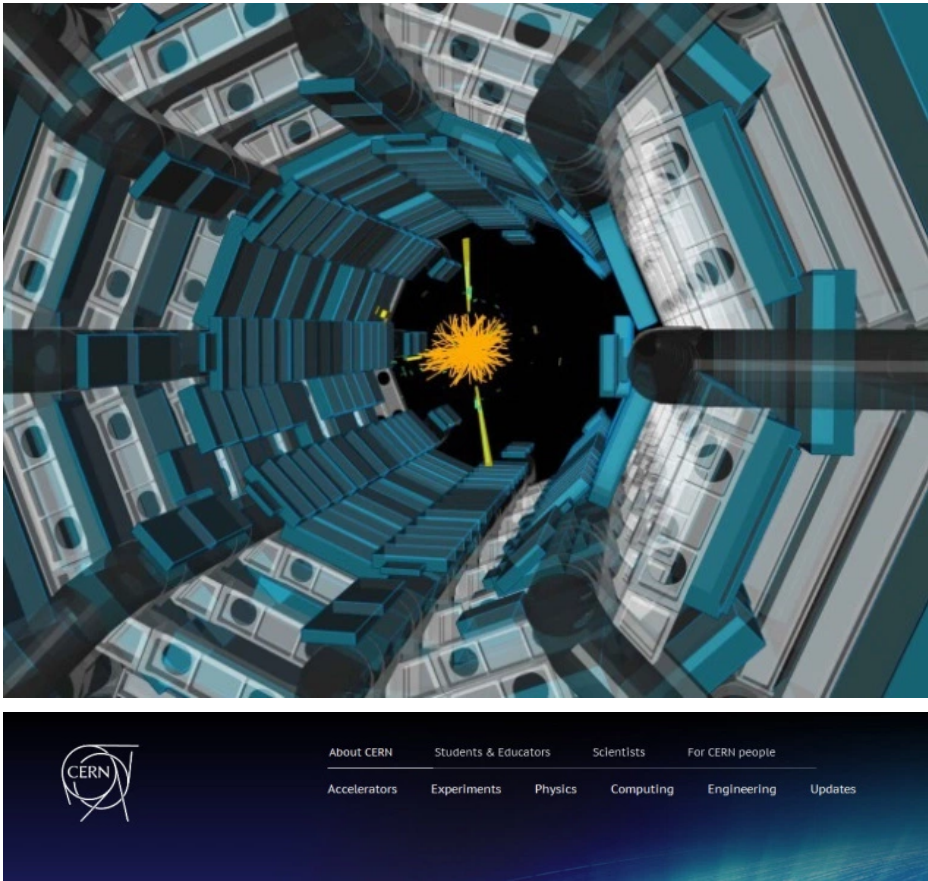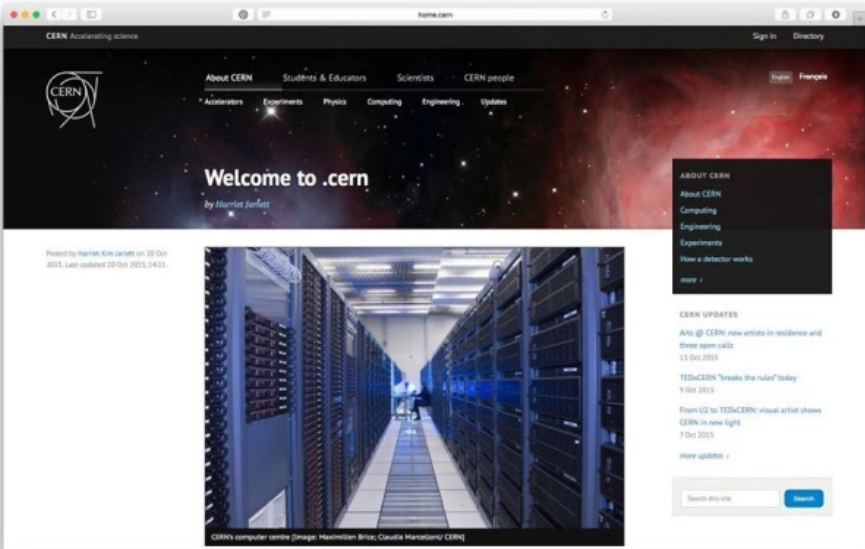# Science Springs

