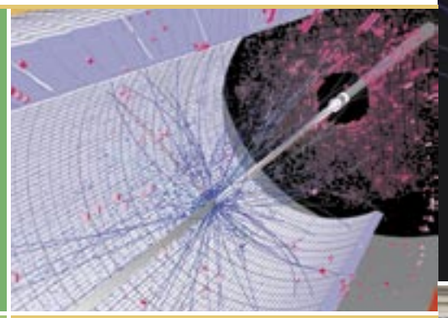
The Grid at CERN

CERN is the place 'where the Web was born'. Now, it is leading some of the most ambitious Grid projects in the world.

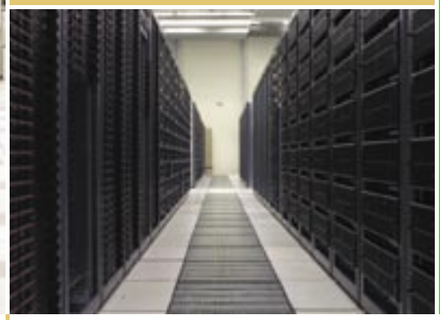
The LHC Computing Grid (LCG), which was launched in 2002, aims to integrate thousands of computers worldwide into a global computing resource to store and analyse the huge amounts of data that the experiments at CERN's Large Hadron Collider (LHC) will be collecting from 2007.



CERN's Computer Centre can now cope with an electric power demand of up to 1 MW, equivalent to the 5000 PCs that are planned to be in place by 2006 (1500 PCs today). Five Petabytes of data storage are available on tape and disk, with planned upgrade to 15 Petabytes by 2006. This includes primary tape backup of all LHC data.

To build and operate a permanent European Grid infrastructure that can serve a broad spectrum of applications, CERN is leading the pan-European consortium Enabling Grids for e-Science in Europe (EGEE). Grid software developed as part of EGEE will be used for the LCG, but its Grid infrastructure will extend beyond particle physics, to fields such as biomedical and geological applications.

A team at the CERN IT Department, using 45 parallel tape drives capable of writing to tape at 30 Megabytes/s, achieved storage-to-tape rates of 1.1 Gigabytes/s for periods of several hours. The average sustained over a three day period was 920 Megabytes/s.



The impact of cutting edge IT technologies on the Grid is assessed in the CERN openlab for DataGrid applications. Through close collaboration with leading industrial partners (Intel, Enterasys Networks, HP, IBM, and Oracle), CERN acquires early access to technology that is still several years from the commodity computing market. In return, CERN provides demanding data challenges to push these new technologies to their limits, and provides a neutral environment for integrating solutions from different partners to test their interoperability.

CERN
European Organization
for Nuclear Research
CH-1211 Geneva, Switzerland

Communication Group, June 2006
CERN-Brochure-2006-006-Eng

LCG

eGee
Enabling Grids
for E-sciencE

CERN
openlab

The Grid

Sharing resources owned by many different organizations to access remote computers, software, and data efficiently and automatically

Secure access to establish the identity of a user or resource, after defining conditions under which sharing occurs

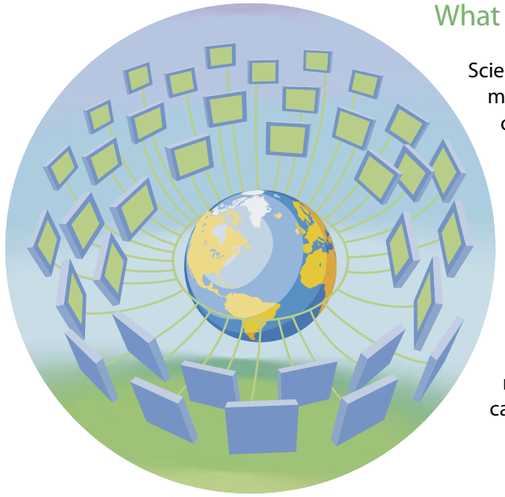
Bridging distance using high-speed connections between computers to create a global Grid

Open standards to allow applications designed for one Grid to run on all others

The Grid is a service for sharing computer power and data storage capacity over the Internet. It goes beyond simple communication between computers, and aims ultimately to turn the global network of computers into one vast computational resource.

