This work is supported by projects Research infrastructure CERN (CERN-CZ, LM2015058) and 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

FZU (Prague) Site Report

Jiří Chudoba on behalf of the FZU Computing center team

15. 5. 2018

Institute of Physics (FZU) of the Czech Academy of Sciences

FZU Tier2 site overview

- 7600 cores in HTCondor
  - mostly HT, always >= 2 GB RAM/core
  - SL6 on all WNs
  - SL6, C7 for services
  - will be reduced to ~6000 this year (old hw removal)
- 6.1 PB on disks (DPM, xrootd) (2.2 PB on a new hw)
- 4x10 Gbps external connectivity
  - 2x10 Gbps to LHCONE, 10 Gbps to Internet, 10 Gbps to Czechlight
- 18 racks, 380 kVA UPS's (180 kVA backed up by diesel agr.)
- 5 administrators (2.6 FTEs)

New hardware – worker nodes

- 51 servers Huawei RH1288 V3
  - *2x 12-core Intel Xeon **E5-2650 v4**, 2.2GHz*
  - *128GB RAM - 8x 16GB DDR4 ECC, 2400MHz*
  - *2x 600GB HDD, SAS, 10krpm*, controller LSI SAS2308
  - *2x **10 Gbps** + 1 Gbps*
- Performance: 510.6 HS06/server (HT on), 10.64 HS06/core, 350 W/server
- 128/48 = 2.67 GB/jobslot
- Mostly for ATLAS, ALICE, NOvA, Auger
- Delivered in Oct 2017

- 2 data switches Huawei CE6810-32T16S4Q-LI
  - 32*10GE BASE-T ports, 16*10GE SFP+ ports, 4*40GE QSFP+ ports
- 2 ”management” switches Cisco SF300-24
  - 24 ports 10/100

New generation of Intel CPUs Skylake competitive only when all 6 memory channels used (192 GB RAM)

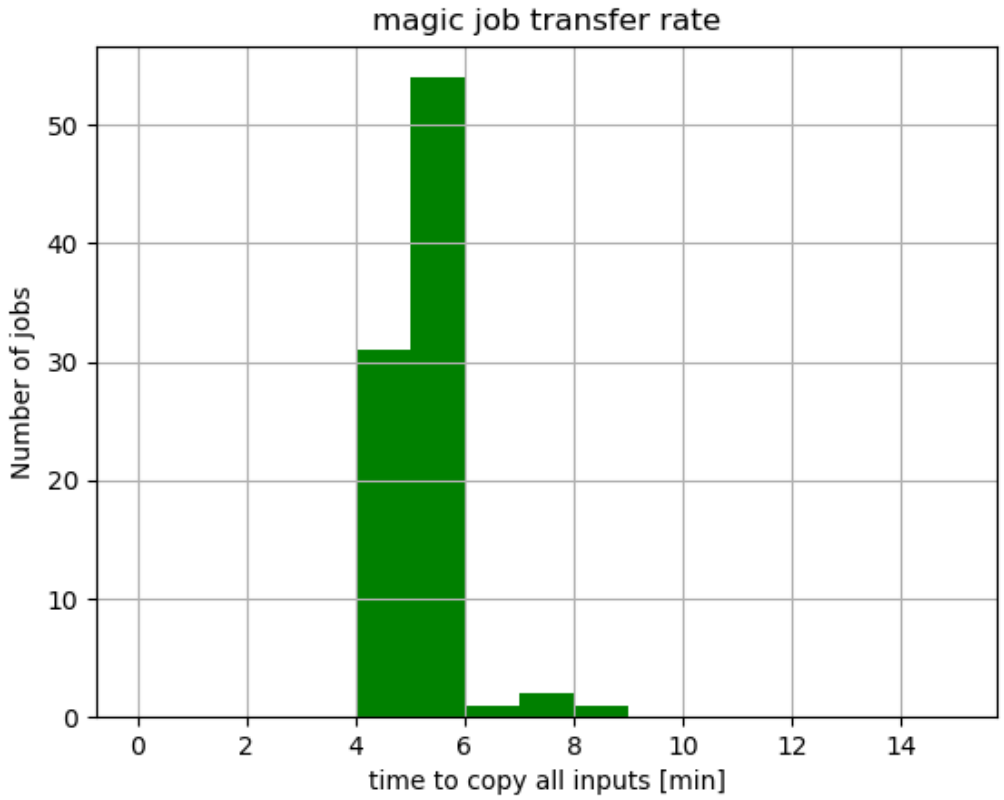
Comparison with an older cluster

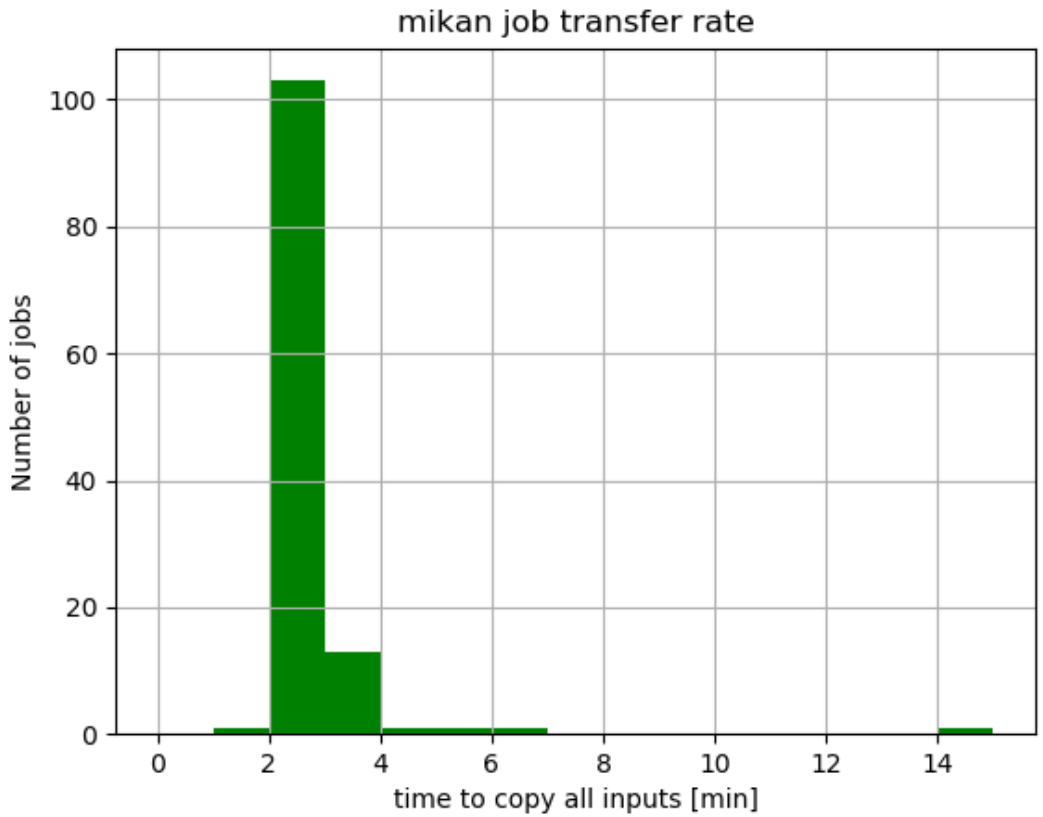
- 36 servers Intel HNS2600KP
  - *2x 10-core Intel Xeon **E5-2630 v4**, 2.2GHz*
  - *128GB RAM - 8x 16GB DDR4 ECC, 2400MHz*
  - *2x 600GB HDD, SAS, 10krpm*,
  - *1 Gbps*
- Performance: 417 HS06/server (HT on), 10.43 HS06/core
- 128/40 = 3.2 GB/jobslot
- Delivered in May 2016

- We expect jobs on Huawei cluster 2% faster (computing part)
  - 10.64/10.43 = 1.02

