



[Home](#) [Industry Insights](#) [Customer Voices](#)

How Will CERN Handle Exabytes of High-Energy Physics Data?

2017-11-09

CERN, established in 1954, is the world's largest particle physics laboratory. Its aim is to study the fundamental structure of the universe, as well as promote science as a means for peace and train the next generation of physicists and engineers. CERN consists of 22 member states, mostly from Europe. Some 12,000 visiting scientists from over 70 countries and with 105 different nationalities — half of the world's particle physicists — come to CERN for their research. In 2012, the Higgs boson, was discovered by the experiments on CERN's Large Hadron Collider (LHC).

At the recently held Huawei Connect 2017, Jan Van Eldik, Resource Provisioning team leader in CERN's IT Department, presented some of the ways CERN is investigating how to handle the rapidly growing data volumes produced by its experiments.



Figure 1. Jan van Eldik presented at Huawei Connect 2017

Reasons for Massive Data

CERN is home to the LHC, the world's largest particle accelerator. Sitting beneath the France-Switzerland border near Geneva, Switzerland, the LHC consists of a 27-kilometre ring of superconducting magnets 100 meters below ground, which creates a magnetic field 100,000+ times more powerful than that of the earth. In the beam pipes with vacuum systems, the particles can be accelerated close to the speed of light and collide within the particle detectors of the four large experiments: ATLAS, CMS, ALICE, and LHCb.

Up to about 1 billion particle collisions can take place every second inside the LHC experiment's detectors. It is not possible to read out all of these events. A 'trigger' system is therefore used to filter the data and select those events that are potentially interesting for further analysis.

The data generated by these operations is then distributed from the CERN data centre to 170 computing centres in 42 countries, through a system known as the Worldwide LHC Computing Grid (WLCG). At the time of writing, over 200 PB of data has been archived on tape at the CERN data centre, with several petabytes of new data expected to be added per month.

