INRIA at the heart of Grid Computing research

- september 2003 -

INRIA
Media contact
Tel. : + 33 1 39 63 57 29
Email : Vincent.Coronini@inria.fr

SUMMARY

Towards Grid Computing

Will computing power, digital data and storage ever become accessible in a decentralized way, just like electricity is accessible in the home in developed countries? Many people believe it will be possible to set up large networks in the not too distant future that will put together geographically distant computer resources. A lot of researchers are currently working on this endeavor. Such computing or data "grids" will make it possible to perform computations and data processing on an unprecedented scale. These grids are one of the key topics of the major "Supercomputing 2002" conference-exhibit to be held in Baltimore, (Maryland) U.S.A., from November 16 to November 22, 2002.

Grids, collective infrastructures for very large scale computing

Tasks that require huge computations and process colossal quantities of data are now numerous and diverse. Such is the case of meteorological or climate prediction, computing the aerodynamic behavior of a new model of aircraft, deciphering the genome of a living organism or detecting the elementary particles produced by an accelerator, to name but a few. These tasks are also becoming increasingly ambitious, and thus more and more demanding in terms of computing power, data flow and memory capacity. How can computer infrastructure meet these continuously growing needs?

Globalization and dematerialization of computer resources

The performance of the hardware and software available in each computing center or to each individual user is rising very sharply. This trend is not however sufficient to meet the many challenges that face science, technology and industry. The computing power in hardware doubles every 18 months or so, on the average, whereas storage capacity doubles every 12 months and the performance of network connections doubles every 9 months. Thus, the performance of computers improves less rapidly than that of networks. Therefore, a potentially revolutionary concept has been developing for six to seven years. The idea is to link geographically distant equipment together, especially via the Internet, to constitute a network that combines the computing power, storage capabilities and so forth of all its members. Each of these members will thus be able to use the sum of available resources in terms of computing power, memory, software and data, put in by all the other members of the network. This is the basic idea behind computing grids. It means that computer resources are simultaneously globalized and dematerialized.

Why the term grid? Because in a way, the concept consists in distributing computer resources the same way as electricity is distributed to homes. Electricity is supplied to consumers without them having to think about where and by whom it was produced. So in relation to power grids, researchers started talking about computing grids.

Overcoming technical and sociological obstacles

The idea of connecting and sharing distributed computer resources was already there in the 60s. However, it is only recently that technological advances have made these prospects relatively concrete. Computer grids come up against many difficulties. First of all there are technical difficulties since the point is to make distant devices that differ in terms of functioning and performance, communicate and work together. It is also to write software that efficiently manages and distributes the network cumulated resources, and to devise programming tools that are adapted to the diffuse and parallel character of the tasks entrusted to the grid, among others. There are also sociological, and even economic and political difficulties, since setting up a grid assumes that separate entities—public institutions, private enterprises, and individuals—can be convinced to put their own resources at the disposal of a collective entity. This aspect has technical repercussions. For example, each one of the nodes of the grid may possess data or software that are deemed to be confidential and not to be communicated to anyone else. Hence the need to guarantee the security of exchanges within the grid through appropriate techniques.

Still embryonic grids

Actual computing grids that make their different nodes cooperate transparently, easily and on an equal basis, currently only exist as prototypes and are still far from being used in production. Quite a few experimental initiatives and research have already been launched. In the United States, the Globus system is linking American supercomputers together to obtain a parallel, virtual hypercomputer that should make it possible for each center to submit jobs using the combined power of all other centers. Grid infrastructures are in the process of being built by several scientific communities. Such is the case for particle physics in Europe in the framework of the European Data Grid project, that must prepare to store and analyze the many petabytes of data (1 petabyte = 10^{15} bytes, the equivalent of some 100 billion book pages) that will be produced every year by the Cern LHC starting in 2007.

In addition, networks devoted to so-called global or peer-to-peer computing on popular problems have appeared. One of these networks is the SETI@home initiative, to look for possible clues for extraterrestrial civilizations among the signals received by the Aricebo radiotelescope in Puerto Rico. The project is currently distributing the signal analysis work to over half a million volunteers who are willing to let their personal computer work for SETI@home when it is otherwise idle. Other similar examples where rather simple but extremely large computations are distributed over a large number of volunteer personal computers, like the Great Internet Mersenne Prime Search, looking for very large prime numbers, or the Decrypthon in France that gathered approximately 75,000 Internet users for a few months, up to May 2002, to compare the sequences of the 500,000 proteins known in living organisms.