For more information on the Grid see: www.gridcafe.org

www.cern.ch



What is the Grid?

Scientists are facing increasingly complicated problems that require more and more computing power and data storage capacities. It is often difficult, expensive, and sometimes impossible to achieve certain scientific goals with current computer technology. The Grid tries to tackle this problem by using computers all over the world and making them act as a single, huge and powerful computer.

Referring to Grid computing as 'the Grid' is a convenient shorthand but can lead to confusion, since there is not one single 'Grid'. Instead, many Grids are evolving—private, public, regional, global—which may be dedicated to one particular problem, or all-purpose. These Grids all have restricted capabilities for the moment, but are gradually growing and becoming more sophisticated.

Has this been done before?

There are already programs that use widely distributed computers linked through the Internet. For example SETI@home is a screensaver running on over 500 000 private PCs to analyse the data of the Arecibo radio telescope in Puerto Rico, searching for signs of extraterrestrial intelligence. This is a form of 'public distributed computing'. Different computers work simultaneously on different parts of the problem and pass the results to a central system for post-processing. Although there are significant technical differences, SETI@home is often used as an analogy for what the Grid aims to do. Whereas SETI@home runs on idle private PCs, the Grid uses dedicated resources in major computer centres and can thus handle problems of far greater complexity.

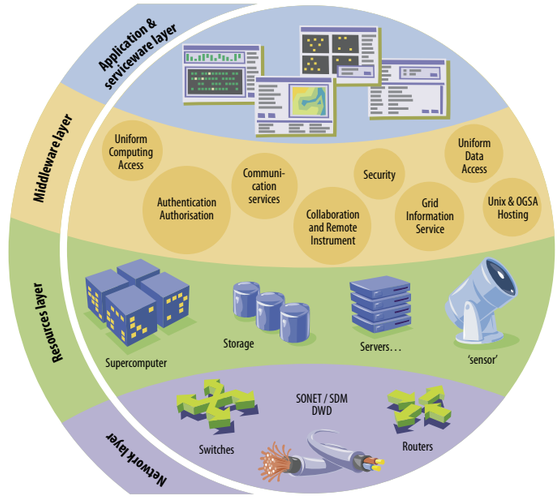
The Grid takes its name from an analogy with the electrical 'power grid'. The vision is that, as when plugging in an appliance, users should not have to worry about where the computing resources are coming from. In other words, users should treat the Grid as a utility from which they can tap computing power and access storage on demand.

Who is developing the Grid?

The Grid is a 'work in progress', with the underlying technology being developed by hundreds of researchers and software engineers around the world. Although its future is still uncertain, it has potentially revolutionary implications for Information Technology. This has attracted considerable interest not only from experts in computer science, but also from scientists in many fields and the business community.

How does it work?

Although the Grid depends on the underlying hardware of the Internet—computers and communications networks—it is novel software that enables the user to access computers distributed over the network. This software is called 'middleware', because it is conceptually between the operating systems software of the computer and the applications software that solves a particular problem. The middleware's task is to organize and integrate the disparate computational resources of the Grid into a coherent whole.



Who will use the Grid?

The first major users of the Grid will probably be scientists with challenging applications, who need large computing power or who have to deal with vast amounts of data. For example:

→ **Biologists** want to simulate thousands of molecular drug candidates to see how they would interact with specific proteins. To unlock the secrets of the **human genome** massive amounts of data have to be analysed, including the sequence of the three billion chemical units that comprise human DNA.

→ **Earth-scientists** keep track of the level of atmospheric ozone with satellite observations, downloading about 100 Gigabytes of raw images per day. Simulating the future of the Earth's climate is a task that requires huge processing power, which can be distributed over thousands of computers.

A driving motivation behind the Grid was the need for collaborative and multi-disciplinary applications in science and engineering, but similar types of application will become fundamental in many other fields, such as entertainment, commerce, finance, industry, and government.

The **World Wide Web** was invented at CERN for the use of scientists who needed it to share information. Nobody involved could have imagined the myriad of practical and commercial uses it has today. Similarly, though it is hard to imagine now, the Grid may be part of people's daily lives in a decade.