



Contribution ID: 294

Type: poster

An Inconvenient Truth: file-level metadata and in-file metadata caching in the (file-agnostic) ATLAS distributed event store

Monday, September 3, 2007 8:00 AM (20 minutes)

In the ATLAS event store, files are sometimes “an inconvenient truth.” From the point of view of the ATLAS distributed data management system, files are too small—datasets are the units of interest. From the point of view of the ATLAS event store architecture, files are simply a physical clustering optimization: the units of interest are event collections—sets of events that satisfy common conditions or selection predicates—and such collections may or may not have been accumulated into files that contain those events and no others.

It is nonetheless important to maintain file-level metadata, and to cache metadata in event data files. When such metadata may or may not be present in files, or when values may have been updated after files are written and replicated, a clear and transparent model for metadata retrieval from the file itself or from remote databases is required. In this paper we describe how ATLAS reconciles its file and non-file paradigms, the machinery for associating metadata with files and event collections, and the infrastructure for metadata propagation from input to output for provenance record management and related purposes.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS)

ATLAS

Primary authors: Dr SCHAFFER, Arthur (LAL Orsay); Dr MALON, David (Argonne National Laboratory); Dr VAN GEMMEREN, Peter (Argonne National Laboratory); Dr HAWKINGS, Richard (CERN)

Presenter: Dr MALON, David (Argonne National Laboratory)

Session Classification: Poster 1

Track Classification: Software components, tools and databases