

Computing Grid Helps Get to the Heart of Matter

In November, when physicists at CERN in Switzerland begin their grand experiment using the world's largest particle accelerator—the Large Hadron Collider, or LHC—computer scientists there and across the globe will also put the world's largest scientific computing grid through its paces.

The success of the experiment—intended to answer such questions as what other particles exist in the universe that we don't know about—will rely in large part on a worldwide, high-speed network that will allow scientists to harness the power of 100,000 computers—mostly PCs—to process the tons of data generated by the experiment.

The universe consists of particles of matter, and scientists currently know only a tiny fraction of them. Gaining greater insight into what else makes up the universe will give them a greater understanding of the universe itself.

The network's 10G bps backbone, linking to 11 scientific data centers, forms the core of the world's largest international scientific grid service, which was created to enable scientists to handle the huge amount of data that will come out of the experiment.

"The LHC is a 27-kilometer ring underground that accelerates protons to high energy and smashes them together in the ring to produce a fireworks of particles," said Francois Grey, director of IT communications at CERN, in Geneva. "Huge underground detectors will pick up the signals [from the collisions] using millions of channels that will read out every 25 nanoseconds. The rate at which [the data] will come out [of the four detectors in place] to be stored is in the hundreds of megabytes per second."

Along with lessons about what the universe comprises, the LHC Computing Grid project will teach network engineers valuable lessons about what it takes to run and manage one of the largest 10G-bps networks in the world.

"Everyone is looking to see who's installing a large backbone on that scale. We've become a reference for other people waiting to see what happens," Grey said. "We have no choice because we need that speed. We're also learning a lot about shipping data at high rates and how to optimize a grid between 10G bps and slower links."

About 200 institutions in 80 countries—some with their own large data centers—will participate in the grid to help process an expected 15 petabytes of data per year generated by the LHC.

"We realized early on there was no way we could store all that data and analyze it here at CERN," Grey said. "The idea was to pull those resources together in a grid."

The grid is organized in a three-tiered hierarchy, with CERN serving as the Tier 0 "fountainhead" from which data subsets will be dispersed to 11 Tier 1 data centers in Europe, North America and Asia, according to Grey.

Tier 2 data centers, located mostly at more than 250 universities around the globe, will serve as the locations where physicists analyze the data subsets they receive.

The LHC Computing Grid rides on dark fiber used in national and international research networks to interconnect each of the 11 Tier 1 sites at 10G bps for continuous paths to the different locations. Commercial links are used to connect participants in Canada, Taiwan and the United States.

In North America, Tier 1 sites include two in the United States—Fermi National Accelerator Lab, in Batavia, Ill., and Brookhaven National Laboratory, in Long Island, N.Y.—as well as the Triumph Laboratory, in Vancouver, British Columbia.

Because of the nature of the computing task, PCs used in the grid don't have to communicate at very high speeds with one another so they are linked via grid middleware for "trivially parallel" processing, according to Grey.

Detectors read out images of the collisions, which are analyzed for particular patterns. "Each collision is independent from the next one, which is why trivially parallel processing works," Grey said.

At CERN, the PCs, CPU servers and disks are linked on a 1G-bps network provided by Hewlett-Packard ProCurve switches. CERN itself will contribute 10 percent of the total necessary processors for the job, including 3,500 PCs and the rest single- or dual-core processors all running a version of Linux called Scientific Linux CERN. CERN will contribute about 8,000 processors to the computing tasks.

The PCs used at CERN are commodity systems from a mix of smaller vendors, including Elonex, of Bromsgrove, England, and Scotland Electronics, of Moray, Scotland.

"We buy them cheap and stack them high," said David Foster, communications systems group leader in CERN's IT department. "The physics applications can run in parallel, but independently on separate boxes, so any PC which fails can be replaced and just that job restarted."

"Our typical workhorses" are dual-processor PCs in a one-rack unit "pizza box" form factor stacked in 19-inch racks, according to Helge Meinhard, technical coordinator for server procurements at CERN.

Although most of the roughly 8,000 PCs are single-socket machines that run single-core chips, about 750 are two-socket systems that use dual-core processors.

Administering all the PCs is a batch scheduler, which identifies available units and assigns a job to them.

