annual report

02

CERN

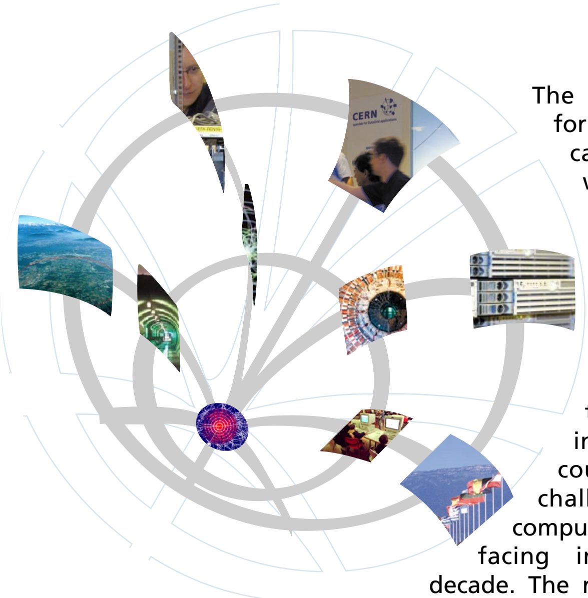
openlab for DataGrid applications

Table of Contents

Second Annual Report of the CERN openlab for DataGrid Applications
Editor: François Grey
Design and layout: Fabienne Marcastel
Photography: CERN photo service
Production: ETT Division CERN

ISBN 92-9083-212-6

A word from the Head of CERN openlab



The CERN openlab for DataGrid applications is now well on the way to fulfilling the ambitions of its founders. The goal was to create a new type of partnership between CERN and industry, which could tackle the challenges of Grid computing that CERN is facing in the coming decade. The results described in this second annual report testify to a solid activity that is already producing record-breaking achievements, and that is far from exhausting its potential for growth.

Since the first annual meeting in March 2002, much has happened. At that time, the original partners and CERN were still reeling from the aftermath of the telecoms crash and severe budgetary constraints. However, new life was injected in the partnership when HP joined in September 2002. Close collaboration between HP and the existing partners, Intel and Enterasys Networks led to the CERN opencluster project. This was a significant breakthrough, as it created a spirit of collaboration between partners which was precisely what CERN openlab was meant to achieve. At about the same time, CERN also boosted its efforts substantially, creating a dedicated and talented team to move the CERN openlab agenda forward.

In December 2002, Manuel Delfino, head of the CERN openlab and CERN IT Division Leader, stepped down to resume his academic duties in Spain, where he is now playing a leading role in national Grid activities. I took over his position officially from January 1st 2003, having been Deputy Division Leader prior to that. I would like to take this opportunity to express my sincerest appreciation to Manuel Delfino, on behalf of all partners, for having had the vision to help foster CERN openlab, and the determination to see it through a challenging first year.

What does the future hold for CERN openlab? With IBM joining in April 2003, the partnership will extend the CERN opencluster into the area of advanced storage – which is crucial for the CERN's DataGrid ambitions. And beyond that, the increasing emphasis on Grid-like solutions throughout the IT industry means that there are many opportunities for extending the scope of the partnership – I look forward to the exciting times ahead for CERN openlab!

Finally, I would like to thank all the industrial partners for their generous sponsorship and unwavering support of CERN openlab. Your commitment is the key to our common success.

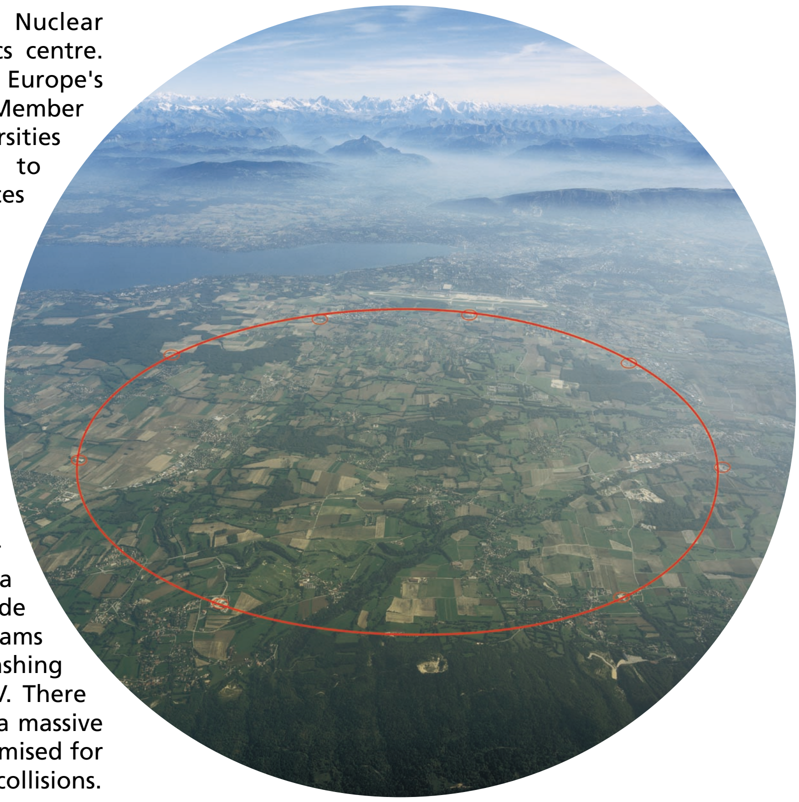
von Rüden

Wolfgang von Rüden
Head of the CERN openlab for DataGrid applications

THE CONTEXT

CERN and the LHC: probing the Universe

CERN is the European Organization for Nuclear Research, the world's largest particle physics centre. Founded in 1954, the laboratory was one of Europe's first joint ventures and includes today 20 Member States. Here, some 6000 physicists from universities and laboratories around the world come to explore what matter is made of and what forces hold it together.



CERN exists primarily to provide the physicists with the necessary tools. These are accelerators, which accelerate particles to almost the speed of light and detectors to make the particles visible.

Building the biggest ring

The world's largest and most powerful particle accelerator, the Large Hadron Collider (LHC), is now being constructed at CERN, in a 27km circular tunnel. It will ultimately collide beams of protons at an energy of 14 TeV. Beams of lead nuclei will be also accelerated, smashing together with a collision energy of 1150 TeV. There will be four detectors on the LHC ring, each a massive instrument several stories high, and each optimised for probing different aspects of the high-energy collisions.

The accelerator is planned to start operation in 2007. It will be used to answer some of the most fundamental questions of science.

- What is the origin of the different masses of fundamental particles?
- What is "dark matter" made of?
- Can the electroweak and strong forces be unified?
- What is the origin of the asymmetry between matter and antimatter in the universe today?

