



Contribution ID: 57

Type: **Oral**

LHC ATLAS users analysing data on the Grid

Wednesday 14 April 2010 14:40 (20 minutes)

In this talk we will describe the experience of the experiment ATLAS in commissioning the system (large scale user-analysis exercises in late 2009) and the first experience with real data.

First of all, we will describe the tools to commission and control the sites contributing to user analysis.

We will describe the support structure and the user experience (study of monitoring data and direct feedback from users). We will also show examples of analysis running on new data collected after the first collisions starting December 2009.

Conclusions and Future Work

The effort ATLAS dedicated in preparing for distributed analysis is demonstrating to be a necessary investment in the preparation for the data taking and the analysis of the first LHC data. This experience is also an useful example for the entire Grid community.

The future direction is to extend the system in view of the increasing user activity (and more sophisticated use cases). Since most of the development are of general interest we are considering to team up with other VO to support and further develop our tools.

Detailed analysis

Analysis in High-Energy Physics is characterised by a very large number of users. This is true both in absolute terms, due to the size of the HEP experiments (2000+ physicists in the case of ATLAS) and as a percentage of the research community.

The most important feature is that each user submits custom applications (normally built on top of the experiment-specific framework) instead of using pre-loaded services like in case of common portal applications.

In general user analysis tend to be I/O limited and iterative (sub data samples read many times by sets of users), which poses new constraints on site performance. The need to minimise the latency for users (compared to maximizing the throughput as in detector simulation and data reconstruction) is fundamental.

To prepare for the data taking, we have contributed (and we will report) in the following areas:

- Development and support of end-user tools (e.g. Ganga)
- Integration in the experiments framework (e.g. ATLAS Distr. Data Management (DDM) and in the experiment scheduling system (gLite WMS on EGEE, PanDA on EGEE, OSG and NDGF, ARC in NDGF)
- Setting up a user support and tutorials
- Organising the commissioning of sites

Impact

Sharing the ATLAS experience in the field of analysis is beneficial for all the grid communities (users, operations and middleware):

- 1- New user communities can consider to the ATLAS experience in order to assess the global impact of a pervasive grid adoption. Solutions like the distributed analysis shift system for supporting users will be illustrated.
- 2- ATLAS put in production a sophisticated system to commission and control grid sites for analysis. The specific characteristics of this activity (short jobs, I/O contention, multiple users) should be taken into account by the sites. Some of the tools developed have an interest beyond the ATLAS community.
- 3- The commissioning of sites for analysis produces performance data which can be used to compare different set up and different usages mode (in particular from a data access point of view). These data (like read efficiency, error rates) are collected and made available also to the middleware community. We plan to extend the error analysis in order to help to pin down common error sources. Similar data are important since they are extracted from the live system and can track the system evolution across HW and SW upgrades.

Keywords

ATLAS analysis

URL for further information

<https://twiki.cern.ch/twiki/bin/view/Atlas/AtlasComputing>

Author: Dr LAMANNA, Massimo (CERN)

Co-authors: Dr VAN DER STER, Dan (CERN); Dr ELMSHEUSER, Johannes (LMU Munich)

Presenter: Dr LAMANNA, Massimo (CERN)

Session Classification: High Energy Physics

Track Classification: Scientific results obtained using distributed computing technologies