the Availability Digest

www.availabilitydigest.com
@availabilitydig

Large Hadron Collider Running Out of Disk

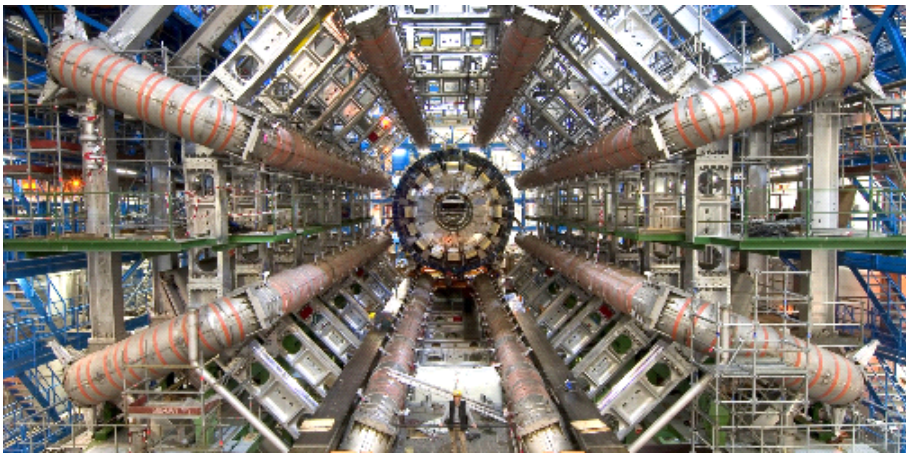
October 2016

The Large Hadron Collider (LHC) is a massive machine that accelerates protons speeding in opposite directions to study the effects of proton collisions. The results of collisions are recorded on hard disk for later analysis. The LHC was originally outfitted with enough disk to handle its anticipated data storage needs for years. However, collision rates have been far more than expected, and the LHC is now running out of disk space.

The Large Hadron Collider

The LHC is by far the most expansive machine ever implemented by man. It is arguably one of the most important scientific tools humanity has ever come up with. It allows us to glimpse what happens at the subatomic level.

The LHC is located at CERN, the European Organization for Nuclear Research, in Geneva, Switzerland. Constructed between 1983 and 1988, it is a circular superconducting tunnel seventeen miles in circumference straddling the Swiss/French border. It is twelve feet wide, and its depth ranges from 164 feet to 574 feet underground.



The Large Hadron Collider

The largest machine in the world, it is designed to smash protons together at nearly the speed of light. The LHC is actually a series of looped accelerators, each designed to speed protons over a certain range. Each accelerator is equipped with more and more powerful magnets to accelerate protons in opposite directions first up to 99% of the speed of light and ultimately up to 99.9999% of the speed of light.

There are four laboratories spaced around the LHC that have batteries of sensors to monitor the collisions. The sensors monitor the subatomic rubble that results from the collisions to scrutinize it for novel particles and the forces that hold them together.

The LHC is administered by thousands of scientists from hundreds of countries around the world. Their activities are coordinated by CERN.

Worldwide LHC Computing Grid

The LHC generates tens of petabytes of data each year. This mass of data is processed by the Worldwide LHC Computing Grid. This is an international collaborative project supporting a grid-based computer network comprising 170 computing facilities in 36 countries. It is the world's largest computing grid.

The LHC Computing Grid was designed by CERN to handle the prodigious amount of data produced by the LHC. The Grid comprises 200,000 processing cores and 150 petabytes (150×10^{15}) of disk space.

The LHC generates about 300 gigabytes (300×10^{9}) of data per second, which is filtered down to about 300 megabytes per second. This amounts to about 27 terabytes (27×10^{12}) of raw data per day.

Overall, about 25 petabytes of data are generated each year and must be stored and processed.

The Unanticipated Data Rates

The LHC is running out of disk space. It has been very stable and very reliable, and that is the problem.

When the LHC was designed, scientists expected that it would be running a third of the time. The rest of the time would be used for maintenance. Instead, the LHC has been running 70% of the time, more than twice what was expected. This may be the first time in history that technicians made a bad estimate of uptime in the wrong direction.

Coupled with the extended running time and an increase in the LHC power leading to a higher rate of proton collisions, the amount of data that is being generated is far in excess of what was originally anticipated. The LHC is currently producing about a billion proton collisions per second. This year alone, the LHC had produced roughly the same number of collisions as it did during all of the previous years of operation put together.

Since April, the LHC has produced roughly 2.4 quadrillion (2.4×10^{15}) particle collisions, and this rate is scheduled to increase. The LHC is scheduled to be operational until 2037, and the 2.4 quadrillion collisions represents just 1% of the total number of collisions planned during its lifetime. The LHC is scheduled to undergo several rounds of upgrades to further increase its collision rate.