The LHC aims to provide answers to these questions, and in the process change our view of the Universe.

LHC and Grid computing: a grand challenge

Grid computing is the computer buzzword of the decade. Not since the World Wide Web was developed at CERN, over ten years ago, has a new networking technology held so much promise for both science and society. Once again, CERN is set to play a leading role in making the technology a reality.

The philosophy of the Grid is to provide vast amounts of computer power at the click of a mouse, by linking geographically distributed computers and developing "middleware" to run the computers as though they were a monolithic resource. Whereas the Web gives access to distributed information, the Grid does the same for distributed processing power and storage capacity.

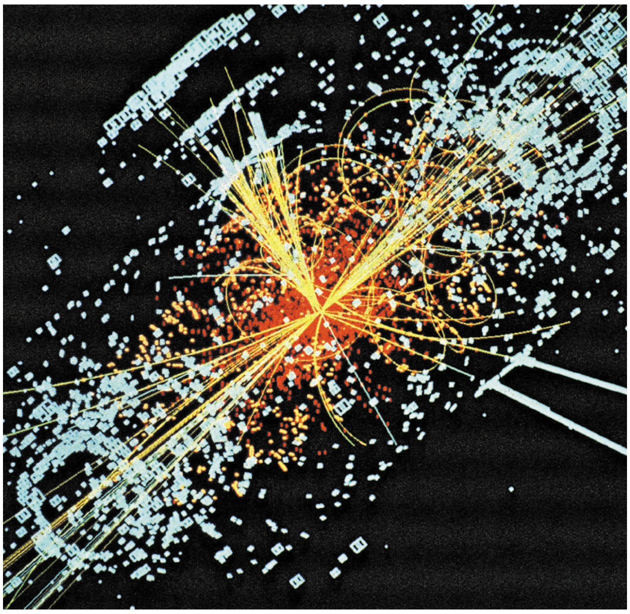
There are many varieties of Grid technology. In the commercial arena, Grids that harness the combined power of many workstations within a single organisation are already common. But CERN's objective is altogether more ambitious: storing petabytes of data from the LHC in a distributed fashion, and making it easily accessible to thousands of scientists around the world. This requires much more than just spare PC capacity – a network of major computer centres around the world must provide their resources in a seamless way.

Petabyte performance

A single particle collision, or "event", can generate a megabyte of data in the detectors, in the form of traces through the detector of the paths of particles created by the collision. In a typical experiment at the LHC, there will be some 40 million collisions per second. After on-line filtering, this huge data flow will be reduced to some 100 events of interest per second. This means a recording rate of up to one gigabyte/sec. Thus, over a year – and taking account the need for backup data – the total data stored from all four detectors on the LHC is expected to reach over 10 Petabytes. Since the lifetime of the experiment is expected to be at least 10 years, and researchers will need to analyse the totality of the data with increasingly sophisticated programs, the data challenge will grow with time.

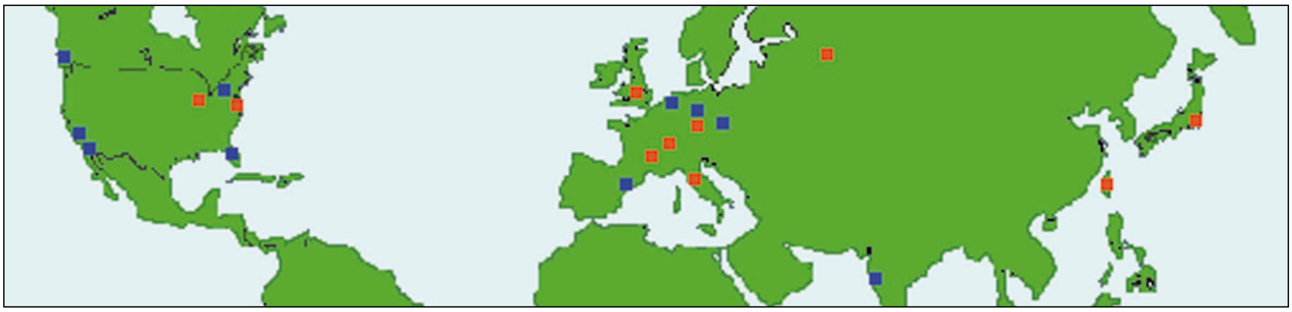
Among all measured tracks, the presence of "special shapes" is the sign for the occurrence of interesting interactions. The figure below shows a simulation of raw data from the LHC. Buried in the raw data are tracks that reveal an elusive particle known as the Higgs boson. Events such as this are expected to be extremely rare, happening in only 1 in 10^{13} collision, so the analysis algorithms must be very sophisticated. In parallel with data analysis, simulation of events plays a key role: starting from theory and detector characteristics, physicists will compute what the detector should register. Only by building up a similar quantity of simulated data can theory and experiment be compared on a statistical basis.

Simulation of a collision producing the Higgs boson, a particle predicted theoretically but that has so far eluded experiment.



The LHC Computing Grid (LCG)

The goal of the LCG project is to deploy a worldwide grid solution to meet the unprecedented computing needs of this virtual organisation.



Deployment of LCG-1.
Red squares: Main computer centres involved in the launch of LCG-1 in July 2003. The first production Grid will involve Fermilab, Brookhaven, RAL, IN2P3, CERN, INFN, FZ Karlsruhe, Moscow State Uni., Academica Sinica and Tokyo Uni. Blue squares indicate other centres expected to join LCG-1, during the course of the first 18 months of operation.

The year 2002 was a turning point for the construction of a computing Grid for the LHC. The LHC Computing Grid (LCG) project was officially launched, with a mission to integrate thousands of computers at dozens of participating centres worldwide into a global computing resource. This technological tour-de-force will rely on software being developed in the European DataGrid (EDG) project, led by CERN, which is the largest Grid development project ever funded by the EU. It will benefit from hardware developments initiated in the CERN openlab.

