



HEPIX SPRING 2015 WORKSHOP

Contribution ID: 1

Type: **not specified**

Accelerating Scientific Analysis with SciDB

Monday 23 March 2015 14:00 (25 minutes)

SciDB is an open-source analytical database for scalable complex analytics on very large array or multi-structured data from a variety of sources, programmable from Python and R. It runs on HPC, commodity hardware grids, or in a cloud and can manage and analyze terabytes of array-structured data and do complex analytics in-database.

We present an overall description of the SciDB framework and describe its implementation at NERSC at Lawrence Berkeley National Laboratory. A case study using SciDB to analyze data from the LUX dark matter detector is described. LUX is a 370 kg liquid xenon time-projection chamber built to directly detect galactic dark matter in an underground laboratory 1 mile under the Black Hills in South Dakota, USA. In the 2013 initial data run, LUX collected 86 million events and wrote 32 TB of data of which only 160 events are retained for final analysis. The data rate for the new dark matter run starting in 2014 is expected to exceed 250 TB / year. We describe how SciDB is used to dramatically streamline the data collection and analysis, and discuss future plans for a large parallel SciDB array at NERSC.

Primary author: GERHARDT, Lisa (LBNL)

Co-author: Mr YAO, Yushu (LBNL)

Presenters: GERHARDT, Lisa (LBNL); Mr YAO, Yushu (LBNL)

Session Classification: End-user Services and Operating Systems

Track Classification: End-User IT Services & Operating Systems