



Contribution ID: 6

Type: Poster

Large-scale ATLAS simulated data on EGEE

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).

ATLAS is one of the four LHC (Large Hadron Collider) experiments at CERN, is devoted to study proton-proton and ion-ion collisions at 14TeV (center of mass energy). ATLAS collaboration is composed of about 2000 scientists spread around the world. The experiment requirements for next year is of about 1,2PB of storage and 26 MskI2k of CPU, and is relying on GRID philosophy and EGEE infrastructure. Simulated events are produced and distributed over EGEE by the Atlas production system.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.

ATLAS is using the services provided by the EGEE middleware. Event simulation jobs are sent to the LCG (LHC Computing Grid) GRID by glite-WMS (Workload Management System) and Condor-G and using the dispatching tools of the CE's. Event simulation jobs perform the Data Management as well, request the inputs and stores the outputs on the desired SE's, file location and information is managed with distributed LCG File Catalogues (LFC) while the

ATLAS Distributed Data Management system (DDM) stays on a top level and takes care of the asymmetric file movement on top of the FTS services.

Services which are causing problems are basically the Storage Elements, as the system is strongly dependent on the inputs for the event simulation jobs and failing to retrieve it produces job failures, while failures in storing the outputs due to SE's instabilities leads to the loss of the CPU consumed by the job and the consequent failure.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications

Data has to be processed and must be accessible by a huge number of scientists for analysis. The throughput of data for Atlas experiment is expected to be of 320 MB/s with an integrated amount of data per year of O(10)Pb. The processing and storage need a distributed share of resources, spread worldwide and interconnected with GRID technologies as the requirements from the LHC are has no precedents.

Event production is the way to produce, process and store data for analysis before the experiment startup, and is performed in a distributed way. Tasks are defined by physics coordinators and then are assigned to Computing

Elements spread worldwide. Some of the jobs that build up the tasks need input data as well to produce

new output, this means the jobs may need input from external sites and store remotely. For that reason sites are connected by File Transfer Service (FTS) channels that links the Storage Elements (SE) interface for each site. The GRID allow this distributed infrastructure

Primary author: Dr ESPINAL, Xavier (PIC/IFAE)

Co-authors: Dr WALKER, Rodney (TRIUMF); Dr CAMPANA, Simone (CERN)

Presenter: Dr ESPINAL, Xavier (PIC/IFAE)

Track Classification: Demo and Poster session