Gearing up

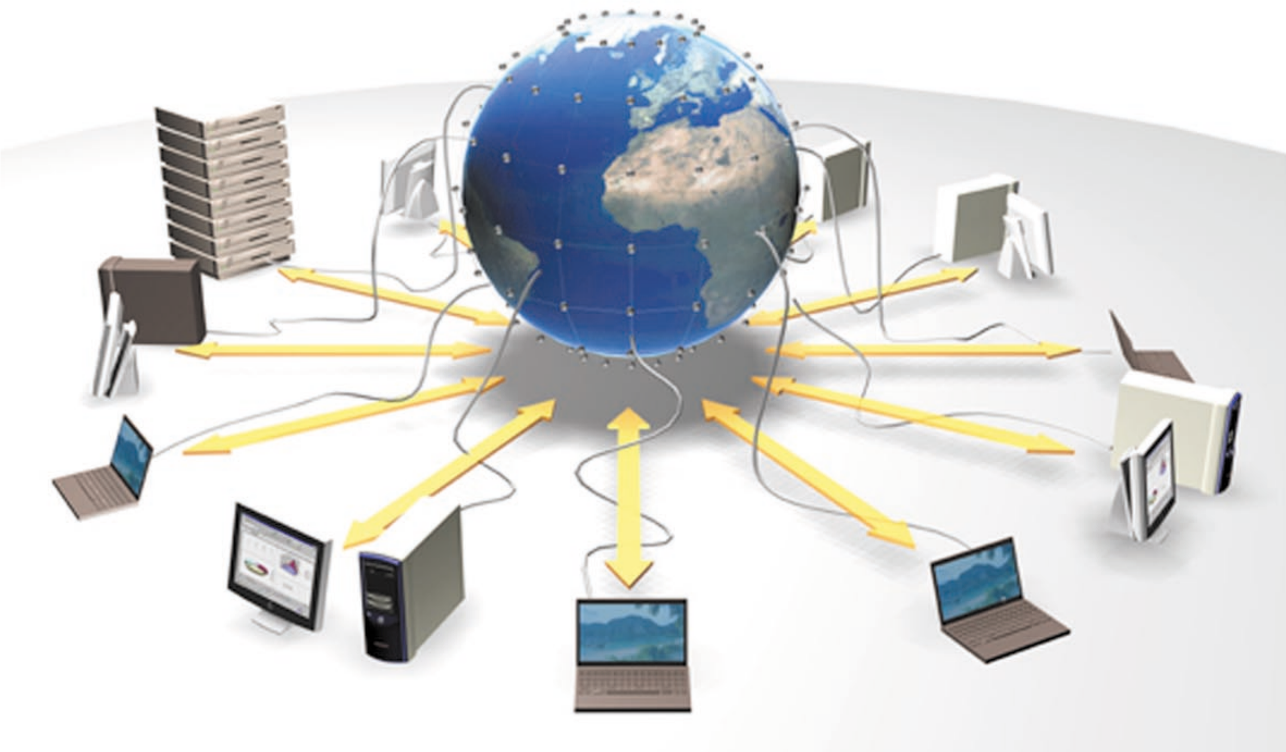
In 2002, LCG began rapidly gearing up for this challenge, with over 50 computer scientists and engineers from partner research centres around the world joining over the year. The focus in 2002 was on defining the stringent data storage and processing requirements of the experiments. For example, a key technical requirement is to ensure data "persistency", which is how the Grid maintains data availability at all times, even as the Grid environment evolves. A first version of the LCG, called LCG-1, is scheduled to start running on a restricted number of sites in July 2003.

The European Data Grid Project (EDG)

The European DataGrid is a three year project funded by the European Union that aims to enable access to geographically distributed computing power and storage facilities belonging to different institutions. This will provide the necessary resources to process huge amounts of data coming from scientific experiments in three different disciplines: High Energy Physics, Biology and Earth Observation.

Testbed success

EDG builds on a software toolkit for Grid technology known as Globus, developed in the US, as well as other software packages, and uses these to build a functioning Grid testbed. The project brings together over 100 computer engineers, who have generated some 250 K lines of code already. In 2002, EDG middleware managed to connect computing resources at some 20 major centres. In collaboration with the LHC experiments CMS and ATLAS, a number of demanding computational challenges were successfully processed. This proved that many components of the EDG software are ready for use in the LCG project.



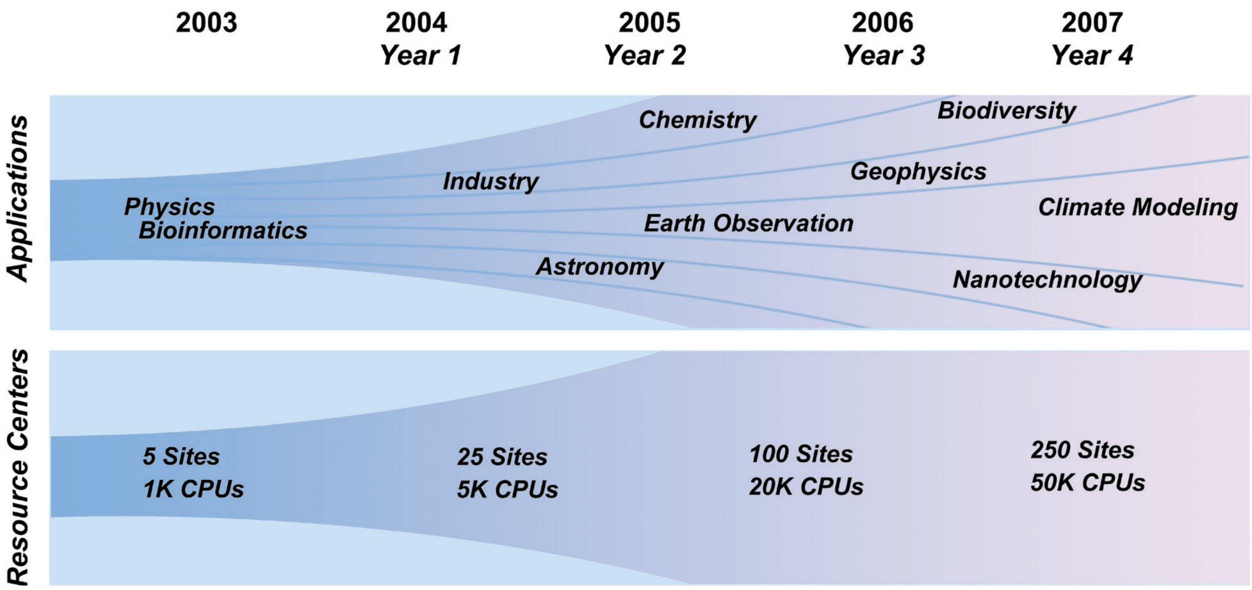
Enabling Grids for E-Science in Europe (EGEE)

The success of EDG has generated strong support for a follow-up effort to build a permanent European Grid infrastructure that can serve a broad spectrum of applications reliably and continuously. Providing such a service will require a much larger effort than setting up the current testbed. So CERN has established a pan-European consortium, EGEE (Enabling Grids for E-science in Europe) to build a production Grid infrastructure, providing round the clock Grid service to scientists throughout Europe. A proposal for such a project has been submitted to the EU 6th Framework Programme in 2003.