Such distributed networks do not constitute grids strictly speaking. In effect, the computations are dispatched by a central authority and participating computers merely carry them out without having the possibility of using the network for their own needs. They nonetheless illustrate the large computing power and the benefits that can be expected from grids. Performance of several dozen teraflops (1 teraflop = 10^{12} floating point operations per second) were obtained in these initiatives, something that was not imaginable just ten years ago.

Research in France and INRIA

Designing and standardizing grid infrastructure requires a lot of effort on the part of the world computing community. In France, an important part of computing grid research is done at INRIA. At least six of the Institute's teams (projects APACHE, OASIS, PARIS, ReMap, RESO and SARDES) are deeply involved, in collaboration with various academic or industry partners. In addition to this, many other INRIA teams are doing work that is more or less closely related to the grid topic, either in the framework of Institute projects, in that of the incitative concerted initiative GRID (Globalization of computer resources and data) launched in 2001 by the Ministry of Research, in that of the National Network of Research in Telecommunications (RNRT, set up in 1997) or in that of the National Network for Research and Innovation in Software Technology (RNTL, set up in 1999).

INRIA at the heart of grid research

Ensuring efficient communication between software components

Using computing grids, more and more scientists and engineers will be able to perform complex numerical simulations involving specific codes for several distinct phenomena that are coupled together. Think about simulating the behavior of a satellite in space, which must simultaneously take into account kinematic, thermal, mechanical and optical aspects. Such simulations will have to be based on parallel technology in order to meet performance requirements and on distributed technology to meet the needs in computer resources. This assumes that both parallel and distributed aspects are sufficiently compatible.

The PARIS team (Programming parallel and distributed systems for large scale numerical simulation) is working in this direction and is striving to develop a high performance software component model adapted to grids. The PARIS team is starting from the CORBA component model (*Common Object Request Broker Architecture*, a software system that is now standard to link applications implemented on heterogeneous platforms) and extending it to develop a software component model that integrates parallel codes. This new system is called GridCCM (Grid CORBA Component Model). It must in particular guarantee efficient communication between parallel software components, which was not the case of previously existing models. PARIS thus designed a communication management "framework" called PadicoTM. This framework can be used to make parallel distributed services work without conflicts without having to modify the applications. PadicoTM has been on display at Supercomputing 2002. It is intended for code coupling applications based on the parallel COBRA objects concept. Even though COBRA is generally regarded as rather slow, PadicoTM allows communication at up to 240 MB/s and latency times of 20μs. This level of performance is comparable to that of MPI (Message-Passing Interface, a message exchange programming executable in parallel programming). Still in the context of code coupling, PARIS is also studying the problem of data globalization. The problem is to store the data needed or produced by the simulation in a distributed fashion over the grid.

Guaranteeing rapid access to data, coupling computing codes

In a grid, one of the bottlenecks can be the speed of access to data files, whose size will certainly be huge. Part of the work of the APACHE team (Parallel algorithmics, programming and load sharing) is on this aspect. APACHE has been interested in parallel architecture programming for several years. The team is now developing parallelization—thus acceleration—methods for data access, through PC clusters. It is also designing methods to couple together computing codes executed on different nodes of the grid in such a way that the various tasks carried out by these codes follow on optimally. APACHE has developed its programming techniques using a 200 i-Vectra PCs cluster supplied by Hewlett-Packard. This cluster ranked 385th in the TOP500 of the most powerful machines worldwide in June 2001. On these subjects, Apache has tight collaborations with companies such as Bull, HP, Mandrake, CS, etc.
The team is validating its techniques on concrete applications such as modeling the crossing of cellular membranes by proteins, in collaboration with chemists, biologists and mathematicians.