Dedicated to spreading the Good News of Basic and Applied Science at great research institutions world wide. Good science is a collaborative process.
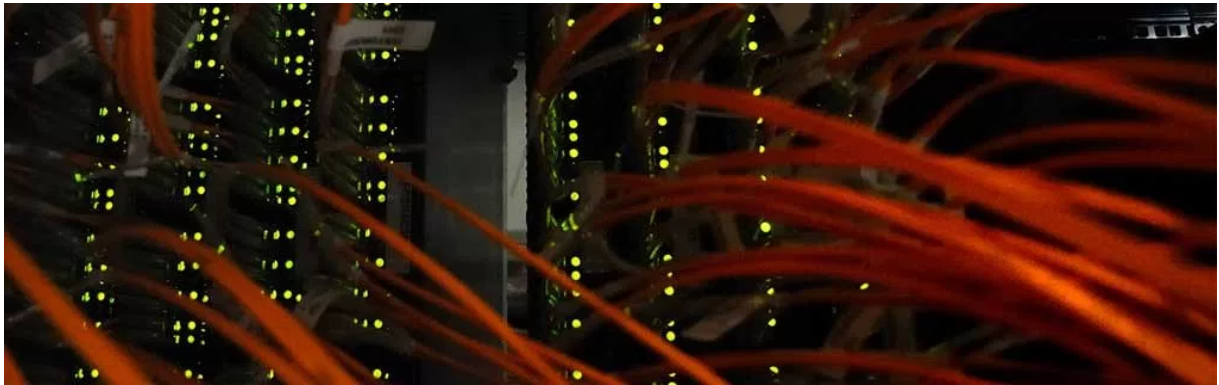
richardmitnick 6:10 pm on November 22, 2017

Permalink (https://sciencesprings.wordpress.com/2017/11/22/from-cern-fermilab-joins-cern-openlab-works-on-data-reduction-project-with-cms-experiment/)

Tags: Accelerator Science (623), Basic Research (8,214), CERN CMS (63), CERN openlab (4), FNAL (585), HEP (914), Intel, Particle Accelerators (757), Particle Physics (1,099), Physics (897)

From CERN: "Fermilab joins CERN openlab, works on 'data reduction' project with CMS experiment"
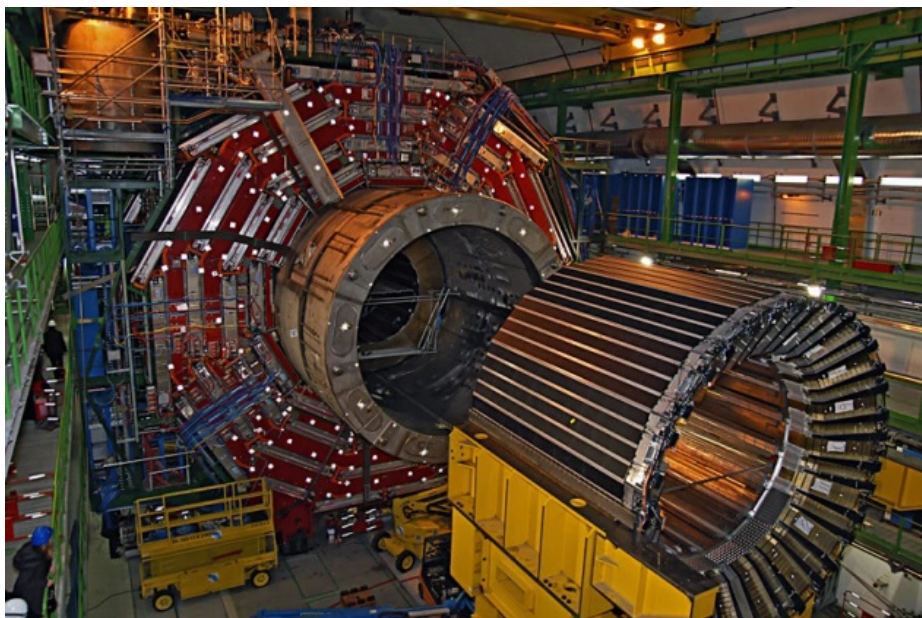