The big picture

As illustrated below, the potential benefits of such an infrastructure for Europe will extend well beyond the LHC, to nearly all areas of science and commerce where large amounts of data must be managed in a distributed way. However, the benefits of Grid technology are not just regional but global. Reflecting this, CERN has been playing an active role in the Global Grid Forum, which helps to set worldwide standards in this rapidly evolving field.

Schema of the evolution of the European Grid infrastructure from two pilot applications in high energy physics and biomedical Grids, to an infrastructure serving multiple scientific and technological communities, with enormous computer resources. The applications and resource figures are purely illustrative. The EGEE project covers Year 1 and 2 of a planned four year programme.



Other EU-funded Grid projects at CERN

The year 2002 also saw the launch of several other EU-funded Grid projects in which CERN plays a significant role. CERN is leading DataTAG, which aims to provide high-speed connections between Grids in Europe and the US.

In 2003, DataTAG set a significant land record, transferring a Terabyte of data from the Stanford Linear Accelerator on the West Coast of the United States to CERN in just over an hour, a factor of 2.5x faster than previous land records.

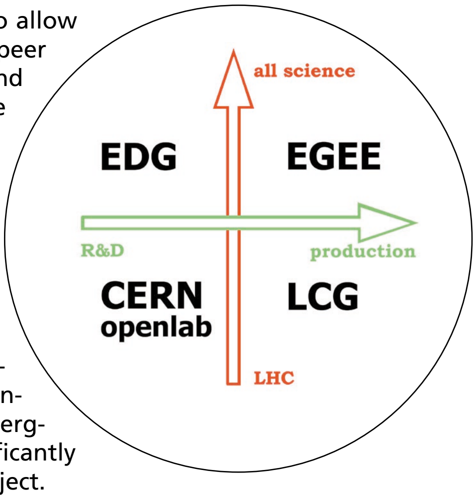
Project portfolio

CERN is also collaborating in CrossGrid, a project that aims to extend EDG functionality to advanced applications such as real-time simulations. The GRACE project develops a higher-layer of software for Grids, including concepts such as semantic Grids, which provide contextual meaning to information stored on the Grid. CERN is also a partner in MammoGrid, a project dedicated to building a Grid for hospitals to share and analyse mammograms in an effort to improve breast cancer treatment. Finally, GridStart aims to coordinate the efforts of the major Grid initiatives in Europe, including EDG, and disseminate information about the benefits of Grid technology to industry and society.

THE STATUS

CERN openlab: off to a fine start

The CERN openlab was conceived to allow CERN and its industrial partners to peer into the technological crystal ball and test technologies that may well be commercially competitive when the LHC is up and running. The industrial partners view this as a great opportunity to develop and test new technologies, which are still far from the commodity computing market, under the rigorous and demanding conditions that CERN's advanced computing environment provides. CERN gets the benefit of advanced knowledge of emerging technologies, which could significantly affect the roadmap for the LCG project.



Schematic illustration of the relationship between CERN openlab and other major Grid projects led by CERN.

A theme of the CERN openlab, as the name suggests, is to pursue open standards and open source solutions. This is of vital importance to LCG, where interoperability between different hardware and software platforms is a necessity for the Grid to work efficiently. And open standards are increasingly viewed as a necessity by many industry leaders, in order to be able to deliver Web services, computing-on-demand, and other Grid-related solutions to customers. Thus, for the industrial partners, discovering how their different technologies perform together, and adapting open source software to these technologies, are added benefits of the partnership.

Cluster creation

The first project launched by CERN openlab is called CERN opencluster, and aims to build a cluster based on technologies that are at the cutting edge of what is available on the market today. In particular, the CERN opencluster will be linked to the EDG testbed, to see how these new technologies perform in a Grid environment. The results are also be closely monitored by the LCG project, to determine how the potential impact of the technologies involved.

Marc Collignon and Jean-Michel Jouanigot with the Enterasys Networks equipment.

The CERN openlab provides the LHC with a new and vital source of industrial sponsorship for long-term technology development and integration. Membership of the CERN openlab is set at 2.4 MCHF over a three year period. The equipment for the CERN opencluster, as well as the contribution of top engineers to develop it, is provided by the industrial partners as part of CERN openlab membership requirements. The concept has proved very popular, with other major computer and software manufacturers eager to join.

Beyond using CERN as a testbed for cutting edge Grid software and hardware, and building an industry consortium for Grid-related technologies of common interest, a further goal of CERN openlab is to provide a training ground for a new generation of engineers to learn about Grid technology. Already, HP is co-funding two CERN fellows (post-doctoral level researchers) a trend that will hopefully be adopted by other partners.

One node in the CERN opencluster.

Partner profiles

In September 2002, HP joined Intel and Enterasys Networks in the CERN openlab for DataGrid applications. Together with CERN, these companies launched the ambitious project called CERN opencluster, which combines 64-bit processor technology and 10Gb/s network cards from Intel, compute servers in a cluster from HP, and a 10 gigabit switching environment from Enterasys Networks. In April 2003, IBM joined the CERN openlab, and brought 28TB of high-end storage to the project, as well as its Storage Tank data management technology.

CERN Rationale

The choice of the industrial partners of CERN openlab reflects CERN's objective to work with industrial leaders in all aspects of the Grid computing challenge. 64-bit technology is on the technological horizon, so testing Intel® Itanium® 2 processors was a clear must. Intel and HP are already actively collaborating on Itanium 2 processor-based cluster solutions, so CERN is benefiting from the fruits of the pre-existing alliance. CERN is also part of the Gelato Federation – an open source activity for 64-bit technology promoted by HP. Enterasys Networks has traditionally competed at the high-end of the switching and routing technology, bringing to the partnership the cutting edge in these technologies. IBM's commitment to computing-on-demand made it a natural partner for a Grid project, and Storage Tank is viewed as a data management system that can rise to the challenge of handling the petabytes of data CERN will produce.

Enterasys Networks

is a leading worldwide provider of broadline intelligent data networking infrastructures for enterprise-class customers. Enterasys' networking hardware and software offerings deliver the innovative security, availability and mobility solutions required by Global 2000 organizations coupled with the industry's strongest service and support. For more information on Enterasys and its products, visit www.enterasys.com.

"The aggregate data throughput for LHC will exceed one terabit per second. Enterasys is confident that its 10-Gigabit Ethernet Technology will enable CERN to unlock the full potential of its DataGrid." John Roese, Chief Technology Officer for Enterasys Networks.

HP

is a leading global provider of products, technologies, solutions and services to consumers and businesses. The company's offerings span IT infrastructure, personal computing and access devices, global services and imaging and printing. HP completed its merger transaction involving Compaq Computer Corp. on May 3, 2002. More information about HP is available at www.hp.com.

