

California is the primary US stop for LHC's ALICE data

July 30 2010

For approximately one month a year, the nuclei of lead atoms traveling near the speed of light will collide in the Large Hadron Collider's (LHC) ALICE experiment, generating a fireball about 100,000 times hotter than the core of our Sun. At these temperatures protons and neutrons dissolve into a "particle soup" of quarks and gluons, known as the quark-gluon plasma -- a state of matter that first occurred in nature at the birth of our Universe almost 14 billion years ago, a few millionths of a second after the Big Bang. By watching this "soup" cool, physicists hope to better understand the nature of matter, which makes up everything from galaxies to humans.

For the month that these lead ions collide, the ALICE experiment will collect over 10 terabytes of data per day, equivalent to the amount of information that could be stored on 20,000 DVDs. About 10 percent of this data will travel from the LHC in Switzerland to the National Energy Research Scientific Computing Center (NERSC) at the Lawrence [Berkeley National Laboratory](#) (Berkeley Lab) and to [Lawrence Livermore National Laboratory](#) (LLNL) in Northern California via the Department of Energy's (DOE's) Energy Sciences Network (ESnet). These facilities will provide the primary computing and storage resources for the ALICE collaboration in North and South America.

A Dedicated Link Between Europe and California

ALICE data will flow from Europe to California on ESnet's Science

Data Network (SDN), which is optimized for large data transfers. This strategy allows the ALICE researchers to leverage ESnet's On Demand Secure Circuits and Reservation System (OSCARS) protocol to set up multi-domain, virtual circuits for guaranteed end-to-end transfers.

"ESnet has been facilitating large, distributed scientific collaborations like ALICE since its inception and has the finely tuned infrastructure and a knowledgeable staff eager to assist with the transfer of these massive datasets," says Steve Cotter, who heads ESnet. "Our high speed networking architecture is finely tuned to guarantee the performance of these temporal data flows. The OSCARS system which manages bandwidth allocations on the Science Data Network is specifically designed for this purpose."

In the last year, DOE's Office of Nuclear Physics funded a significant expansion of NERSC's PDSF platform to handle the ALICE data processing and analysis. The U.S. ALICE collaboration also secured a storage allocation of about 600 TB on the NERSC's High Performance Storage System (HPSS) for 2010. The team recently made these computing and storage resources available to more than 1,000 collaborators worldwide via the ALICE Grid.

"Storage is very important for the ALICE experiment. Thousands of scientists from around the globe will be looking for different things in the same dataset and building on the results of other collaborators; so in addition to duplicating and storing raw data, we also have to copy and archive it at many different stages of the analysis," says Jeff Porter, computing project manager for the U.S. ALICE project and member of the ALICE Grid Task Force. He is also a member of the NERSC Outreach, Software & Programming Group.

The international ALICE Grid connects computing and mass storage resources across 31 countries, and facilitates science by allowing

researchers all over the world to submit computing jobs to a single queue. Supercomputing centers connected to this grid run software that locates the data needed to complete a job, determines which systems can perform the job's tasks and where the results will be archived in order to submit and run the computing job. All of these complex procedures are transparent to individual researchers. Later this year LLNL's Green Linux Collaborative cluster will be added to the ALICE Grid. This system will also be funded by DOE's Nuclear Physics office.

The U.S. portion of the ALICE Grid leverages the existing Open Science Grid, which relies heavily on the ESnet networking infrastructure and operations.

"The LHC computing problem is so vast that CERN alone cannot handle all of this data, it requires an international collaboration," says Peter Jacobs, a senior scientist in Berkeley Lab's Nuclear Science Division. He also leads the team of ALICE collaborators based at Berkeley Lab.

He notes that the nature of ALICE data analysis is "embarrassingly parallel," meaning each task requires its own CPU to reconstruct independent collision events in addition to disk storage so it can efficiently access the very large event samples. Because the PDSF and HPSS systems at NERSC have a track record of providing these types of resources to similar experiments, the center was a prime candidate for serving as a Tier-2 facility to handle ALICE data.

"The Lawrence Berkeley and Livermore Laboratories have a history of supporting ALICE science. Researchers from both labs have already collaborated to design and build the experiment's electromagnetic calorimeter detector; and the partnership between NERSC and Green Linux Collaborative to provide computing resources is an extension of that tradition," says Ron Soltz, a researcher at LLNL and coordinator of the ALICE U.S. Computing Project.

According to Soltz, the different strategies that NERSC and LLNL use to procure hardware will be highly beneficial to the ALICE collaboration, providing both optimal pricing through large volume purchases and continuous upgrading of the combined system to keep pace with the growth of ALICE computing requirements with time. This is accomplished by LLNL purchasing all of its new hardware approximately every three years with NERSC upgrading its PDSF system in a more incremental fashion every year.

"Computing is a very important part of this scientific process. You can build a world-class accelerator like ALICE, but you can't do the science without computing," says Soltz.

Provided by Lawrence Berkeley National Laboratory

Citation: California is the primary US stop for LHC's ALICE data (2010, July 30) retrieved 22 September 2024 from <https://phys.org/news/2010-07-california-primary-lhc-alice.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.