CERN

Fermilab Wilson Hall

Fermilab, the USA's premier particle physics and accelerator laboratory, has joined CERN openlab as a research member. Researchers from the laboratory will collaborate with members of the CMS experiment and the CERN IT Department on efforts to improve technologies related to 'physics data reduction'. This work will take place within the framework of an existing CERN openlab project with Intel on 'big-data analytics'.
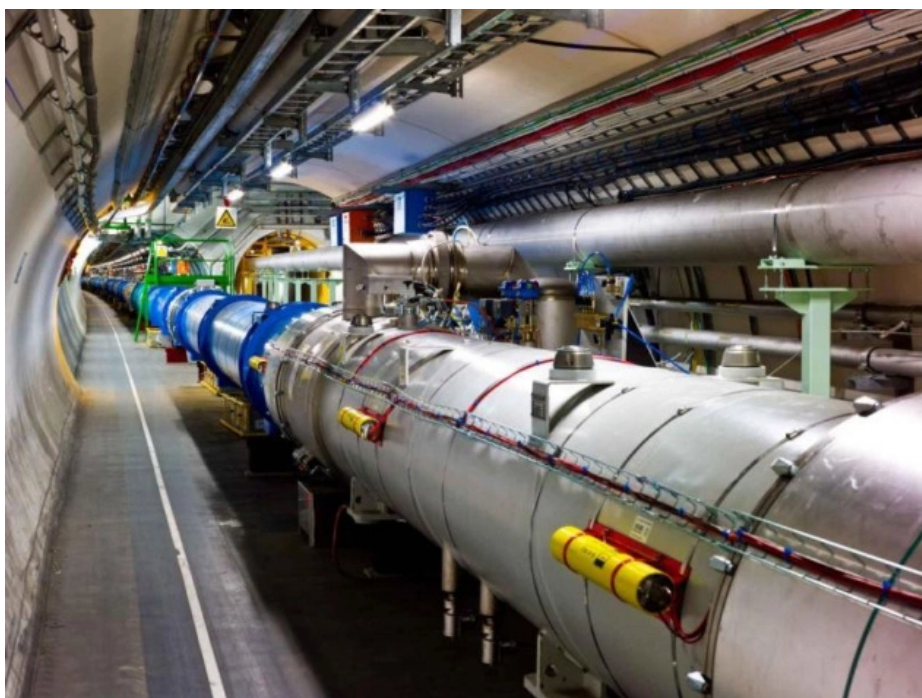
CERN/CMS Detector

'Physics data reduction' plays a vital role in ensuring researchers are able to gain valuable insights from the vast amounts of particle-collision data produced by high-energy physics experiments, such as the CMS experiment on CERN's Large Hadron Collider (LHC).

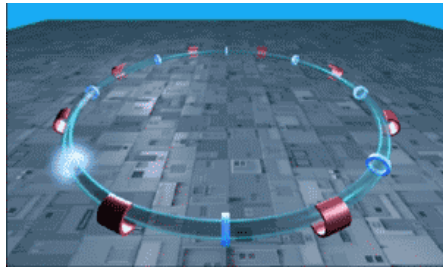LHC

CERN/LHC Map

CERN LHC Tunnel

CERN LHC particles

The project's goal is to develop a new system — using industry-standard big-data tools — for filtering many petabytes of heterogeneous collision data to create manageable, but rich, datasets of a few terabytes for analysis. Using current systems, this kind of targeted data reduction can often take weeks; but the aim of the project is to be able to achieve this in a matter of hours.

