst 4500-X 16 Port 10G)

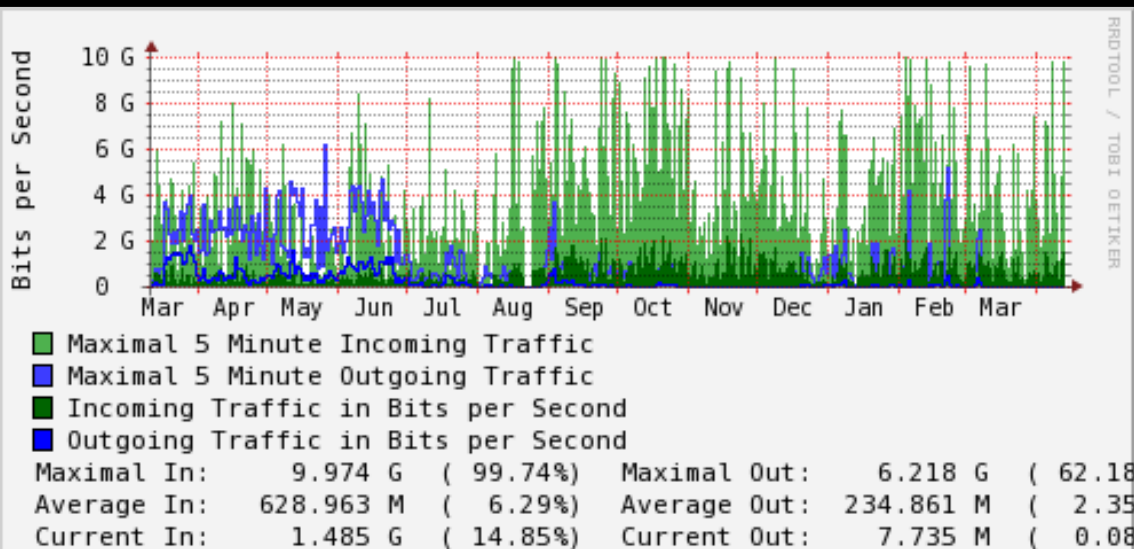
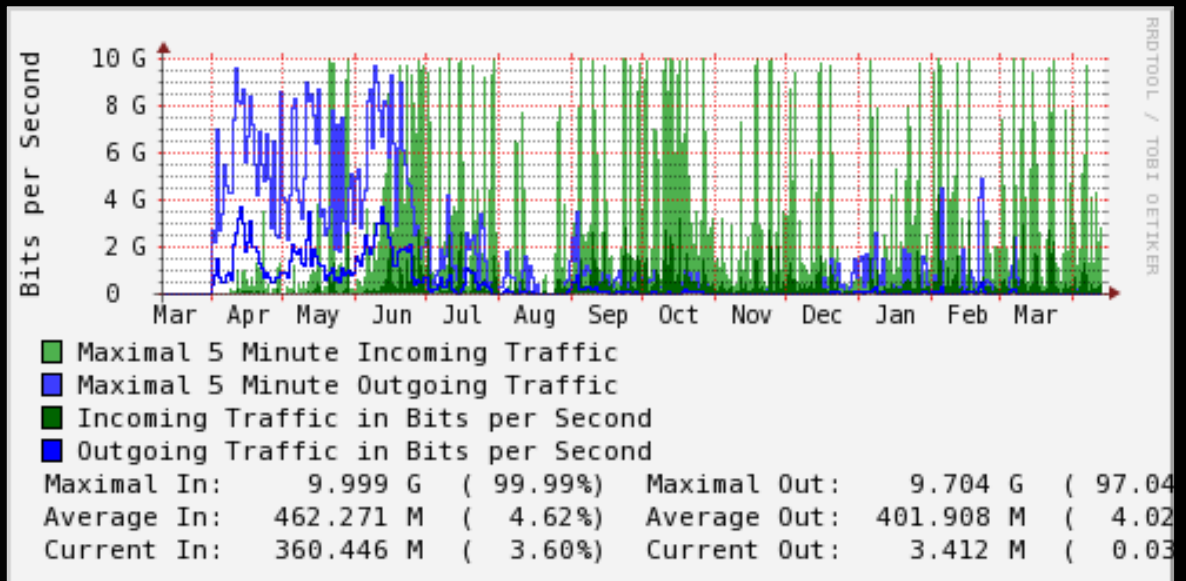


swL014 - traffic to/from Worker Nodes clusters and storage clusters



swL014 - traffic to/from XRootD storage nodes xrootd5 and xrootd4

xrootd 5



xrootd4



Czech Republic resource share

MANDATORY ACCORDING TO THE ALICE CONSTITUTION: 2.1% of the total required T2 (+T1?) resources