"Through this collaboration with the CERN DataGrid, HP's researchers and engineers will be put to the test to truly push the envelope in developing advanced Grid computing technologies." Jim Duley, Director for Technology Programs, HP University Relations.

IBM

is the world's largest information technology company, with 80 years of leadership in helping businesses innovate. Drawing on resources from across IBM and key Business Partners, IBM offers a wide range of services, solutions and technologies that enable customers, large and small, to take full advantage of the new era of e-business. Additional information about IBM is available at www.ibm.com/us/

"This is the perfect environment for us to enhance our Storage Tank technology to meet the demanding requirements of large scale Grid computing systems." Jai Menon, IBM Fellow and co-director of IBM's Storage Systems Institute

Intel

the world's largest chip maker, is also a leading manufacturer of computer, networking and communications products. Additional information about Intel is available at www.intel.com/pressroom.

"CERN's DataGrid project is an ideal application for Intel's most powerful processor yet, the Intel® Itanium® 2 processor. The awesome computer power required will find a formidable engine in the Itanium." Steve Chase, Director, Business and Communication Solution Group of Intel.

Organisation and human resources

CERN openlab now represents a significant investment in manpower for the IT Division, with seven technical staff working on the CERN opencluster project at CERN, and significant technical contributions from technical liaisons in the partner companies.

In September 2002, an openlab Management Unit was created, and as of April 2003, two HP-funded CERN fellows (postdoctoral level researchers) joined this team. In addition, there is active participation from CERN liaisons to the main Grid projects, LCG and EDG, as well as CERN liaisons to the individual partner companies.

CERN openlab Board of Sponsors

Luciano Maiani [head of Board]	CERN Director General
Wolfgang von Rüden	Head of the CERN openlab
John Roese	Enterasys Networks
Jim Duley	HP
Tom Hawk	IBM
Tom Gibbs	Intel

CERN openlab Management Unit

Wolfgang von Rüden	Head of the CERN openlab
François Fluckiger	Associate Head
François Grey	Development and Communication
Sverre Jarp	Chief Technologist

CERN openlab Technical Team

Sverre Jarp	Chief Technologist
Andreas Hirstius	Opencluster
Andreas Unterkircher	Opencluster
Jean-Michel Jouanigot	Networking
Marc Collignon	Networking
Rainer Többicke	Storage
Jean-Damien Durand	Storage

Industry Partner Technical Liaison with CERN

Markus Nispel	Enterasys Networks
John Manley	HP
Brian Carpenter & Jai Menon	IBM
Herbert Cornelius	Intel

Industry Partner Organizational Liaison with CERN

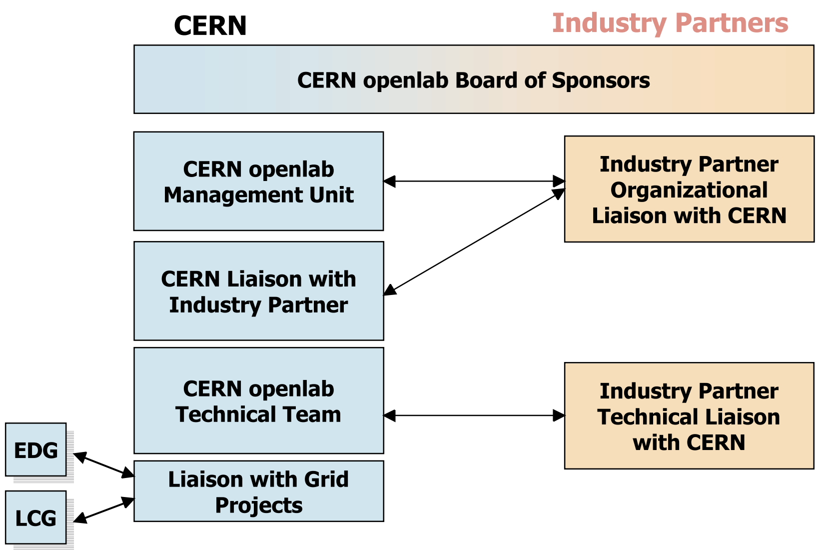
Christoph Kälin	Enterasys Networks
Michel Benard	HP
Pasquale Di Cesare	IBM
Pierre Mirjolet	Intel

CERN Technical Liaison with Industry Partners

Sverre Jarp	HP and Intel
Jean-Michel Jouanigot	Enterasys Networks
Rainer Többicke	IBM

CERN Grid Project Liaison

David Foster	LCG
Fabrizio Gagliardi	EDG



The CERN openlab facilities

The CERN opencluster area is a central area of the Computer Centre that has been reserved for the CERN opencluster equipment, which benefits from optimum visibility from the visitors gallery.

The openlab openspace is a new VIP meeting room which has been built with access directly on the Computer Centre. The aim is that CERN and industrial partners should be able to use it for key technical and promotional meetings. It provides professional, high-tech surroundings for meetings, with information about CERN openlab and the industrial partners in evidence.

Offices for the CERN openlab team have been established in building 513, just along the corridor from the openlab openspace. The technical team and most of the openlab management unit is installed there already.

Meeting in the openlab openspace.

CERN opencluster:

pushing the limits

The major focus of the technical work in CERN openlab over the past year has been to get the CERN opencluster up and running, an objective which has been successfully achieved.

The opencluster consists of thirty-two Linux-based Hewlett Packard rx2600 rack-mounted servers equipped each with two 1 GHz Intel® Itanium® 2 processors. In addition, seven development systems, plus two which have been assigned to the experiment Alice, were installed in 2002. Three desktop workstations have been delivered, and are used by the opencluster technical team for their development work.

The latest equipment

A number of 10-GbE Network Interface Cards (NICs) have been delivered by Intel and installed on the HP rx2600 computers. These allow high-speed transfer tests to be carried out, which are described in more detail below.

Enterasys Networks has delivered a switch equipped to operate at 10 Gbps and with additional port capacity for 1 Gbps, provisionally on loan. This comes in addition to two switches delivered in 2002. In summary, there are 3 switches, 20 10-GbE ports and 76 1-GbE ports.

Interface to CERN opencluster for remote management. Nodes that are unavailable are colour-coded.

Andreas Hirstius with the CERN opencluster.

After IBM joined the CERN openlab in April 2003, IBM and CERN established procedures to support the teamwork between the two partners, and a detailed schedule for the next 6 months work. IBM is sponsoring 28TB of high-end storage, which was delivered and installed during the second quarter of 2003.

The opencluster nodes were installed using the fully automatic kick-start procedure which is also in use for CERN's production-oriented PCs. In addition to the Linux operating system, the nodes were also equipped with openAFS, the LSF batch system, and three sets of compilers (Intel, GNU, and the Open Research Compiler)

Thanks to the preinstalled management card in each node, automation was further developed to allow remote system restart and remote power control. This development confirmed the notion that - for a modest hardware investment - large clusters can be controlled remotely. This result is highly relevant to the LHC Computing Grid, which will need to deploy such automation at a large scale.

