How Intel Is Helping CERN to Unlock the Biggest Mysteries in Science

Intel® technology will help to filter ever-increasing amounts of data from the famous Large Hadron Collider

By Libby Plummer, Technology Writer

KEY TAKEAWAYS

- CERN's Large Hadron Collider is currently being upgraded and will restart in 2021 when it will once again produce huge amounts of data.

- Intel® technology will help researchers at CERN to comb through the data deluge to filter and flag only the significant information.

- This research could help experts to gain a better understanding of the origins of the universe and answer some of the most baffling questions in physics.

The world of scientific research produces massive amounts of data at an ever-increasing pace. One of the biggest challenges for researchers is how to store and process it. CERN*, the European Laboratory for Particle Physics, has already used Intel® technology to help cut through the enormous amounts of data it produces to look for the most rewarding nuggets and is looking to do so again in future as the amount of data produced continues to grow.

> *"We look forward to future collaboration with our colleagues at CERN to leverage open source tools and Intel hardware and engineering expertise to solve challenging data problems and enable exciting new research opportunities"*

The particle physics lab on the Franco-Swiss border hosts a giant complex of particle accelerators —the oldest of which has recently celebrated 60 years in operation[1]. CERN's largest and most powerful particle accelerator is the Large Hadron Collider* (LHC), which sits in a tunnel 100m underground near the Swiss city of Geneva. The biggest and most complex machine ever built, the LHC takes the form of a massive ring, measuring 27km long and made up of thousands of superconducting magnets. Inside the gigantic machine, particles are propelled at high speeds, smashing either into targets or into particles circulating in the opposite direction.[2]
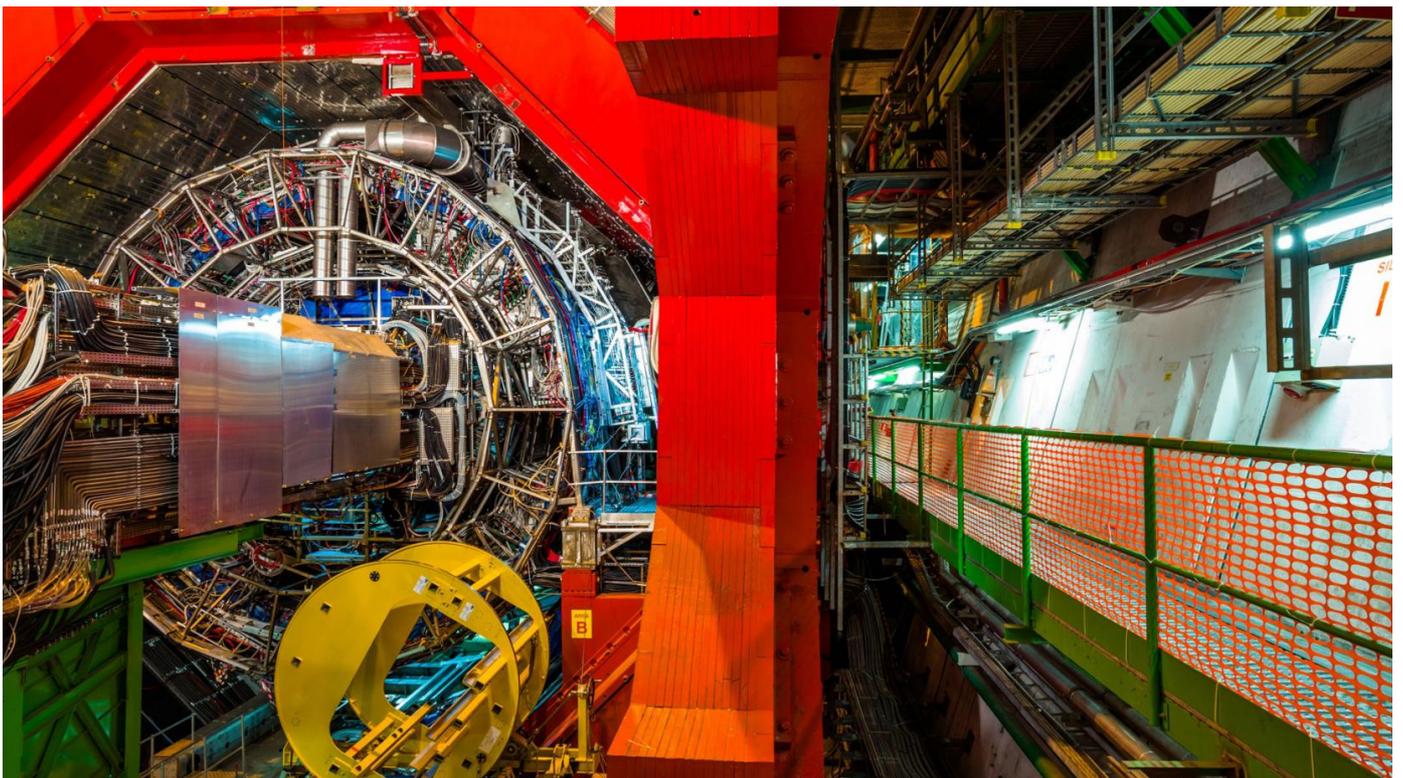
These collisions enable physicists to get a better understanding of the tiny particles that make up atoms, the building blocks of all chemical elements. This in turn enables us to learn more about matter and the origins of the universe, helping scientists to answer some of the most baffling mysteries in physics.

