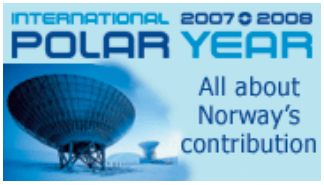

Press Releases Printable version

For further information, please contact:

IOP Institute of Physics

Please register to view contact details

03 November 2008

What to do with 15 million gigabytes of data

When it is fully up and running, the four massive detectors on the new Large Hadron Collider (LHC) at the CERN particle-physics lab near Geneva are expected to produce up to 15 million gigabytes, aka 15 petabytes, of data every year. Andreas Hirstius, manager of CERN Openlab and the CERN School of Computing, explains in November's Physics World how computer scientists have risen to the challenge of dealing with this unprecedented volume of data.

When CERN staff first considered how they might deal with the large volume of data that the huge collider would produce when its two beams of protons collide, in the mid-1990s, a single gigabyte of disk space still cost a few hundred dollars and CERN's total external connectivity was equivalent to just one of today's broadband connections.

It quickly became clear that computing power at CERN, even taking Moore's Law into account, would be significantly less than that required to analyse LHC data. The solution, it transpired during the 1990s, was to turn to "high-throughput computing" where the focus is not on shifting data as quickly as possible from A to B but rather from shifting as much information as possible between those two points.

High-performance computing is ideal for particle physics because the data produced in the millions of proton-proton collisions are all independent of one another - and can therefore be handled independently. So, rather than using a massive all-in-one mainframe supercomputer to analyse the results, the data can be sent to separate computers, all connected via a network.

From here sprung the LHC Grid. The Grid, which was officially inaugurated last month, is a tiered structure centred on CERN (Tier-0), which is connected by superfast fibre links to 11 Tier-1 centres at places like the Rutherford Appleton Laboratory (RAL) in the UK and Fermilab in the US. More than one CD's worth of data (about 700 MB) can be sent down these fibres to each of the Tier-1 centres every second.

Tier 1 centres then feed down to another 250 regional Tier-2 centres that are in turn accessed by individual researchers through university computer clusters and desktops and laptops (Tier-3).

As Andreas Hirstius writes, "The LHC challenge presented to CERN's computer scientists was as big as the challenges to its engineers and physicists. The computer scientists managed to develop a computing infrastructure that can handle huge amounts of data, thereby fulfilling all of the physicists' requirements and in some cases even going beyond them."

Also in this issue:

•President George W Bush's science adviser, the physicist John H Marburger, asks whether Bush's eight years in office have been good for science in the US.

•Brian Cox may be the media-friendly face of particle physics, but how does the former D:Ream pop star, now a Manchester University physics professor, find the time for both research and his outreach work?

•Beauty and the beast: in his 100th column for Physics World, Robert P Crease asks whether CERN's Large Hadron Collider, the biggest experiment of all time, can be dubbed "beautiful".

Notes for editor

Please register to view editor notes

Reference URL

http://www.physicsworld.com

Keywords (click on a keyword for related items)

Computing, Physics