Comparison of Intel (I) and Huawei (H) servers

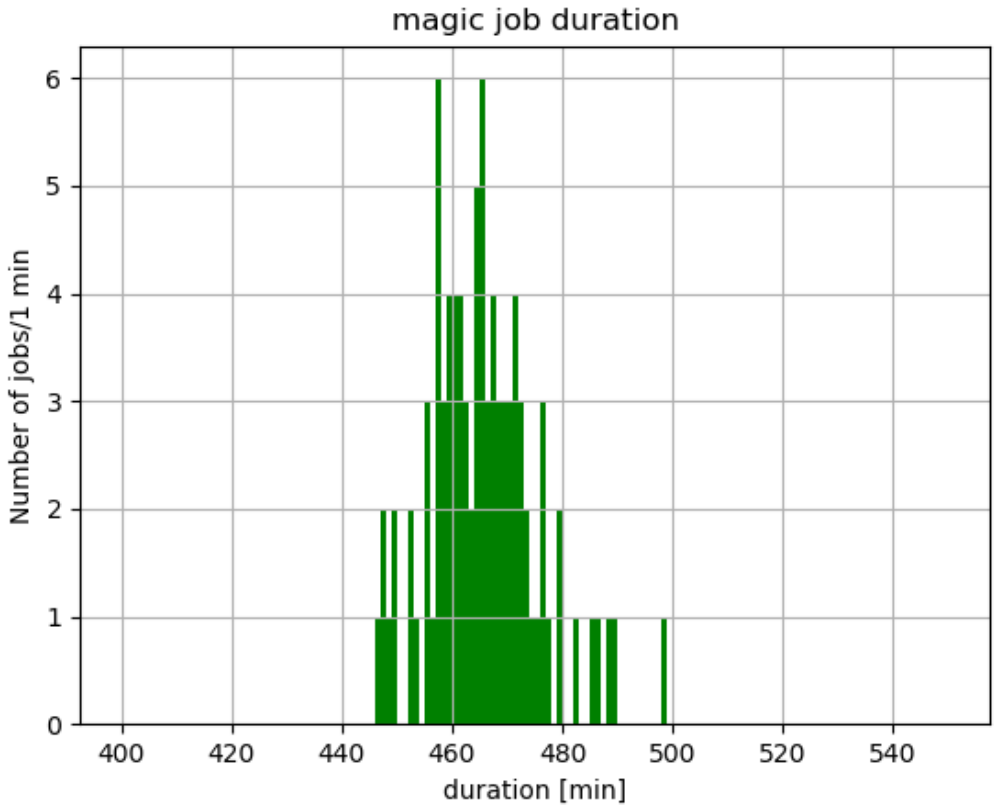
ATLAS task 14027559, 89 (I) vs 121 jobs (H), input size about 30 GB
Transfer rate around 100 MB/s (I) and 200 MB/s (H)

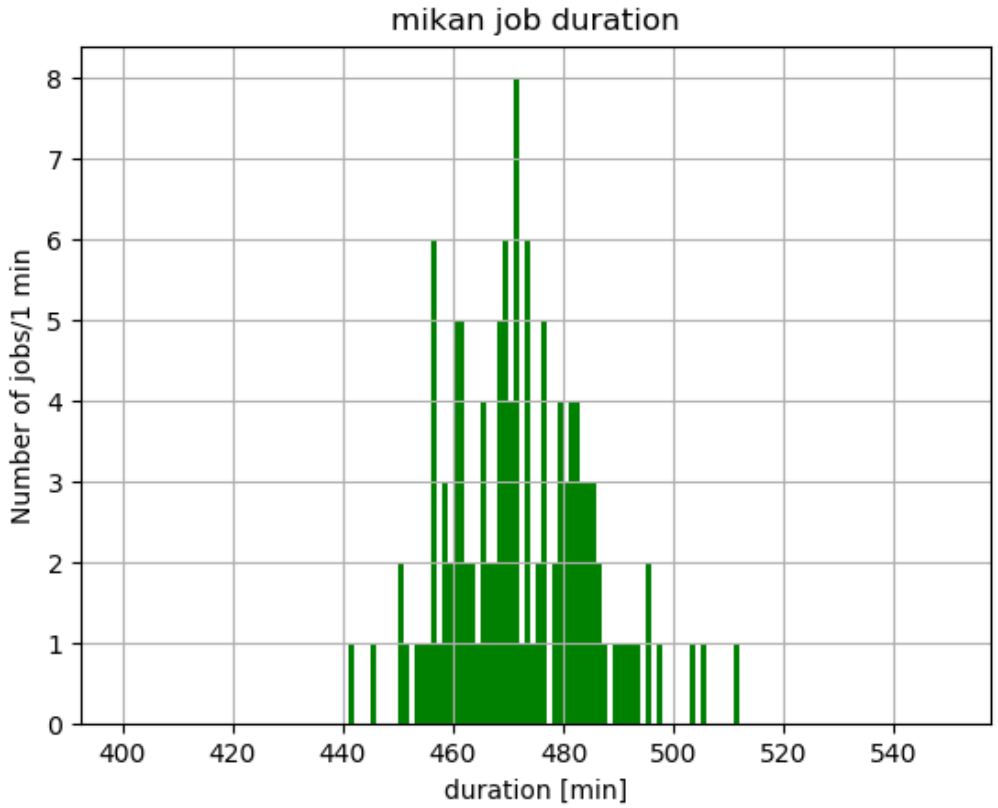




Comparison of Intel (I) and Huawei (H) servers

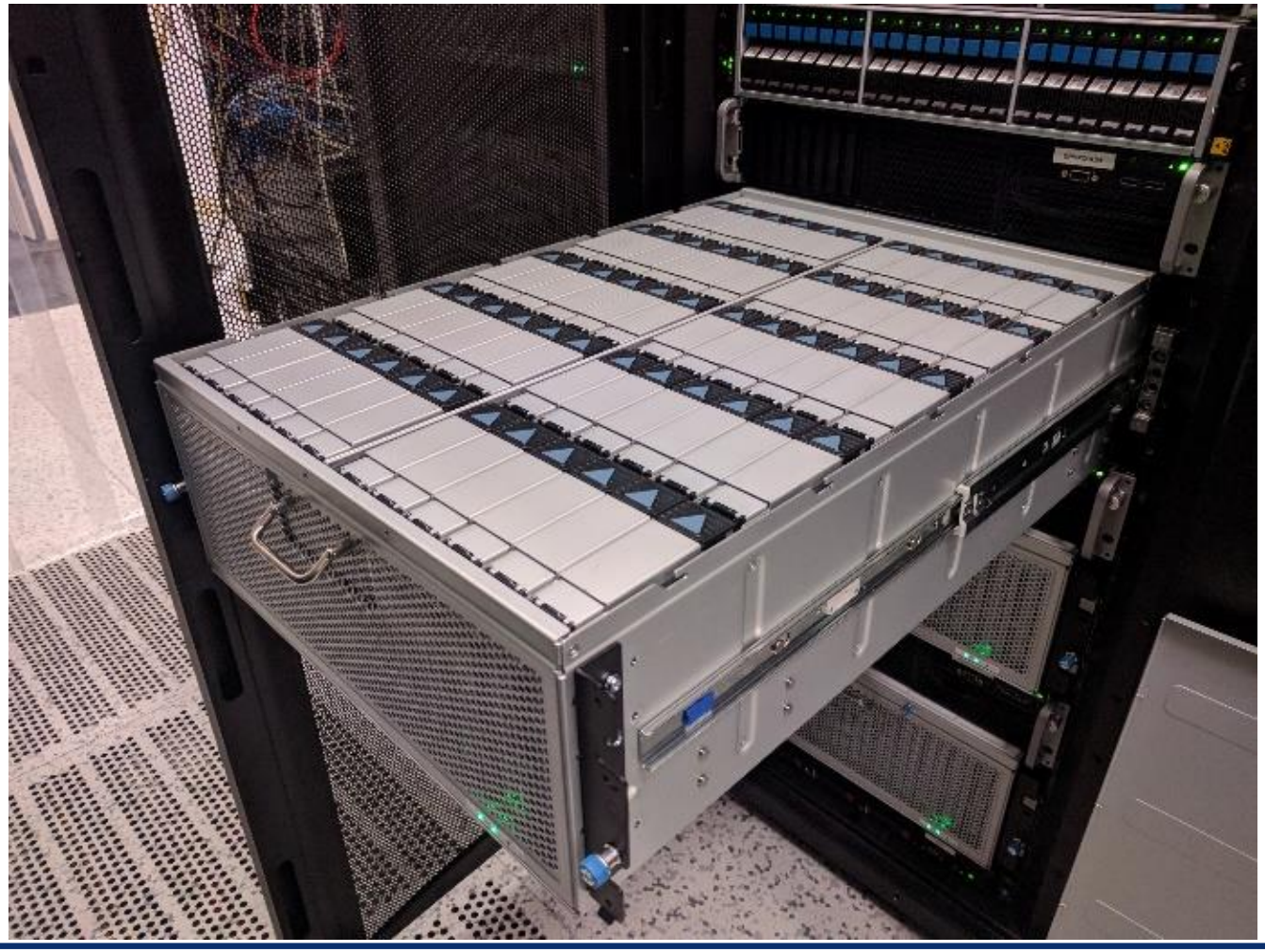
ATLAS task 14027559, 89 (I) vs 121 jobs (H)
avg time 27936 s vs 28333 s – diff 397 s = 1% - but I faster than H





