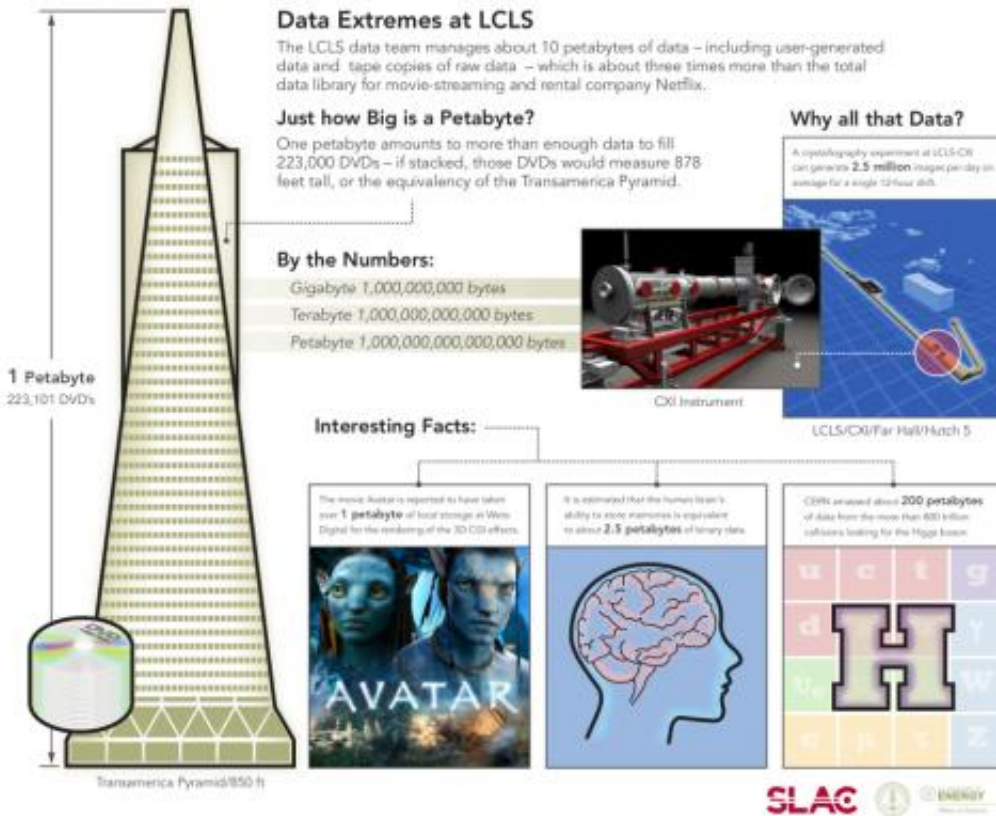


SLAC's X-ray laser explores big data frontier

June 14 2013, by Glenn Roberts Jr.



Managing large data is the topic for an upcoming National User Facility Organization annual meeting that will be held from June 19-21 at Lawrence Berkeley National Laboratory. SLAC's Amedeo Perazzo is scheduled to present a talk on "Data Management at the LCLS" at 1 p.m. June 20.

(Phys.org) —It's no surprise that the data systems for SLAC's Linac Coherent Light Source X-ray laser have drawn heavily on the expertise

of the particle physics community, where collecting and analyzing massive amounts of data are key to scientific success.

With its detectors collecting information on atomic- and molecular-scale phenomena measured in quadrillionths of a second, LCLS stores data at a rate and scale comparable to experiments at the world's most powerful [particle collider](#), the [Large Hadron Collider](#) in Europe.

The LCLS data team manages about 10 [petabytes](#) of data from experiments – about three times more than the total data library for movie-streaming and rental company Netflix. That's enough data to fill about 2.2 million DVDs, which if stacked would stand about 8,800 feet high. An experiment at one LCLS instrument produces about 10 million X-ray images in about 48 hours, on average, with the largest experiments generating 150-200 terabytes (about 154,000 to 205,000 gigabytes) of data.

Big data gets bigger

Coupled with such head-spinning statistics is the unavoidable reality that this "big data" frontier in science is ever-growing, which presents constant challenges: At LCLS, more sensitive detectors, more complex experiments. [Multiple simultaneous experiments](#), a possible increase in the laser [pulse rate](#) and other planned upgrades will require advances in data acquisition, storage and delivery systems.

"Pretty soon we will be taking a factor of 20 more data than we are taking today," said Amedeo Perazzo, who leads the Photon Controls and Data Systems Department at SLAC, which manages LCLS data. "At that point we will not be able to operate in the same way we do now."

Scientific diversity presents unique challenges

Many of the people on Perazzo's staff have worked on [particle physics](#) experiments such as [BaBar](#) and [ATLAS](#). And while the data demands at LCLS are similar to those of the high-energy physics community, the data team at LCLS has adapted to the varying needs of LCLS users, who come from a wide range of scientific backgrounds.

"You have biologists and chemists and materials scientists and different kinds of physicists," said Igor Gaponenko, a research software developer for LCLS data systems. "It's a new world. It's very diverse."

In high-energy physics, scientists have a common vocabulary and standardized data systems, and individual experiments can run for years. But at LCLS, experiments typically run for only a few days, and scientists need immediate access to their data so they can decide whether to change samples or X-ray energies in the middle of an experiment.

LCLS users want "reliability, flexibility and immediacy," said Sebastien Carron Montero, an engineering physicist who works on data systems for the Atomic, Molecular and Optical Science instrument at LCLS. "To have all of them at the same time is very demanding. And each one of them is using different tools."

An equal playing field for data access

Carron Montero added that in particle physics it's common to immediately and aggressively filter the data to single out particular types of events. But at LCLS, "We tend to collect most of the data, which puts an enormous burden on our data system. That makes our system here quite big and difficult to manage."

The amount of data produced by experiments can vary widely, Perazzo said, noting that his team has upgraded the [data acquisition](#) capabilities to capture up to 10 gigabytes per second, enough to handle the load from

simultaneous experiments and larger detectors.

Some teams choose to store and analyze their data at SLAC, while others transfer data over high-speed scientific computing networks. Larger teams may bring their own data experts to LCLS for experiments.

There is a push to improve the user interface to make LCLS data tools more accessible to scientists, offer more real-time data during experiments, train staff to work more closely with users on learning the data systems and continue to work toward common data standards.

"Our job is to make it so all of the groups have the ability to extract science from the data. We need to find a language that works for the entire community and we need to lower the threshold that is required to start out," Perazzo said. "Something we have learned is that we really need to sit with them. We want to establish a closer relationship on the data-analysis side."

The next generation

As data demands increase, a likely solution will be the routine high-speed transfer of data to other data storehouses, where it can remain accessible for longer periods of time, freeing up LCLS to accept data from new experiments.

Perazzo said his department's system for handling LCLS data is scalable to meet these challenges, noting that other X-ray laser facilities launching over the next several years, including the [European X-ray Free-Electron Laser](#) project in Germany, are considering LCLS as a model for their own data systems.

"I do believe we are in the right spot," said Perazzo. "The next generation is actually looking at our systems."

Provided by SLAC National Accelerator Laboratory

Citation: SLAC's X-ray laser explores big data frontier (2013, June 14) retrieved 24 April 2024 from <https://phys.org/news/2013-06-slac-x-ray-laser-explores-big.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.