



# CERN Multi-site T0 & Job efficiency

**Status Report**

**Manuel Guijarro / CERN IT-PES**

- Background
- Network to/in Wigner
- Results so far
- Low Job efficiencies
- Conclusion

- +1000 Worker nodes in WIGNER
- Aim at understanding Batch Job efficiency
  - Geneva vs Wigner
  - Job efficiency = CPU time/Wall Clock time
  - Focusing on SLC6 virtual worker nodes
    - Not yet virtual vs physical worker nodes
    - Not yet SLC6 vs SLC5
- Contributors:
  - IT experiment representatives for Atlas/CMS/LHCb/Alice
  - CERN IT groups: CS, DSS, PES, SDC
  - **Alessandro Di Girolamo, Dirk Duellmann, Stefan Roiser, Edoardo Martelli, Valentina Mancinelli, Edward Karavakis, Ulrich Schwickerath, Nicolo Magini, Jan Iven, Jerome Belleman, Maarten Litmaath, Bernd Panzer-Steindel, Spyros Lalos, Manuel Guijarro**

**Two issues impacted the network in the past, now solved:**

**100G WAN link flapping:** one Geneva-Budapest link going down for 5 seconds every 2 days for several months. Cause eventually identified in the Carrier's transmission devices. Last faulty item replaced on the 11<sup>th</sup>/4/2014.

**10G NIC-switch incompatibility:** random packet drops because of incompatibility of certain NICs connected with Twinax cables. Workaround: identified a switch firmware that fix the issue. Never impacted production machines.

## **perfSONAR installations:**

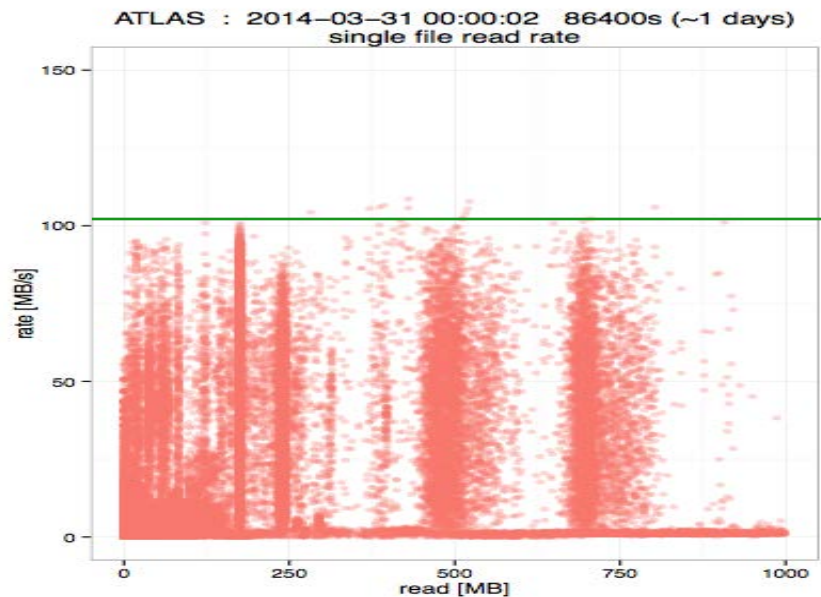
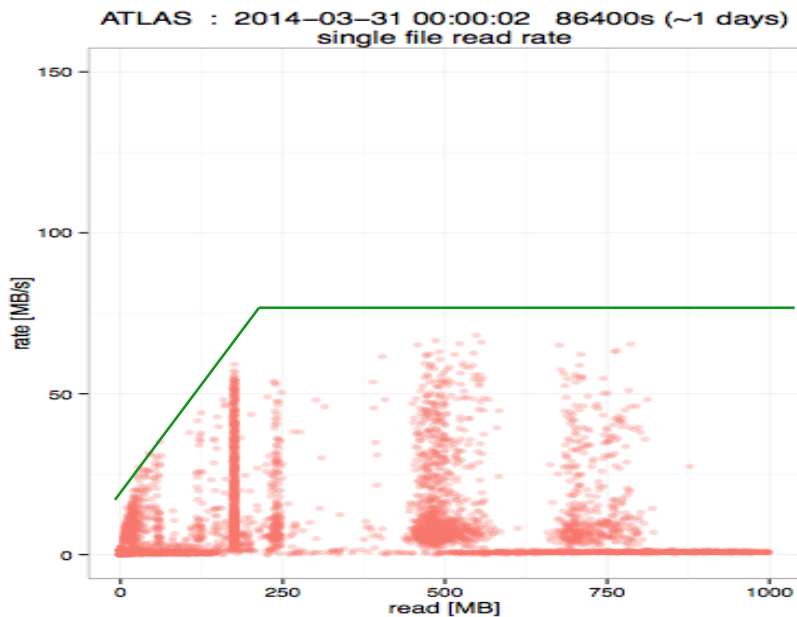
**WLCG:** one bandwidth and one latency probe available in Geneva for WLCG. Recently upgraded to the latest perfSONAR-ps version

**Wigner-Geneva:** four probes being installed. Will run SLC6 with standard TCP parameters. Full mesh of bandwidth measurements: 1G-1G 1G-10G, 10G-10G, Wgr-Gva, Wgr-Wgr, Gva-Gva.

**Availability: end of April 2014**

- Data was not always reliable
  - collected during crisis time
  - covering more than virtual and SLC6
- Atlas production (Jan 2014):
  - Meyrin 93% vs Wigner 87% ; 6% loss
- Atlas analysis (Jan 2014) : 20% less efficient
  - Probably due to not using XROOT TTreeCache
- LHCb: Meyrin 97% vs Wigner 94%
  - Montecarlo; 100 evts; No input; Output to EOS
  - 20k jobs @ Meyrin and 5k jobs @ Wigner in Jan 2014

- Courtesy of Dirk Duellmann
- Atlas jobs accessing EOS storage
- Default TCP parameters
  - Some tuning potential
- For well behaved jobs, minimal impact



- Low job efficiencies
  - Zombie pilot jobs
    - Block job slots: block resources
    - Only freed once passed batch queue limit
    - Working with experiment representatives on resolution
  - Some users transfer data from/to remote locations
    - Awful efficiency
    - Yet batch end user may be happy
  - Automatic monitoring, detection of low efficiency jobs in place and automatic notification to user concerned
- Some batch applications optimised for Intel CPUs
  - CERN Batch contains Intel and AMD
  - Wigner is mainly AMD



- No indication of any significant difference of job efficiency between CERN Geneva and Wigner
- Monitoring significantly improved
- Next steps:
  - Intel vs AMD
  - Virtual vs Bare Metal
  - SLC5 vs SLC6