New hardware – disk servers

Compact solution: 5 x (server + JBOD) = 5 x (2 + 4 U)



New hardware – disk servers

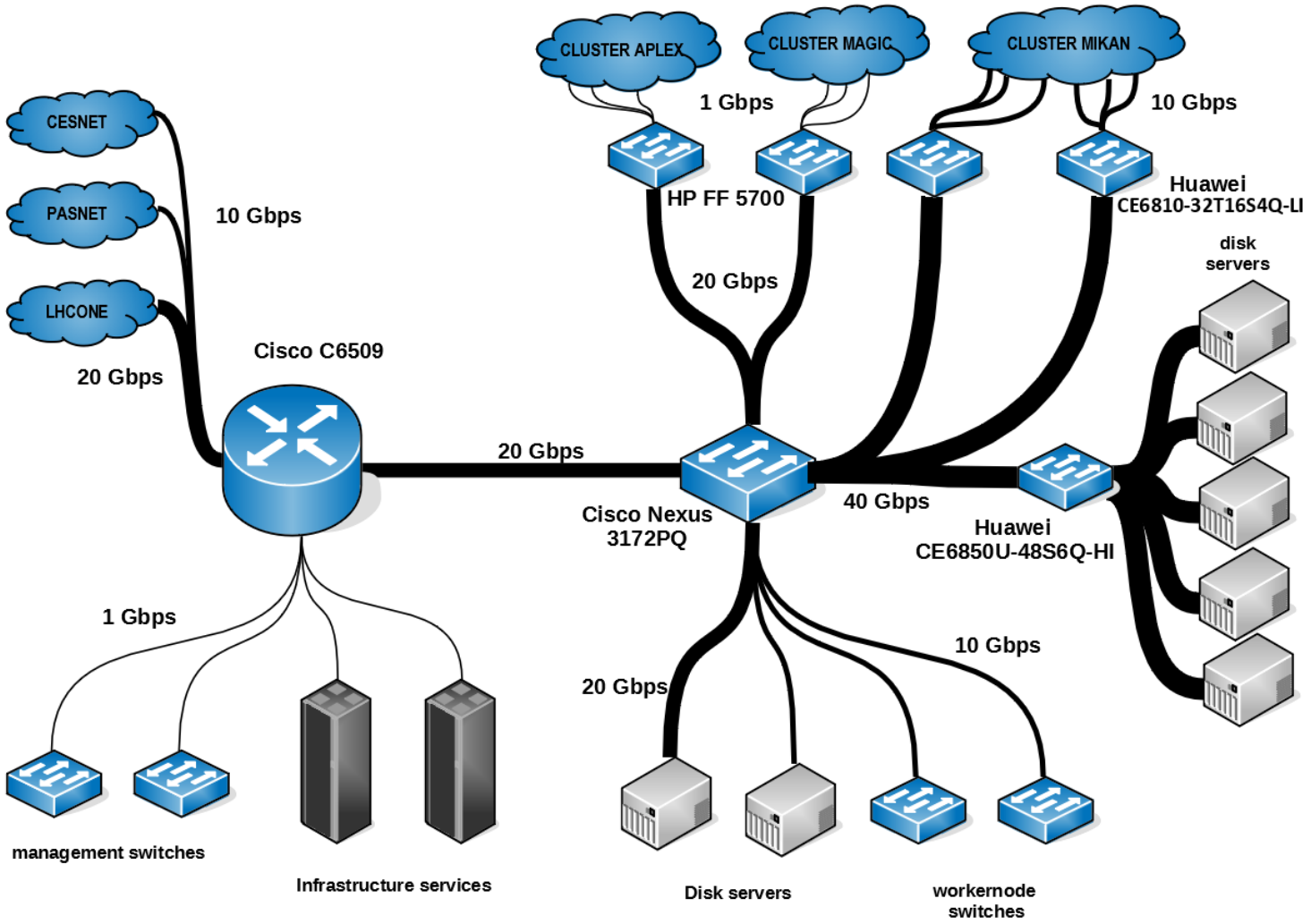
- 5 x server + JBOD
- servers Intel R2208WFTZS
  - 2 x 8 core CPUs Intel Xeon Silver 4108
  - 96 GB RAM, 2x Intel DC SSD 480GB
  - 4x 10GbE SFP+
- JBOD HGST 4U60 G2 Storage Platform
  - 58x HDD HGST Ultrastar 10TB
  - 450 TB usable, RAID6, 2 hotspares
- Switch Huawei CE6850U-48S6Q-HI
  - 48*10GE SFP+ ports, 6*40GE QSFP+ ports

- Performance
  - rebuild took 17 hours (fully occupied FS)
  - iozone: above 11 MB/s per thread (450 threads)
- SL6 (3 servers, fast deployment), C7 (2 servers)
- 4 DPM pools, 1 xrootd server

Old hardware

- Old (5 – 7 years) disk servers
  - new HBAs instead of more expensive controllers
  - new disks (in some servers)
  - enough spare disks
- zfs instead of xfs
  - applied on several old servers
  - later problems with
    - kernel updates
    - failed disk identification
    - manpower
- new servers use xfs

Network - topology



Network – IPv4 vs IPv6 protocols - DPM

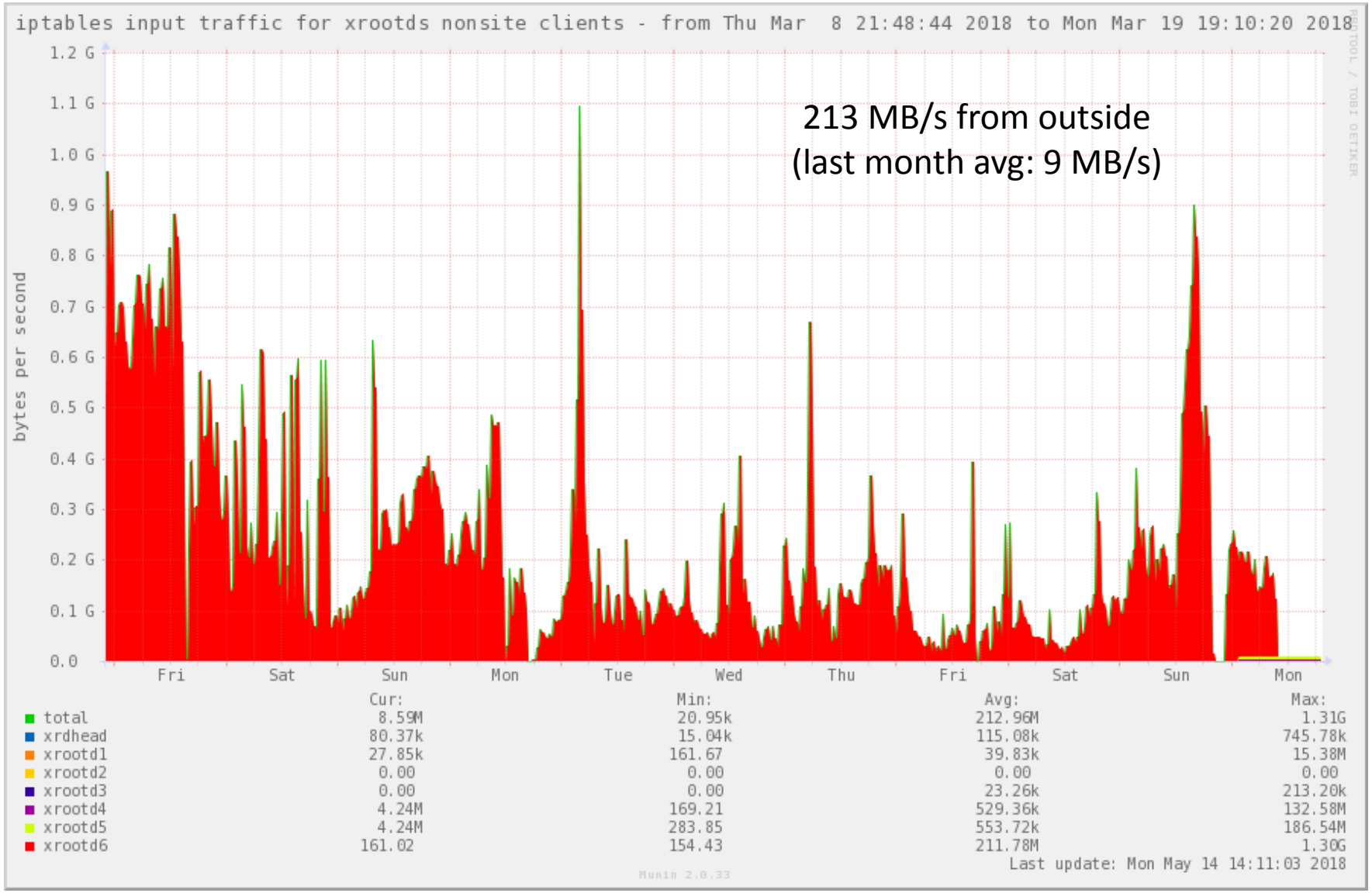
- Local traffic to/from DPM mostly via IPv6
  - Monthly averages:

	IPv4 [MB/s]	IPv6 [MB/s]
DPM in, local	6	30
DPM in, non-local	65	63
DPM out, local	15	485
DPM out, non-local	98	35

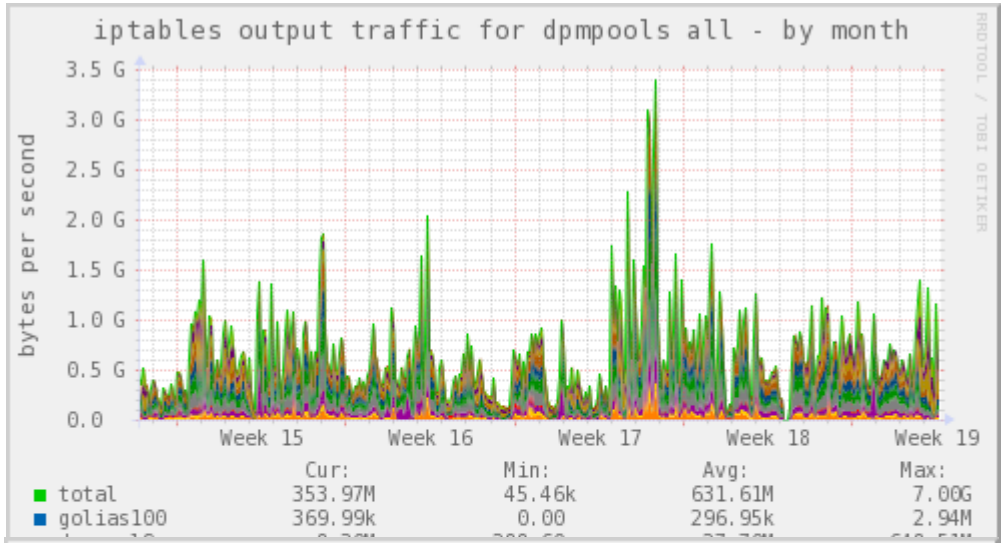
- Traffic to/from xrootd (ALICE) mostly via IPv4

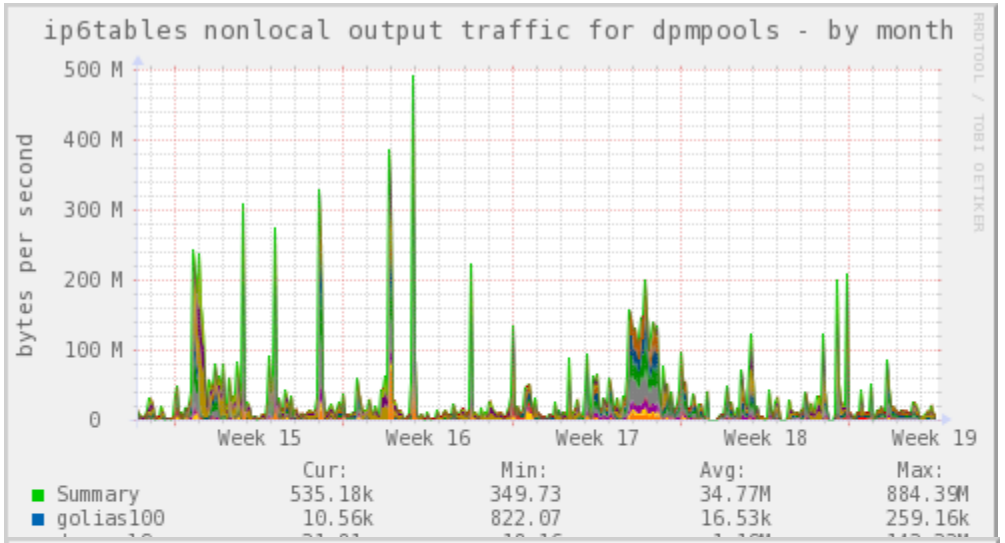
	IPv4 [MB/s]	IPv6 [MB/s]
xrootd in, local	2	0
xrootd in, non-local	9	0
xrootd out, local	354	0
xrootd out, non-local	28	0

