COMPUTING FOR ALICE IN THE CZECH REPUBLIC in 2018/2019

D. Adamová
with the help and support of :
M. Adam, J. Chudoba, A. Mikula, P. Vokáč, P. Horák and J. Hampl

Bucharest 2019

Outline

Status of the WLCG Tier-2 site in Prague

ALICE operations in 2018/2019

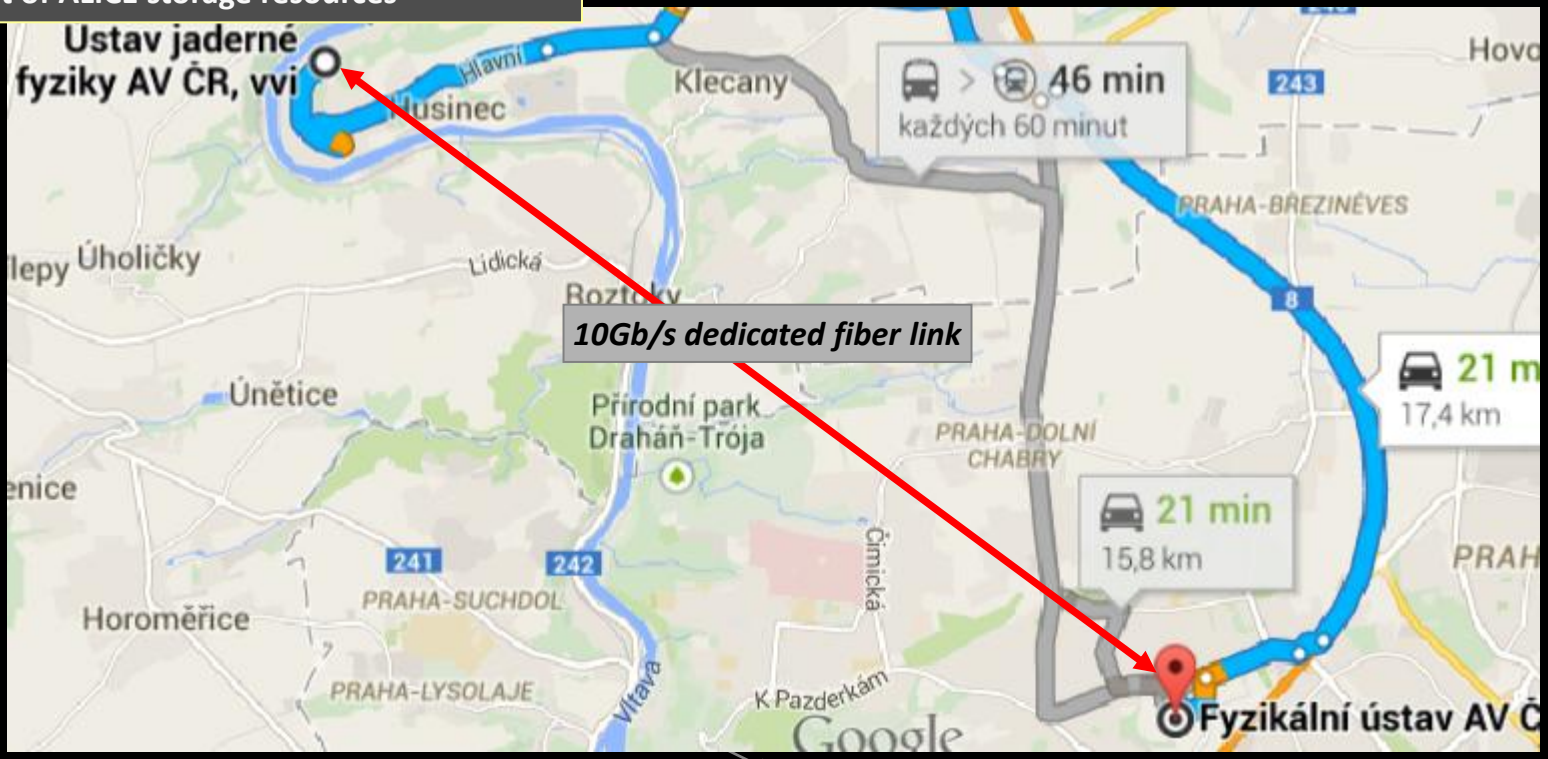
Report on the delivery of mandatory resources

A couple of various issues

Summary and Outlook

The geographical layout

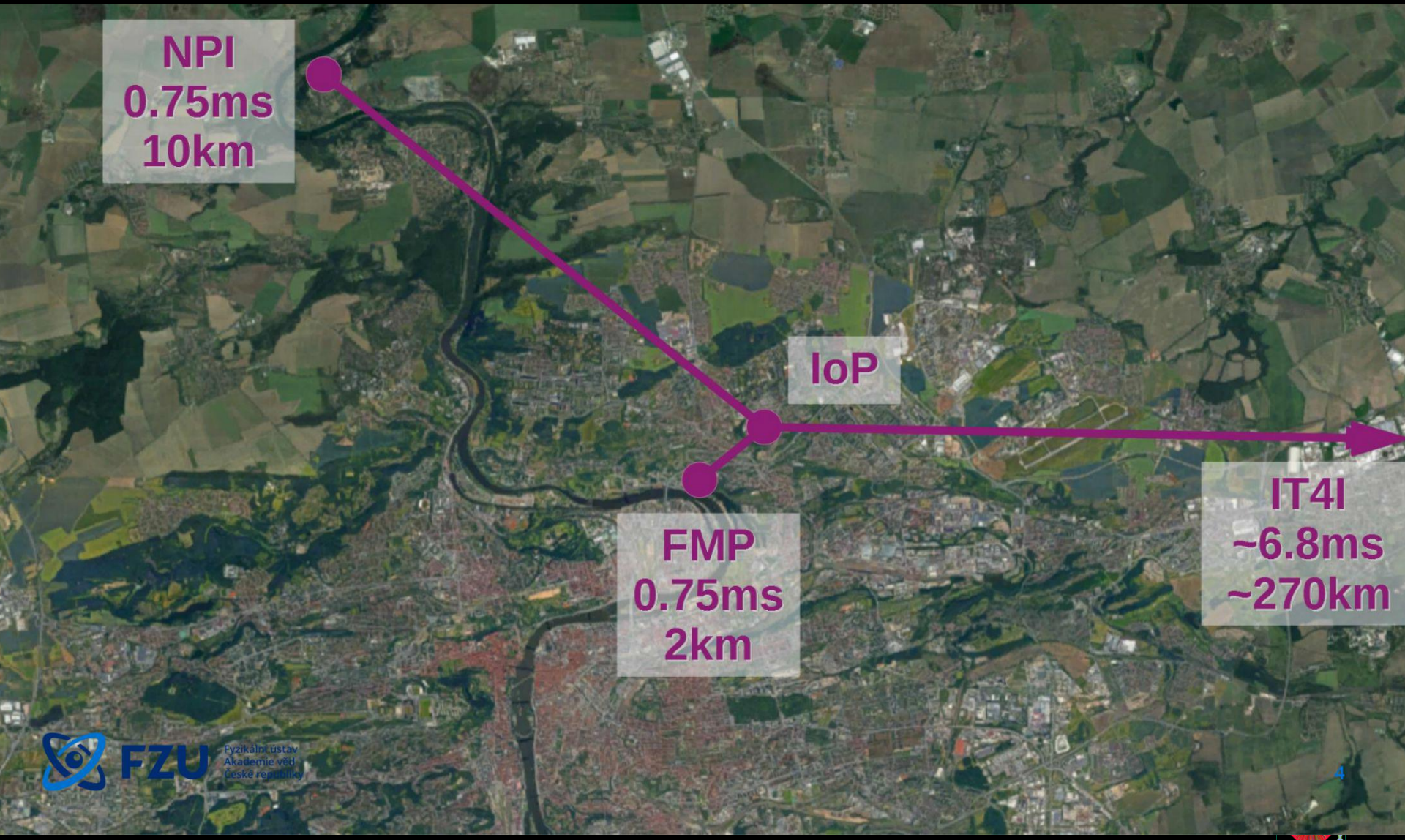
Nuclear Physics Institute AS CR (NPI)
a large ALICE group, no ATLAS involvement
A part of ALICE storage resources



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 and ATLAS
One part of storage resources for ALICE
Quite small ALICE group, much larger ATLAS community

An alternative geographical layout



FZU Fyzikální ústav Akademie věd České republiky

HEP Computing in Prague: WLCG 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 various Grid projects.

