CASE STUDY

Spectra helps CERN advance the boundaries of human knowledge providing storage solution for LHC physics data

> One of the great challenges of our experiments is the management and processing of the vast amounts of data generated. Our Spectra TFinity tape systems help us to address this unprecedented multi-petabyte data-processing challenge by ensuring data is preserved and accessible in the long term. This system is anticipated to be able to withstand the data production rates foreseen for the next LHC restart (Run 3) and beyond.

Vladimír Bahyl, Senior Data Storage Engineer, CERN

CERN

supplier

About CERN

The European Organization for Nuclear Research, known as CERN, was established in 1954. It was one of Europe's first joint ventures and now has 23 member states. Physicists and engineers at CERN use the world's largest and most complex scientific instruments to study the basic constituents of matter – fundamental particles. Their goal is to advance the boundaries of human knowledge by delving into the smallest building blocks of our universe.

The Challenge

To gain insight into how sub-atomic particles interact, CERN employs purpose-built particle accelerators and detectors to boost beams of particles to high energies, observe the beams as they are made to collide with each other or with stationary targets and record the results of these collisions. CERN operates the world's largest and most powerful particle accelerator, the Large Hadron Collider (LHC). The LHC consists of a 27-kilometre ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles along the way.

Collisions in the LHC generate the myriad of sub-atomic particles. Electronic circuits record the passage of each particle through a detector as a series of electronic signals, and send the data to the CERN Data Centre for digital reconstruction. The digitized summary is recorded as a 'collision event'. Up to about 1 billion particle collisions can take place every second inside the LHC experiments' detectors, so

In June 2020, the new CERN Tape Archive (CTA) solution officially entered production after 83 petabytes of ATLAS data initially stored in obsolete CASTOR system were successfully migrated to CTA. Pictured: The CERN Data Center.

CASE STUDY: CERN (European Organization for Nuclear Research)

the data is filtered, and a 'trigger' system is used to select those events that are potentially interesting for further analysis.

The CERN Data Centre processes on average one petabyte (one million gigabytes) of data per day. The LHC experiments produce more than 90 petabytes of data per year, and an additional 30 petabytes of data are produced per year for data from other (non-LHC) experiments at CERN. This involves not only scientific data but also many other types, including photographs, videos, minutes, memoranda, web pages, etc. Data formats and the tools to access them change constantly, and constant effort is required to tackle the issue. Archiving the vast quantities of data is an essential function at CERN. Magnetic tapes are used as the main long-term storage medium and data from the archive is continuously migrated to newer technology, higher density tapes.

The Solution

Extensive consolidation of the data storage infrastructure culminated in 2020 when the new CERN Tape Archive software (CTA) entered production. During Runs 1 and 2 of the LHC, the software system used to manage the archival storage of physics data was CASTOR, the CERN Advanced Storage manager, which was conceived as a system to manage both disk and tape storage. CTA is an evolution of the CASTOR system that removes the necessity to maintain a second disk-management system because it is tightly integrated with CERN's existing disk solution EOS.

CTA is a high-performance archival storage system developed at CERN for LHC physics data. CTA will replace CASTOR as the CERN archival solution for LHC Run 3 and beyond. To get ready for Run 3 and the vast amounts of data expected to be produced by the detectors in the LHC experiments and elsewhere in the accelerator complex, CERN installed two fourteen-frame Spectra TFinity ExaScale Tape Libraries in the CERN Data Centre. The tapes are stored in tape libraries, where they are retrieved by robotic arms. Their TFinity library is leveraging the open format LTO (Linear Tape-Open) tape technology,

CERN's first Spectra TFinity ExaScale Tape Library was installed in August 2020.

has 15,000 physical cartridge slots and can accommodate 48 tape drives in total. Each TFinity system can store up to 540PB of data in its current configuration using LTO-9 tapes. The TFinity ExaScale is engineered for superior reliability, performance and flexibility. The library contains dual High Performance Transporters (HPTs) – two robotic pickers optimized for performance and throughput. Furthermore, it offers data encryption and a suite of features that allow end users to actively check data, ensuring tapes are usable and verifying the integrity of data written.

Environment Snapshot

- Two fourteen-frame Spectra® TFinity® ExaScale Tape Libraries
- LTO-8 and LTO-9 full-height tape drives
- BlueScale® Software Standard Encryption
- CERN Tape Archive (CTA) data management software
- IBM Spectrum Protect backup software

Why Spectra?

- Long-term relationship
- Excellent professional services
- Support for multiple tape media formats
- Smallest footprint in the market

Solution Recap

Spectra TFinity ExaScale Tape Library – With unsurpassed storage density packaged in the smallest footprint of any enterprise library on the market, the Spectra TFinity ExaScale offers industry-leading scalability with the speed necessary to meet requirements of the most demanding environments. Deployed by some of the most recognized organizations in the world, the Spectra TFinity ExaScale provides maximum flexibility by allowing you to select the tape technology that is the perfect fit for your business. In addition to LTO tape technology, the Spectra TFinity ExaScale is also compatible with IBM® TS11X0 enterprise tape technology and Oracle T10000x enterprise tape technology, enabling all three in the same library.

Toll Free: 800-833-1132 • 303-449-6400 • 6285 Lookout Road • Boulder, CO 80301 USA www.spectralogic.com