

Achievements of the ATLAS Distributed Analysis during the first run period

Wednesday, 10 April 2013 16:00 (30 minutes)

Impact

The WLCG approach has the benefit of distributing responsibility in such a way that CERN's role is to generate the raw data along with the additional calibration needed to interpret it while the broad international community accesses and analyzes this data through its own hierarchical network. The main point is that the data analysis, storage and deployment are driven by the requirements of the experimenters and theoretical analysts. Big achievements make the collection, handling and basic analysis of large data feasible. The LHC data intensive science activities have proved that the data volumes are large but they are not unmanageable. This is a very valuable experience for the scientific community that faces the challenge of storing, analysing and accessing large data sets, as well as the need to archive data in a robust and enduring way.

Summary

In the LHC operations era analyzing the large data by the distributed physicists becomes a challenging task. The Computing Model of the ATLAS experiment at the LHC at CERN was designed around the concepts of grid computing. Large data volumes from the detectors and simulations require a large number of CPUs and storage space for data processing. To cope with these challenges a global network known as the Worldwide LHC Computing Grid (WLCG) was built. This is the most sophisticated data taking and analysis system ever built. Since the start of data-taking, the ATLAS Distributed Analysis (ADA) service has been running stably with the huge amount of data. The reliability of the ADA service is high but steadily improving; grid sites are continually validated against a set of standard tests, and a dedicated team of expert shifters provides user support and communicates user problems to the sites. The ATLAS Grid Computing Model is reviewed in this talk. Emphasis is given to ADA system.

Description

The central goal here is to analyze the results of the collisions of high energy particles as a way of probing the fundamental forces of nature. The challenge in a LHC data analysis is the need to explore large data volumes from the detectors and simulation, which require a large number of CPUs for processing. In addition to the complex experiment software structure and a high connectivity to the data that needs to be provided by the large-scale physics analysis activities.

Analysis jobs are routed to sites based on the availability of relevant data and processing resources. The data distribution is optimized to fit the resource distribution, and it is dynamically changed to meet rapidly the evolving requirements for analysis use cases. Distributed analysis tools used to analyze the data are reliable and fast to work with. Both the user support techniques and the direct feedback of users are the goal of improving the success rate and user experience when utilizing the distributed computing environment. The service is actively used: more than 1600 users have submitted jobs in the year 2012 and more than 2 million analysis jobs per week. PanDA is the ATLAS workload management system for processing user analysis, group analysis and production jobs. The reliability of the ADA service is high but steadily improving; grid sites are continually validated against a set of standard tests, and a dedicated team of expert shifters provides user support and communicates user problems to the sites.

The ATLAS Grid Computing Model is reviewed in this talk as to how the ATLAS data is stored, distributed and processed. An emphasis is given on the distributed data analysis services, summarize this year of distributed analysis activity, and present the perspectives for future improvements to the system.

Primary author: Dr FASSI, Farida (IFIC-CSIC-UV)

Co-author: FASSI, Farida (on behalf of the ATLAS collaboration)

Presenter: Dr FASSI, Farida (IFIC-CSIC-UV)

Session Classification: Community Applications

Track Classification: Community Platforms (Track Lead: P Solagna and M Drescher)