PHYSICS IN ACTION

Building the next IT revolution

The Grid is the only way to deal with data from the next generation of particle accelerators

From **Steve Lloyd** in the Department of Physics, Queen Mary, University of London

A new chapter in scientific computing opened earlier this summer when a prototype of the LHC Computing Grid (LCG) was deployed across three continents. The LCG project is designed to meet the unprecedented computing requirements of the Large Hadron Collider (LHC), which is currently being constructed at CERN. The LHC should start producing data in 2007 and will be used by thousands of physicists from universities and laboratories across the world to study fundamental mechanisms in particle physics such as supersymmetry and the search for the Higgs boson.

The giant experimental detectors at the LHC will generate more than 10 million gigabytes of data each year – which is equivalent to the storage capacity of about 20 million CDs – and would require more than 70 000 of today's fastest PC processors to analyse it all. The goal of the LCG project is to integrate the computing resources of the several hundred participating institutes into a worldwide computational "Grid".

Computer power

The Grid gets its name from its analogy with the electrical power grid, which provides electricity via a standard plug-and-socket interface throughout an entire country. The "power stations" in the Grid are clusters, or farms, of computers and the "power lines" are the fibre optics of the internet. The first grids in the late 1990s used several supercomputers in the US as a single resource to run applications that were too large or too complex for a single supercomputer. Today's grids, in contrast, are being constructed from large numbers of off-the-shelf PCs and disks. In addition to being relatively cheap and easy to integrate, the use of standard hardware enables the Grid to be made gradually bigger and more complex. Whereas a PC using the Web provides information or access to services, such as banking or shopping, a PC on the Grid offers its computing power and storage.

Distributed computing has been around for many years. The SETI@home project, for example, allows home computers to analyse datasets in a search for extraterrestrial intelligence. What distinguishes today's grids, however, is the development of special software called middleware. Previously a scientist wishing to run analysis programs over large datasets might have computer accounts in several international institutions. Researchers submitted jobs to a par-

Data deluge – collisions between particles at the Large Hadron Collider will take place 800 million times per second, and will produce more data per year than the amount currently handled by the world's entire telecommunications network.

ticular site, initiated file transfers manually between the sites and basically kept track of everything themselves.

In principle, the Grid middleware allows you to submit a job to the entire Grid. This software – which includes "resource brokers", "replica managers" and information services that run on some of the Grid computers – determines where best to run the job. It then automatically copies or moves the data files as necessary, before returning the results without the user knowing or caring where they came from. To the user, the Grid therefore looks like a very large distributed PC with the middleware acting as its operating system. Security is paramount in such a system, and users are authenticated by a single public-key digital certificate that acts like a passport. Users can also have different levels of authorization, which is administered through membership of "virtual organizations", like having visas in a passport.

The prototype grids that are currently being established use middleware from the US, such as the Globus Toolkit and Condor, and from new developments in Europe through the European DataGrid project

and others. The challenge is whether these grids can be scaled from a small number of computers to the huge numbers required in the future. This is precisely what projects such as the LCG will test. There are also sociological challenges to be overcome, such as how to balance local ownership of resources while making them available to the larger community and how to overcome local security worries about giving access to "anonymous" non-local users.

Enter e-science

In addition to solving the computational demands of particle physicists, grids and grid technology are being developed in almost every branch of science, along with many industries and businesses. In the UK, the use of large widely distributed computing systems such as grids for scientific research by large collaborations has become known as "e-science".

Several grid projects were showcased at the e-science "All Hands Meeting" this September, which was held in Nottingham in the UK. Astronomers, for example, are building virtual observatories that federate data from many sources and wavelengths. Instead of looking at all objects in the infrared at one centre and then all visible objects at another, for example, the Grid would allow users to look at an object at all wavelengths simultaneously.

In healthcare there are several projects designed to allow remote sharing of data, such as mammography images between consultants and radiographers. Chemists are using grids to simulate and visualize molecular structures. Meanwhile, projects like NASA's Information Power Grid initiative, which integrates computers, databases and instruments across its facilities in the US, is enabling large industries to integrate resources between sites. Even in banking, a typical risk calculation that currently takes a few hours on a single PC could be performed in a few minutes using the Grid. This would allow traders to make informed decisions on the spot about whether to buy or sell.

There is not yet, however, a "killer application" for the Grid, and it could be a long time before the public use the Grid as they use the Web. But any application such as gaming or video-streaming that needs much larger amounts of computer power or disk storage than that available on a single PC could benefit. We will have to wait to see if the Grid really is the next IT revolution. But for today's physicists and engineers it appears to be the only cost-effective solution for future computing demands.