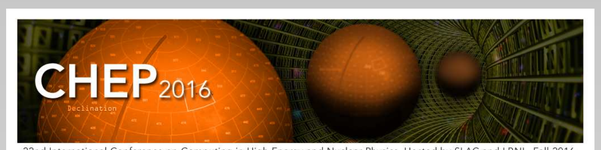
A Multipurpose Computing Center with Distributed Resources

J. Chudoba, M. Adam, D. Adamova, T. Kouba, A. Mikula, V. Rikal, J. Svec, J. Uhlirova, P. Vokac, M. Svatos

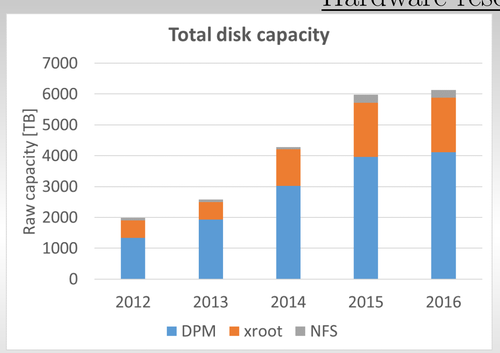


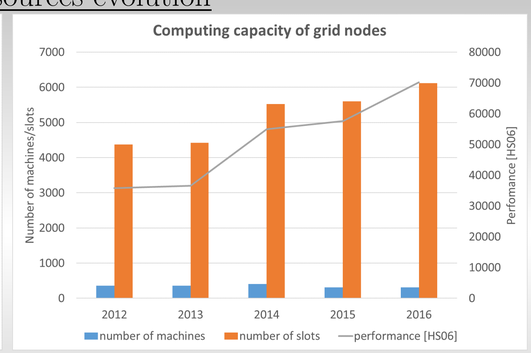
22nd International Conference on Computing in High Energy and Nuclear Physics, Hosted by SLAC and LBNL, Fall 2016

FZÚ

The Computing Center of the Institute of Physics (CC IoP) of the Czech Academy of Sciences serves a broad spectrum of users with various computing needs. It runs WLCG Tier-2 center for the ALICE and the ATLAS experiments; the same group of services is used by astroparticle physics projects the Pierre Auger Observatory (PAO) and the Cherenkov Telescope Array (CTA). OSG stack is installed for the NOvA experiment. Other groups of users use directly local batch system. Storage capacity is distributed to several locations. Computing clusters LUNA and EXMAG dedicated to users mostly from the Solid State Physics departments offer resources for parallel computing. They are part of the Czech NGI infrastructure MetaCentrum with distributed batch system based on torque with a custom scheduler.

