

International Symposium on Grids and Clouds (ISGC) 2015

Abstract ID : 45

Scaling up ATLAS production system for Big Data: development, commissioning and operational experience

Content :

In the ATLAS experiment, the Big Data processing generates a multiplicity of requirements as more data and use cases emerge. For Big Data processing, we adopted the data driven approach, where software applications transform the input data into the output data. In the ATLAS production system, each data transformation is represented by a task – a collection of many jobs submitted by the ATLAS workload management system (PanDA) and executed on the Grid. Over the years, the success of our data transformation approach resulted in the exponential growth both in the number of data transformation applications and in the rate of task submissions. The LHC shutdown presented an opportunity to re-engineer the ATLAS Big Data processing infrastructure, adding extra layers further improving the system scalability and flexibility.

Scaling up ATLAS production system for further challenges, our analysis identified patterns in ATLAS data transformation workflows comprised of many tasks. For example, the Monte Carlo simulations workflow is composed of many steps: generate hard-processes, hadronize signal and minimum-bias events, simulate energy deposition in the ATLAS detector, digitize electronics response, simulate triggers, reconstruct data, transform the reconstructed data into reduced forms for physics analysis. Such patterns provided scalable framework for automated workflow definitions implemented in the Database Engine for Tasks (DEfT). In the novel bi-level production system, the top-level DEfT generates individual workflow tasks for processing by the bottom-level Job Execution and Definition Interface (JEDI). The JEDI is coupled with PanDA to provide dynamic job definition tailored to the computing resource capabilities. Other ATLAS Big Data processing infrastructure upgrades include the new distributed data management system Rucio and the extension of the ATLAS Metadata Interface with capabilities to configure data transformations.

We report on our development, commissioning and operational experience in scaling up the production system for a growing number of requirements from main areas of the ATLAS experiment: Physics, Data Preparation and Trigger.

Primary authors : BORODIN, Mikhail (MEPhI) ; DE, Kaushik (UT Arlington) ; GARCIA NAVARRO, Jose Enrique (IFIC Valencia) ; GOLUBKOV, Dmitry (IHEP Protvino) ; KLIMENTOV, Alexei (BNL) ; MAENO, Tadashi (BNL) ; VANIACHINE, Alexandre (Argonne)

Co-authors :

Presenter : BORODIN, Mikhail (MEPhI) ; DE, Kaushik (UT Arlington) ; GARCIA NAVARRO, Jose Enrique (IFIC Valencia) ; GOLUBKOV, Dmitry (IHEP Protvino) ; KLIMENTOV, Alexei (BNL) ; MAENO, Tadashi (BNL) ; VANIACHINE, Alexandre (Argonne)

Track classification : Physics (including HEP) and Engineering Applications

Contribution type : --not specified--

Submitted by : VANIACHINE, Alexandre

Submitted on Saturday 15 November 2014

Last modified on : Saturday 15 November 2014

Comments :

on behalf of the ATLAS Collaboration