ATLAS Computing: From Commissioning to 7TeV Data

Graeme A Stewart[*][†]

For the ATLAS collaboration

E-mail: graeme.andrew.stewart@cern.ch

In this paper we summarise ATLAS operations from the STEP09 campaign in June 2009 through to ATLAS taking data in the first 7 TeV collisions at the LHC in 2010. We describe the lessons which were learned from the STEP09 challenge, both in proving which parts of the system were in good shape, but also in highlighting those areas which required improvement. We then describe the experience of ATLAS computing operations during the first LHC data taking era. The ATLAS experiment has successfully recorded, reconstructed, distributed and analysed millions of collision events delivered by the LHC at an unprecedented centre-of-mass energy of 7 TeV. The involved large-scale data processing operations, both the prompt reconstruction at Tier0 and the subsequent reprocessing campaigns in the Tier1 sites in the Grid, worked remarkably well.

---

[*]Speaker.

[†]University of Glasgow, Glasgow G12 8QQ, Scotland.

http://pos.sissa.it/

1. STEP09

In the run up to LHC restart, ATLAS undertook a large offline computing exercise, called STEP09, run in parallel with other LHC experiments. This involved ramping up activities in the areas of data distribution, reprocessing at Tier-1s, simulation production and user analysis to very high rates and critically examining ATLAS offline and site performance.

The results of STEP09 showed that: (a) data distribution ran generally well, but some sites needed to improve their peak bandwidth to help manage backlogs; (b) simulation production was essentially a solved problem; (c) reprocessing worked well at most Tier-1s, but some sites needed to address issues with their tape systems; (d) massive user analysis was a serious problem for many sites and much optimisation was required.

These issues were tackled in the run up to LHC restart, with significant improvements in analysis rates at Tier-2s occurring.

2. Data Preparation and Calibration

As data is collected from the online system by the Tier-0, prompt reconstruction of the express and calibration streams is done, using initial online calibration constants and LHC beam conditions data, which have been uploaded to COOL (this data has a per-lumi block resolution, currently 2 minutes). The express stream represents some 10% of the data, including high p_t and lepton triggers, and the calibration stream has partial events from one sub-detector.
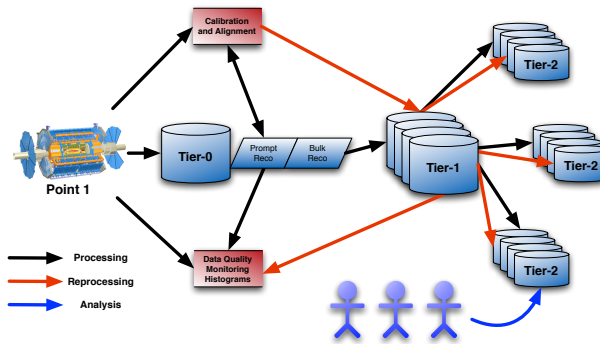

Figure 1: ATLAS offline data flow. Calibration and alignment loops feed back to first pass processing at the Tier-0 and any subsequent reprocessing at Tier-1s.

The Tier-0 populates histograms every 10 minutes, to allow further data quality assessments to be made efficiently during the run (in total over 20k histograms are available on demand).

This initial data quality and alignment procedure then provides the calibration constants that are used for bulk reprocessing, which is run some 36 hours later, after a final, manual, data quality sign-off (illustrated in Figure 1).

An example of calibration is shown in Figure 2, where the result of beam spot determination changes the initially misaligned distance of closest approach into a well aligned distribution, uniform in ϕ.
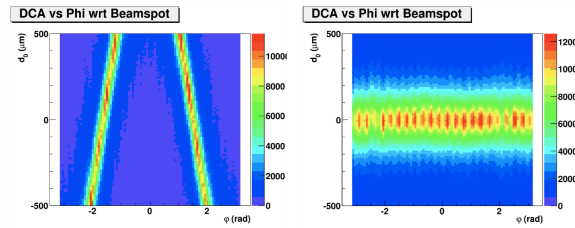
Figure 2: Misaligned beamspot results in a sinusoidal distance of closest approach (left); then is corrected in the calibration loop to a uniform distribution with ϕ (right).