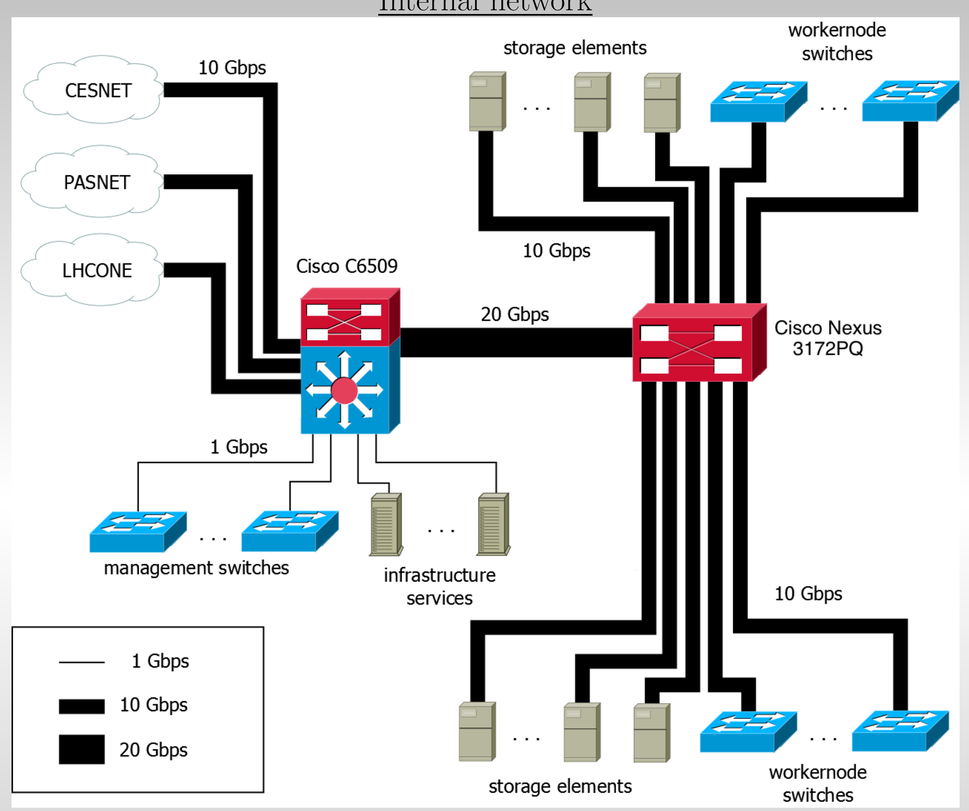
Hardware resources evolution





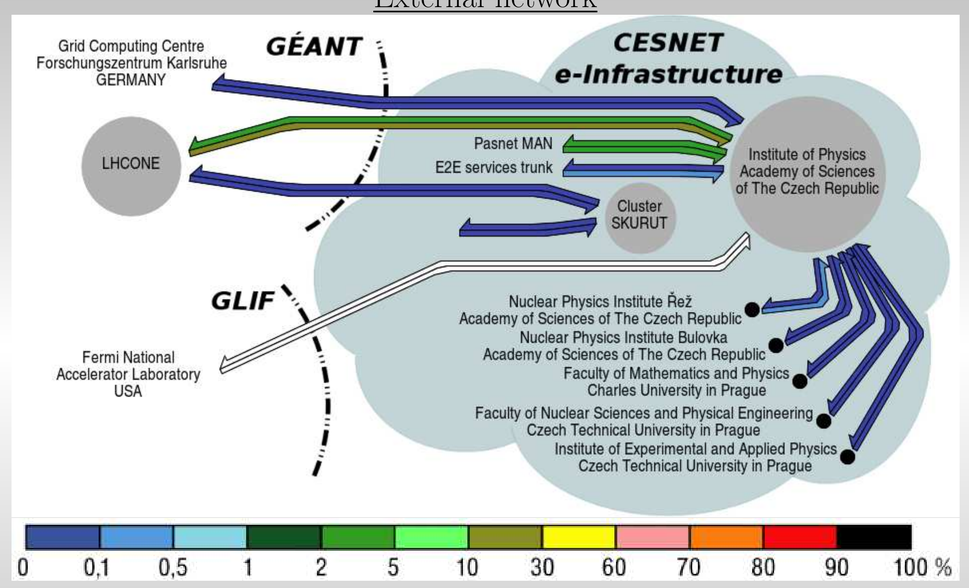
Resources are almost annually upgraded using budget of projects. All subclusters connected to the EGI grid are based on x86_64 technology, using the same OS (SL6), and jobs are distributed from single Torque server.

Internal network



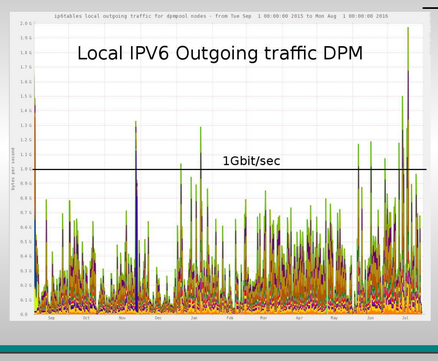
All disk servers are connected via 10 Gbps to the central CISCO Nexus 3172PQ switch. Worker node connection is 1 Gbps to local node switches which have 10 Gbps connection to the central switch. Clusters for parallel computing have independent internal network with outside connection just 1 Gbps. Parallel jobs use QDR Infiniband (not shown).

External network



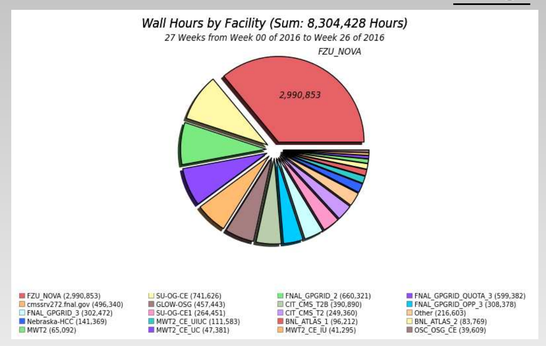
Major load on external connection comes from LHC Experiments ATLAS and ALICE. The 10 Gbps LHCONE link is often saturated and will be upgraded to 2×10 Gbps in 2016.

IPv6 networking

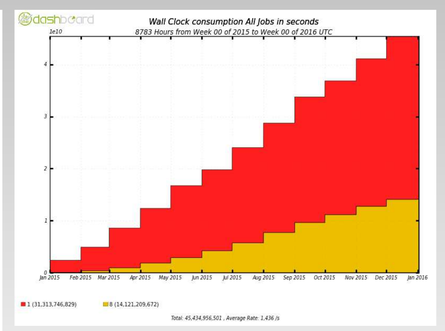


All worker and storage nodes have IPv4 and IPv6 addresses. All local transfers via gridftp protocol utilise preferentially IPv6.

