



Contribution ID: 17

Type: Oral

ATLAS Distributed Analysis tests in the Spanish Cloud

Wednesday 4 March 2009 11:20 (20 minutes)

ATLAS distributed analysis challenges need to be performed in order to validate site and cloud readiness for the full-scale user load. Breaking points and bottlenecks which result from the site/cloud design or configuration need to be identified. For that test, we are using a real analysis code from physicists, the Ganga LCG/EGEE backend, using its data-based brokering and splitting and both Posix I/O and copy mode for accessing to the data. A set of metrics and error classification will be provided.

Impact

The ATLAS detector is going to take real data in Autumn 2009. Spanish sites have to be validated in order to be ready for the physics analysis. With these tests we will identify breaking points and bottlenecks which result from the site design or configuration taking into account that needs of analysis jobs differ from those of production. The essential impact is ensuring the robustness and effectiveness of the complex system of GRID Analysis.

URL for further information

<http://gangarobot.cern.ch/st/>

Conclusions and Future Work

Distributed analysis tests are necessary to stress the facilities at a simulated full user load. Very useful first test that allows to identify "big" problems (missing software, missing/bad information published) and already to point out some limitation (e.g.: network connection in some sites). In the future, we would like to test new parameters to improve access to the data using DCAP and LUSTRE and reproduce the exercise with more jobs to really hammer the sites in order to stress the system.

Keywords

Distributed Analysis, GRID, LCG/EGEE; ATLAS, GANGA, LUSTRE, DCAP

Detailed analysis

The needs of analysis jobs differ from those of production, so though a site may function for MC, it may have a poor performance for analysis. With these tests we are trying to identify what works well, and what doesn't work well. The distributed analysis challenge framework is built with Ganga LCG/EGEE jobs, which submits an application to the sites, tracks the job statuses periodically as they run, and allows us to easily develop reports after the complete. The jobs read directly from the Storage Element the data using posix I/O and some performance metrics, for instance success/failure rate, CPU/walltime, events per second and error classification, e. g.: different I/O errors are recorded. The results led to discovery network and switch problems and software bugs. The same application and input data were used in all sites.

Author: Dr GONZALEZ DE LA HOZ, Santiago (IFIC-Valencia/CERN)

Co-authors: Ms LAMAS, Alejandro (IFIC-Valencia); Mr FERNANDEZ, Alvaro (IFIC-Valencia); Ms OLIVER, Elena (IFIC-Valencia); Dr FASSI, Farida (In2p3-Lyon); Dr AMOROS, Gabriel (IFIC-Valencia); Mr SANCHEZ, Javier (IFIC-Valencia); DEL CANO, Laura (UAM-MAdrid); Dr MARCH, Luis (UAM-MAdrid); MUÑOZ, Luis (UAM-MAdrid); Dr KACI, MOhammed (IFIC-Valencia); Mr VILLAPLANA, Miguel (IFIC-Valencia); FERNANDEZ, Pablo (UAM-MAdrid); Mr VIVES, Roger (IFIC-Valencia); Dr ESPINAL, Xavier (PIC/IFAE); PACHECO, andreu (IFAE-Barcelona); BORREGO, carlos (IFAE-Barcelona); NADAL, jordi (IFAE-Barcelona); DEL PESO, jose (UAM--MAdrid); Dr SALT, jose (IFIC-Valencia); PARDO, juan (UAM-MAdrid); CAMPOS, marc (IFAE-Barcelona)

Presenter: Dr GONZALEZ DE LA HOZ, Santiago (IFIC-Valencia/CERN)

Session Classification: High Energy Physics

Track Classification: Scientific results obtained using grid technology