Bringing ATLAS production to HPC resources
A use case with the Hydra supercomputer of the Max Planck Society

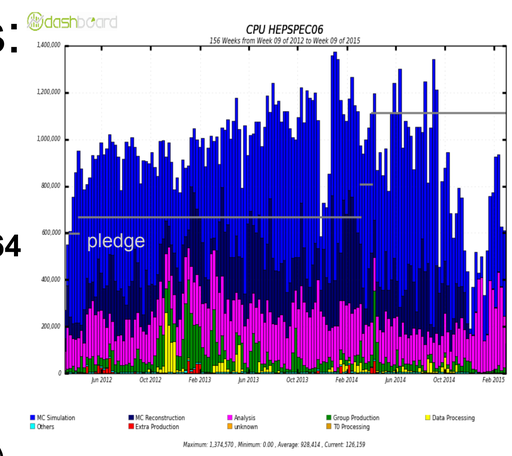
On the behalf of *ATLAS Collaboration*
Authors:
Dr. Stefan Kluth (Max Planck Institut für Physik)
Dr. John Alan Kennedy (Rechenzentrum Garching)
Dr. Luca Mazzaferro (Max Planck Institut für Physik)
Dr. Rodney Walker (Ludwig Maximilians Universität München)

Objectives:

1

- Demonstrate the possibility to use **HPC** resources for **ATLAS** grid jobs.
- Opportunistic usage of **HYDRA** Supercomputing resources

Why **HPC** for **ATLAS** jobs:



- **RUN2:** expect that required CPU will **exceed** the computing resources **pledged** to **ATLAS**, as observed in Run1.
- Many modern **HPC** systems have linux **x86_64** OSes installed
  - **Compatibility** with SL6
  - **Simulation jobs** preferred
- Possibility to use resources in **opportunistic way** (backfilling otherwise unused resources.)

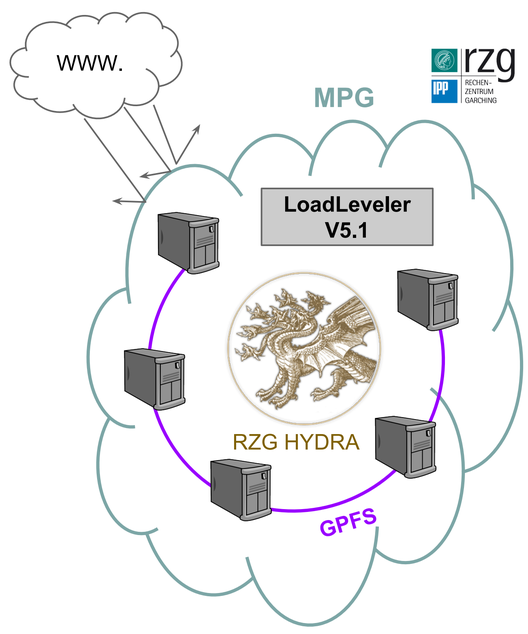
Technical Challenges:

2

- **No outbound IP** connectivity from WNs
- **No local** compute **node disk**
  - Shared File System for workdir and software install
- Grid-services/Gatekeeper crude or missing:
  - HPC users use ssh to login, move data and submit jobs
- Scheduling usually for many nodes/cores.

The **Hydra** Supercomputer:

3



- **Intel-based IBM iDataPlex** HPC supercomputer hosted at the **Rechenzentrum Garching**
- Used to run parallel jobs with **high number** of **cores** and **memory** requirements.
- For security reasons the **access to HYDRA is generally only allowed from within the MPG network**.
- Batch System: **IBM LoadLeveler**
- **No local disk => GPFS** parallel file system is used to share data between the nodes.

Specifications:

- ~ 83000 Cores organized as follows:
  - 3500 nodes Intel Ivy Bridge processors: 20 cores @ 2.8 GHz each
    - 350 nodes are equipped with accelerator cards (NVIDIA K20X GPGPUs and Intel Xeon Phi cards each).
  - 610 nodes Sandy Bridge-EP processors: with 16 cores @ 2.6 GHz each
- Main Memory: 280 TBytes
- Network: InfiniBand FDR14
- Aggregated Peak performance: 2.8 PFlop/s
- OS: **SLES 11 sp3**

Conclusions:

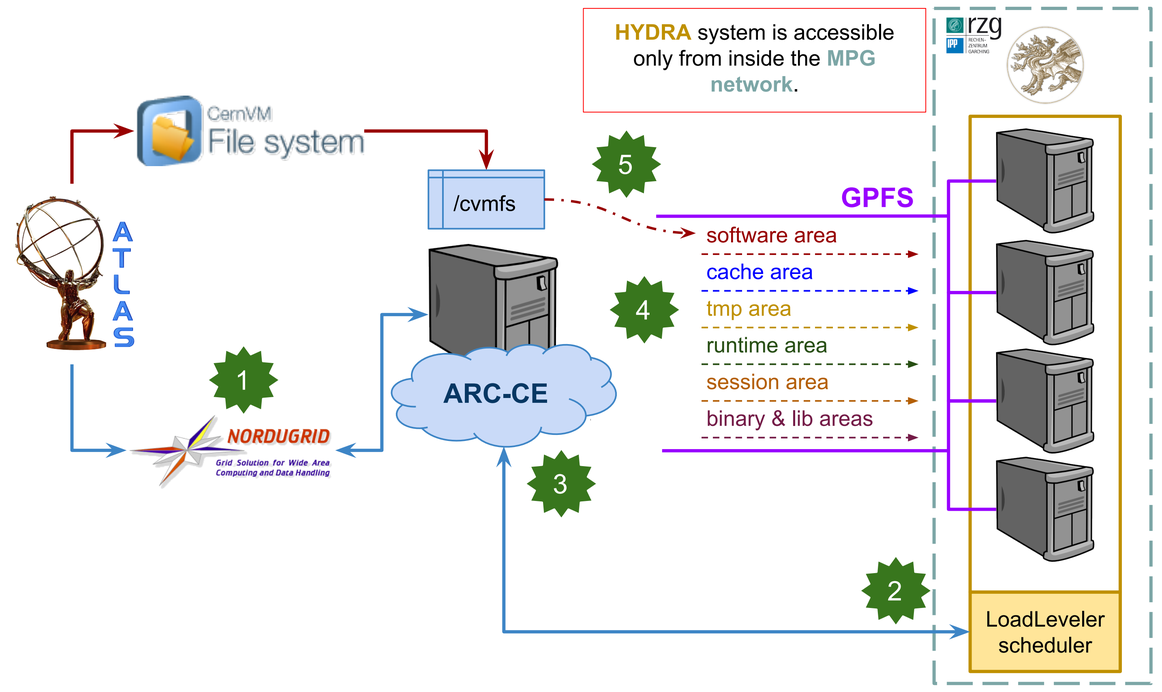
- The **ATLAS**-**HYDRA**-system is **running** in stable production since February 2014
- Its **contribution** to the ATLAS computing reaches a significant level compared with MPPMU Tier2 (~24% of CPU consumption).
- Mainly running MC simulation jobs.
- Good job efficiency (~68%) and low failure rate (~12%).

Room for improvement:

- **ATLAS MonteCarlo code improvements** for multi-core jobs.
- **backfill/preemptive** job deletion.
  File system usage **optimization**.