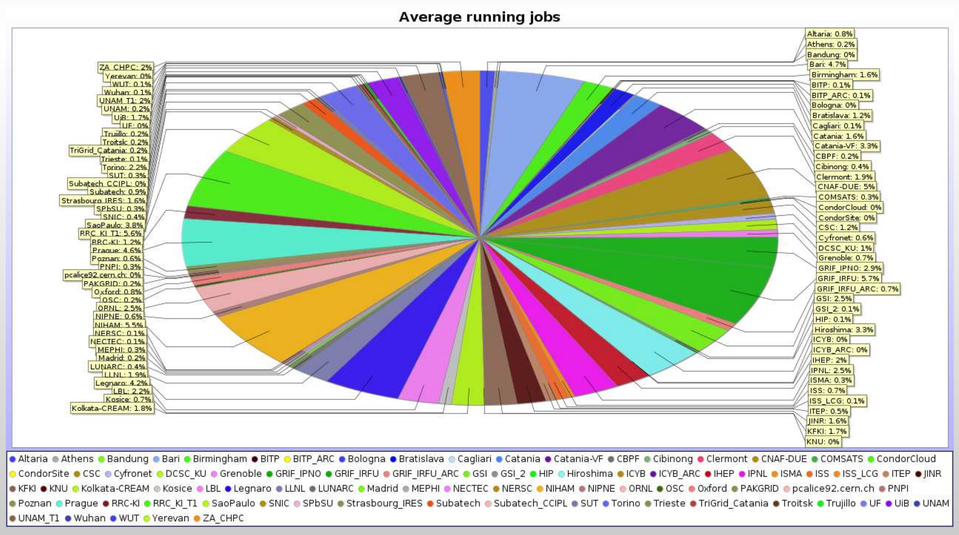
Usage



Institute of Physics (FZU) is a major contributor to NOvA simulations on OSG.



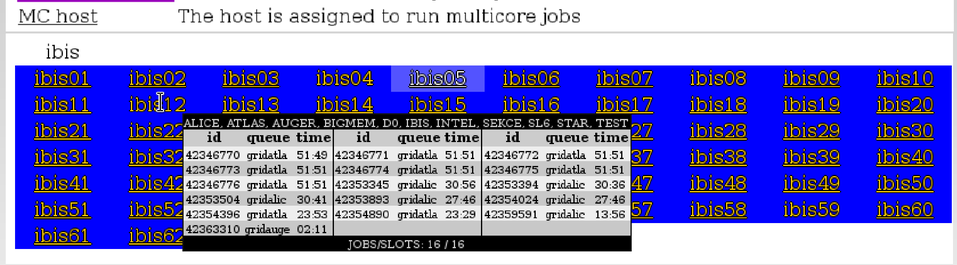
The ATLAS experiment submits single core or 8 core jobs. Local users use only single core jobs for analyses.



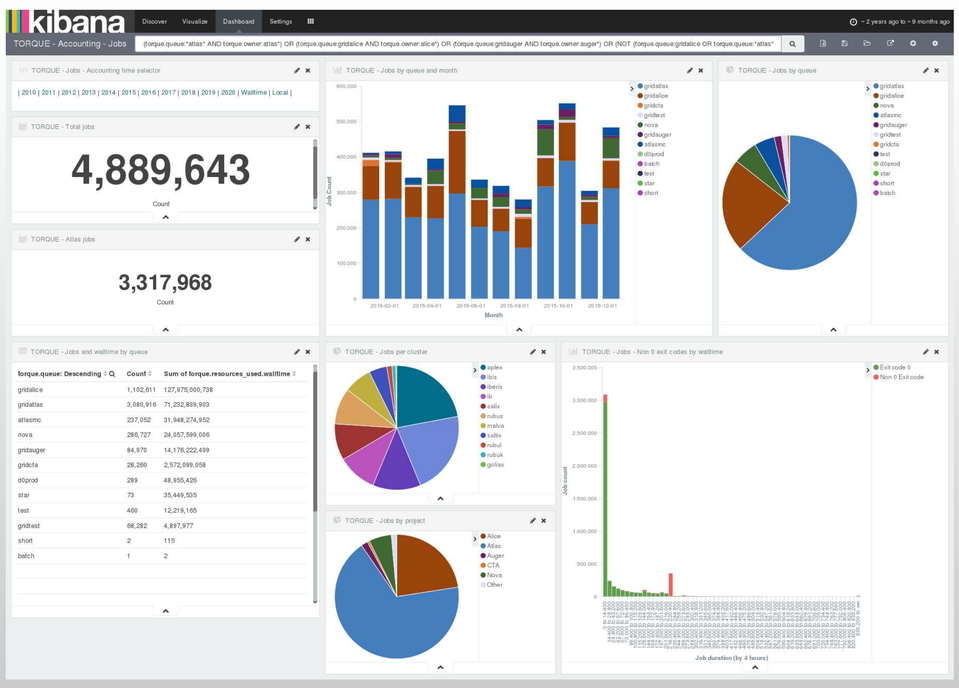
Contributions of ALICE Tier-2 sites to the total CPU time consumption in 2015. The Prague site delivered 4.6%.

Monitoring

State	Description
Down	Node is inaccessible by the torque server
Offline	Node is intentionally out of production
Offline with jobs	Offline with some jobs still running
Empty	Node is ok, but with no jobs
Empty Full	Node has jobs
Unknown	State is unknown - see output of pbsnodes
MC host	The host is assigned to run multicore jobs



Many different tools are used to monitor the whole infrastructure. Nagios results are checked via web browser and summary of issues is regularly sent via emails. Munin, Ganglia and Observium store plots for later debugging and performance comparison. A custom script displays occupancy of individual worker nodes and lists jobs running there.



Elastic Stack is used to process all relevant log files. The figure shows an example with accounting graphs and figures from Torque log files.