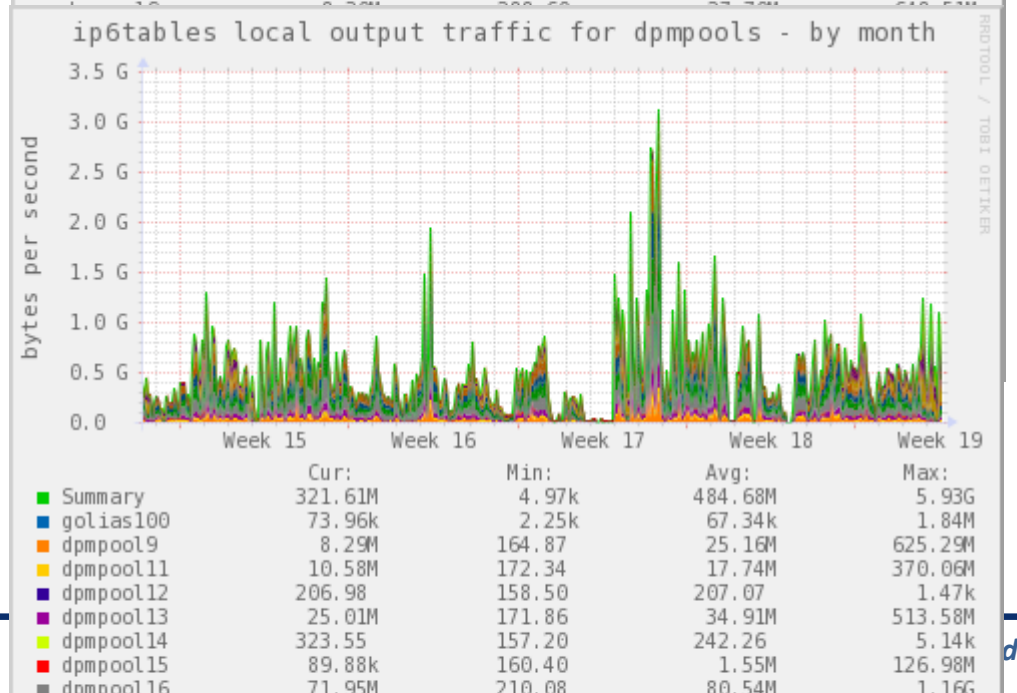
New xrootd server added

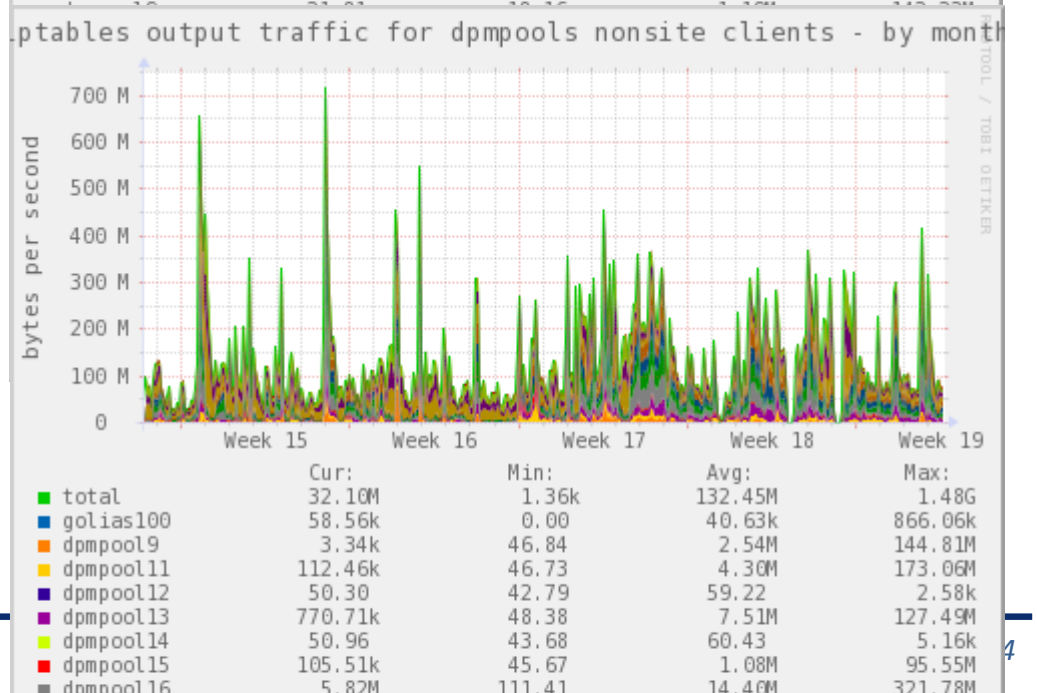


Network – IPv4 vs IPv6 protocols - DPM

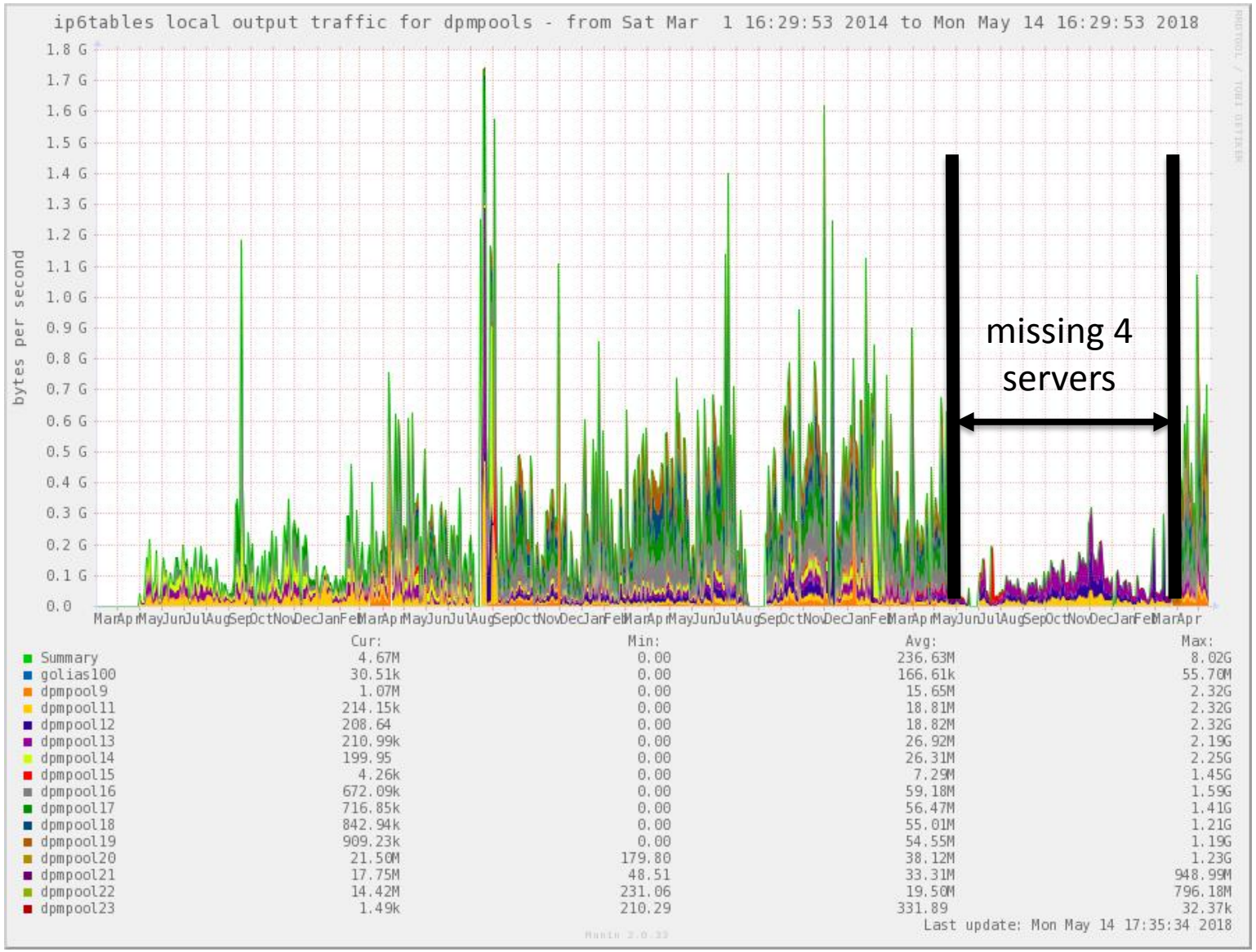




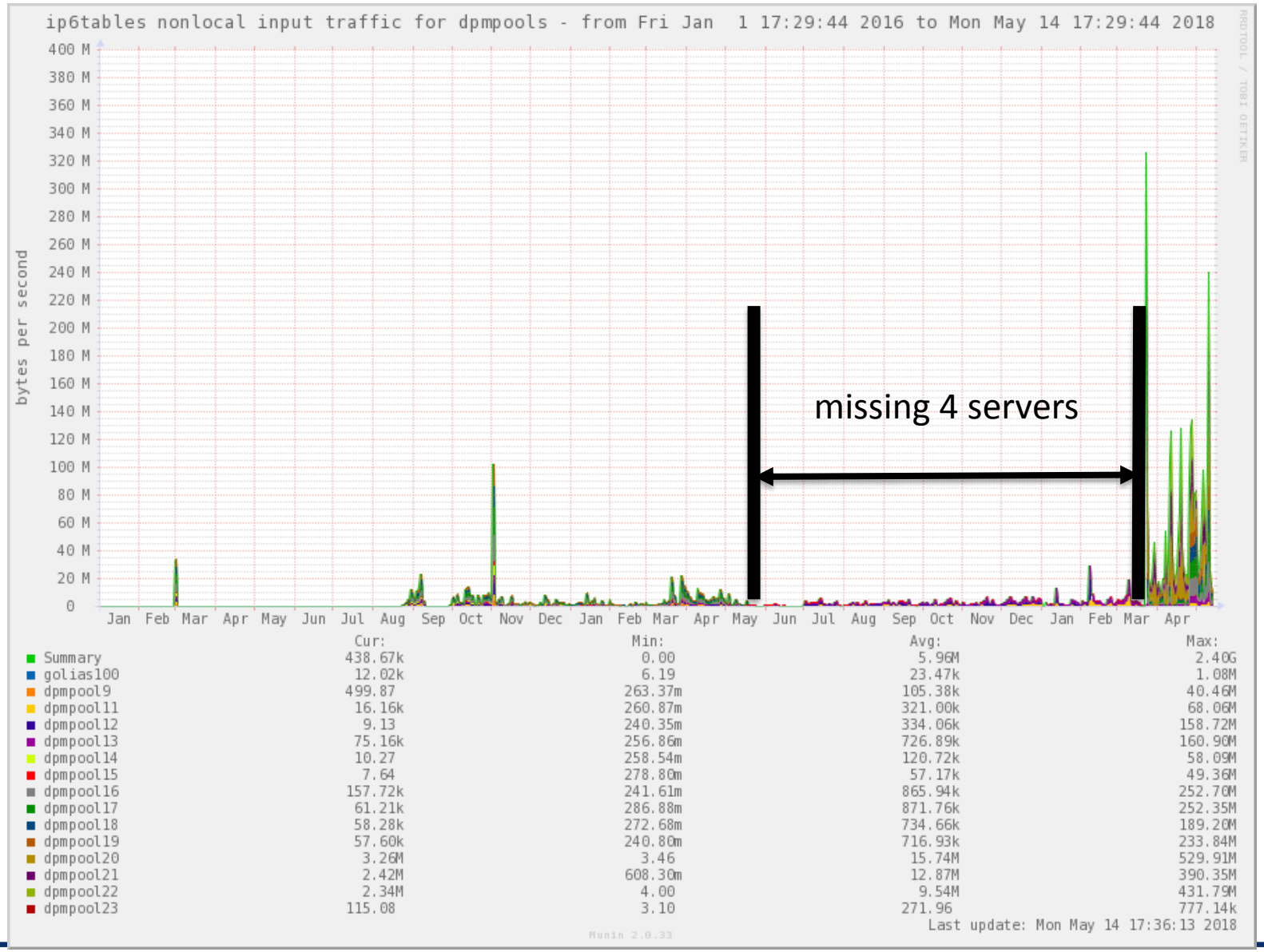




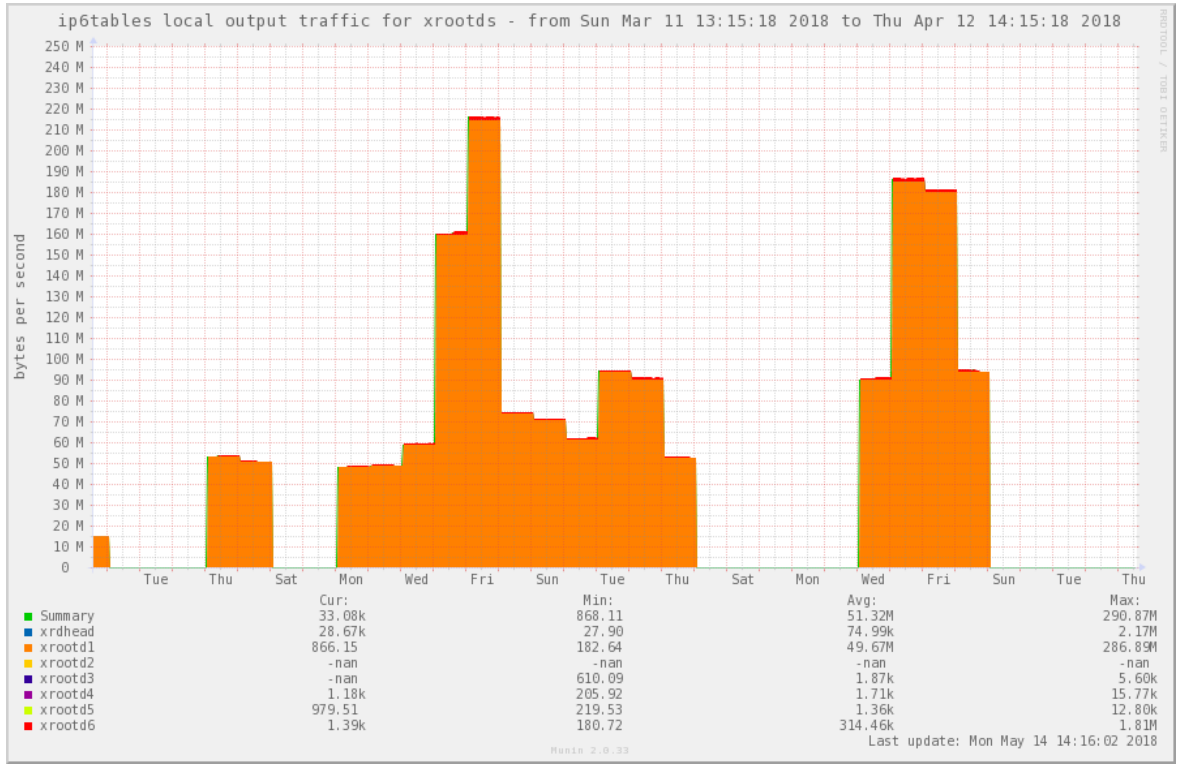