Resources delivery in 2015:

CPU (kHepSpec06)

ALICE requirement:	7,2
REBUS info: pledged	3,5
REBUS info: delivered	8,83

Resources planning for 2016:

CPU (kHepSpec06)

ALICE requirement:	7,2	(11,4)
REBUS info: pledged	5,0	

Resources delivery in 2015:

Disk (PB):

ALICE requirement:	0,6
REBUS info: pledged	1,030
REBUS info: delivered	1,591

Resources planning for 2016:

Disk (PB):

ALICE requirement:	0,92	(1,3)
REBUS info: pledged	1,4	

For the ALICE Czech group easier to scale up the storage capacity than to pile up CPUs.

ALICE pledges in 2015 of computing and storage resources were satisfied

(for storage to > 200%), although our CPU pledges were low :

- by sharing the hardware between various projects***
- by stable operations with almost no outage***
- by using the servers out of warranty***

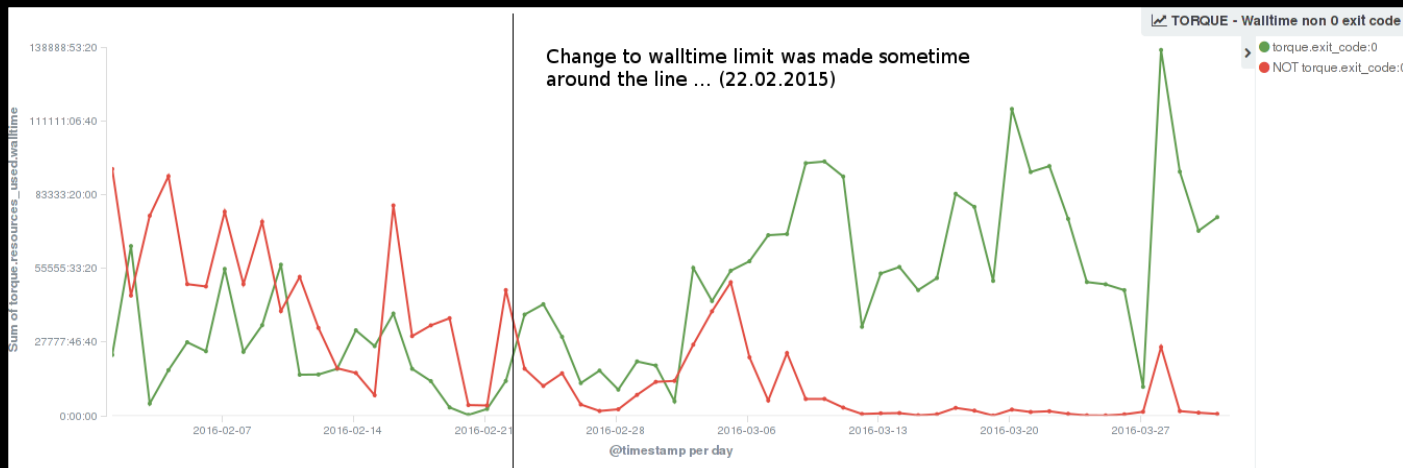


A couple of various issues

- In December 2014, some new hardware for ALICE was installed which upscaled the resources for ALICE into 2015 (a computing cluster at FZU, one storage server at FZU with ~ 420 TB of disk capacity and one storage server at NPI with ~120 TB of disk capacity)
- Most of the nodes of the Prague/Rez storage cluster for ALICE have been upgraded to the XRootD version v. 4.3.0.
- New redirector machine was put into operation, without the client functionality due to overload of the original redirector+client. We were the first users of the new installation method released in July 2015.
- In 2016, there was an arguing about financing of the operation of (old) computing servers for ALICE.
- One 6 years old cluster was switched off for high power consumption costs:
→ minus 320 cores, 2.7 kHS06.
- Another cluster from 2009 will be switched off during 2016 (144 cores, 1.7 kHS06).
- → *It might happen that we will be strongly underperforming in 2016.*
- ? *But we will not be really criticized, because there is no MoU on computing resources delivery inside the ALICE collaboration?*

A couple of various issues (contd)

- *Fortunately, we are in the process of preparation of a grant application to obtain money for new hw for ALICE and ATLAS. If successful, new capacities should come in the period 2017 – 2019.*
- Change – improvement in the ALICE jobs exit codes in Torque after increasing the limit on the wall time from 60 to 120 hours. JAs TTL is 48 hours



A couple of HW interventions were needed through 2015/2016, but the interruption of operations was negligible.