Benchmarking activities

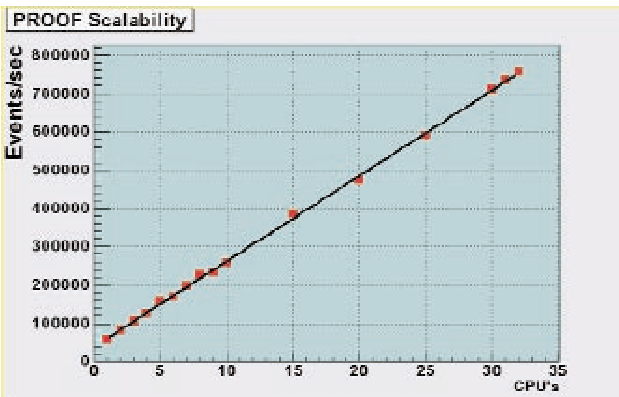
Major physics packages (CLHEP, GEANT4, ROOT) have been successfully ported and tested on the CERN opencluster. This work was done together with the authors and maintainers of the various packages in collaboration with the newly formed software group in the EP Division of CERN. The CERN data management package, CASTOR, was also ported and certified. Benchmarking of the physics packages has begun and first results are promising.

For example, PROOF (Parallel ROOT Facility) is a version of the popular CERN-developed software ROOT, an object-oriented data analysis framework. It is being developed for interactive analysis of very large ROOT data files on a cluster of computers. The scalability of PROOF has been tested using the CERN opencluster, and proves to be linear up to this cluster size. The ROOT team at CERN is very interested to pursue tests on even larger cluster sizes (see figure below).

Rainer Többicke with the IBM storage system.

Getting better all the time

A compiler review was performed together with Intel, detailing both short term and medium term development opportunities on the Intel® Itanium® 2 processor. This close collaboration between the CERN openlab technical team and Intel staff should lead to significant advancements of compiler performance – illustrating nicely the sort of technical benefits that industrial partners in the CERN openlab enjoy.



Results for scalability of PROOF, the Parallel Root Facility. Each node of the cluster has one copy of the data set (4 files, total of 277 MB), 32 nodes. In total, this amounts to 8.8 Gbyte in 128 files, equivalent to 9 million events. On one node, the total processing time was 325 s, on 32 nodes in parallel it was 12 s.

CERN openlab data challenges

The CERN openlab technical team has defined the concept of CERN openlab data challenges – tangible objectives that may be jointly targeted by some or all of the sponsors.

A major openlab challenge is the 10 Gigabit challenge: a common effort by Enterasys Networks, HP, Intel and CERN to achieve data transfer rates within the opencluster of 10Gbps. This involves multiple technological issues such as network switch tuning, Linux Kernel and TCP/IP parameter tuning.

A dedicated meeting took place at CERN in March 2003 with technical representatives from all three industrial sponsors, to chart out a path for this challenge. A joint 10Gb Operational Group was set up as a result of the meeting at CERN. Minuted operational meetings of the group are organized on a bi-weekly basis via phone-conferencing. Initial results are described below.

First results of the 10Gbps challenge

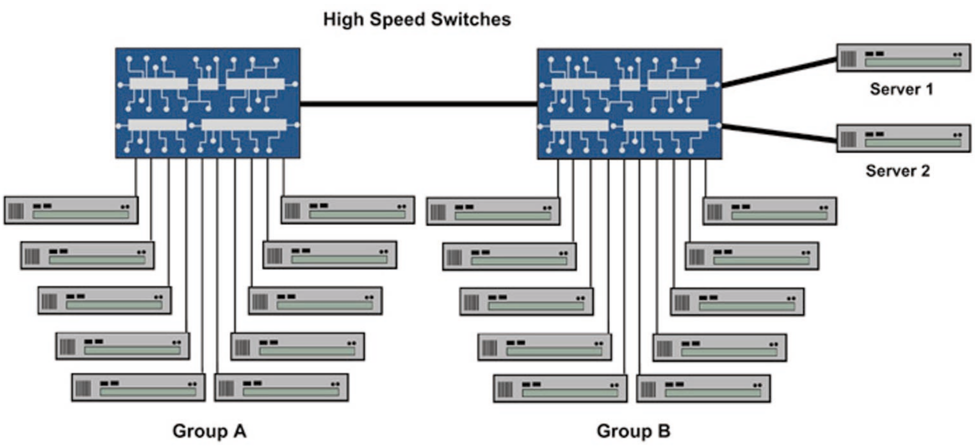
One of the major challenges of the opencluster project is to take maximum advantage of the partners' 10 Gbps technology. In April 2003, a first series of tests was conducted between two Linux-based Hewlett Packard rx2600 computers equipped each with two 1 GHz Intel® Itanium® 2 processors. The two computers were directly connected ("back-to-back" connection) through 10 GbE Network Interface Cards (NICs) from Intel (see recent news). The transfer reached a data rate of 755 MBps, that is, approximately 3/4 of the maximum attainable bit rate of the interfaces.

The transfer took place over a 10 km fibre. It used very big frames (16 KB) in a single stream, as well as the regular suite of Linux Kernel protocols (TCP/IP). When using "jumbo frames" (9 KB), the same test reached 693 MBps.

The best results were obtained when aggregating the 1 Gbit bi-directional traffic involving the 10 nodes in each group. The peak traffic between the switches was then measured to be 8.2 Gbps.

The next stages of the test campaign will include continued tuning of the parameters, and evaluating the next version of processors.

Schema of the setup was used for testing the Enterasys switches. Thick lines indicate 10 Gbps connections, thin ones are 1 Gbps.



The GBps storage-to-tape challenge

In May, CERN announced the successful completion of a major data challenge aimed at pushing the limits of data storage to tape. While this was not a CERN openlab data challenge per se, this data challenge involved in a critical way several components of the CERN opencluster.

Using 45 newly installed StorageTek 9940B tape drives, capable of writing to tape at 30 MBps, Bernd Panzer and his team at the IT Division of CERN were able to achieve storage-to-tape rates of 1.1 GBps for periods of several hours, with peaks of 1.2 GBps – roughly equivalent to storing a whole movie on DVD every four seconds. The average sustained over a three day period was of 920 MBps. Previous best results by other research labs were typically less than 850 MBps.

LHC compatible

The significance of this result, and the purpose of the data challenge, was to show that the CERN's IT Division is on track to cope with the enormous data rates expected from experiments on the LHC. These experiments will produce data at rates in excess of 100 MBps, and one experiment alone, called Alice, is expected to produce data at rates of 1.25 GBps.