IPv6 local traffic - DPM



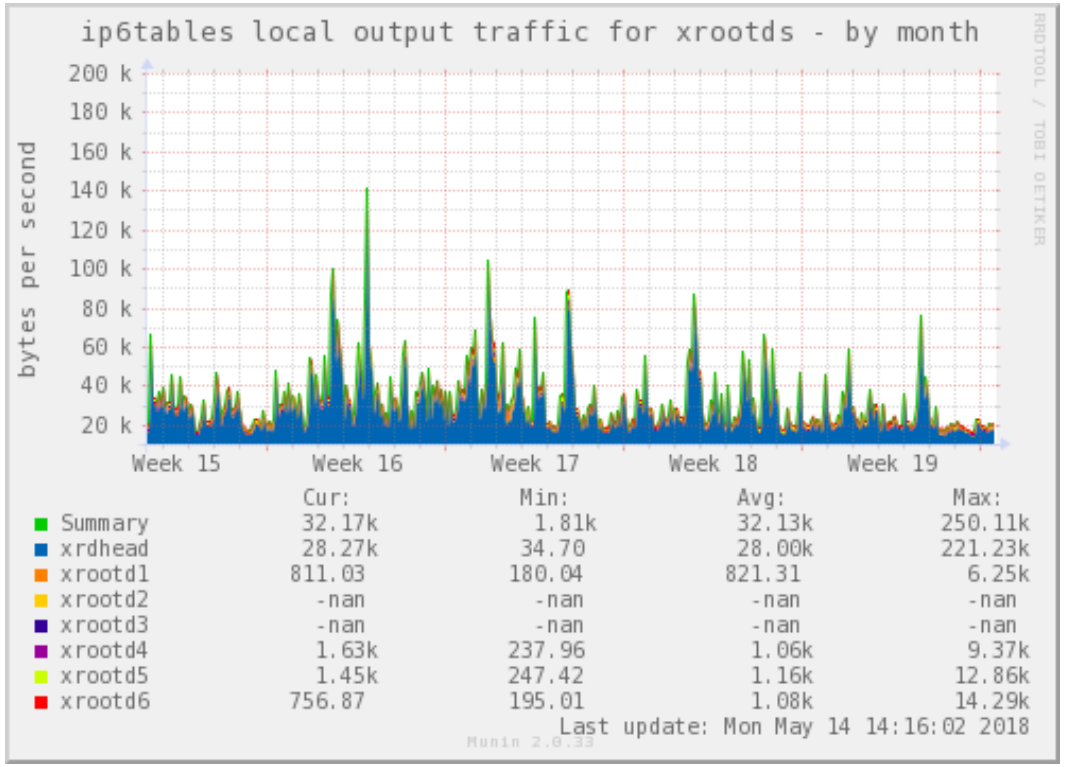
IPv6 non-local traffic - DPM



Network – IPv4 vs IPv6 protocols - xrootd



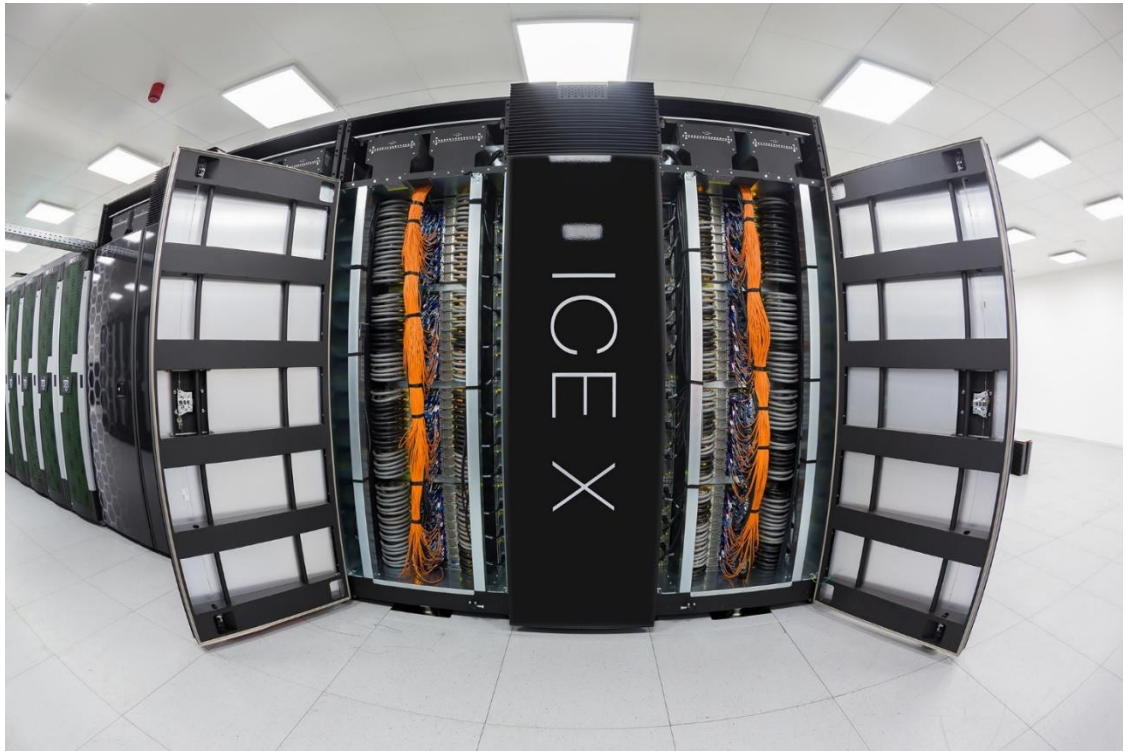
IPv6: rsync copy from xrootd1
(prepare to decommission]