- **April 2008, WLCG MoU signed by Czech Republic** (ALICE+ATLAS).

- **Very good network connectivity:** Multiple dedicated 10 – 100 Gb/s connections to collaborating institutions, 100 Gb/s connection to LHCONE (upgraded from 2x10 Gb/s in April this year).

- **Provides computing services** for ATLAS + ALICE, Dune, Auger, NOVA, CTA, Fermilab, Astrophysics ...

Current numbers

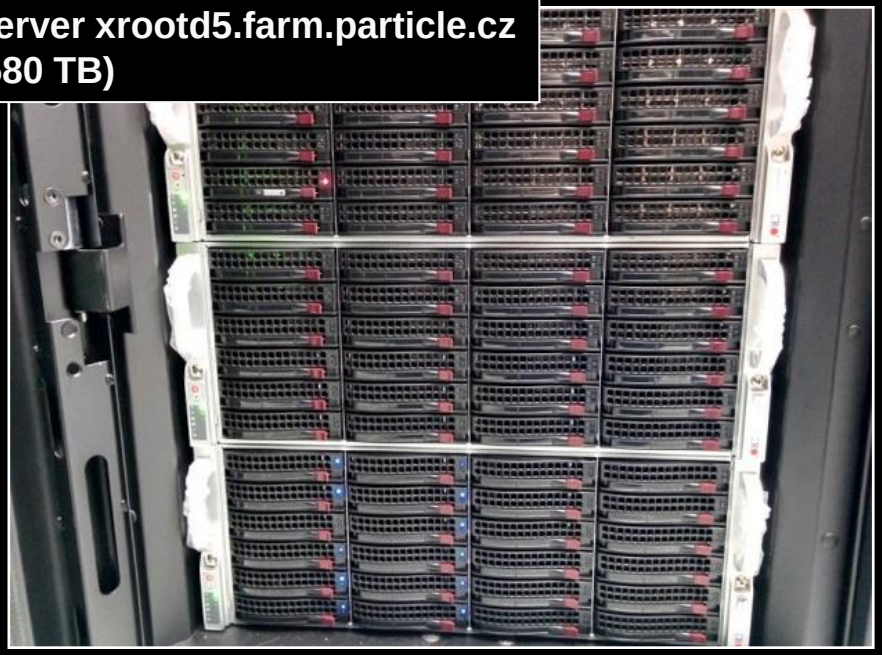
- 1 batch system (HTCondor)
- 2 main WLCG VOs: ALICE, ATLAS
  - Other users: Auger, NOVA, CTA, Dune, Fermilab user group, astrophysics group
- ~7600 job slots on site + 350 job slots at Charles University
- ~7 PB in total on disk storage on site and at NPI (DPM, XRootD, NFS)
- Regular yearly upscale of resources on the basis of various financial supports, mainly the academic grants.
- Monitoring: Nagios, Munin, Observium, Ganglia.
- Configuration management by Puppet
- Provisioning and SW management by Foreman

About 40 VMs provide core and grid services on 4 KVM hypervisors .

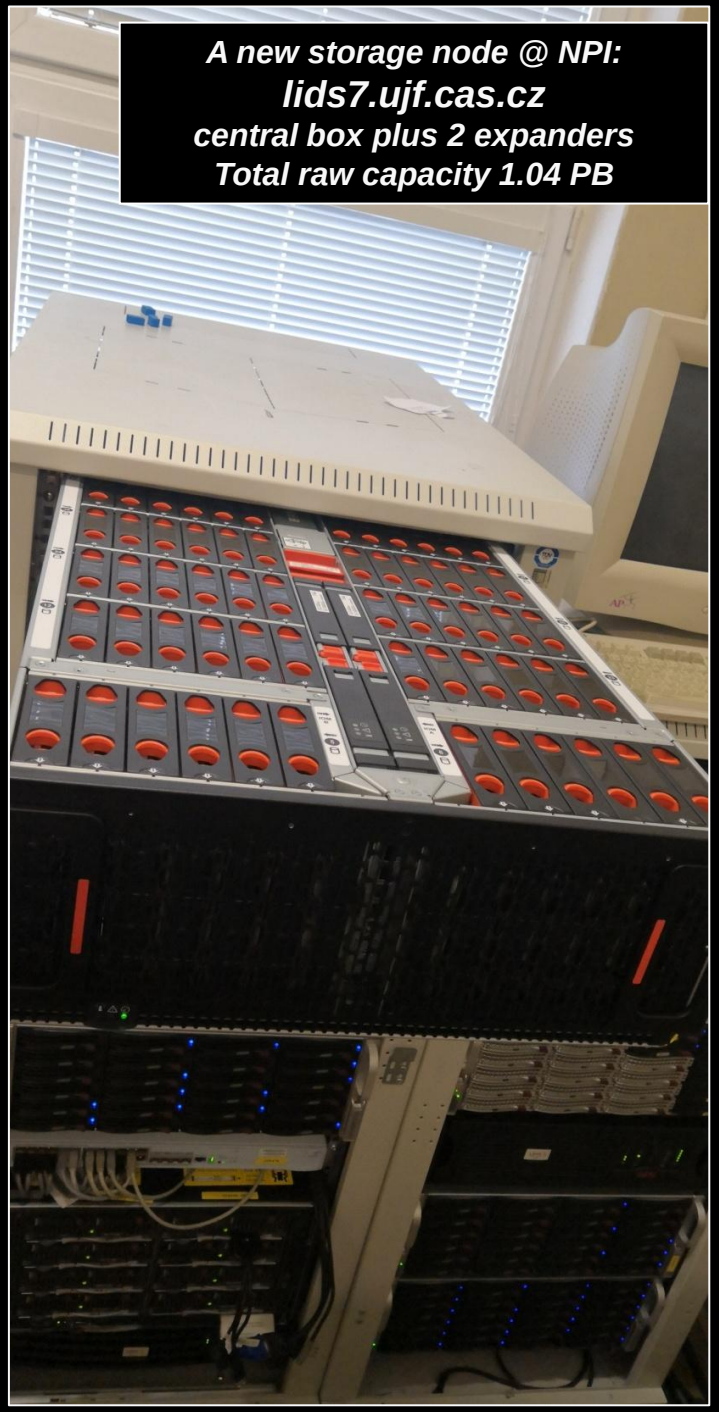
ALICE disk XRootD Storage Element
ALICE::Prague::SE

- ~ 2.2PB of raw disk capacity in total
- a distributed storage cluster:
  - Redirector + 2 clients @ FZU,
  - 4 clients @ NPI Rez
- 10 Gb/s *dedicated* connection from FZU to the NPI ALICE storage cluster

Server xrootd5.farm.particle.cz (580 TB)



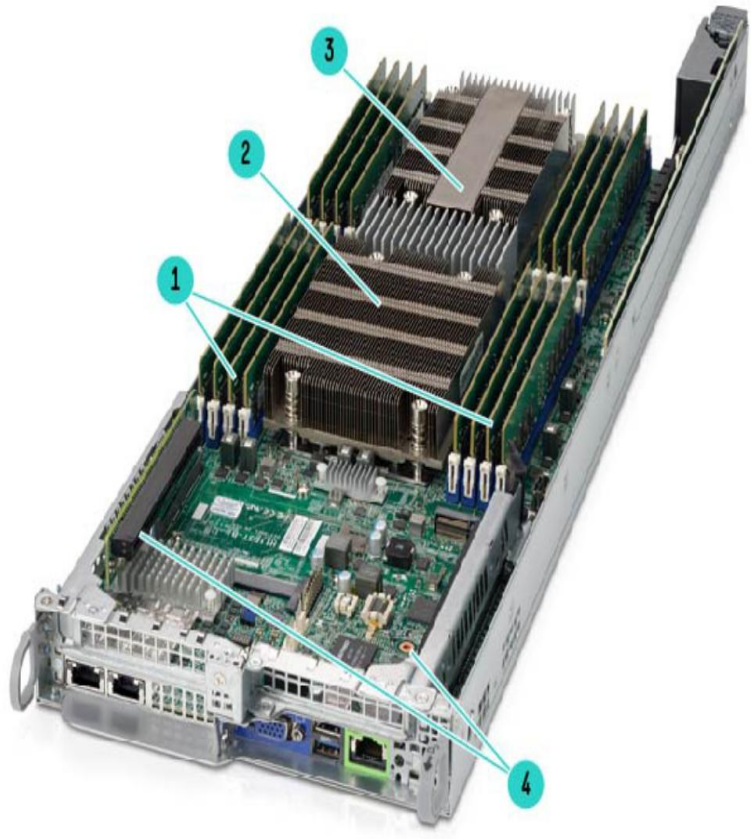
A new storage node @ NPI:
lids7.ujf.cas.cz
central box plus 2 expanders
Total raw capacity 1.04 PB