- Problems to find students: we need IT students with some experience – interested but expect better financing
- At the moment only one Master student. Work as a sysadmin at the Prague Tier-2 focusing on ALICE support. Collaboration with C. Cordeiro from CERN IT on development of monitoring tools of VM performance in the commercial clouds (continue a Summer Student work).

Aspects of future upscale

The future LHC runs will require a huge upscale of computing, storage and network resources which will be impossible for us to achieve with the expected flat or even declined budget.

Activities towards getting access to external resources, HPC centers or computational clouds are ongoing at all LHC experiments.

On the local level, there are lengthy discussions with the HPC center IT for Innovations (IT4I) in Ostrava.

Another possibility: using services of the OpenNebula IaaS cluster/cloud of CESNET (MetaCloud, FIWARE lab).

The showstopper is the absence of tools for submission of ALICE jobs to these sites.

Summary and Outlook

Czech Republic was performing well during 2015 concerning delivery of computing and storage resources

- Mandatory requests satisfied to > 100%**
- High-level accessibility, reliable delivery of services, fast response to problems**

Into the upcoming years, we will do our best to keep up the reliability and performance level of the services and deliver the pledges

Crucial is the high-capacity network infrastructure :

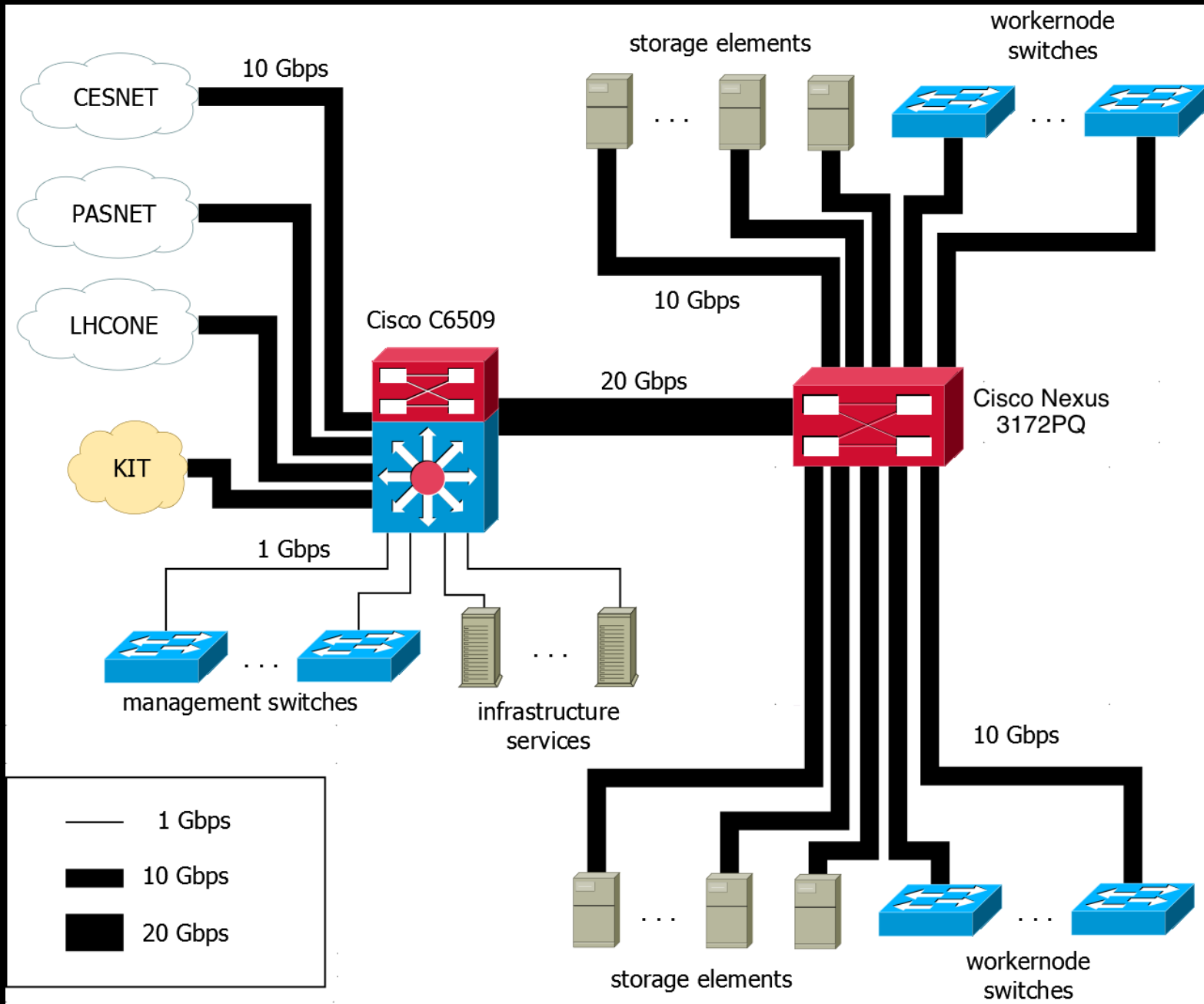
- Endeavour to upgrade the NPI storage nodes for higher traffic performance**
- Possible upgrade of the network system at the Tier-2 site at FZU**