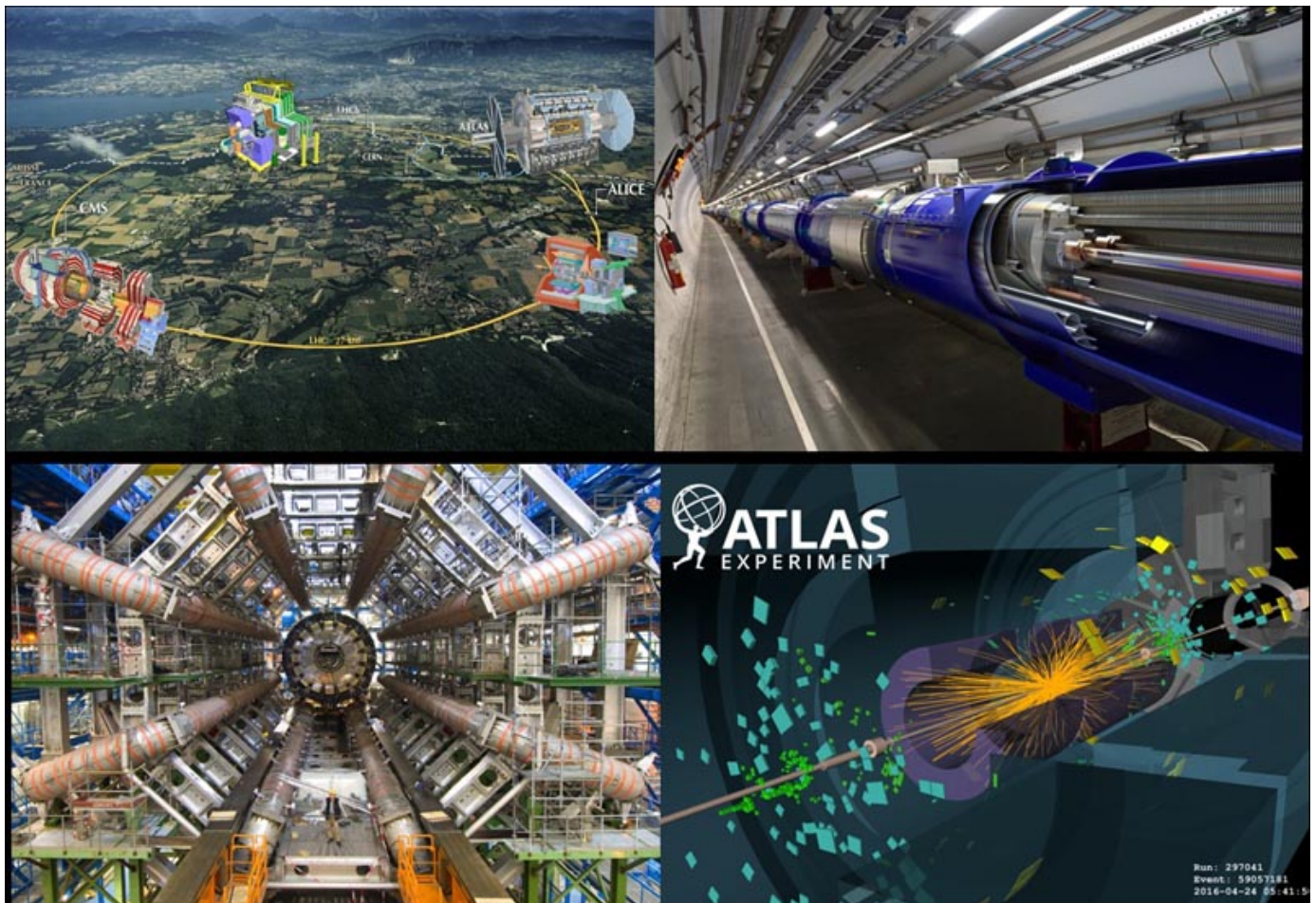


Figure 2. LHC and its experiment scenarios

Why Public Cloud?

Since 2013, CERN has deployed an OpenStack-based private cloud to manage resources across its main data centre in Meyrin, Switzerland, and a remote extension in Budapest, Hungary. Today, this includes approximately 10,000 dual-CPU servers, with approximately 300,000 processor cores. These resources serve around 3600 projects, including high-density computing, physics data analysis, and virtual-service setup.

Planned upgrades to the LHC and the experiments at CERN mean that computing and storage requirements are set to significantly increase. For instance, by the time the successor to the LHC, the High-Luminosity LHC, comes online in 2026, the ATLAS and CMS, experiments expect collect and reconstruct five to ten times more collision events than today. Using current software, hardware, and analysis techniques, it is estimated that the computing capacity required would be around 50-100 times higher than today. The data storage needs are expected to be in the order of exabytes by this time.

There is no expectation that budgets will increase in the manner required to close this 'resource gap' by simply increasing the total ICT resources available. For this reason, and to ensure maximum efficiency of resource use, it is vital to explore new technologies and methodologies. One approach under investigation is the use of commercial cloud resources in a hybrid model that would enable CERN to dynamically extend its in-house resources with those from commercial providers.

Teaming up with Open Telecom Cloud to Address Future Challenges

This approach is being investigated through a project known as [Helix Nebula](#). Earlier this year, three consortia were selected — through an open competitive tender — to enter a 'prototype phase'.

One of these consortia is based on Open Telekom Cloud (OTC). OTC is a public cloud platform co-built by Deutsche Telekom and Huawei, in which Huawei provides an OpenStack-based architecture and support for other solutions. In this prototyping phase, OTC has delivered HPC public cloud solutions to CERN in the field of scientific computing, deployed thousands of HPC nodes, and analyzed the results of high-energy particle collisions.

"The biggest advantage of OTC is that it is based on OpenStack, with an architecture the same as that of the CERN's existing private cloud. This means it could be used in a flexible manner to dynamically extend CERN's in-house resources," Mr. Eldik said.

Both CERN and Huawei are contributors to the OpenStack project, with Huawei a platinum member of the OpenStack Foundation. In the future, Huawei is looking forward to more in-depth technical

cooperation with CERN using OpenStack.

"Through a [CERN openlab](#) project, we have started to work jointly on the improvements for OpenStack, particularly to run OpenStack at large scales... And this will allow everybody in the OpenStack community to benefit from these community efforts. I'm very excited about these particular projects and I look forward to seeing the results of it in production in our clouds in the coming years," added Mr. Eldik.

[Share](#)

Information For

About HUAWEI

News & Events

Industry Insights

Quick Links

Follow us

[Terms of use](#) | [Privacy](#) | [Contact](#)

