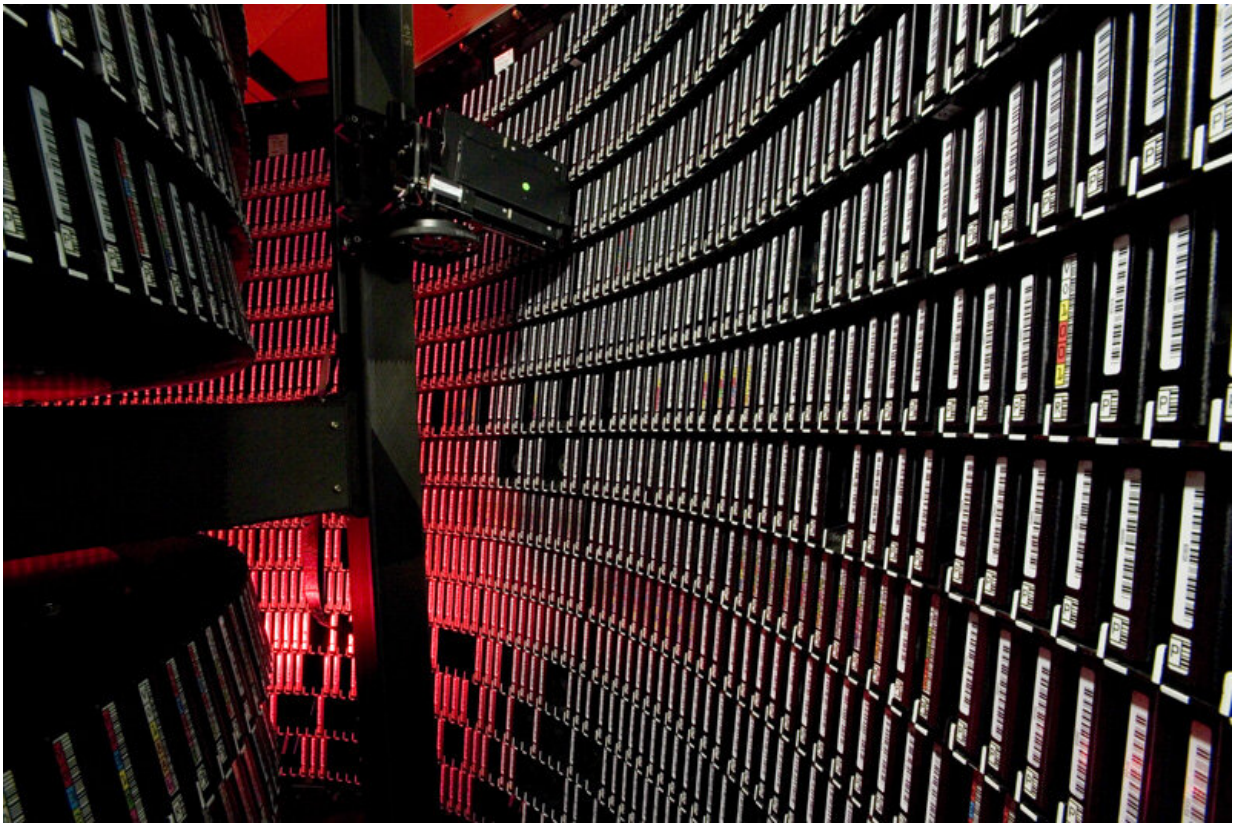


Solid-state technology for big data in particle physics

October 27 2020, by Jerald Pinson



When scientists need to access the stored files to perform analyses, a long robotic arm descends from the ceiling, selects a tape, and transfers the data it stores to a hard drive. Credit: Reidar Hahn, Fermilab

At CERN's Large Hadron Collider, as many as 40 million particle

collisions occur within the span of a single second inside the CMS particle detector's more than 80 million detection channels. These collisions create an enormous digital footprint, even after computers winnow it to the most meaningful data. The simple act of retrieving information can mean battling bottlenecks.

CMS physicists at the U.S. Department of Energy's Fermi National Accelerator Laboratory, which stores a large portion of LHC data, are now experimenting with the use of NVMe, or nonvolatile memory express, solid-state technology to determine the best way to access stored files when scientists need to retrieve them for analysis.

The trouble with terabytes

The results of the CMS experiment at CERN have the potential to help answer some of the biggest open-ended questions in physics, such as why there is more matter than antimatter in the universe and whether there are more than three physical dimensions.

