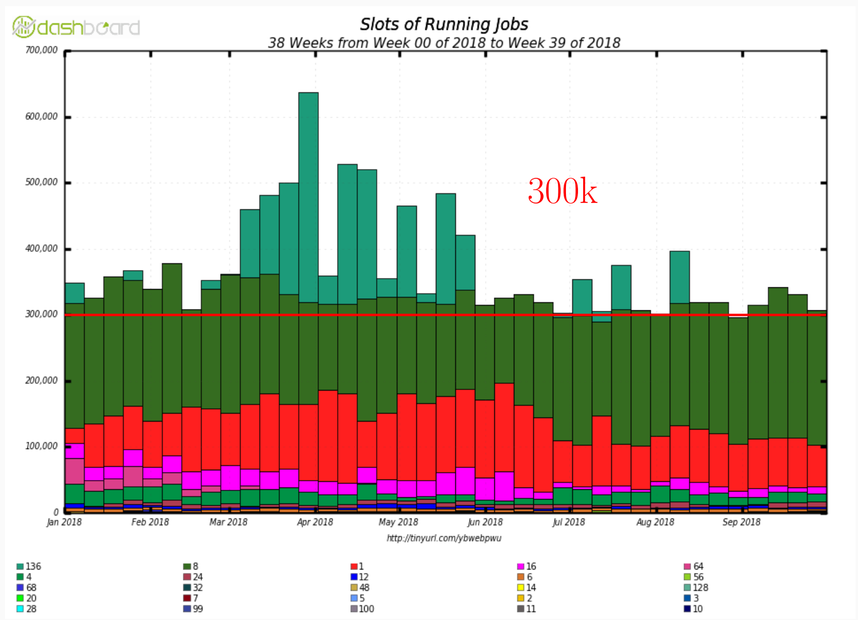
Usage of Salomon HPC in the ATLAS Distributed Computing