New hardware from the OP RDE grant

- Recently delivered, just being installed
- 6 systems Apollo 35 with 24 servers in total
- AMD EPYC 7301 (16c, 2.2 GHz)
- RAM DDR4 ECC 2666MHz 16x8 = 128 GB (2 GB / HT core)
- 1.44 TB SSD disks
- 768 cores, 1536 job slots
- 18000 HS06 (estimated)
- Expected 22% higher performance per unit currency (CZK) in (estimated) HS06 than for cluster from 2017

New cluster – HPE Apollo 35



Internal View

Item	Description	Item	Description
1	8 Memory DIMM slots per CPU	3	CPU 2
2	CPU 1	4	Two low profile HH/HL PCIe 3.0 x16 slots

New hardware
from the OP RDE grant (cont.)

New disk server installed at the NPI

lids7.ujf.cas.cz

Delivered in Dec2018/ Jan2019

TOTAL RAW DISK CAPACITY ~ 1.04 PB

- On the picture from top to bottom:
- first expander lids7 (60 disks 10TB)
- central server lids7.ujf.cas.cz
- ups
- ups
- second expander lids7 (44 disks 10TB)

Issues with storage

In April 2018, the total raw disk capacity for ALICE in the CR was ~ 2 PB.

This complied with the ALICE requirements.

3 servers at the main site had repeated problems and were often put offline.

During the following months, these 3 storage servers had to be decommissioned (were old, problems), as well as 1 of servers at the NPI.

Altogether it was almost minus 600 TB

In the beginning of 2019, a new storage node, lids7.ujf.cas.cz, was put into operation at the NPI, with the raw disk capacity of ~ 1.04 PB.

Immediately after it started functioning, began migration of data from one storage server at the main site, which was temporarily left for ALICE use but in fact belongs to ATLAS. The capacity of this server is 500 TB. Migration was finished yesterday.

In summary, the disk capacity for ALICE is approximately the same as last year, although a quite promising new server was put into operation. But still the ALICE requirements are met. Also, the Prague storage cluster got rejuvenated compared to last year.

Recent issue with local network

To improve traffic on the line to LHCONE it was decided to upgrade to 100 Gb/s.

For this purpose it was necessary to upgrade the main router. The current C6500 was replaced by C9500. The choice of the new router was made based on a recommendation of some network colleagues. As it turned out, this recommendation was bad for our situation.

After the replacement of the router by C9500 and additional re-configuration of the local network, it was impossible to run jobs of several projects including ALICE.

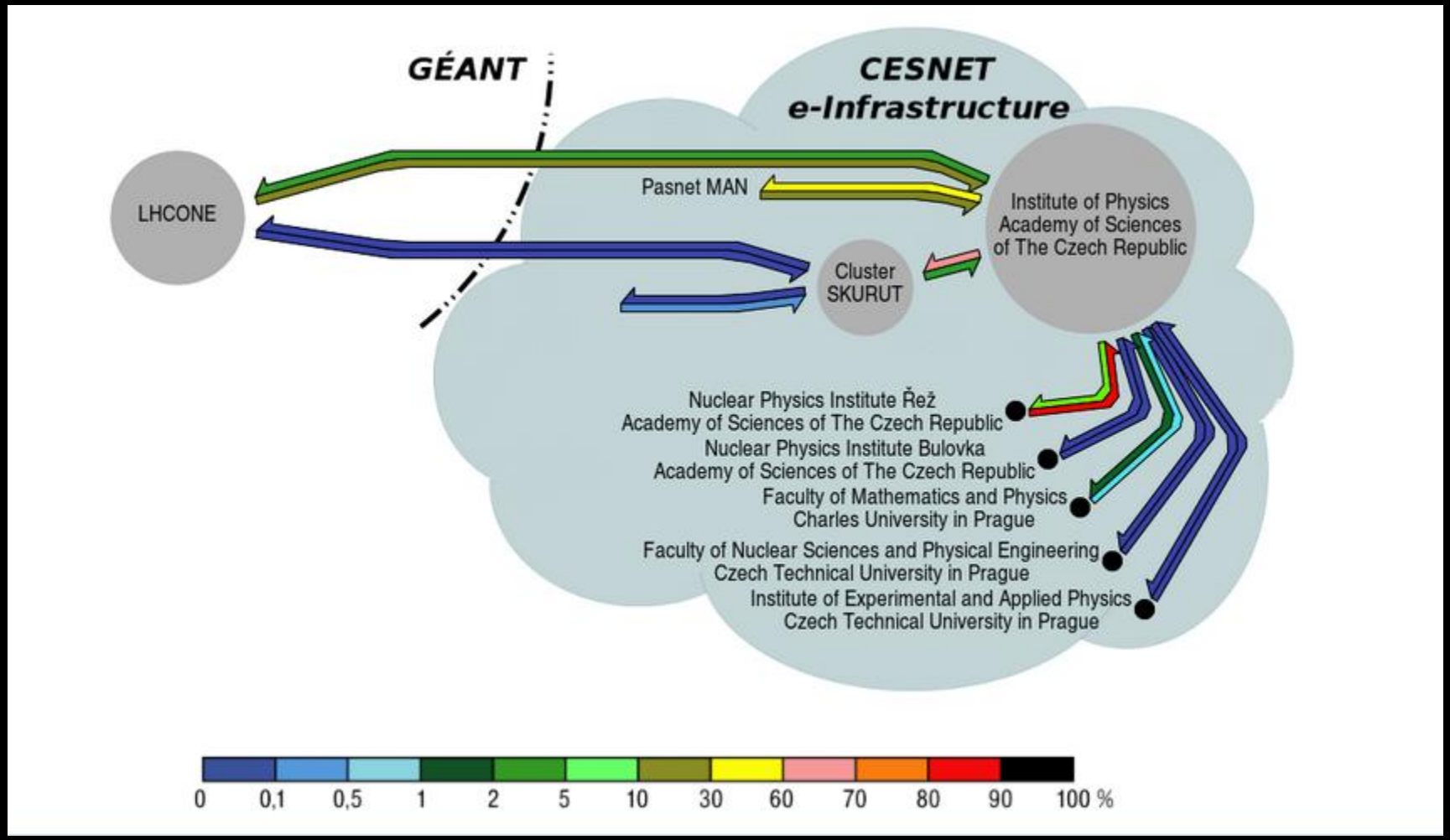
After careful investigation it was found that the problem was in bad NAT performance of C9500.

As a rough patch, an older server was re-configured as NAT server and included in the network.

