DOT-COMing the GRID: Using Grids for Business

Wolfgang Gentzsch

Sun Microsystems Inc, Palo Alto, California, USA

Abstract: In this presentation, a short outline of the history of past and present Grid projects in research and industry is given, followed by some near- and long-term Grid scenarios and visions on how data and compute Grids will complement current Internet services and thus change our working and living environments and habits. In essence, implementation and professional exploitation of the complex and highly sophisticated Grid technologies will still take a couple of years and give us time enough to adapt to the dramatic changes and potential opportunities Grids will create in the future.

1. Grids in Research

In the early Nineties, research groups all over the world started exploiting distributed computing resources over the Internet: scientists collected and utilized hundreds of workstations for highly parallel applications like molecular design and computer graphics rendering. Other research teams glued large supercomputers together into a virtual metacomputer, distributing subsets of a meta-application (e.g. the computer simulation of multi-physics applications) to specific vector, parallel and graphics computers, over wide-area networks.

The scope of many of these research projects was to understand and demonstrate the actual potential of the networking, computing and software infrastructure and to develop it further. This led us to Internet infrastructure projects like Globus and Legion, which enable users to combine nearly any set of distributed resources into one integrated metacomputing workbench to allow users to measure nature (e.g. with microscope or telescope), process the data according to some fundamental mathematical equations (e.g. the Navier-Stokes equations), and provide computer simulations and animations to study and understand these complex phenomena.

These projects created a new era in distributed computing, according to the book The Grid: Blueprint for a New Computing Infrastructure. Generally speaking, a computational Grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to computational capabilities. These Grids, in the near future, will be used by computational engineers and scientists, associations, corporations, environment, training and education, states, consumers, etc. They will be dedicated to on-demand computing, high-throughput computing, data-intensive computing, collaborative computing, and supercomputing, potentially on an economic basis. Grid communities, among others, are national Grids (like ASCI), virtual grids (e.g. for research teams), private grids (e.g. a BMW CrashNet for the car manufacturer BMW and its suppliers, for collaborative crash simulations), and public grids (e.g. consumer networks).

Today, we see the first attempts to more systematically exploit Grid computing

resources over the Internet. Distributed computing projects like SETI@home, Distributed.Net, and Folderol, let Internet users download scientific data, run it on their own computers using spare processing cycles, and send the results back to a central database. Recently, Compute Power Market project has been initiated to develop software technologies that enable creating Grids where anyone can sell idle CPU cycles, or those in need can buy compute power much like electricity today.

2. Grids in Industry

Encouraged by the response of thousands of participants on these research initiatives, new Internet startup companies like Popular Power, Entropia, Distributed Science, and United Devices are trying to turn this idea into real business that resell untapped resources for a profit, hoping that computer users will be interested to donating their extra computing power to projects that crunch a lot of data, such as the search for a new cancer drug or patterns in the human genome. Other potential candidate applications are complex financial analysis and generation of intensive graphics.

While this kind of global (and 'wild') Internet computing will probably be successful in the future where privacy and security are only minor issues (i.e., mostly in research-oriented projects), global industries might have some real concerns in using this Internet computing technology for their strategic businesses. Beside security of information and data, these companies need guarantees for the availability and utilization of dedicated resources, high-level quality of services, easy, fast and authenticated computing portal access to hardware and software, and tools for accounting, reporting, monitoring, and planning.

Just recently, industry started to experiment with more commercially oriented ebusiness models for high-performance and data-intensive computing via the Internet. For example, debis Systemhaus, a DaimlerChrysler company in Germany, offers its NEC SX-5 supercomputer power through an Internet e-commerce gateway using a public web server, a secure web server and a discussion server. The web pages are based on JAVA applets, CGI scripting and JAVA servlets. In addition an LDAP customer database is used for the management of security and encryption certificates. A user can register using HTML forms; the secure Web site requests certificates to identify user; a hummingbird UNIX desktop from the browser redirects application to customer desktop; and Pegasus, (the application dependent job submission GUI), submits the job to the batch system.

3. Grid Resource Management

Most of the underlying sophisticated technologies are currently under development. Large research communities like the GridForum and EGrid are coordinating all kinds of Grid research, prototype Grid environments exist like public-domain Globus and Legion, research in resource management is underway in projects like EcoGrid, and the basic building block for a commercial Grid resource managers exists with Sun's Grid Engine software. Grid Engine is a new generation distributed resource management software which dynamically matches users' hardware and software requirements to the available heterogeneous resources in the network according to predefined policies usually prescribed by the management in the enterprise. The Grid Engine acts much like our body's central nervous system (sometimes called The Body's Internet'). The Grid Engine Master ('the brain') with its sensors in every computer (comparable to the sensations of touch, sound, smell, taste, and sight) dynamically acts and reacts, according to set policies (comparable to move, eat, drink, sleep) to allow for full control and achieve optimum utilization and efficiency. Grid Engine has been developed as an enhancement of Codine from former Gridware Inc, according to well defined requirements from the Army Research Lab in Aberdeen, and BMW in Munich, where today Grid Engine manages over 800 powerful compute servers in each of these local Grids. Average usage increased from well under 50% to over 90%, in both environments.

4. Future Grid Economies

The next step is to enhance Grid Engine, which currently is restricted to manage local computer resources, towards 'The GRID Broker', which will be able to match the user's compute jobs with the available resources in the network, including invoicing users for the CPU power they consume, very much like todays electric power consumption, telephone usage or water supply. The Grid broker will match the user's requirements to the best fitting Application Service Provider (ASP) in the universe which optimally fulfills the user's hardware, software and service needs.

This GRID Broker belongs to the enabling technologies of the next Internet Age. The Internet, for a long time, has been used only for information. Only recently, enabled by several important improvements in hardware infrastructure, security, authentication, and ease of access, it is used for electronic commerce. And just now, the next revolutionary step complementing the Internet can be foreseen: The Grid Computing Infrastructure, i.e. all kinds of dedicated GRIDs used for collaboration and collaborative computing in industry and research, for application simulation and animation, for real-time video, on-demand virtual reality presentations, and other services for consumers and producers.

This high-quality and economically oriented usage of the Internet will be enabled by several new technologies and achievements made recently. For example, CORBA offers a standard interface definition to interconnect any distributed object in the world. JAVA provides a common platform for distributed resources and thus guarantees full cross-platform portability, and JINI allows to interconnect any electronic device in a scalable way. And the chaos which potentially can arise with this wealth of interconnected devices, clusters, subgrids, and grids, will be removed and brought into a well-organized and well-functioning 'organism', by the GRID Resource Broker, supported by intelligent agents which, through the network or wireless, report to the Central Grid Engine the details on available resources, and the consumers' habits and needs for specific resources in the GRID.

Then, eventually, in a next (and final?) step, the central Grid Engine will disappear, partly as an integrated component of the local operating systems, and partly being replaced by intelligent mobile agents, which enable a universal and self-healing environment with potentially infinite compute power available on-demand, and as easily accessible as our today's electricity, telephony, roads and water infrastructures. For URLs to projects referenced the paper see: gridcomputing.com.