The LHC, which started initial operations in 2008, is capable of accelerating protons to 99.9999991% of the speed of light. When the LHC is in operation, up to 1 billion particle collisions occur every second, generating immense heat. These particle collisions generate up to a massive 1 petabyte (or a million gigabytes) of data every single second. This is too much data for even the world's biggest research institutes to store, so there needs to be a way to filter it, keeping only the useful information.

The LHC has been undergoing scheduled upgrade and maintenance work since 2018, following a three-year run known as Run 2. Data gathering will restart again in 2021 for Run 3 before another shut down in preparation for a major upgrade ahead of Run 4, which is planned to start in late 2027. As the name suggests, the HL (High-Luminosity) LHC project will see a major boost in the collider's luminosity. With higher luminosity, more data can be gathered, enhancing existing research and raising the potential for new discoveries.
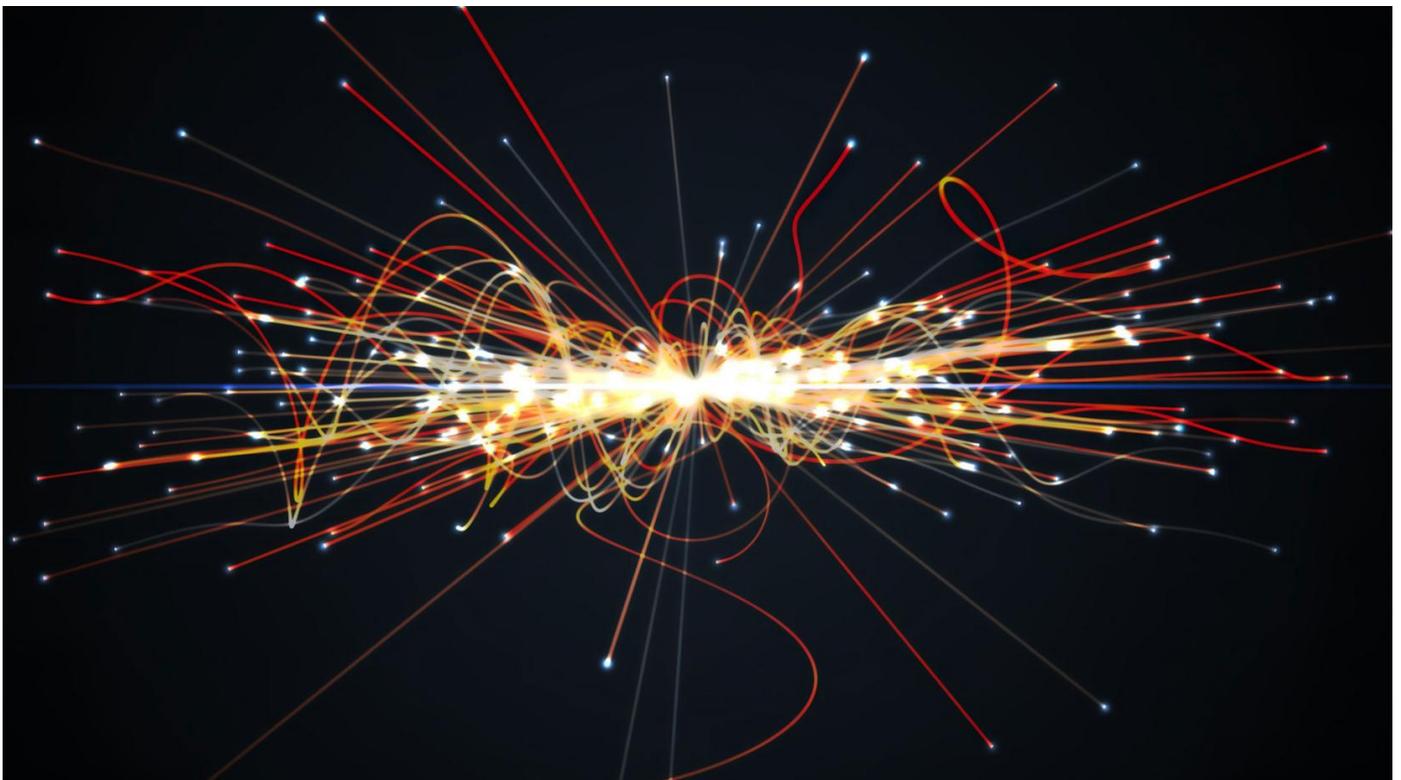

The key challenge will be dealing with the gigantic increase in data, with the HL-LHC set to accumulate a staggering ten times more collision events than its predecessor . There will be a huge increase in ICT demands in the coming years, exceeding what is expected to be achievable within a constant budget by several factors, for both storage and computing capacity.

CERN uses specialised software to filter the data produced, selecting particle collisions of interest for further analysis. It is important that such a system reduces the data volume efficiently, without filtering out particle-collision events that could hide new secrets. Researchers at the four main LHC experiments are always looking for new ways to improve their filtering systems. At the CMS experiment, researchers are investigating deep neural networks – sets of algorithms that are inspired by the structure of the human brain and designed to spot patterns.