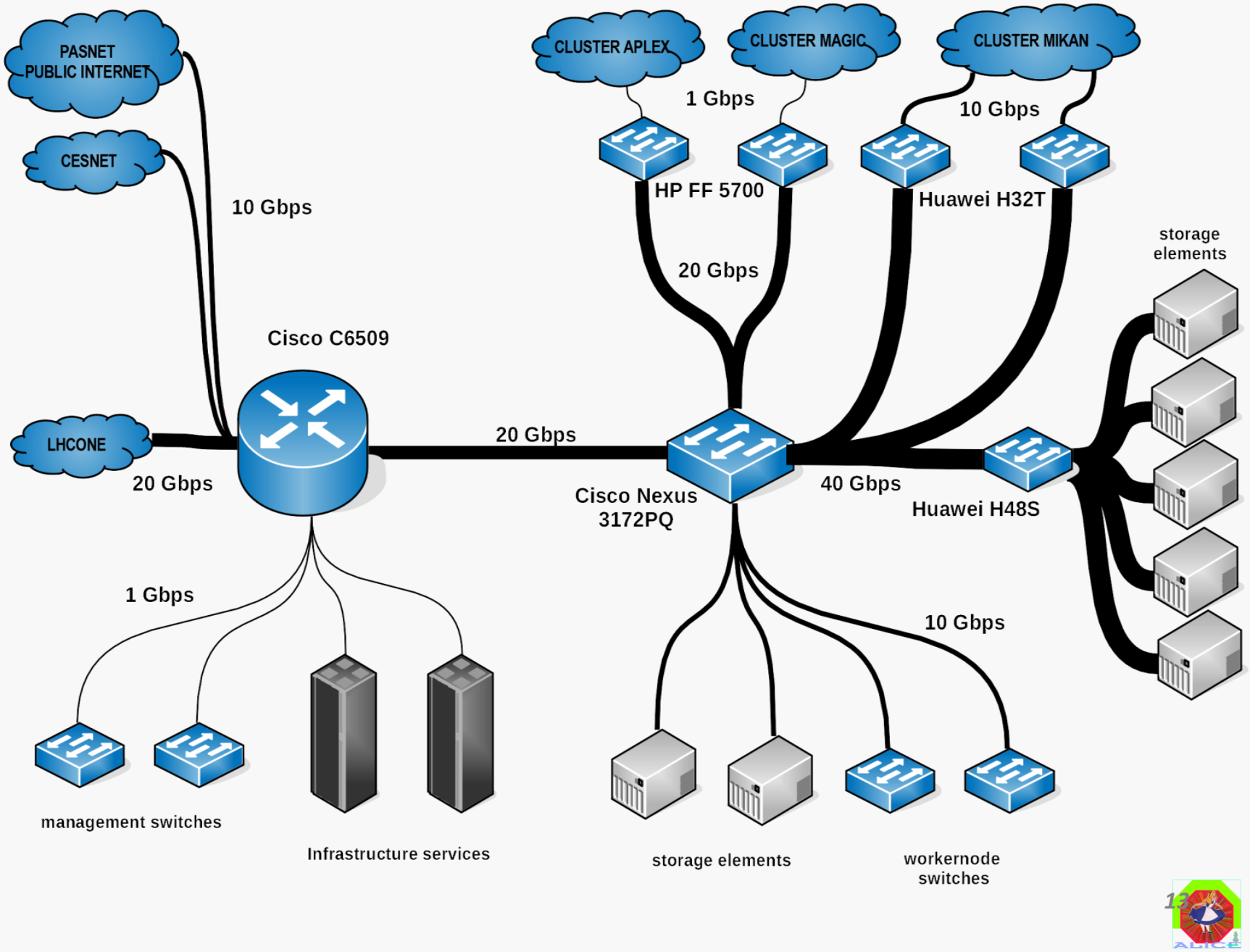
Since then situation seems back to normal except that the performance is not good enough.

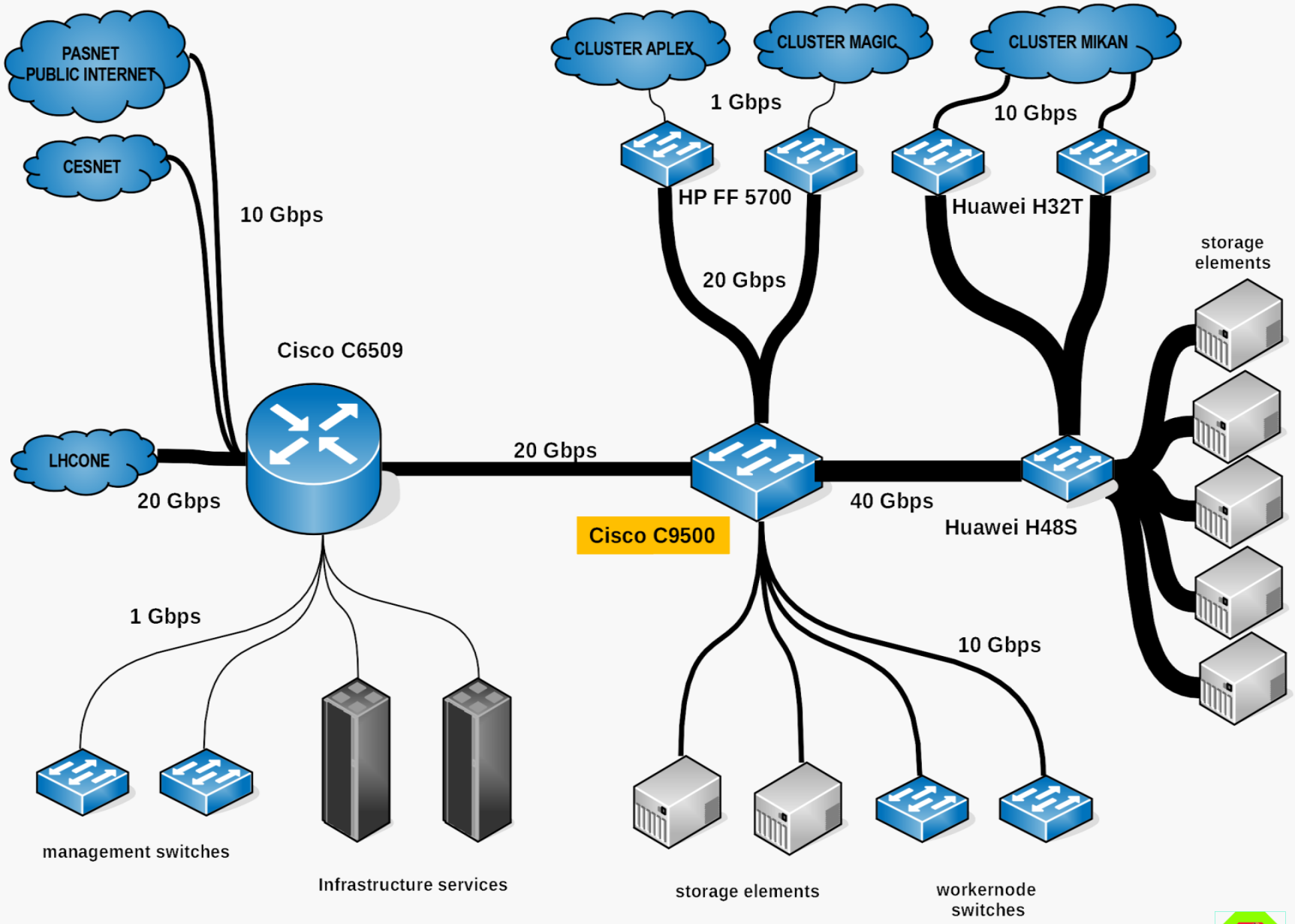
The right solution of the problem should be to purchase an appropriate (and expensive) router but this decision is up to the management of the farm.