In order to simulate the LHC data acquisition procedure, Dr. Panzer's team generated an equivalent stream of artificial data, using 40 compute servers. This data was stored temporarily to 60 disk servers, which included the CERN opencluster servers, before being transferred to the StorageTek tape servers as illustrated in the figure, right. A data compression factor of 1.3 was deliberately chosen during this data challenge, as this is characteristic of the compression that can be achieved with real experimental data.

A key contributing factor to the success of the data challenge was a high performance switched network from Enterasys Networks with 10 GBps Ethernet capability, which routed the data from PC to disk and from disk to tape. As indicated in the figure, once the LHC is running, data will come directly from the experiments, and be stored on disks on the DataGrid, as well as archived onto tape at CERN.

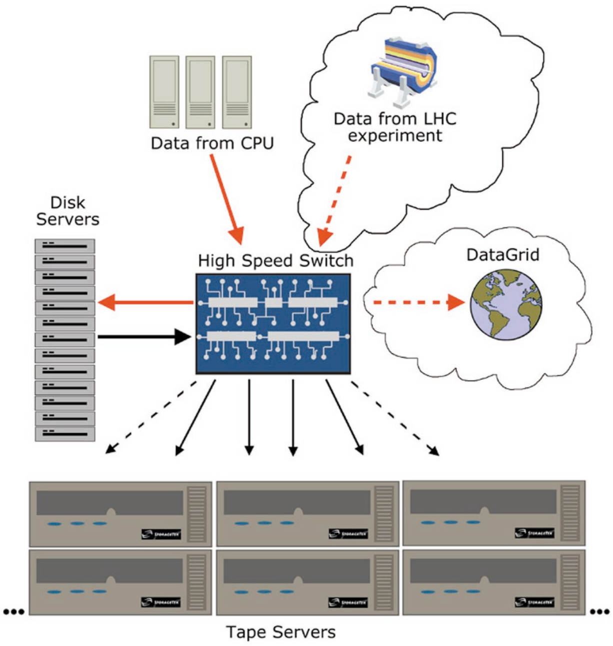


Illustration of the setup used for the 1GBps storage to tape challenge. Simulated data from CPUs is stored temporarily on disk servers, then transferred to the 45 tape servers through the Enterasys switched network. Once the LHC is running, a similar configuration will be used, with data coming from the experiments, and being copied to the Grid as well as stored on tape (bubbles indicated).

CERN openlab dissemination: spreading the news

While many of the benefits for industrial partners of CERN openlab are technical, there is also a strong emphasis in CERN openlab's mission on the public relations opportunities that this novel partnership provides. Therefore, the CERN openlab Management Unit has developed over the last year a spectrum of activities that promote CERN openlab in different contexts, and provide added benefits for the partners.

The openlab Technical Team is ensuring dissemination of CERN openlab results through traditional technical channels such as conference presentations, technical publications and thematic workshops. In addition, the cutting-edge nature of the CERN openlab data challenges lends itself well to Press releases and other non-technical publications, as well as to presentations to visiting VIPs at CERN.

Publications

Two publications in the Proceedings of Computing for High Energy and Nuclear Physics (CHEP '03, La Jolla, California) (to be published, see web information below)

Conference contributions

Two talks at CHEP '03, La Jolla, California featured highlights from CERN opencluster. These presentations can be viewed at:
http://chep03.ucsd.edu/files/139.ppt
http://chep03.ucsd.edu/files/262.ppt

Press releases and non-technical articles

Joint Press Release with HP (25/9/2002)
Joint Press Release with IBM (2/4/2003)
Joint Press release with StorageTek (28/5/2003)
CERN bulletin 39/2002 p5
CERN bulletin 12/2003 p6
CERN courier (to be published Oct. 2003)
CERN annual report 2002

Thematic workshops

As part of the technical program, Thematic Workshops are proposed to sponsors and open to CERN staff. The standard format of Thematic Workshops is as follows: on day one, CERN experts present specific CERN requirements and the status of their technical development; on day two, CERN meets sponsors on a one-to-one basis.

The first Thematic Workshop took place on 17-18 March 2003, dedicated to Data and Storage Management. The second Workshop is scheduled the 8 and 9 July 2003, on Fabric Management. Future Thematic Workshops on Total Cost of Ownership and on Energy Consumption in Data Centers are being planned.

Website

The website structure has been expanded and is now formed of a top level public website (openlab style, stable content) and a new public working site (IT division style, frequently updated)

Public events

CERN openlab has hosted two First Tuesday events at CERN – details listed below – in collaboration with the organization "First Tuesday Suisse Romande". These public events attracted some 250 persons each, and involved speakers from CERN and industrial partners, as well as other guest speakers. The public is primarily from the regional business and investor communities. The events were also webcast live and are archived at www.rezonance.ch. A First Tuesday on Grid technology is scheduled for Q3 2003.

- 1st FT @ CERN: Grid Technologies: scientific and industrial prospects (27 September 2002).

- 2nd FT @ CERN: New trends in industrial partnership and innovation management at European research laboratories (19 March 2003).

UIP Presentations

- European Commissioner for Enterprise and Information Society, Erkki Liikanen (February 21st 2003)

- Swiss State Secretary for Economic Affairs, David Syz (March 3rd 2003)

- European Parliament Technology Assessment delegation (March 18 2003)

- Norwegian State Secretary Bjorn Haugstad, Ministry of Education and Research (30 April 2003)

- National Technological University (USA), Management of Technology Programme (May 14th 2003)

- Finnish Parliamentary Committee (May 27 2003)

Student activities

Based on a pilot programme run last year in collaboration with the Helsinki Institute of Physics, a CERN openlab student programme will run in summer 2003, involving 11 students from seven European countries. The idea is to have the students work in international teams, on projects related to the applications of Grid technology. In 2003, one student team will work on issues related to the CERN opencluster, another on Grid solutions for data acquisition from the Athena experiment, and finally, one team will work on the Grid café website, described below. The objective is to get the students to continue their work in their home institute, and to attract talented students back to CERN to do M.Sc. and Ph.D. projects.

CERN openlab is co-sponsor of an application by CERN for an EU grant from the Marie-Curie programme. If successful, this application will lead to funding for 1-2 Ph.D. students associated with CERN openlab.

Outreach activities

CERN openlab is actively supporting the establishment of a Grid Café for the CERN Microcosm exhibition. The idea is to create a web café with a focus on Grid technologies, including a dedicated website that will give links to key Grid technology websites, and access to a demonstration testbed developed by EDG. This activity is also planned to be part of CERN's presence on conference stands such as one currently being planned for Telecom'03 in Geneva in October, and in the framework of the RSIS (Role of Science in the Information Society) conference to take place at CERN the 8th and 9th December 2003 as a side event to the WSIS (World Summit on the Information Society).