Optimally distributing tasks and data for computation

One of the problems tackled by computer scientists is the design of an appropriate algorithmics for computing on a grid. The various tasks in a given computation must be ordered and the data must be placed in such a way that the computation executes optimally in terms of the hardware configuration and the current state of the grid. This question is part of the ReMaP team concerns (Regularity and Massive Parallelism). The team is looking for heuristic methods for task scheduling and data placement, that are validated via simulation. The simulation is done using the SIMGRID simulator designed and developed by the University of California at San Diego, with the participation of ReMaP. Project ReMaP is also developing software layers to let a client of the grid export certain parts of a computation to be done to other grid nodes or servers. Such software layers must for example choose the most appropriate server at any given instant, i.e., the less loaded, the fastest, or the most adapted server to the task in question. They must also choose the data that must be supplied to the servers. ReMaP relies on CORBA to develop their toolbox called DIET (Distributed Interactive Engineering Toolbox). DIET will be demonstrated at Supercomputing 2002 on an Ethernet network. The toolbox has been developed with support from the RNTL and is being tested on various applications (digital terrain models, simulation of electronic circuits) in the framework of project ASP of the GRID concerted initiative.

A Java program library for distributed parallel computing

Part of the activities of project OASIS (Active Objects, Semantics, Internet and Security) has similar objectives and is concerned with programming tools for distributed applications, either on a local Intranet network, on a workstation cluster or on Internet grids. The team is developing in particular a program library entirely written in Java for distributed parallel computing in the framework of the ObjectWeb consortium founded by France Télécom R&D, Bull and INRIA. This library, called ProActive, can be used to perform mobile computations, that is to say computations initiated on one machine of the grid and continued on another machine. It also includes security tools such as data exchange encryption and user authentification. ProActive has many more attractive features: dynamic and transparent code loading, online documentation, ease of installation and use, visualization and graphic control of program execution. ProActive will be demonstrated at Supercomputing 2002. OASIS is validating its ProActive development on electromagnetism computations (aircraft radar image computations). The ProActive library is available on the Internet under LGPL license. It has already been downloaded by numerous academic and industry users.
The OASIS team very recently succeeded in executing an application to solve 3D Maxwell's equations in electromagnetism on a 64 processor cluster, showing an practically optimal acceleration. The application was developed in collaboration with project CAIMAN and entirely written in ProActive Java. Executing the same application in P2P Intranet on desktop machines and standard INRIA production network made it possible to solve a 150 to the cube mesh, that is to say over 100 million facets, on 252 processors.

Adapting communication protocols to heterogeneous infrastructures and very high speed

Network connections, a central element of the future grids, are capable today of delivering considerable throughputs. The American network TeraGrid that links the National Science Foundation computing centers reaches 40 gigabits per second. The European network GEANT that connects the main European capitals reaches 10 gigabits per second. It is however crucial that such capabilities are used to their maximum. This entails the design of communication protocols and performance measurement and prediction tools that are adapted to very high speeds and heterogeneous infrastructure. This is one of the main tasks of the RESO team (Protocols and Software Optimized for Heterogeneous High Speed Networks). The IP and TCP protocols used on the Internet are already old and are neither adapted to very high speed nor to the grid concept. RESO is studying the possible evolution of these protocols. One of the developments proposed by the team consists in introducing a differentiation of service, that is to say in modulating the degree of priority of information packets, and optimized transport protocols. The idea is to allow the very heavy flows attached to a grid to take advantage of the periods when regular Internet traffic is low, in order to fully exploit the capabilities offered.

RESO is carrying out this work in collaboration with national projects (the VTHD network of the RNRT, the e-Toile platform of the RNTL) and the international projects DataGrid and DataTAG. DataGrid is a European project that aims at designing a platform and software for a data grid at the service of particle physics, Earth monitoring and biology. It relies on European networks such as GEANT or national networks like RENATER. The DataTAG project (Data TransAtlantic Grid) objective is to interconnect European and American grids via very high speed links.

I-Cluster 2: a shared experimental platform

INRIA teams carry out their research activities using experimental platforms. I-Cluster 2, currently being installed in the INRIA Rhône-Alpes research unit, provides the institute with its most powerful supercomputer yet.

Its architecture is based on Itanium 2 dual processors communicating through a Myrinet network. A total of 104 dual-processors at 900MHz, 312 Go RAM, are arranged as 10 racks of

10 nodes and 1 rack of 4 nodes with additional disk storage. I-cluster 2 is connected to the VTHD network and is running Linux OS (RedHat Advanced Server). First Linpack experiments at INRIA (Aug. 2003) have reached a 560 GFlop/s performance.