IPv6: xrootd – only to xrdhead

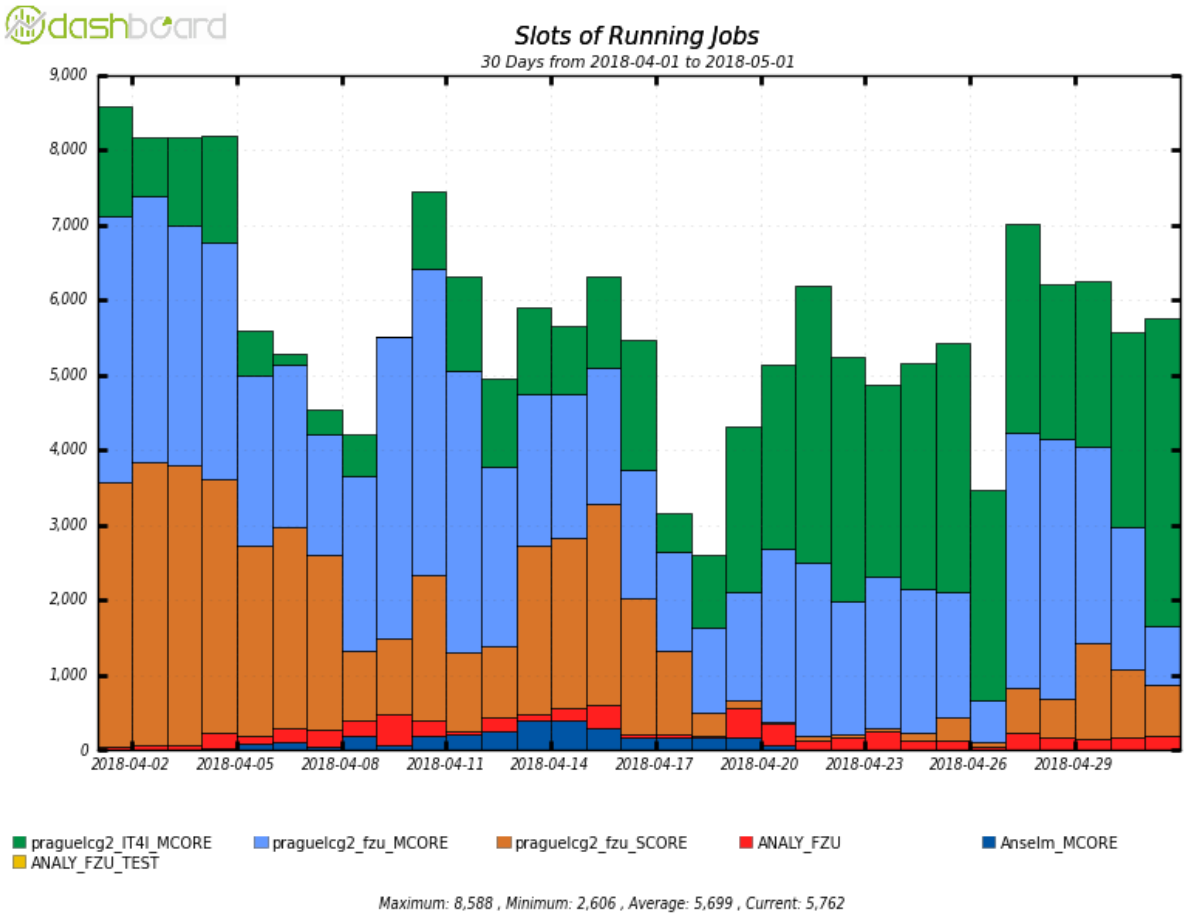
External HPC resources: IT4I

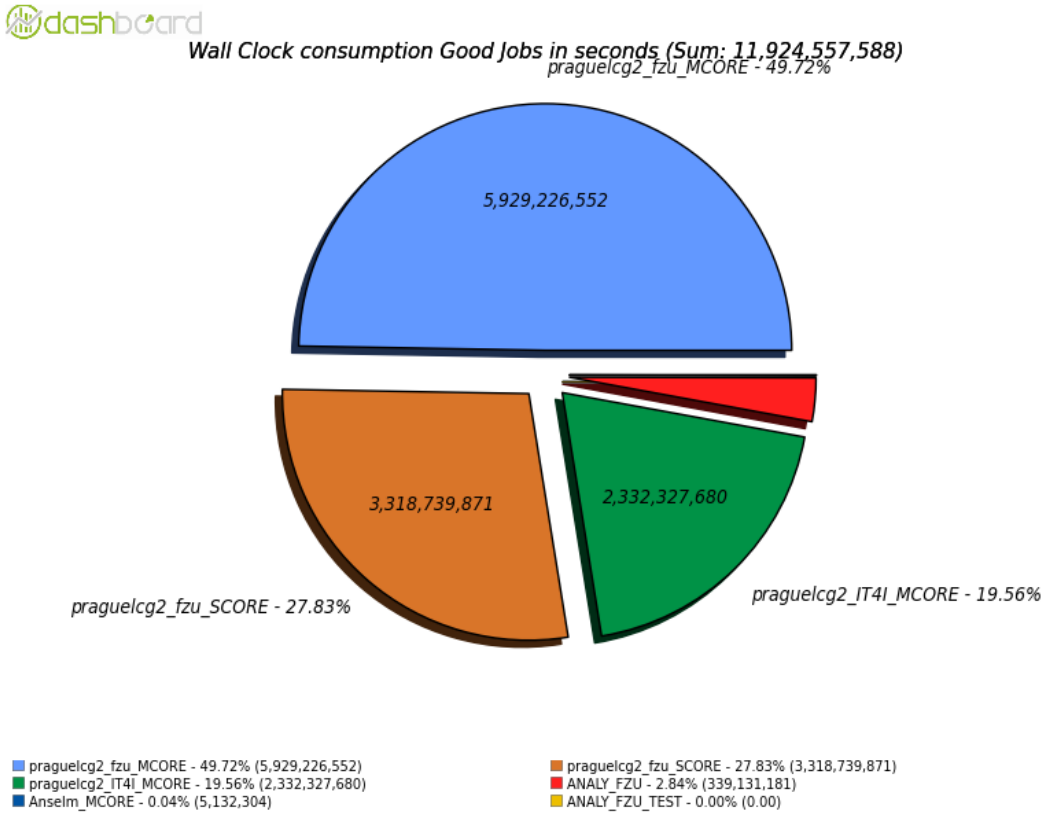
- IT4I – IT4Innovations
  - Czech National Supercomputing Center located in Ostrava (300 km from Prague)
  - Founded in 2011, first cluster in 2013
- Cluster Salomon – 2015
  - 2 PFLOPs peak perf – nr. 87 in 2017/11
  - 1008 compute nodes
- ATLAS jobs via ARC CE at FZU
  - Sw installed by rsync with the site CVMFS
  - A special Panda queue on praguelcg2



We greatly appreciate the possibility to use IT4I resources and very good support from IT4I team.

