

The ATLAS TAGS Database distribution and management - Operational challenges of a multi-terabyte distributed database system

Monday 23 March 2009 08:00 (20 minutes)

The TAG files store summary event quantities that allow a quick selection of interesting events.

This data will be produced at a nominal rate of 200 Hz, and is uploaded into a relational database for access from websites and other tools.

The estimated database volume is 6TB per year, making it the largest application running on the ATLAS relational databases, at CERN and at other voluntary sites.

The sheer volume and high rate of production makes this application a challenge to data and resource management, on many aspects.

This paper will focus on the operational challenges of this system. These include:

uploading the data from files to the CERN's and remote sites' databases;

distributing the TAG metadata that is essential to guide the user through event selection; controlling resource usage of the database, from the user query load to the strategy of cleaning and archiving of old TAG data.

Summary

Proposal for a poster presentation illustrating the data flow of the TAGS data, making a focus on the solutions tested and challenges faced in its management.

Primary authors: MALON, David (Argonne National Laboratory); VIEGAS, Florbela (CERN); CRANSHAW, Jack (Argonne National Laboratory)

Co-authors: WONG, Andrew (TRIUMF - Canada's National Laboratory for Particle and Nuclear Physics); NAIRZ, Armin (CERN); GAMBOA, Carlos (Brookhaven National Laboratory); VINEK, Elisabeth (Universitaet Wien); GAL-LAS, Elizabeth (University of Oxford); DIMITROV, Gancho (Lawrence Berkeley National Laboratory); GOOSENS, Luc (CERN); NOWAK, Marcin (Brookhaven National Laboratory)

Presenter: VIEGAS, Florbela (CERN)

Session Classification: Poster session

Track Classification: Distributed Processing and Analysis