Contribution ID: **255**   Type: **poster**

Job optimization in ATLAS TAG based Distributed Analysis

Monday 23 March 2009 08:00 (20 minutes)

The ATLAS experiment is projected to collect over one billion events/year during the first few years of operation.

The efficient selection of events for various physics analyses across all appropriate samples presents a significant technical challenge.

ATLAS computing infrastructure leverages the Grid to tackle the analysis across large samples by organizing data in a hierarchical structure and exploiting distributed computing to churn through the computations. This includes the same events at different stages of processing: RAW, ESD (Event Summary Data), AOD (Analysis Object Data), DPD (Derived Physics Data).

Event Level Metadata Tags (TAGs) contain a lot of information about all events stored using multiple technologies accessible by POOL and various web services. This allows users to apply selection cuts on quantities of interest across the entire sample to compile a subset of events which are appropriate for their analysis.

This paper describes new methods for organizing jobs to using the TAGs criteria to analyze ATLAS data using enhancements to ATLAS POOL Collection Utilities and ATLAS distributed analysis systems.

It further compares different access pattern to the event data and different ways to partition the workload for event selection and analysis, where analysis is intended as a broader event processing, including also events selection and reduction operations known as skimming, slimming and thinning, and DPD making.

Specifically it compares analysis with direct access to the events (AODs, ESDs, …) to access mediated by different TAG base event selections.

We then compare different ways of splitting the processing to maximize performance.

Author: MAMBELLI, Marco (UNIVERSITY OF CHICAGO)

Co-authors: MALON, David (Argonne National Laboratory); CHRANSHAW, Jack (Argonne National Laboratory); NOWAK, Marcin (Brookhaven National Laboratory); MAENO, Tadashi (Brookhaven National Laboratory)

Presenter: MAMBELLI, Marco (UNIVERSITY OF CHICAGO)

Session Classification: Poster session

Track Classification: Distributed Processing and Analysis