External HPC resources

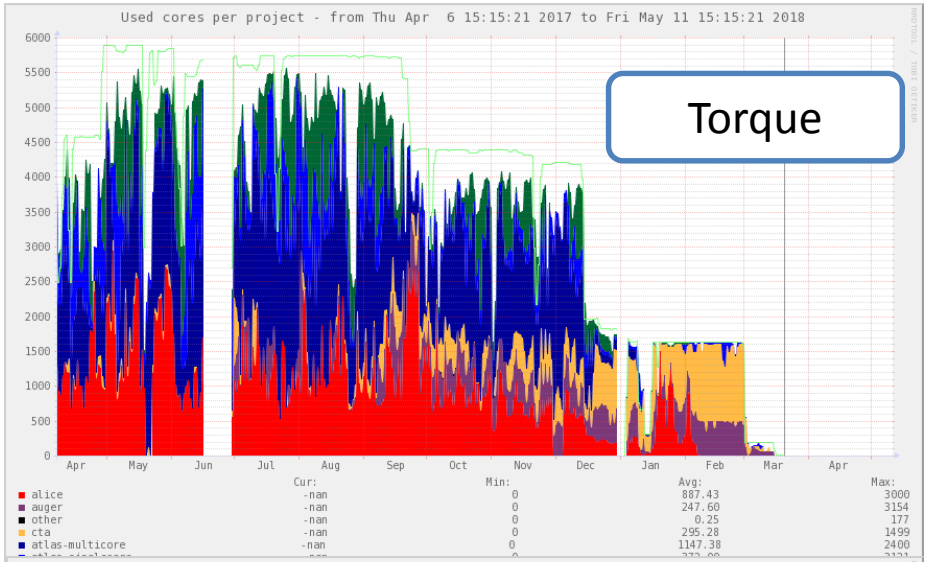


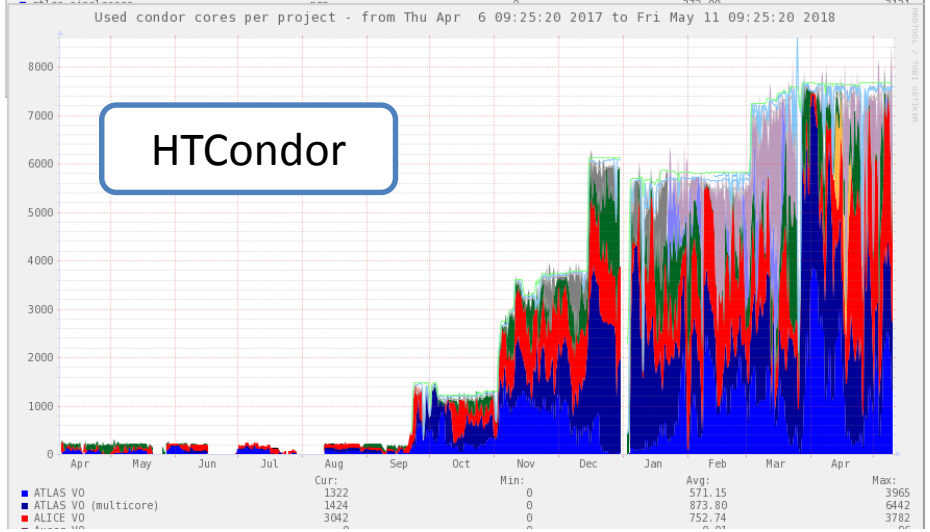


Almost 20% of the ATLAS CPU capacity delivered by IT4I (April 2018)

Torque/Maui -> HTCondor

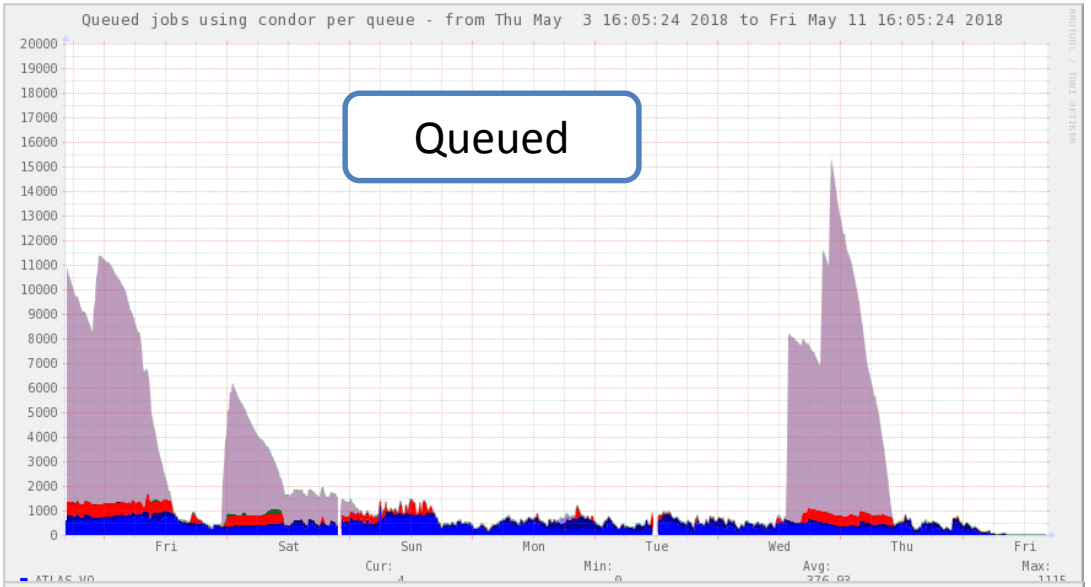
- Torque/Maui to HTCondor migration
  - Gradual migration during the last year
  - No major problems for local users
  - Also Cream CE -> ARC CE migration

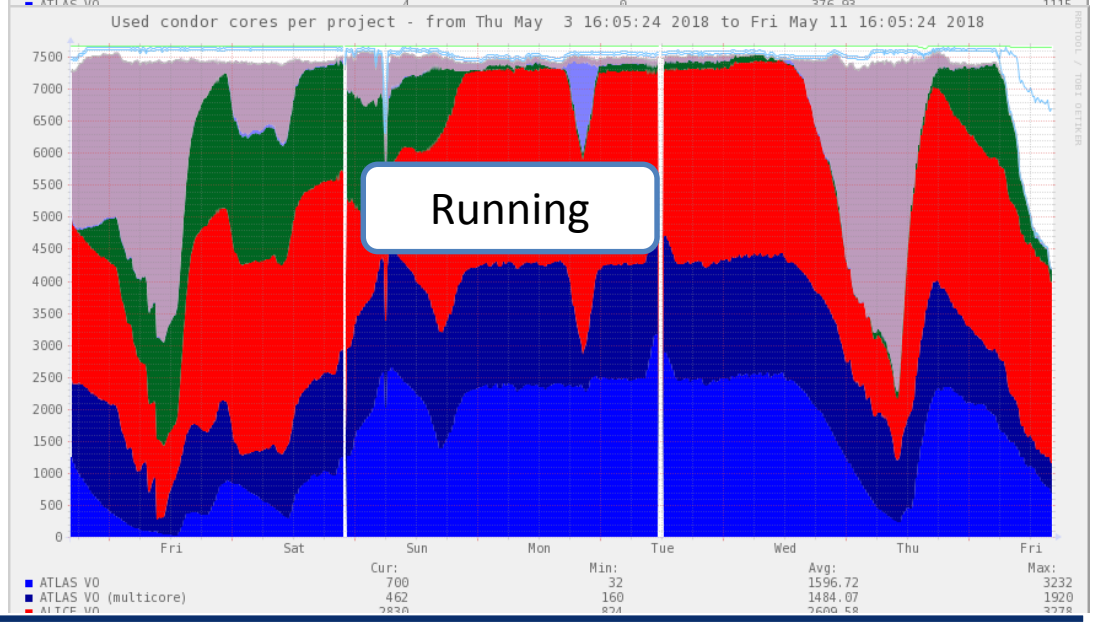




Torque/Maui -> HTCondor

- Torque/Maui to HTCondor migration
  - Still difficulties with fairshare settings





- Thanks to my colleagues:
  - From FZU: Martin Adam, Petr Horak, Alexandr Mikula, Michal Svatos, Vaclav Strachon (left), Jana Uhlirova
  - from Czech Technical University: Petr Vokac
  - from Nuclear Physics Institute: Martin Adam, Dagmar Adamova

Thank you for your attention!