

Petaminer: Using ROOT for Efficient Data Storage in MySQL Database

Monday, 23 March 2009 08:00 (20 minutes)

High Energy and Nuclear Physics (HENP) experiments store petabytes of event data and terabytes of calibrations data in ROOT files. The Petaminer project develops a custom MySQL storage engine to enable the MySQL query processor to directly access experimental data stored in ROOT files.

Our project is addressing a problem of efficient navigation to petabytes of HENP experimental data described with event-level TAG metadata, which is required by data intensive physics communities such as the LHC and RHIC experiments. Physicists need to be able to compose a metadata query and rapidly retrieve the set of matching events, where improved efficiency will facilitate the discovery process by permitting rapid iterations of data evaluation and retrieval. Our custom MySQL storage engine enabled the MySQL query processor to directly access TAG data stored in ROOT TTrees. As ROOT TTrees are column-oriented, reading them directly provides improved performance over traditional row-oriented TAG databases. Leveraging the flexible and powerful SQL query language to the data stored in ROOT TTrees, the Petaminer approach will enable rich MySQL index-building capabilities for further performance optimization.

We studied feasibility of using the built-in ROOT support for automatic schema evolution to ease handling of large volumes of calibrations data of the large working experiment stored in MySQL. Over the lifecycle of calibrations, their schema may change. Support for schema changes in relational databases requires efforts. In contrast, ROOT provides support for automatic schema evolution. Our approach has a potential to ease handling of the metadata needed for efficient access to large volumes of calibrations data.

Presentation type (oral | poster)

oral

Summary

We report on the development of a custom MySQL storage engine to directly access high energy and nuclear physics experimental data stored in ROOT files. Since ROOT data storage features both column-oriented access and automatic support for schema evolution, we expect that Petaminer software will facilitate efficient handling and access to large volumes of the events and calibrations data of the HENP experiments and other data intensive sciences.

Primary authors: VANIACHINE, Alexandre (Argonne National Laboratory); MALON, David (Argonne National Laboratory); CRANSHAW, Jack (Argonne National Laboratory); LAURET, Jérôme (Brookhaven National Laboratory); HAMILL, Paul (Tech-X Corporation); FINE, Valeri (Brookhaven National Laboratory)

Presenters: VANIACHINE, Alexandre (Argonne National Laboratory); MALON, David (Argonne National Laboratory); CRANSHAW, Jack (Argonne National Laboratory); LAURET, Jérôme (Brookhaven National Laboratory); HAMILL, Paul (Tech-X Corporation); FINE, Valeri (Brookhaven National Laboratory)

Session Classification: Poster session

Track Classification: Software Components, Tools and Databases