"Time is critical in analysing the ever-increasing volumes of LHC data,"says Oliver Gutsche, a <u>Fermilab</u> scientist working at the CMS experiment. "I am excited about the prospects CERN openlab brings to the table: systems that could enable us to perform analysis much faster and with much less effort and resources." Gutsche and his colleagues will explore methods of ensuring efficient access to the data from the experiment. For this, they will investigate techniques based on Apache Spark, a popular open-source software platform for distributed processing of very large data sets on computer clusters built from commodity hardware. "The success of this project will have a large impact on the way analysis is conducted, allowing more optimised results to be produced in far less time," says Matteo Cremonesi, a research associate at Fermilab. "I am really looking forward to using the new open-source tools; they will be a game changer for the overall scientific process in high-energy physics."

The team plans to first create a prototype of the system, capable of processing 1 PB of data with about 1000 computer cores. Based on current projections, this is about 1/20th of the scale of the final system that would be needed to handle the data produced when the <u>High-Luminosity LHC</u> comes online in 2026.

Using this prototype, it should be possible to produce a benchmark (or 'reference workload') that can be used evaluate the optimum configuration of both hardware and software for the data-reduction system.

"This kind of work, investigating big-data analytics techniques is vital for high-energy physics — both in terms of physics data and data from industrial control systems on the LHC," says Maria Girone, CERN openlab CTO. "However, these investigations also potentially have far-reaching impact for a range of other disciplines. For example, this CERN openlab project with Intel is also exploring the use of these kinds of analytics techniques for healthcare data."

"Intel is proud of the work it has done in enabling the high-energy physics community to adopt the latest technologies for high-performance computing, data analytics, and machine learning — and reap the benefits. CERN openlab's project on big-data analytics is one of the strategic endeavours to which Intel has been contributing," says Stephan Gillich, Intel Deutschland's director of technical computing for Europe, the Middle East, and Africa. "The possibility of extending the CERN openlab collaboration to include Fermilab, one of the world's leading research centres, is further proof of the scientific relevance and success of this private-public partnership."

See the full article <u>here</u>.

Please help promote STEM in your local schools.

<u>Stem Education Coalition</u>

About <u>CERN openlab</u>

CERN openlab is a unique public-private partnership that accelerates the development of cutting-edge solutions for the worldwide LHC community and wider scientific research. Through CERN openlab, CERN collaborates with leading ICT companies and research institutes.

Within this framework, CERN provides access to its complex IT infrastructure and its engineering experience, in some cases even extended to collaborating institutes worldwide. Testing in CERN's demanding environment provides the ICT industry partners with valuable feedback on their products while allowing CERN to assess the merits of new technologies in their early stages of development for possible future use. This framework also offers a neutral ground for carrying out advanced R&D with more than one company.

CERN openlab was created in 2001 (link is external) and is now in the phase V (2015-2017). This phase tackles ambitious challenges covering the most critical needs of IT infrastructures in domains such as data acquisition, computing platforms, data storage architectures, compute provisioning and management, networks and communication, and data analytics.