Using Intel-developed open-source tools BigDL* and Analytics Zoo*, along with big-data processing framework Apache Spark* and with Jupyter* notebooks, the researchers worked with a team of Intel engineers to develop a deep-learning pipeline, running on Intel® Xeon® Scalable processors. This pipeline was designed to work on existing IT infrastructure and was aimed at filtering the data from their experiment more effectively, ensuring that uninteresting collisions are not stored or processed. Not only could this help them to improve their data analysis, it could also help to save on compute and storage resources. The team at the CMS experiment is now aiming to test this new system further, with a view to expanding both scalability and performance on this Intel-powered infrastructure.
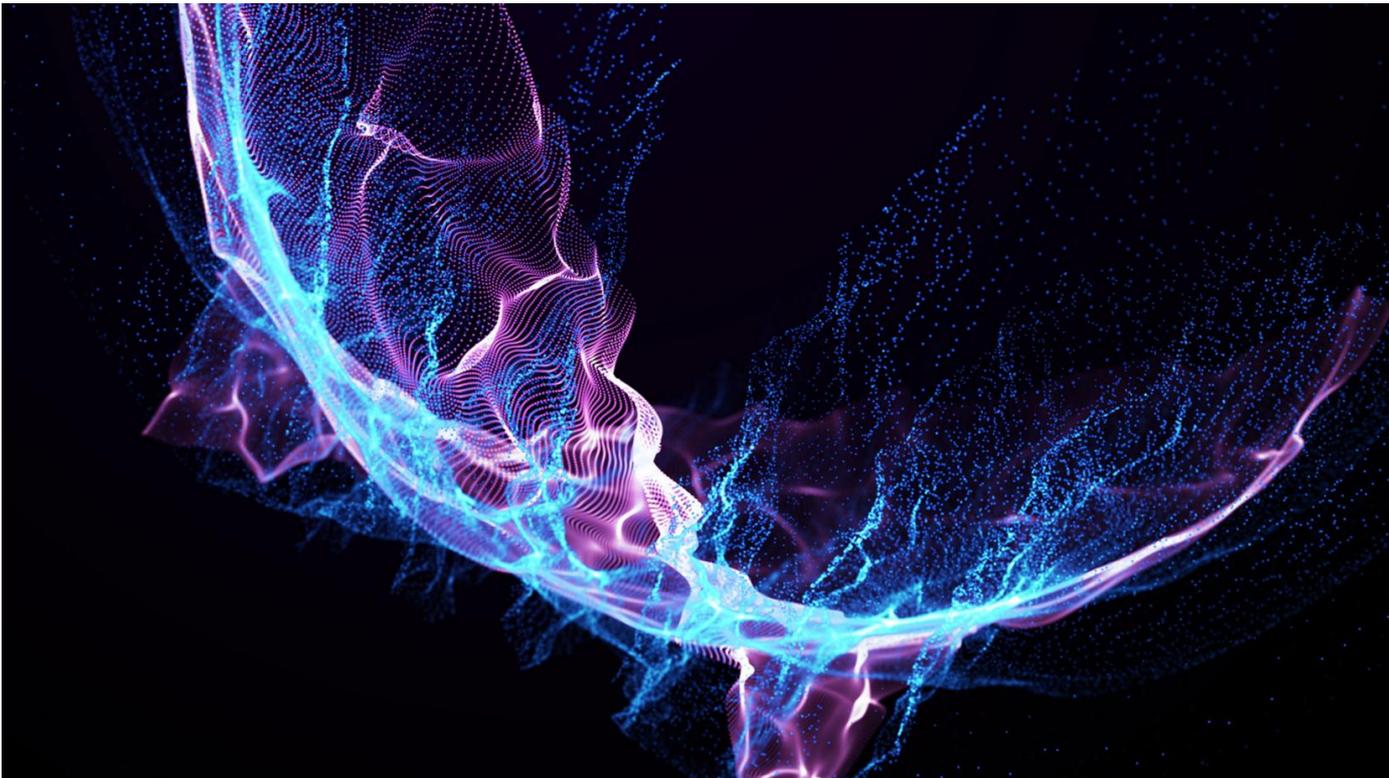
"We've seen BigDL deliver value in industries like financial services and security, but it's particularly rewarding to see BigDL be used at CERN to investigate some of the ultimate mysteries of science," said Sajan Govindan, Solutions Architect, Data Analytics Technologies at Intel in a blog post. "We look forward to future collaboration with our colleagues at CERN to leverage open source tools and Intel hardware and engineering expertise to solve challenging data problems and enable exciting new research opportunities."

In addition, researchers at CERN are investigating Intel® oneAPI to help with the HL-LHC upgrade. This toolkit is designed to deliver uncompromising performance for diverse workloads across multiple architectures. This has the potential to help researchers at the laboratory manage the huge computing demands of the HL-LHC.

"The relationship that Intel has with CERN is simply great," said Claudio Bellini, Account Executive at Intel. "Intel is part of the CERN openlab collaboration, through which CERN scientists get early access to the latest Intel® software tools and