



Contribution ID: 264

Type: Oral presentation to parallel session

ATLAS Job Transforms: A Data Driven Workflow Engine

Tuesday 15 October 2013 16:29 (22 minutes)

The need to run complex workflows for a high energy physics experiment such as ATLAS has always been present. However, as computing resources have become even more constrained, compared to the wealth of data generated by the LHC, the need to use resources efficiently and manage complex workflows within a single grid job have increased.

In ATLAS, a new Job Transform framework has been developed that we describe in this paper. This framework manages the multiple execution steps needed to 'transform' one data type into another (e.g., RAW data to ESD to AOD to final ntuple) and also provides a consistent interface for the ATLAS production system.

The new framework uses a data driven workflow definition which is both easy to manage and powerful. After a transform is defined, jobs are expressed simply by specifying the input data and the desired output data. The transform infrastructure then executes only the necessary substeps to produce the final data products. The global execution cost of running the job is minimised and the transform can adapt to scenarios where data can be produced along different execution paths. Transforms for specific physics tasks which support over 60 individual substeps have been successfully run.

As the new transforms infrastructure has been deployed in production many features have been added to the framework which improve reliability, quality of error reporting and also provide support for multi-threaded and multi-process jobs.

Author: STEWART, Graeme Andrew (CERN)

Co-authors: SARRAZIN, Bjorn (Universitaet Bonn (DE)); MADDOCKS, Harvey Jonathan (Lancaster University (GB)); SANDHOFF, Marisa (Bergische Universitaet Wuppertal (DE)); Mr HARENBERG, Torsten (UNIVERSITY OF WUPPERTAL); BREADEN MADDEN, William Dmitri (University of Glasgow (GB))

Presenter: STEWART, Graeme Andrew (CERN)

Session Classification: Distributed Processing and Data Handling B: Experiment Data Processing, Data Handling and Computing Models

Track Classification: Distributed Processing and Data Handling B: Experiment Data Processing, Data Handling and Computing Models