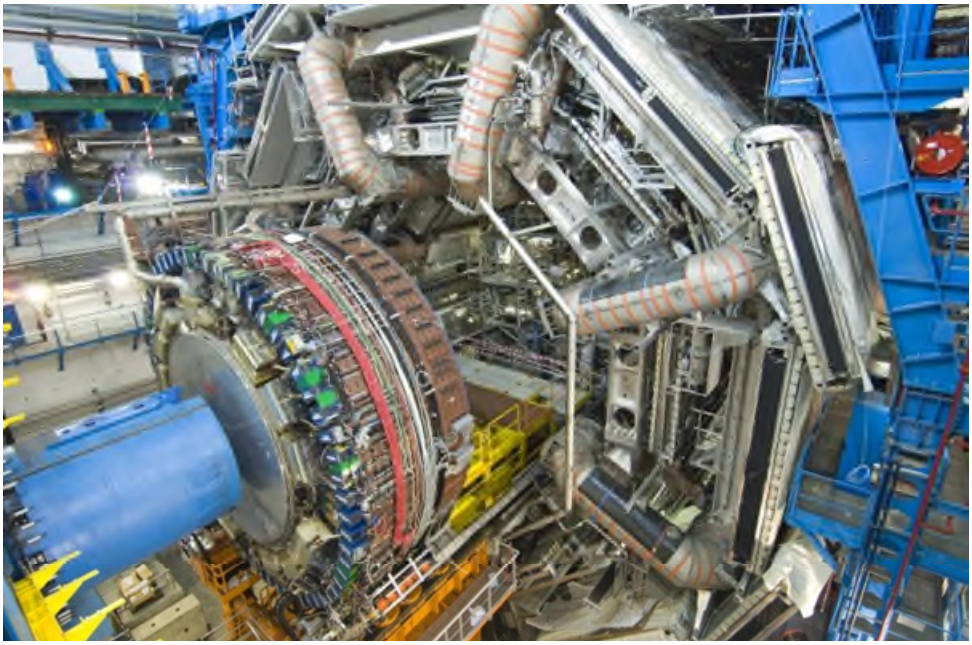
Lessons from Proton Collisions

As the speed of proton collisions increases, scientists can learn more about the rare processes and particles that make up the subatomic world. It was the faster collisions that led scientists to finally detect the Higgs boson, the particle (or actually the energy packet) that provides mass to other particles. The Higgs boson had been predicted many years ago (by Dr. Peter Higgs and five other scientists in 1964). A Higgs boson is created about once for every billion proton collisions.

There are a number of different experiments going on with the LHC, and they all have to jockey for time on the accelerator.

Great Expectations from Fewer Collisions

At the other end of the speed scale is the Atlas experiment. It is studying low speed, low energy collisions of protons. This occurs when two protons survive their collisions – they bounce and head off in slightly different directions without creating any new particles. The experiment hopes to use this study to understand why cosmic rays split into secondary particle when they strike the Earth's atmosphere.



The Atlas Experiment

Another low energy experiment is TOTOM. This experiment hopes to measure the probability that two protons can pass directly through one another without interacting at all. After all, a proton is mostly empty space, being make up of three quarks held together by gluon.

The TOTEM and Atlas experiments run detectors on both sides of two of the big LHC detectors. They study the protons as they come out of elastic scattering interactions at slightly changed angles

Running Out of Disk Storage

The LHC is running out of disk space to store the results of all of its experiments. It is recording collisions 70% of the time, twice its expected rate when the LHC was originally designed.

The number of drives that were purchased were determined years in advance based on projected luminosity (the collision rate). The LHC is outperforming all of its original estimates. Technical staff has had to consolidate old simulations and data to make room for data for new collisions.

The current hard drives have a 250 petabyte capacity. LHC is buying additional drives to handle the overflow. Its technical staff is also exploring other options such as a hybrid system that incorporates cloud computing with storage and data analytics.

All LHC experiments are planning to upgrade their experimental infrastructure during the winter of 2017.

Summary

The LHC is the most important tool scientists have to study the subatomic world. It has already discovered the Higgs boson and several other particles that result from proton collisions. As its energy is increased, it is yet unknown what other discoveries it will make.

The amount of data that is generated and that must be stored and processed is extraordinarily large. This has led to an underestimate of the amount of data that must be stored. The LHC is being outfitted with petabytes of additional storage to manage this data explosion.

Acknowledgements

Information for this article was taken from the following sources:

How does the Large Hadron Collider work?, *Extreme Tech*; July 25, 2015.
Pop! Goes the Weasel! *Availability Digest*; June 2016.
http://www.availabilitydigest.com/public_articles/1106/cern_weasel.pdf
Great expectations from fewer collisions, *CERN*; September 22, 2016.
LHC smashes old collision records, *Symmetry*; September 29, 2016.
The Large Hadron Collider is running out of disk space, *Extreme Tech*; October 3, 2016.
Physics Has a Problem – The Large Hadron Collider Is Running out of Disk Space, *Futurism*; undated.
Worldwide LHC Computing Grid, *Wikipedia*.