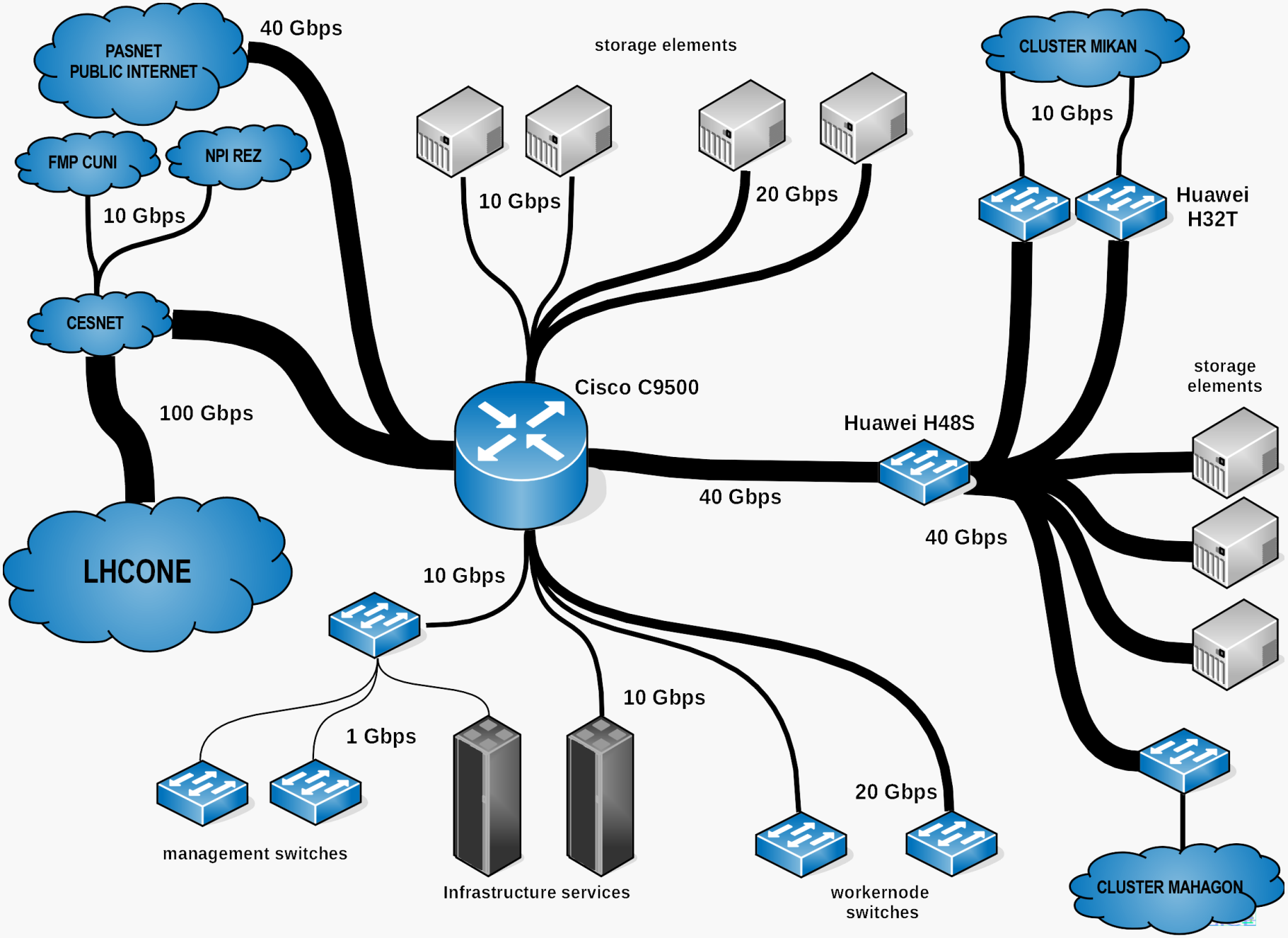
External connectivity, by CESNET



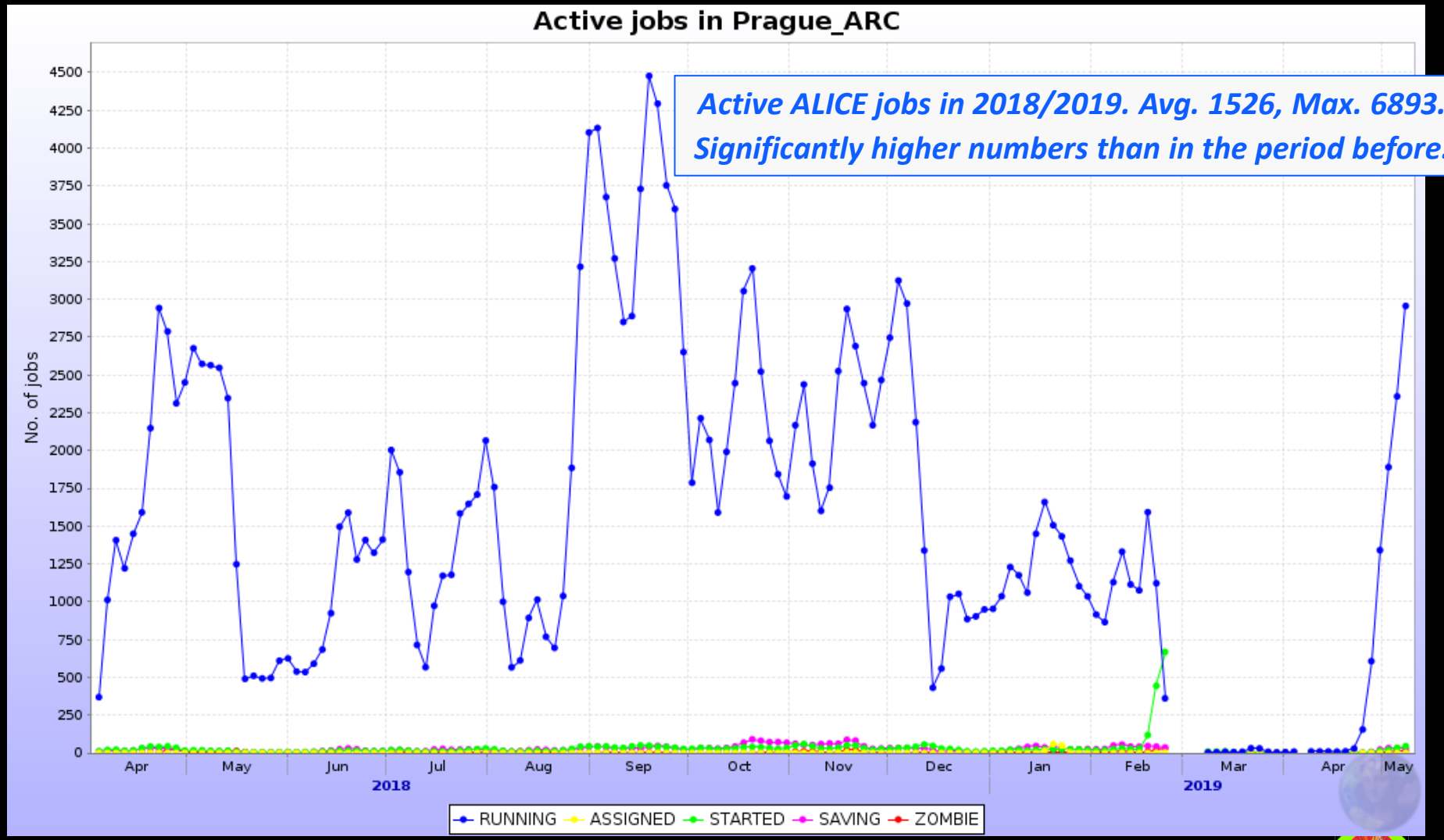
Core load on external connections comes from LHC Experiments ATLAS and ALICE.


PASNET
PUBLIC INTERNET
CESNET
10 Gbps
LHCONE
20 Gbps
Cisco C6509
1 Gbps
management switches
Infrastructure services
20 Gbps
Cisco Nexus
3172PQ
CLUSTER APLEX
CLUSTER MAGIC
CLUSTER MIKAN
1 Gbps
10 Gbps
HP FF 5700
Huawei H32T
20 Gbps
storage
elements
40 Gbps
Huawei H48S
10 Gbps
storage elements
workernode
switches


PASNET
PUBLIC INTERNET
CESNET
10 Gbps
LHCONE
20 Gbps
Cisco C6509
1 Gbps
management switches
Infrastructure services
CLUSTER APLEX
CLUSTER MAGIC
CLUSTER MIKAN
1 Gbps
10 Gbps
HP FF 5700
Huawei H32T
20 Gbps
storage elements
20 Gbps
Cisco C9500
40 Gbps
Huawei H48S
10 Gbps
storage elements
workernode switches


PASNET
PUBLIC INTERNET
40 Gbps
FMP CUNI
NPI REZ
10 Gbps
CESNET
100 Gbps
LHCONE
storage elements
10 Gbps
20 Gbps
Cisco C9500
CLUSTER MIKAN
10 Gbps
Huawei
H32T
Huawei H48S
40 Gbps
storage
elements
40 Gbps
10 Gbps
1 Gbps
10 Gbps
20 Gbps
management switches
Infrastructure services
workernode
switches
CLUSTER MAHAGON