I-cluster 2 is part of a scientific program financed by the French ministry of Research and Education, the Rhône-Alpes region, INRIA, Ecole Normale Supérieure de Lyon, the Institut National Polytechnique of Grenoble and Joseph Fourier University.

Other INRIA projects

Other INRIA teams are carrying out research that concern data or computing grids one way or another. Thus, the ARES team (Architecture for Service Networks) that is working on problems related to service deployment on radio network infrastructures, is developing a platform called DARTS (Deployment and Administration of Resources, Processing and Services). In the framework of grids, this platform can be used for administering and instrumenting the computing resources available on different points of the grid. It also supplies services to simplify the interaction between the various components administered (asynchronous messaging, naming, application dynamic loading). The platform also offers facilities for application administration (deployment, porting).

Researchers in project SARDES (Constructing Software Infrastructures for Large Scale, Heterogeneous, Distributed Systems) are studying the architecture and design of distributed software infrastructures for global information processing environments, by systematically using reflection and component building techniques (a reflexive system can be defined as a system offering an explicit, operable and causally connected representation of itself).

Project CARAVEL (Information Mediation Systems) is concerned with the problem of integrating information in networks that contain heterogeneous, autonomous information sources, as is the case for grids. The question is to offer a uniform mode of access to a set of information sources through an integrated view, to facilitate the construction and maintenance of coherent data warehouse and to offer modes of navigation in an information network that is adapted to different categories of users.

Project ScAlApplix (High Performance Schemes and Algorithms for Complex Scientific Applications) brings together several scientific skills for a multidisciplinary study of high performance computing and its applications to complex scientific computations—chemical reactions simulations, unsteady fluid flows simulations, host-parasite systems simulations and so on—that require massive computing power of the order of the teraflops and very large volumes of data of the order of a terabyte. In addition to modeling and simulation techniques and high performance algorithms, project ScAlApplix is also working on the visualization and steering of distributed numerical simulations using a virtual reality code.

The GRID concerted initiative

In February 2001, the French Ministry of Research launched a scientific program called GRID (Globalization of Computer Resources and Data). This ACI (Incitative Concerted Initiative) is funded by the Science National Fund. Its objective is to reinforce French research in computing and data grids, so that it is able to rank well in international competition. Through this initiative, already active research teams are consolidated, new groups can be made aware of the field of grids, and computer scientists and users of these new technologies can meet in multidisciplinary teams. GRID relies on the high speed network platforms supplied by RENATER and the VTHD and on national or regional centers to develop several research directions that are necessary to build experimental grids (software tools, systems and environments for distributed computing or the exploitation of very large volumes of data, modeling, algorithmics, code coupling, visualization, and so on). Another objective of the GRID ACI is to foster company foundation in the field of ASP (Application Service Providers, i.e. companies that sell high added value computer services for specific applications such as financial analysis, simulation, virtual reality or data mining).

The GRID ACI grants financial support to research team consortiums following a call for proposals. The amounts distributed were 2.5 million euros in 2001 and 3 million euros in

2002, essentially in the form of operating funds and fixed term engineer contracts. GRID thus launched 18 projects in 2001 and 12 projects in 2002. As of today, there are therefore thirty such projects. INRIA teams participate in fourteen of them:

- ARGE (Eastern Networks Group) gathering research teams in networks and distributed systems from the eastern part of France.

- GRID^2 (Meeting, Information and Discussion Group on the Globalization of Computer Resources and Data) training young researchers, information diffusion, meeting organization for researchers involved in work on grids in France.

- CARAML (Coordination and Distribution of Multiprocessor Applications in Objective CAML) development of libraries for high performance globalized computing around the CAML language of INRIA in its Objective Caml or Ocalm dialect, examples of applications in molecular simulation.

- RMI (High Performance Distributed Objects for Computing Grids) development of a programming model for computing grids that combines parallel computing and distributed computing models.

- ASP (Client-Server Approach for Simulation on the Grid) validation of a NES (Network Enabled Servers) type architecture on a national grid for a set of applications in chemistry, physics, electronics and geology.