CERN openlab is also actively involved in the planning of the CERN Globe of Innovation, a stunning architectural structure that the Swiss government has given to CERN, and that will be used as meeting and networking centre around the theme of science, technology and innovation.

THE FUTURE

The technological horizon

The plans for the next two years are to double the number of nodes on the CERN opencluster each year, reaching a cluster size of 128. The high-speed switching environment will be correspondingly upgraded, to ensure that the 10Gbps challenge can be attained with this size of cluster.

In parallel with this hardware expansion, software activities will focus in particular on the gridification of the opencluster, and on establishing a meaningful set of benchmarks to evaluate the performance of such an advanced cluster on a computing Grid.

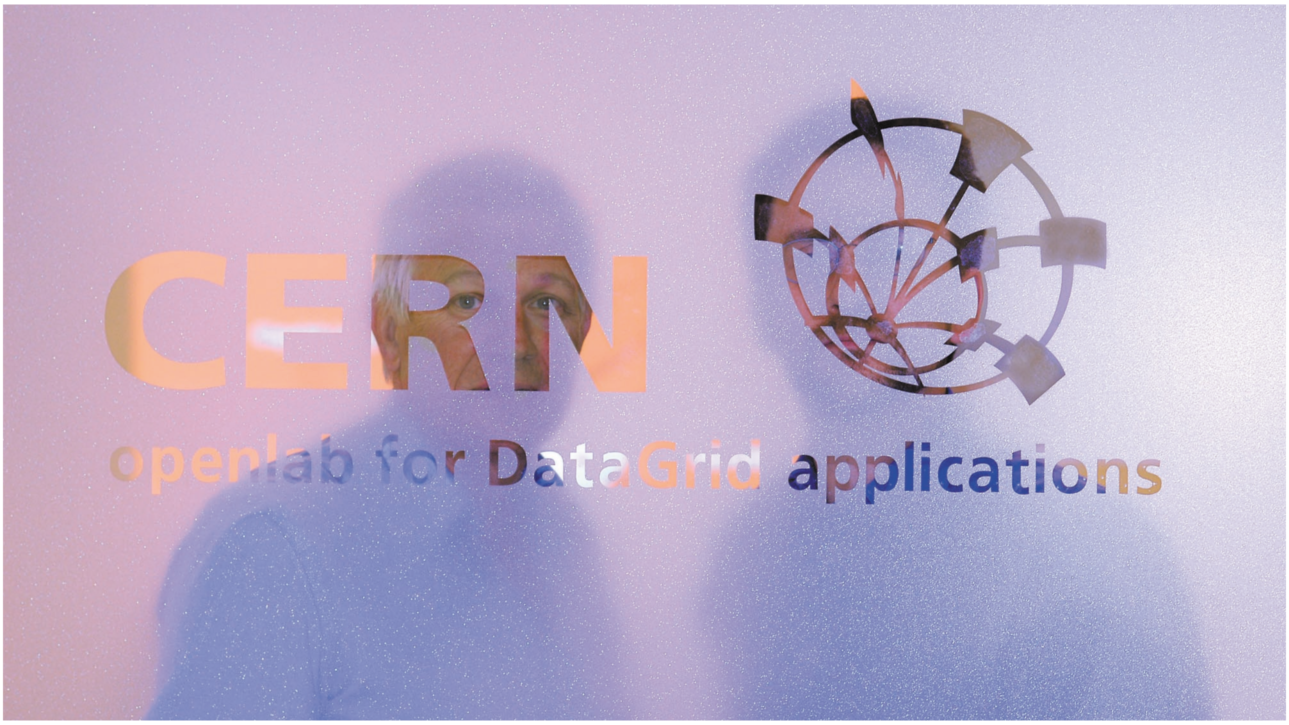
On the storage side, IBM is contributing Storage Tank®, a new technology that is designed to provide scalable, high-performance and highly available management of very large amounts of data using a single file namespace regardless of where or on what operating system the data reside. (Recently, IBM announced that the commercial version will be named IBM TotalStorage™ SAN File System.) IBM and CERN will work together to extend Storage Tank's capabilities in combination with the CERN opencluster, so it can manage the LHC data and provide access to it from any location worldwide.

New project opportunities

Discussions are ongoing in CERN openlab to evaluate other possible areas of technological collaboration, with current or future partners. These could be in terms of added functionality and performance of the existing opencluster, using complementary technologies.

Equally promising are collaborative efforts in the area of open standards and open source software, as well as specific Grid-related issues – such as Grid security and mobility – that have implications for long-term commercial applications of Grids. Detailed technical proposals for developing projects in such areas will be prepared in the coming year.

CERN openlab has potential as a vehicle for industrial participation in the EGEE project and similar European Grid projects. CERN openlab will endeavour to keep industrial partners informed about the progress of these projects, and investigate ways to enhance collaboration with them.



This is the second annual report of the CERN openlab for DataGrid applications. It was presented to the Board of Sponsors at the Annual Sponsors meeting, June 13th 2003.

Participants in the Annual Sponsors Meeting, from left to right: Sverre Jarp (CERN openlab), François Fluckiger (CERN openlab), Pasquale Di Cesare (IBM), Wolfgang von Rüden (CERN openlab), Tom Hawk (IBM), John Manley (HP), Tom Woodget (Intel), Luciano Maiani (CERN), Michel Benard (HP), Brian Carpenter (IBM), Pierre Mirjolet (Intel), Lou Witkin (HP), Jai Menon (IBM), Jim Duley (HP), Les Robertson (CERN LCG), Gérard Vivier (Enterasys Networks), François Grey (CERN openlab), Markus Nispel (Enterasys Networks), Christoph Kälin (Enterasys Networks).

FOR MORE INFORMATION

Executive Contact

Wolfgang von Rüden,
Head of the CERN openlab for DataGrid applications,
CERN Information Technology Division,
Wolfgang.von.Rueden@cern.ch
Tel: +41 22 767 6436

Public Relations

François Grey,
CERN openlab developer,
CERN Information Technology Division,
Francois.Grey@cern.ch
Tel: +41 22 767 1483

Technical Contact

François Fluckiger,
Associate Head of the CERN openlab for DataGrid applications,
CERN Information Technology Division,
Francois.Fluckiger@cern.ch
Tel: +41 22 767 4984

http://www.cern.ch/openlab

CERN

openlab for DataGrid applications

sponsored by

hp invent

IBM

ENTERASYS NETWORKS™

intel®