Integrating **HYDRA** into the **ATLAS** grid

4

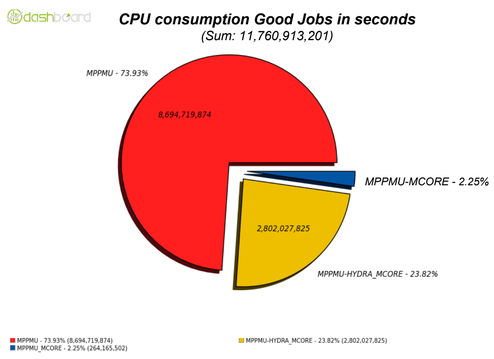


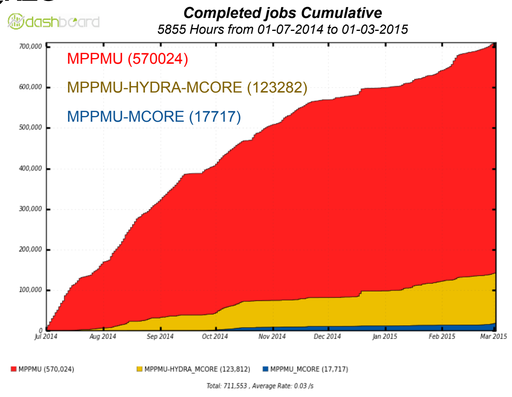
1. **ATLAS** submits jobs via **arcControlTower** which interacts with the ARC Computing Element (**ARC-CE**).
2. **ARC-CE** "translates" the job description file to be processed by LoadLeveler and submits the job.
3. **ARC-CE** also takes care of:
   a. **monitoring** the job status;
   b. **file staging** in and out;
   c. **providing informations** about jobs to the grid.
4. The **GPFS shared storage** area provided for both **ARC-CE** and **HYDRA** WNs.
5. The **ATLAS software** from **CVMFS** is **synchronized** into the **GPFS shared area**.
   a. Required software needed inside the shared area before running jobs.

Results*:

5

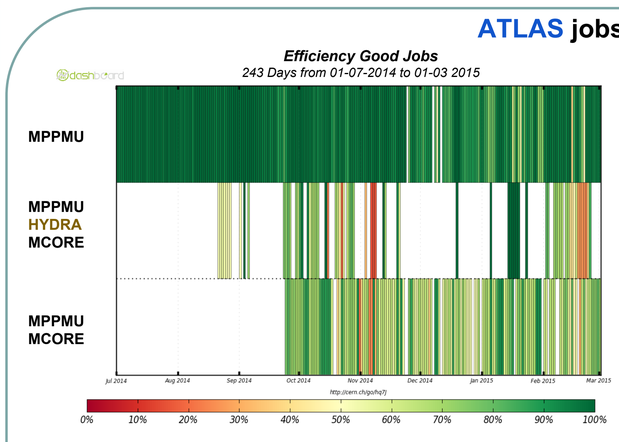
Impact of HYDRA on ATLAS computing compared to the MPPMU ATLAS Tier2@RZG

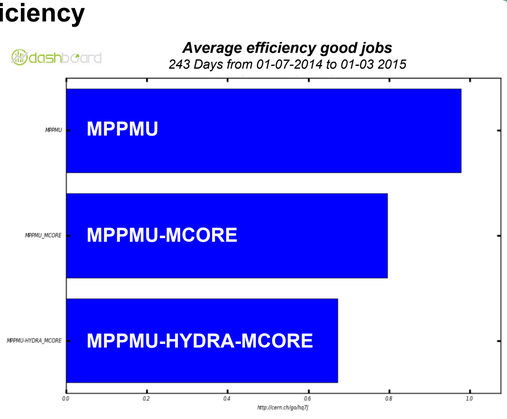




The impact of HYDRA is around 24% in terms of CPU consumption and 20% in terms of jobs completed

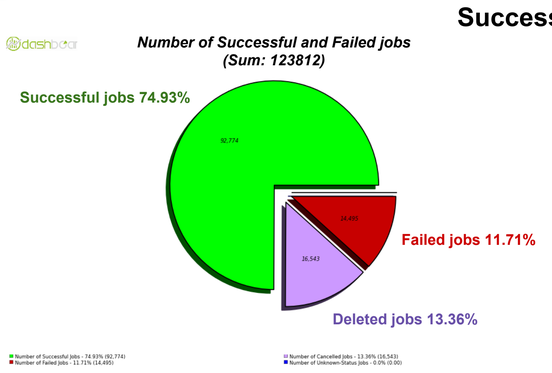
ATLAS jobs efficiency

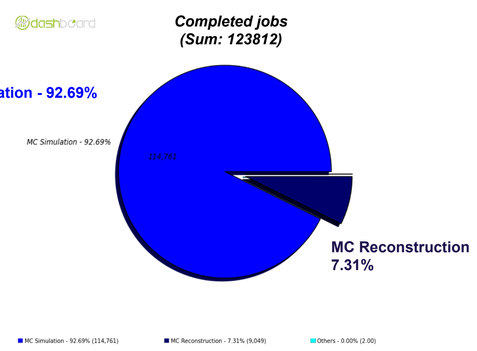




jobs efficiency in **time** (left) and **average** over full period (right) for MPPMU Tier2 and **HYDRA**.
Multicore jobs could benefit from further code optimization

Success Rate





Percentage of **successful**, **failed** and **deleted** jobs.

- **Failed Jobs:** Large contribution is related to jobs exceeding their estimated walltime limit

Job types percentage

**** All the plots have been created using the ATLAS dashboard.***
The period of time considered is 01 July 2014 -> 01 March 2015