HP switches, including 600 ProCurve 3400cl, 400 ProCurve 3500y1 and 20 ProCurve 5400-series devices, link the CERN processors at 1G bps, with 10 Gigabit uplinks into the grid's core backbone. The network uses primarily fiber connectivity, although it also uses some UTP (unshielded twisted-pair) Category 6 copper cabling for 1G bps links.

Sixteen 10G bps routers from Force10 Networks in the core backbone link the CERN network to other participants in the grid.

HP and Force10 Networks were chosen for the LHC grid project because of their feature set, cost-effectiveness and "a great willingness [by HP] to work with us at an engineering level on the challenges," Foster said.

Key features in the ProCurve switches include security and manageability. "We must be able to automate management to run such a large network, and we want to secure it to the level that we allow only authorized MAC [media access control] addresses to access the network," he said.

HP officials simplified management of the switches for CERN by enabling different types of management functions to be executed on the switches using the industry-standard SNMP. HP made it possible to "get the temperature of the switch via SNMP and do configuration of the switch using SNMP," said Pierre Bugnon, account manager for HP, in Geneva.

"We also need to make sure we can work with [HP] in the future and make sure they are open to

collaboration to figure out how to do more. The relationship we established and their technology road map are very good, too," Grey said.

"CERN is a very important customer for us in terms of relationship," said Victor Svensson, business development manager for HP, in Grenoble, France. "We're providing a very high level of support directly to CERN. It goes beyond the sales engagement to include strategic collaboration between the companies."

HP's strict adherence to standards was also a key in its selection. "One thing that really helped is that we worked off standards. That was a key requirement," said Svensson.

"[CERN] didn't want proprietary features. They have a lot of different [networking] firms involved," added Bugnon.

Once the experiments begin, 7,000 scientists will analyze subsets of the data, looking for proof of the elusive Higgs boson particle or theoretical supersymmetric particles not yet proved. The Higgs boson particle has never been seen, but scientists believe that if it can be identified, it could help explain why an electron has a negative charge and a proton has a positive charge.

While scientists using telescopes can perceive 3 percent of what the universe is made up of, the other 97 percent remains a mystery. "There are many candidates for what dark matter could be—one could be supersymmetric particles. The hope is to find them," Grey said. "It's fairly esoteric, but it's also pretty fundamental. It's about understanding the universe."

The glue that holds the project together and makes the data capture and analysis possible is the grid middleware, "a layer of software that allows you to do your analysis without having to worry where the data is or the computing power on the grid," Grey said.

The middleware, which optimizes use of the grid, includes such elements as resource brokers that determine at any given point in time which data centers have the necessary capacity for a task submitted by an authorized physicist and determines where the task will be handled.

The types of jobs that the middleware, developed in-house, will distribute out across the processors fall into three categories. "It is either simulation of physics interaction in the detector; reconstruction of real detector signals or of simulated data; [or] physics analysis, where the outcome of many positions are sampled in a statistical way," Meinhard said.

The middleware also implements authentication and authorization to ensure that research institutions supporting the grid have appropriate access and that "others are not getting a free ride," Grey said.

The grid middleware, which represents more than 1 million lines of code created using the open-source Globus Toolkit, also performs accounting functions to "make sure nobody's hogging the grid," Grey said. "It also implements security and monitoring to ensure the grid is available 24 by 7."

"It is a big engineering effort to make sure the middleware is stable and runs well when the real data comes out later this year," Grey said. "It's being continuously improved and re-engineered and hardened to make sure it is [up-to-date]."

To date, as the accelerator is being completed, CERN and its partners are running simulations across the grid, shifting "gigabytes of files and large amounts of data" to test its mettle, Grey said.

CERN is also leading the charge to create a European multiscience grid that will support a range of scientists and experiments.

"The long-term vision with these grids is like the Web," Grey said. "At some point, they link up and standards develop, so that, as a scientist, you just submit [computing tasks] to the grid and don't ask which one [will complete it]. But we're quite a ways from this."

Once the experiments start in November, the project will gather data for 15 years, although the data could be studied for many years after the LHC shuts down.

Copyright 2007 by Ziff Davis Media, Distributed by United Press International

This document is subject to copyright. Apart from any fair dealing for the purpose of private study, research, no part may be reproduced without the written permission. The content is provided for information purposes only.