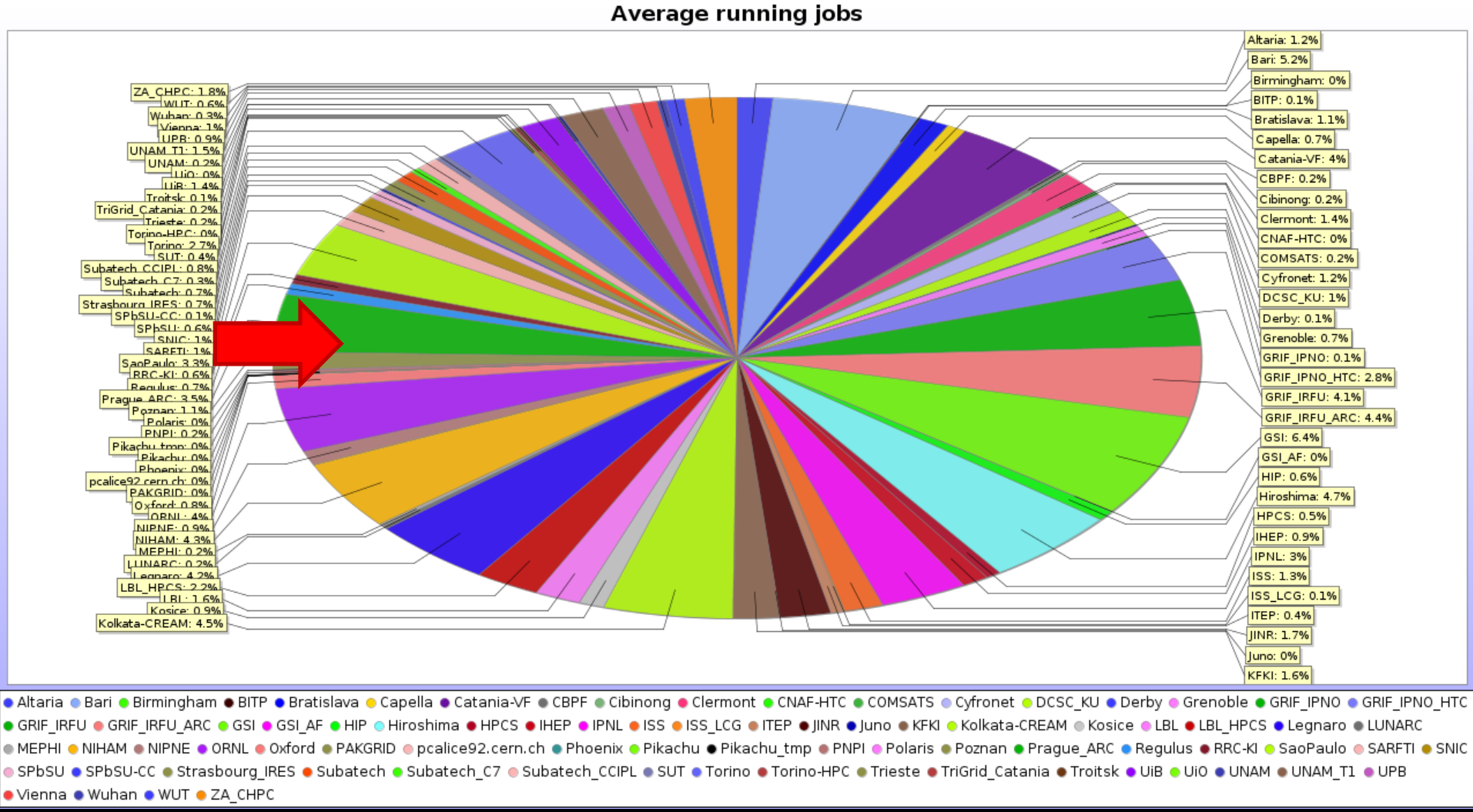
Running jobs profile: 1.4.2018 – 10.5.2019



Active ALICE jobs in 2018/2019. Avg. 1526, Max. 6893. Significantly higher numbers than in the period before.

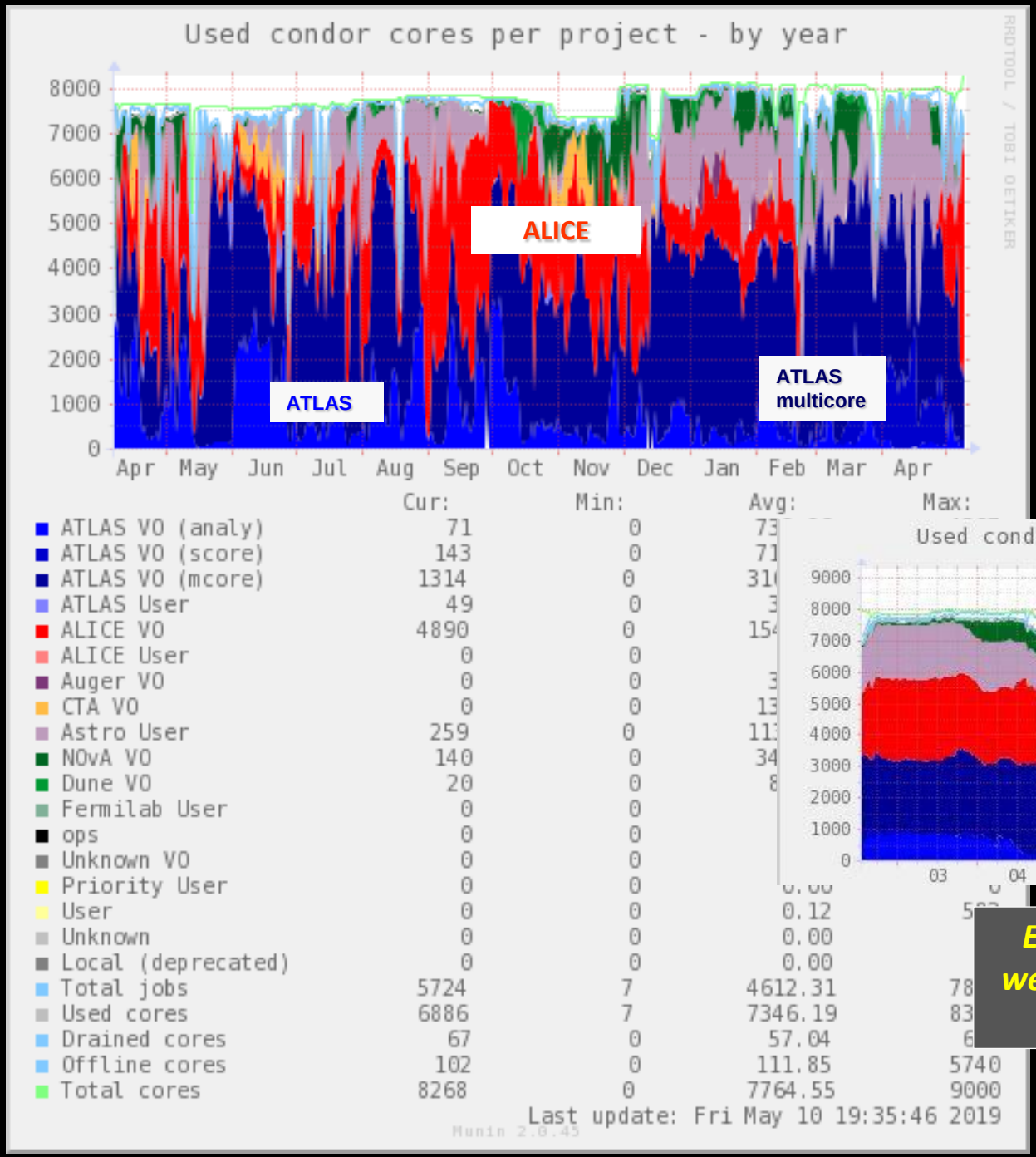
In total, 4,119,866 ALICE jobs was processed in Prague