Work on the development of the local ALICE monitoring tools is planned

Further negotiations with the IT4I HPC center and possible development of job submission tools (the help of ALICE Offline central team needed ...)

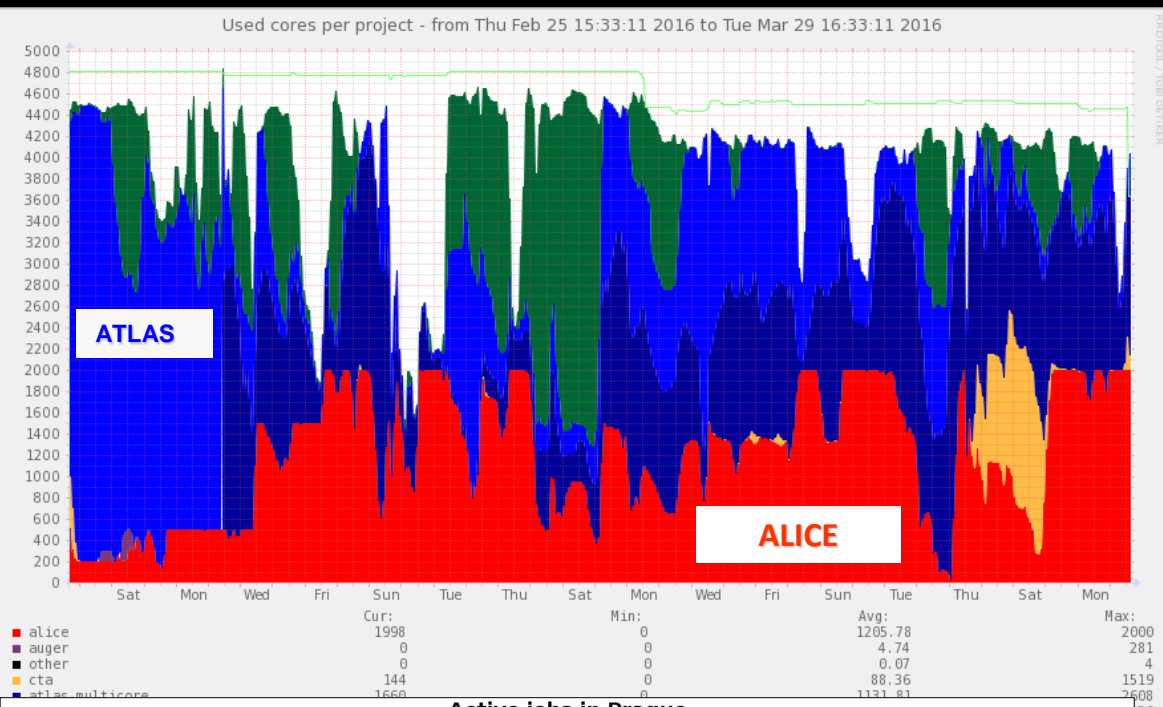
Backups

Network architecture at the Prague site



ALICE operations in 2016

Running jobs profile at the Prague Tier-2 in March 2016. The ALICE share is lower than in 2015.



Running jobs profile as shown in ALICE MonALISA. The number of concurrently running ALICE jobs at FZU in the first months of 2016 was in average ~1300, in peaks ~ 3000. This will very likely drop down later in 2016 due to problems discussed in next slides.

