



Contribution ID: 41

Type: Oral

Toward real-time data query systems in HEP

Tuesday, August 22, 2017 2:00 PM (20 minutes)

Exploratory data analysis must have a fast response time, and some query systems used in industry (such as Impala, Kudu, Dremel, Drill, and Ibis) respond to queries about large (petabyte) datasets on a human timescale (seconds). Introducing similar systems to HEP would greatly simplify physicists' workflows. However, HEP data are most naturally expressed as objects, not tables. In particular, analysis-ready data consists of arbitrary-length lists of particles, possibly containing nested lists of detector measurements. Manipulations of these structures, such as applying quality cuts to particles, not just events, selecting pairs for invariant masses, or matching generator-level data to reconstructed data, are difficult or impossible in SQL-like query languages.

To enable fast querying in HEP, we are developing Femtocode, a language designed for real-time plotting of HEP-scale datasets. We use the same techniques as modern big data query systems, such as performing operations on memory-cached, homogeneous columns of data, rather than each event individually, but adapt them to the scope of manipulations required by HEP. In this talk, I will describe key aspects of the language, how object-oriented expressions are translated into vectorized statements, and how computations are distributed in a fault-tolerant way. I will also show preliminary performance results, which suggest that a thousand-core cluster would be capable of real-time analysis of large LHC datasets. The new capabilities offered by this system may also find application outside of HEP.

This project is being developed in association with FNAL-LDRD-2016-032.

Primary authors: PIVARSKI, Jim (Princeton University); Dr ELMER, Peter (Princeton University (US))

Presenter: PIVARSKI, Jim (Princeton University)

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research