CPU delivery share: 1.4.2018 – 10.5.2019

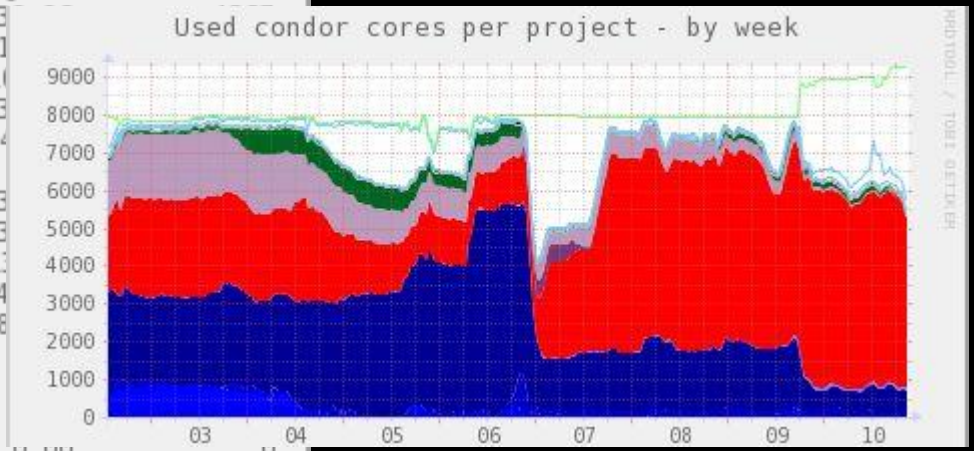


Czech republic share during 1.4.2018 – 10.5.2019 was ~ 3.5%. In 2 periods before it wa ~ 2.4%.

In this respect we improved ☺

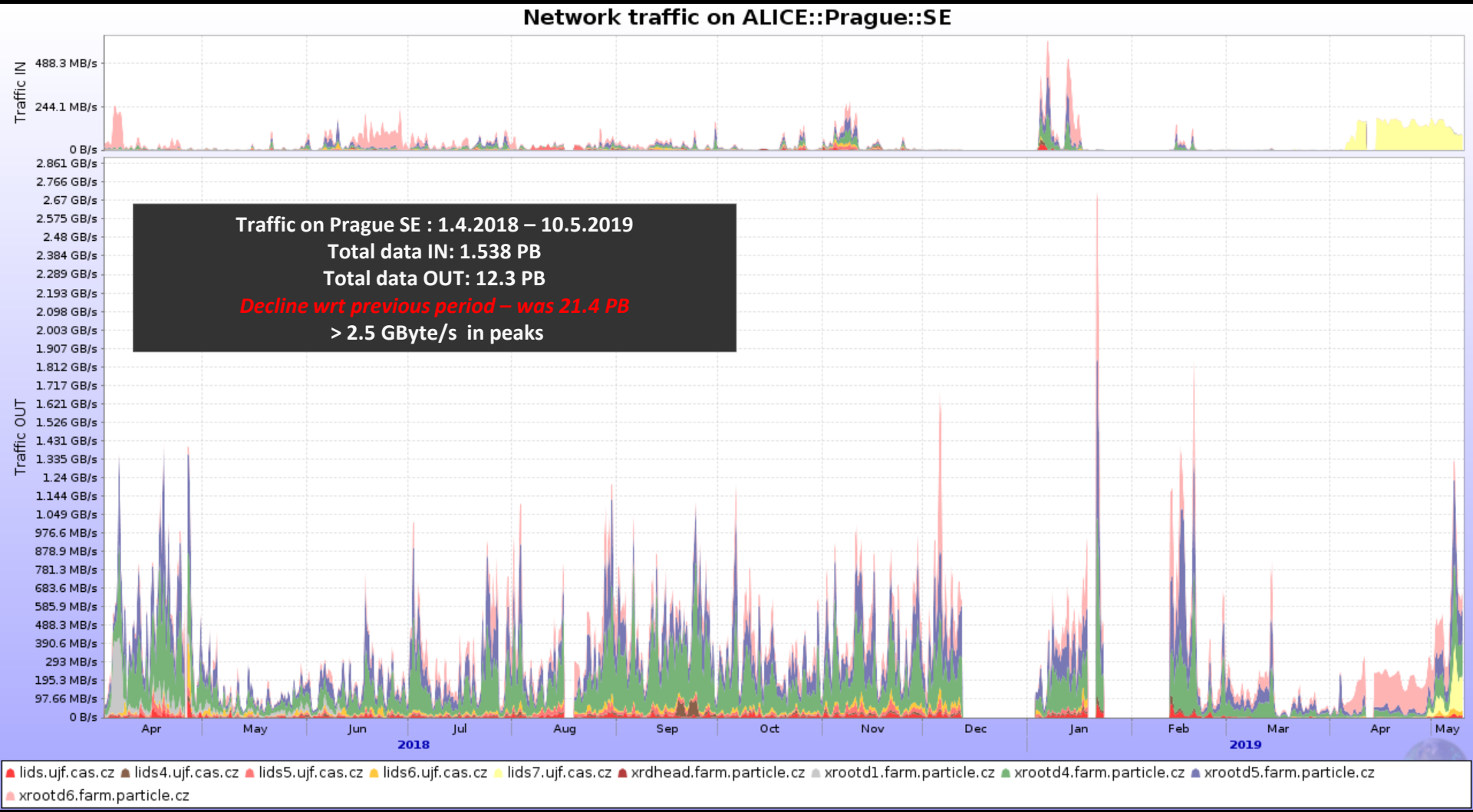


Local monitoring
Various projects
ALICE gets less share than last year



Exceptional large ALICE share last week - fairshare, no ALICE jobs for a couple of weeks before

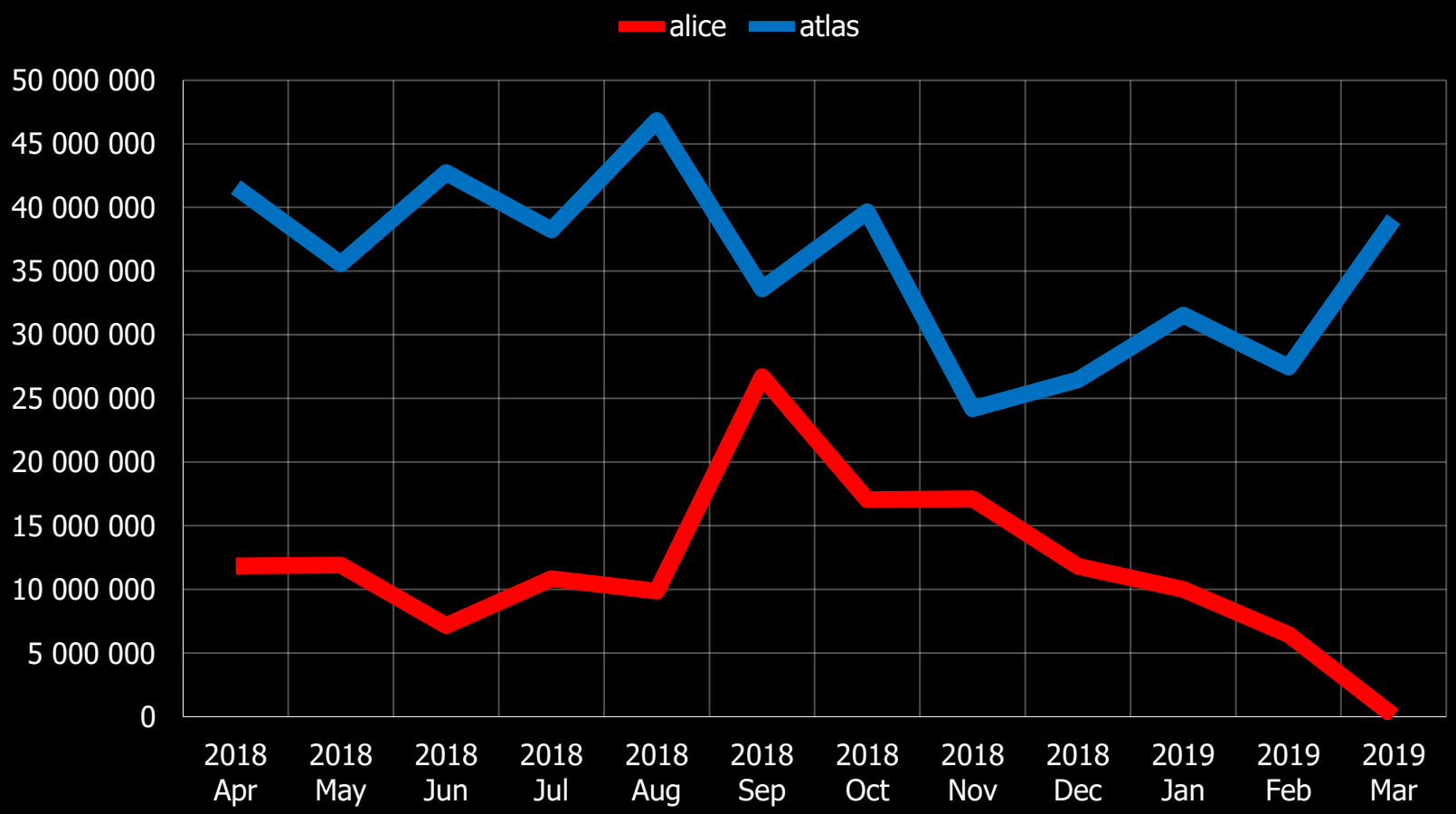
Traffic on ALICE::Prague::SE: 1.4.2018 – 10.5.2019