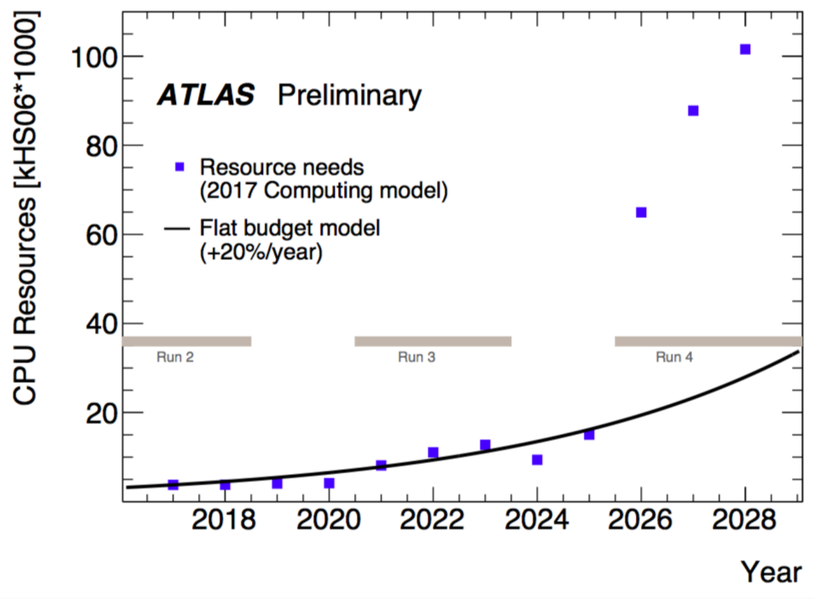
M. Svatoš, J. Chudoba, P. Vokáč

2nd Users Conference of IT4Innovations

7.11.2018

ATLAS Distributed Computing

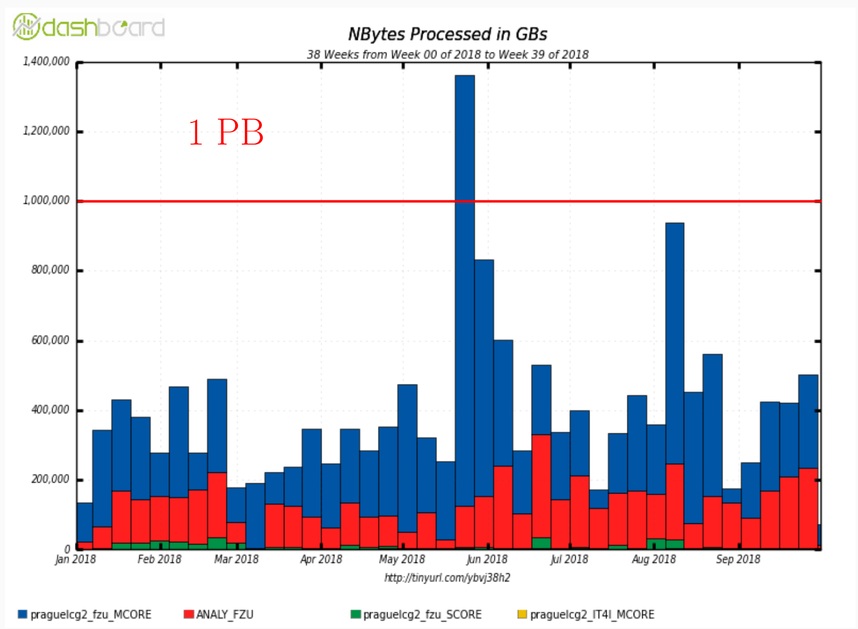


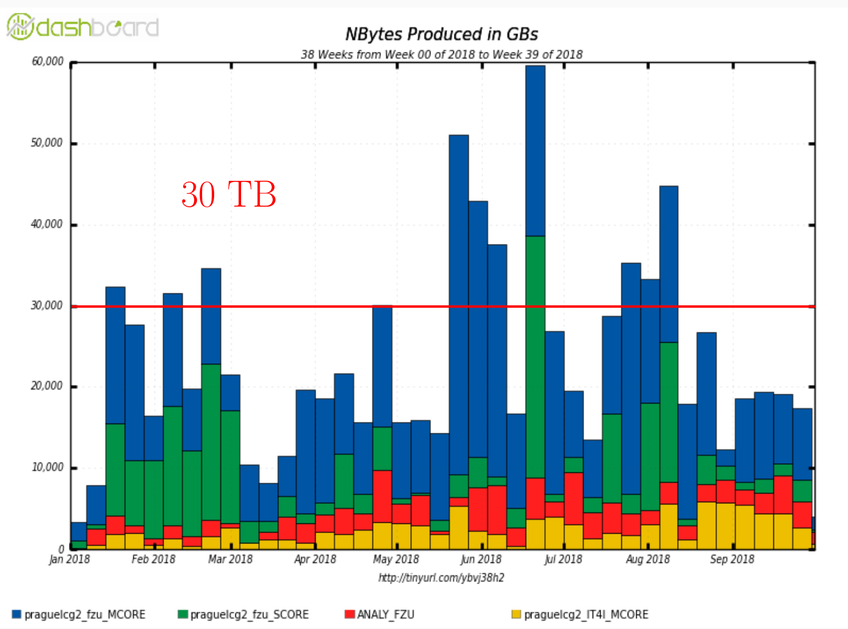


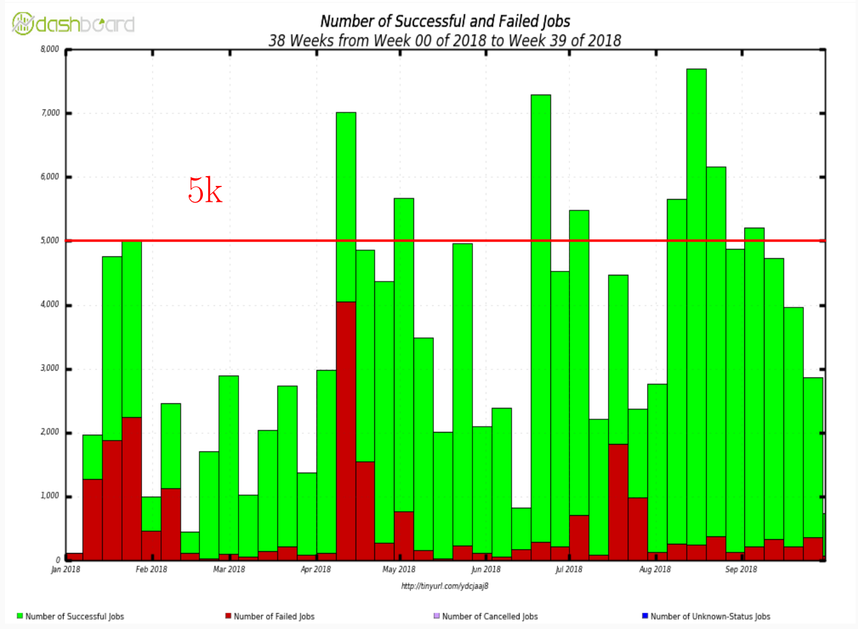
The ATLAS Experiment is located at the Large Hadron Collider (LHC) at CERN near Geneva. Its physics programme consists of analyzing products of elementary particles collisions. One of its greatest achievements was discovery of the Higgs boson.

