



Contribution ID: 161

Type: oral presentation

Building a Scalable Event-Level Metadata System for ATLAS

Wednesday, September 5, 2007 3:20 PM (20 minutes)

The ATLAS TAG database is a multi-terabyte event-level metadata selection system, intended to allow discovery, selection of and navigation to events of interest to an analysis. The TAG database encompasses file- and relational-database-resident event-level metadata, distributed across all ATLAS Tiers.

An oracle hosted global TAG relational database, containing all ATLAS events, implemented in Oracle, will exist at Tier 0. Implementing a system that is both performant and manageable at this scale is a challenge.

A 1 TB relational Tag database has been deployed at Tier 0 using simulated tag data. The database contains one billion events, each described by two hundred event metadata attributes, and is currently undergoing extensive testing in terms of queries, population and manageability. These 1 TB tests aim to demonstrate and optimise the performance and scalability of an Oracle TAG database on a global scale.

Partitioning and indexing strategies are crucial to well-performing queries and manageability of the database and have implications for database population and distribution, so these are investigated. Physics query patterns are anticipated, but a crucial feature of the system must be to support a broad range of queries across all attributes.

Concurrently, event tags from ATLAS Computing System Commissioning distributed simulations are accumulated in an Oracle-hosted database at CERN, providing an event-level selection service valuable for user experience and gathering information about physics query patterns.

In this paper we describe the status of the Global TAG relational database scalability work and highlight areas of future direction.

Summary

An oracle hosted global TAG relational database prototype, deployed at cern as part of multi-terabyte event-level metadata selection system, is discussed.

This is an ATLAS Offline Computing abstract, Distributed data analysis and information management track assigned by the ATLAS speakers committee.

Primary authors: Dr MALON, David (Argonne National Laboratory); Dr TIQUE AIRES VIEGAS, Florbela (CERN); Ms MCGLONE, Helen (University of Glasgow/CERN); Dr CRANSHAW, Jack (Argonne National Laboratory); Dr GOOSSENS, Luc (CERN)

Presenter: Ms MCGLONE, Helen (University of Glasgow/CERN)

Session Classification: Distributed data analysis and information management

Track Classification: Distributed data analysis and information management