Delivered CPU power in the period 1.4.2018 – 30.4.2019
From the EGI Accounting Portal
ALICE requirement was/is *12,744,000 HS06 hours*

Month	ALICE	ATLAS
2018 Apr	11,845,380	41,608,491
2018 May	11,920,690	35,584,947
2018 Jun	7,167,925	42,732,910
2018 Jul	10,845,090	38,237,427
2018 Aug	9,869,627	46,824,098
2018 Sep	26,706,939	33,600,278
2018 Oct	17,054,695	39,679,263
2018 Nov	17,127,243	24,195,024
2018 Dec	11,818,244	26,462,693
2019 Jan	10,010,682	31,550,483
2019 Feb	6,472,486	27,471,058
2019 Mar	68,986	39,078,338
2019 Apr	827,843	34,337,646
Percent	23.50%	76.50%

Delivered CPU power in the period 1.4.2018 – 30.4.2019
From the EGI Accounting Portal
ALICE requirement was/is *12744000 HS06 hours*



Recent and planned compute/storage capacity upgrades

(financing approved/confirmed)

1. New CPU cluster “mahagon” – 6 systems HPE Apollo 35. Delivered in the early May 2019.
   For ALICE and ATLAS. Each with 18000 HS06 (estimated). 1536 jobslots.
   Currently under benchmarking tests.

2. New disk server installed at the NPI, for ALICE only.
   lids7.ujf.cas.cz
   Delivered in Dec2018/ Jan2019
   TOTAL RAW DISK CAPACITY ~ 1.04 PB

3. A new grant proposal is under way for years 2020 – 2022. The focus is exclusively on purchase of new hardware. The planned items include one computing cluster and disk servers both for the main site and for the NPI.

Summary and Outlook

2018 and early 2019 was a demanding time for the ALICE operations at the Prague Tier-2. We were not able to deliver the mandatory CPU resources and had problems with the local network. Also, four storage nodes must have been decommissioned.

On the other hand, we put into operation a new storage server with capacity of > 1 PB and a new computational cluster with 1536 job slots is under testing.

We are preparing a new grant proposal focused on purchase of new hardware for ALICE and ATLAS.

There still remains an open problem with the lack of sysadmins at the Prague site.

Into the upcoming year, we will do our best to keep up the reliability and performance level of the services and deliver the mandatory resources.

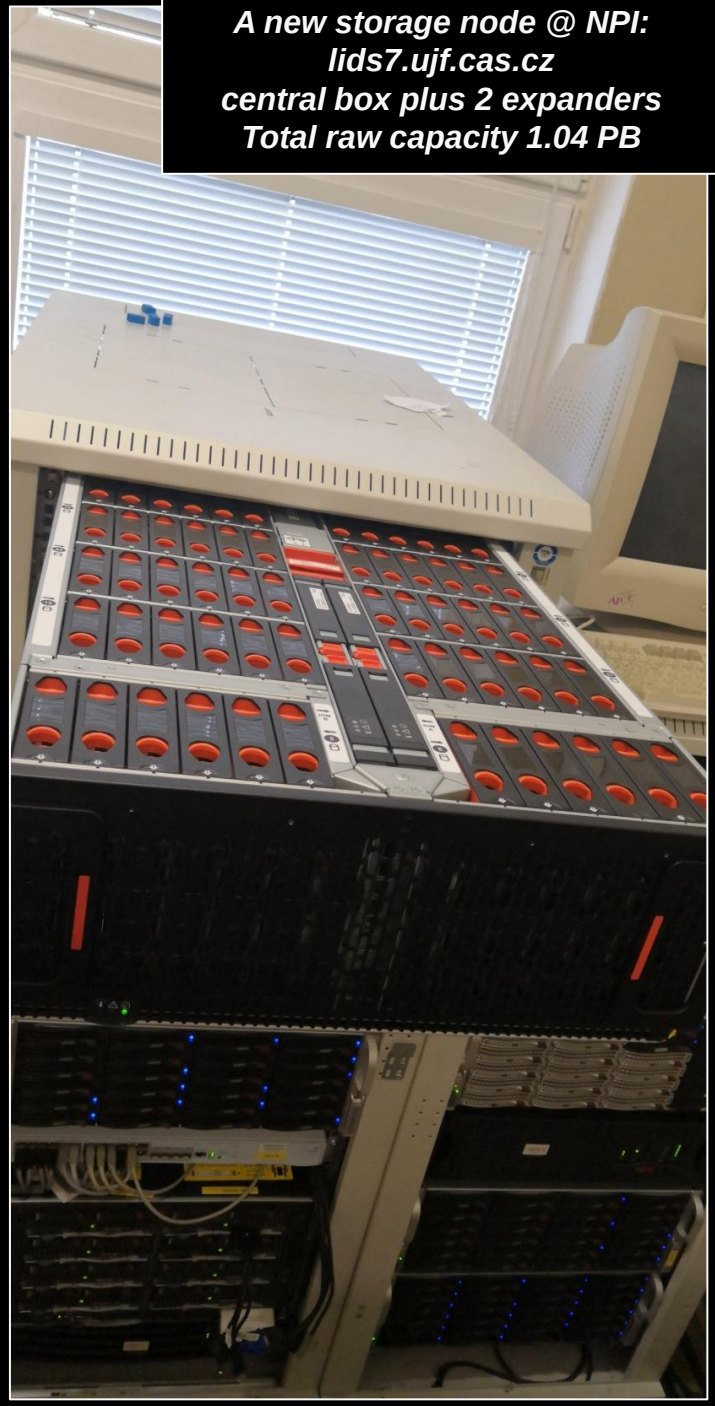
This work was supported by a project OP RDE CERN Computing (CZ.02.1.01/0.0/0.0/1 6013/0001404) from EU funds and MEYS.

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

Backups

A new storage node @ NPI:
lids7.ujf.cas.cz
central box plus 2 expanders
Total raw capacity 1.04 PB



Server xrootd5 (580 TB)

Czech Republic resources delivery

MANDATORY ACCORDING TO THE ALICE CONSTITUTION: 1.96% of the total required T2 resources

Resources delivery in 2016:

CPU (kHepSpec06)

ALICE requirement:	7,2
REBUS info: pledged	5,0
REBUS info: delivered	8,85

Resources delivery in 2017:

CPU (kHepSpec06)

ALICE requirement:	12,22
REBUS info: pledged	5,0
REBUS info: delivered	9,4

Resources delivery in 2016:

Disk (PB):

ALICE requirement:	0,92
REBUS info: pledged	1,4
REBUS info: delivered	1,79

Resources delivery in 2017:

Disk (PB):

ALICE requirement:	1,12
REBUS info: pledged	1, 4
REBUS info: delivered	1,79

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

ALICE requests in 2016 of computing and storage resources were satisfied, though the CPU pledges were low.:

- by stable operations with almost no outage

- by using the servers out of warranty

- by extensive use of other projects resources