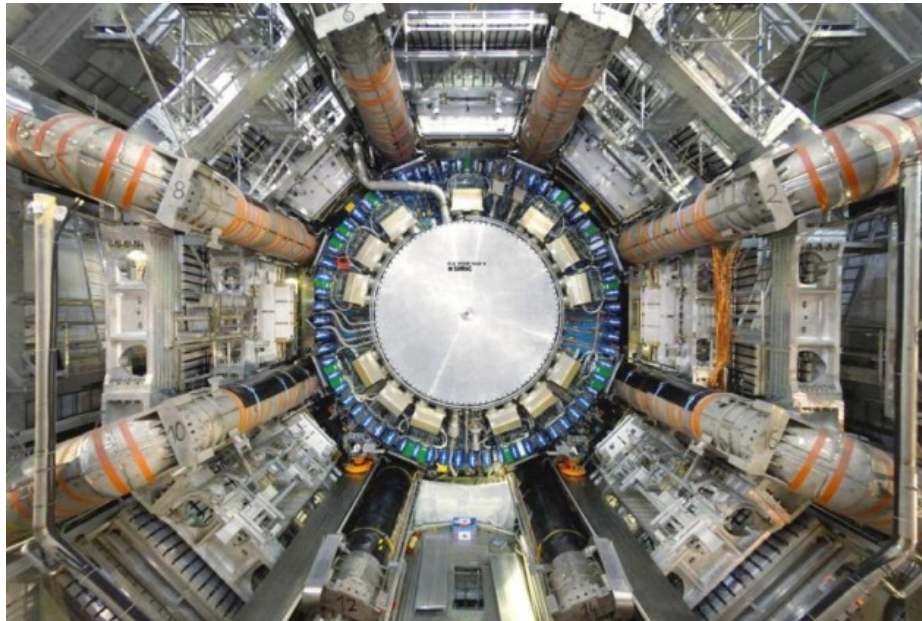Meet CERN in a variety of places:

Quantum Diaries

Cern Courier

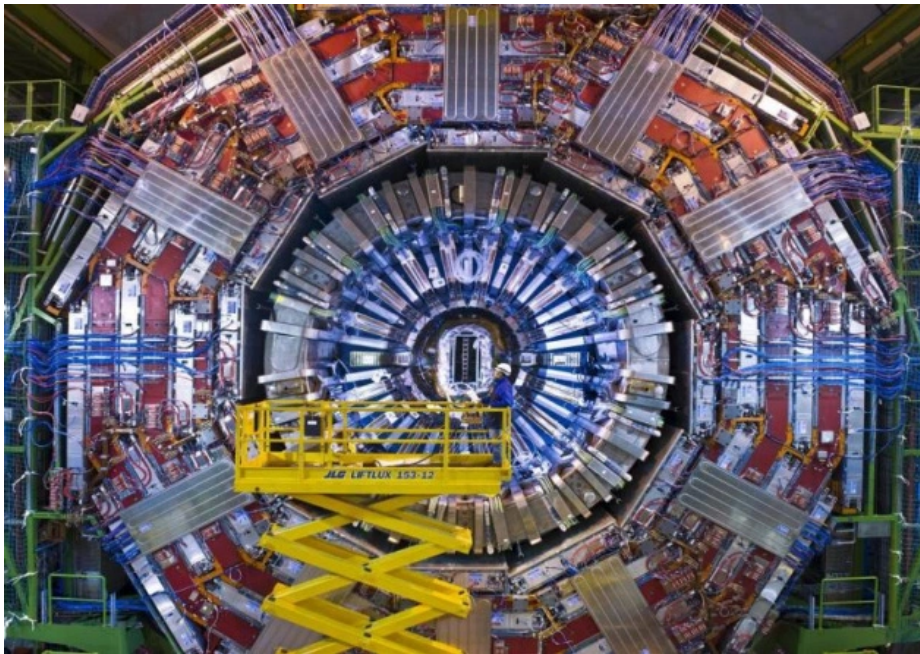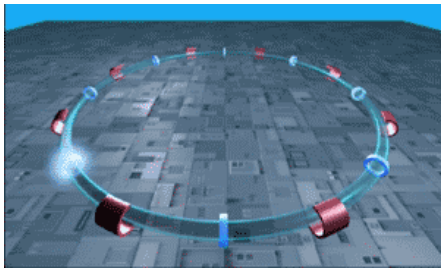
THE FOUR MAJOR PROJECT COLLABORATIONS

ATLAS

ALICE

CMS

LHCb

LHC

Quantum Diaries