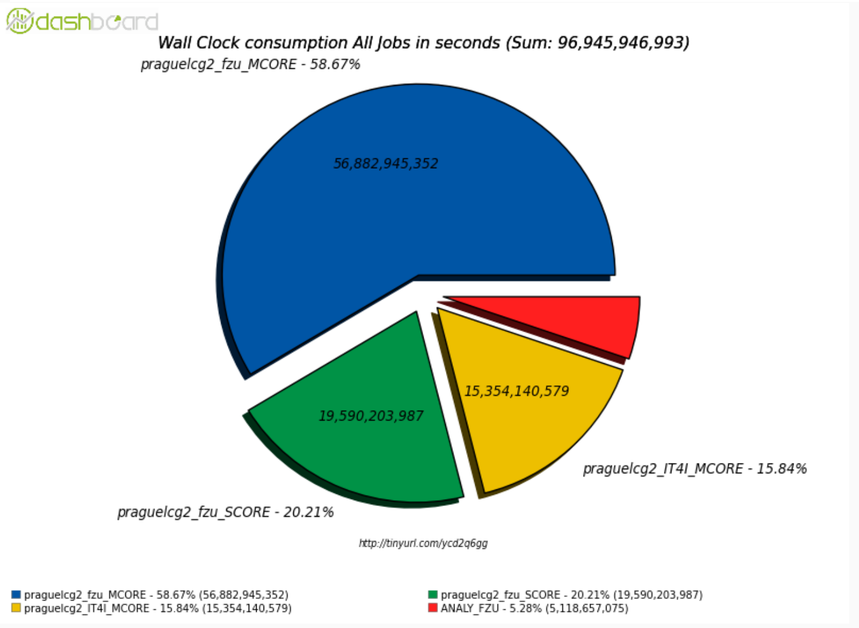
The ATLAS Distributed Computing (ADC) manages more than 400 PB of data spread on more than 700 storage endpoints (on more than 150 sites located around the world). To process the data, the ADC uses heterogeneous computing resources consisting of Worldwide LHC Computing Grid (WLCG) sites, cloud resources, HPCs, volunteer computing (BOINC) resources, etc. (Figure on the left shows amount of CPU cores used by ADC since the beginning of this year as a function on number of cores used by the job). Future upgrade of the LHC will cause increase in amount of data that needs to be processed and analyzed. If the 2017 computing model is used, required computing resources would need to increase by one order of magnitude (Figure on the right).

Job statistics (January-September 2018)





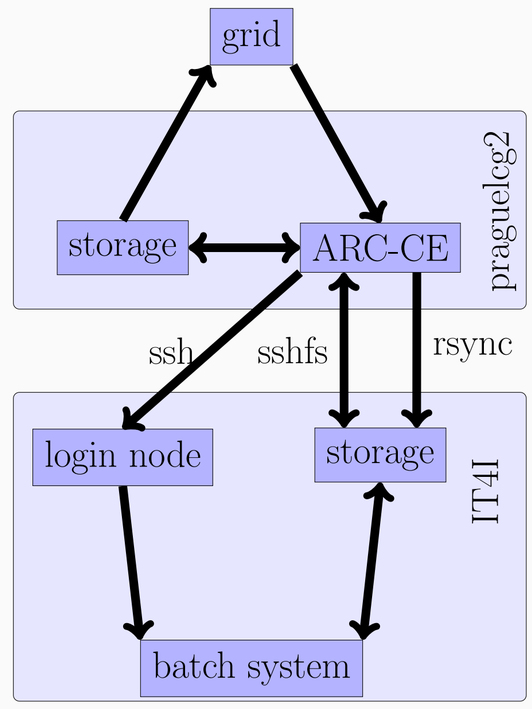




- Amount of input processed at the HPC is usually below or around 1 TB per week. Amount of output produced at the HPC is usually few TB per week. These values are negligible in comparison to other praguelcg2 queues (top Figures: input on the left, output on the right).
- Thousands of jobs finish every week (bottom left Figure shows successful and failed jobs on the HPC queue) - the major causes of job failures at Salomon are: PBS filehandle leak (until first half of February), installation issue (April) which wasted almost no walltime, and testing of file transfer setting (July)
- Comparison of wallclock consumption of all jobs (bottom right Figure) of all active praguelcg2 queues shows that the HPC provides significant amount of resources (about 15%) even with batch system limit of 100 submitted jobs.

Settings

- ARC-CE
  - located at praguelcg2 site
  - receives job description from ARC Control Tower (aCT) submission system
  - translates aCT job description into script that can be run in batch system
  - downloads input files
  - puts all into folders that are shared between ARC-CE and dedicated scratch space via sshfs
  - submits job via ssh connection to a login node
  - uploads output files after the job has finished
- application software is synchronized via rsync between CVMFS available on ARC-CE machine and scratch space
  - current size is slightly over half TB in $\sim$13 million files
- cache of input files located on scratch space allowed faster filling of available slots
  - without cache it took $\mathcal{O}(1)$ h to create 100 jobs
  - with cache it takes $\mathcal{O}(1)$ m to create 100 jobs
  - files which are not used for more than 60 day are deleted
  - size of the cache stabilized at $\sim$3 TB



Acknowledgement

Computing resources as co-financed by projects Research infrastructure CERN (CERN-CZ) and OP RDE CERN Computing (CZ.02.1.01/0.0/0.0/16013/0001404) from EU funds and MŠMT.

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

MINISTRY OF EDUCATION, YOUTH AND SPORTS