- GeoGrid, a proposal for innovative solutions for visualization and complex numerical computation problems, such as those encountered in oil exploration and exploitation.

- Guirlande-fr (Computer management and uses of language resources for the diffusion and study of the French language) setting up a distributed server grid to federate resources on the French language (coded and annotated texts) and tools to process these resources.

- CiGrid (CIMENT GRID) experimenting intensive computing on a grid made of the six platforms of project CIMENT (Intensive Computing, Modeling, Numerical and Technological Experimentation), each of the platforms being attached to a specific field (astrophysics and Earth science, environment and climate, cluster computing, chemistry, biology, physics).

- EPSN (Environment to Steer Distributed Numerical Simulations) design and development of a software platform to steer and conduct a distributed numerical application for visualization, validation on applications in molecular chemistry, environment and acoustics.

- HydroGrid, modeling and simulation of fluid transfer and solution transport in underground geological media.

- MecaGrid, building a regional computing grid based on PC clusters located in the Provence-Alpes-Côtes d'Azur region, for applications of massively parallel computing in heterogeneous fluid mechanics.

- PADOUE (Data Sharing for Use in Environment) data and processing program interoperability in the field of environmental research, document and archiving methods to help in locating and identifying shared resources.

- DataGRAAL (DataGrid for animation and large scale applications) organizing meetings between the database and systems communities, links between the different French projects in the field of data grids, drawing up software projects on data management in large scale through preliminary mock-ups.

- GénoGRID: Setting up a computing grid for researchers in molecular biology, with applications in protein folding, genome and genetic sequence comparison. The grid comes with a simple access portal on the Web that is transparent and secure.

The Gelato consortium

INRIA joined the Gelato Federation in February 2003**.** The Gelato Federation (http://www.gelato.org/) aims at providing the Itanium (IA-64) community with the most current and useful open source resources for running Linux. INRIA is part of the Parallel File Systems, Cluster administration tools, runtime for parallel programming and distributed scheduling focus areas. INRIA contributes to experiment on a cluster built with104 Intel® Itanium® 2 processors at 900 MHz interconnected using a Myrinet network. The INRIA projects involved are APACHE, OASIS, PARIS, ReMaP, RESO, and SARDES.

A few Web sites

- Supercomputing 2002 conference-exhibit : http://www.sc2002.org
- Workshop GRID 2002 of November 18, 2002: http://www.gridcomputing.org/grid2002
- GRID incitative concerted initiative: http://www-sop.inria.fr/aci/grid/public/acigrid.html
- An ERCIM issue on grids:
 http://www.ercim.org/publication/Ercim_News/enw45/
- Project PARIS: http://www.irisa.fr/paris/
- Project APACHE: http://www.inria.fr/recherche/equipes/apache.fr.html
- Project ReMaP: http://www.inria.fr/recherche/equipes/remap.fr.html
- Project OASIS: http://www.inria.fr/oasis/ProActive/
- Project RESO: http://www.inria.fr/recherche/equipes/reso.fr.html
- DARTS platform: http://darts.insa-lyon.fr/
- Project SARDES: http://sardes.inrialpes.fr/
- Project CARAVEL: http://www-caravel.inria.fr/
- Project ScAlApplix: http://www.inria.fr/recherche/equipes/scalapplix.fr.html
- DataGrid: http://eu-datagrid.web.cern.ch/eu-datagrid/
- DataTAG: http://datatag.web.cern.ch/datatag
- e-Toile: https://www.urec.cnrs.fr/etoile/
- *Globus*: http://www.globus.org
- *European Data Grid*: http://grid.web.cern.ch/grid
- *SETI@home*: http://setiathome.ssl.berkeley.edu
- ERCIM News : http://www.ercim.org/publication/Ercim_News/enw45/

Contacts

Scientific coordinators

Michel Cosnard
Email : Michel.Cosnard@inria.fr

Thierry Priol
Email : Thierry.Priol@inria.fr

Writer

Maurice Mashaal
Email : mashaal@club-internet.fr

Press

Christine Genest, Vincent Coronini
Media contact - INRIA
Tel : + 33 1 39 63 57 29
Email : Vincent.Coronini@inria.fr

Yasmina Madafi
Ketchum
Tel : + 33 1 41 34 01 31
Email : ymadafi@ketchum-pr.fr