After bulk processing data quality flags are available per sub-detector and the overall data taking efficiency for ATLAS has been 95% (the largest inefficiencies come from 'warm starts', e.g., turning on high voltages after stable beams are declared).

3. Tier-0 Operations and Workflow

The Tier-0 plays a pivotal role in data processing and is additionally charged with: robustly accepting data from the online system and archiving it to tape; registering data in the ATLAS Distributed Data Management System (DDM); being the export source for data export to Tier-1 and calibration Tier-2 sites.

The Tier-0 consists of 65 disk servers at CERN, with 650TB capacity, but the main design criterion is the i/o capacity of the pool, which must be able to reach 6Gb/s.

The T0 runs a very robust reconstruction environment and had run 1.8 million jobs, processing 1.7PB of data, in the 4 months before ICHEP 2010 (April-July). Persistent crashes were very rare, with 99.997% of events processed successfully.

4. Reprocessing

When ATLAS decides to reprocess data, this is done at our 10 Tier-1 centres. There are two types of reprocessing which can be undertaken: 'fast' reprocessings use software versions already known to be good; 'full' reprocessings use a new version of ATLAS software and require extra validation steps.

As 10 sites are involved there is a considerable operational workload in reprocessing and problems at any single site can hold up the end of the campaign. However, experience has led to improvements such as (a) validation of sites undertaken in parallel with software validation; (b) organising 'vertical slices', with some data going through all steps to final histograms early in the campaign, in order to pick up rare or unusual bugs that were missed in validation; (c) automatically moving jobs which exceed memory or wall clock time limits to special queues with extra resources, which allows automatic recovery for challenging or complex events.

In the two reprocessing campaigns of April and May 2010, 100% of events were successfully reprocessed on the grid.

5. Data Distribution

ATLAS data was distributed on the grid according to the ATLAS computing model, but with some variations for early running conditions: raw data was explicitly put on tape, to help reprocessing and debugging; ESD (event summary data) was moved in bulk to ATLAS Tier-2s as this was a very popular early data format; more copies of AOD (analysis object data) were also made available. This resulted in a steep increase in disk use (Figure 3).
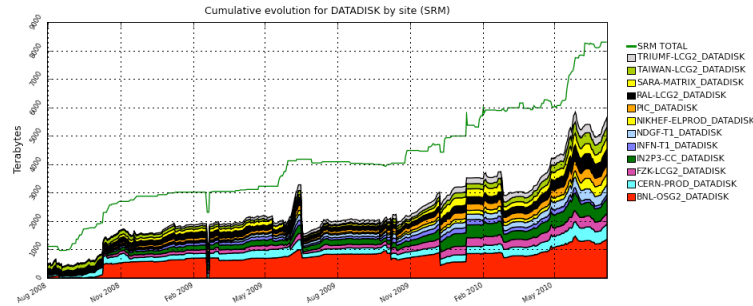

Figure 3: Growth in the utilisation of disk space at ATLAS Tier-1s. The top green line is the total T1 space available. Coloured bands represent the space used at each site.

Also observed was some network saturation, especially when large amounts of reprocessed real and simulated data were ready to be distributed to sites.

6. Grid Analysis

Rates of analysis on the grid have risen sharply, from about 2 000 jobs running simultaneously before data arrived (with bursts of 5 000 jobs in tests), to in excess of 10 000 concurrent jobs once 7 TeV beam data was taken. Efficient support structures have been put in place to help users with software or grid problems and this has helped hundreds of users to successfully tackle distributed analysis. Site stability problems continue, however, and are currently laborious for users and shifters to deal with.

7. Conclusion

ATLAS computing has performed extremely well during the early period of LHC data taking. Data taking efficiency has been excellent and data has been successfully processed at the Tier-0, re-processed at Tier-1s, distributed to Tier-2s and analysed on the grid.

The offline computing system now has plans for improvements to robustness and automation to manage the next years of LHC data taking. Such improvements include: automating data quality checks; better 'slow control' between Tier-0 and other systems to improve stability; more dynamic data placement, responding to user interest in 'hot' data; automation of blacklisting for sites failing jobs.

All these improvements will allow us to provide an even better computing system for the ATLAS collaboration.