Before scientists can answer such questions, however, they need to access the collision data recorded by the CMS detector, much of which was built at Fermilab. Data access is by no means a trivial task. Without online data pruning, the LHC would generate 40 terabytes of data per second, enough to fill the hard drives of 80 regular laptop computers. An automated selection process keeps only the important, interesting collisions, trimming the number of saved events from 40 million per second to just 1,000.

"We care about only a fraction of those collisions, so we have a sequence of selection criteria that decide which ones to keep and which ones to throw away in real time," said Fermilab scientist Bo Jayatilaka, who is leading the NVMe project.

Still, even with selective pruning, tens of thousands of terabytes of data from the CMS detector alone have to be stored each year. Not only that, but to ensure that none of the information ever gets lost or destroyed, two copies of each file have to be saved. One copy is stored in its entirety at CERN, while the other copy is split between partnering institutions around the world. Fermilab is the main designated [storage](#) facility in the U.S. for the CMS experiment, with roughly 40% of the experiment's data files stored on tape.

A solid-state solution

The Feynman Computing Center at Fermilab houses three large data libraries filled with rows upon rows of magnetic tapes that store data from Fermilab's own experiments, as well as from CMS. If you were to combine all of Fermilab's tape storage capacity, you'd have roughly the capability to store the equivalent of 13,000 years' worth of HD TV footage.

"We have racks full of servers that have hard drives on them, and they are the primary storage medium that scientists are actually reading and writing data to and from," Jayatilaka said.

But hard drives—which have been used as storage devices in computers for the last 60 years—are limited in the amount of data they can load into applications in a given time. This is because they load data by retrieving it from spinning disks, which is the only point of access for that information. Scientists are investigating ways to implement new types of technology to help speed up the process.

To that end, Fermilab recently installed a single rack of servers full of solid-state NVMe drives at its Feynman Computing Center to speed up particle physics analyses.



In an attempt to speed up analyses in high-energy physics research, Fermilab recently installed a single rack of servers full of solid state drives called NVMe. Credit: Bo Jayatilaka, Fermilab

Generally, solid state drives use compact electrical circuits to quickly transfer data. NVMe is an advanced type of solid-state drive that can handle up to 4,000 megabytes per second. To put that into perspective, the average hard drive caps at around 150 megabytes per second, making solid-state the obvious choice if speed is your main goal.

But hard drives haven't been relegated to antiquity just yet. What they

lack in speed, they make up for in storage capacity. At present, the average storage limit in solid-state drives is 500 gigabytes, which is the minimum amount of storage you'd commonly find available on modern hard drives. Determining whether or not Fermilab should replace more of their hard drive memory storage with solid-state drives will thus require a careful analysis of cost and benefits.

Undertaking an analysis

When researchers analyze their data using large computer servers or supercomputers, they typically do so by sequentially retrieving portions of that data from storage, a task well-suited for hard drives.

"Up until now, we've been able to get away with using hard drives in high-energy physics because we tend to handle millions of events by analyzing each event one at a time," Jayatilaka said. "So at any given time, you're asking for only a few pieces of data from each individual hard drive."

But newer techniques are changing the way scientists analyze their data. Machine learning, for example, is becoming increasingly common in particle physics, especially for the CMS experiment, where this technology is responsible for the automated selection process that keeps only the small fraction of data scientists are interested in studying.

But instead of accessing small portions of data, [machine learning](#) algorithms need to access the same piece of data repeatedly—whether it's stored on a hard drive or solid-state drive. This wouldn't be much of a problem if there were only a few processors trying to access that data point, but in high-energy physics calculations, there are thousands of processors that are vying to access that data point simultaneously.

This can quickly cause bottlenecks and slow speeds when using

traditional hard drives. The end result is slower computing times.

Fermilab researchers are currently testing NVMe technology for its ability to reduce the number of these data bottlenecks.

The future of computing at Fermilab

Fermilab's storage and computing power are much more than just a powerhouse for the CMS experiment. The CMS computing R&D effort is also setting the foundations for the success of the upcoming High-Luminosity LHC program and enabling the international, Fermilab-hosted Deep Underground Neutrino Experiment, both of which will start taking data in the late 2020s.

Jayatilaka and his team's work will also allow physicists to prioritize where NVMe drives should be primarily located, whether at Fermilab or at other LHC partner institutions' storage facilities.

With the new servers in hand, the team is exploring how to deploy the new solid-state technology in the existing computing infrastructure at Fermilab.

Provided by Fermi National Accelerator Laboratory

Citation: Solid-state technology for big data in particle physics (2020, October 27) retrieved 7 May 2024 from <https://phys.org/news/2020-10-solid-state-technology-big-particle-physics.html>

<p>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.</p>
--