Zeitschrift: Comtec : Informations- und Telekommunikationstechnologie = information and telecommunication technology

Herausgeber: Swisscom

Band: 81 (2003)

Heft: 12

Artikel: Grid Computing : between hype and reality

Autor: Mazza, Giuseppe

DOI: https://doi.org/10.5169/seals-876711

Grid Computing – Between Hype and Reality

Grid Computing is the large-scale virtualisation of computing and storage capacity, letting users access the immense resources distributed all over the Internet. Standards which enable interworking of different hard and software components are fundamental steps towards the implementation of a global Grid Computing network. First finalised Grid standards are expected early next year. However, the technical implementation of Grid Computing networks has to deal with some constraints compared to the original Grid vision. Nevertheless, some of the major IT companies are strongly pushing the development of Grid Computing networks by supporting standardisation and providing Grid-enabling software and hardware solutions.

The programme "Software and Security Technologies" explores new opportunities for Swisscom arising from current software technology trends and assesses the impact these trends may have on the efficiency of the service creation process, as well as on the quality of the services. In particular, the programme focuses on IT-related areas such as Web Services, trends in software architectures, Ontology Languages and Knowledge Representation technologies, and advanced software security issues.
With its Innovation Programmes, Swisscom Innovations follows the objective of assessing the impact of technological developments, finding new business opportunities, promoting technical synergies, and developing concrete innovation proposals. Further, the expertise built up enables active engineering support of business innovation projects.

The idea of Grid Computing dates back to the first half of the 1990's. Due to the enormous computational resources needed for scientific calculations, scientists at CERN developed the idea of a network which bundles the

GIUSEPPE MAZZA

distributed computational resources and fits them to the user's needs.
The vision behind Grid Computing is often explained using the electric power grid metaphor (fig. 1). The electric power grid delivers electric power in a pervasive and stand-ardised way. When using electricity one does not worry where it is produced and how it is delivered. The same principle applied in the computer world means: Millions of computing and storage systems all over the planet provide transparent access to computing power and storage volume in a uniform way.

Based on the deployed computing resources we can discern two basic Grid types: Computing Grids and Data (Storage) Grids.

Grid Computing will provide access to the immense unused computing resources distributed all over the Internet. While standardisation efforts are making good progress, the economics of Grid Computing are still a complex issue and strongly dependent on the development of the relationship of the costs for bandwidth, storage and computing cycles.

A Computing Grid is essentially a collection of distributed computing resources (computing cycles) provided by the processors of the machines comprising the Grid. The processors can vary in speed, architecture, software platform and are aggregated to act as a unified processing resource or virtual supercomputer. However, in today's implementations this aggregation feature is subject to restrictions.
A Data (or Storage) Grid provides wide area, secure access to data. It enables users and applications to manage and efficiently use database information from distributed locations. Data Grids eliminate the need to unnecessarily move, replicate, or centralise data, translating into cost savings. Initial Data Grids are being constructed today, primarily serving collaborative research communities.
The goal of the study presented here was to provide an overview of the current situation in Grid Computing. For this purpose, a research covering standardisation, technical implementation and business aspects was performed. Based on the optained results, conclusions on the future development and impact of Grid Computing in the IT world were drawn.

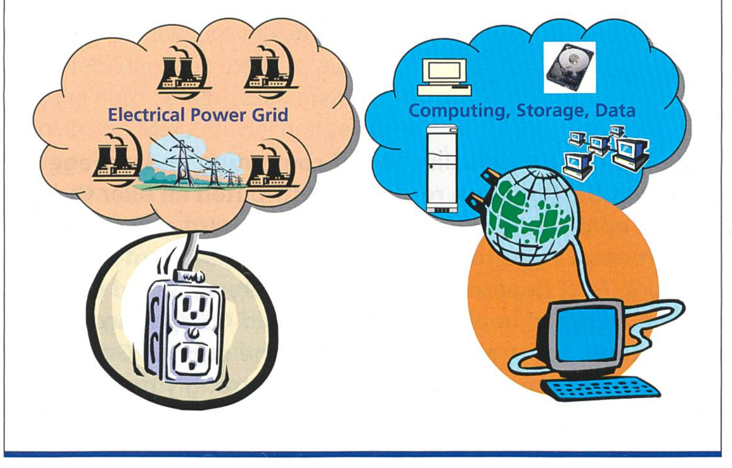


Fig. 1. The Grid vision as a power grid metaphor.

How Grid Computing Works

Today's existing Grid implementations are proprietary, since no finalised Grid standards exist yet. This leads to differing Grid implementations (e.g. EU Datagrid, US Teragrid etc.). However, taking the existing draft standards, a typical standard Grid workflow can be described as follows (fig. 2):

1. The user submits a request via a web service interface to a Resource Broker, specifying high level requirements, for example the kind of application he wants ("job category"), the operating system, resources needed, etc.
2. The Resource Broker finds and allocates suitable resources. He sends back a "Grid Handle Resolver" in which the address of the available resources and XML schemes of the job attributes are specified.
3. The User Application contacts the Distributed Resource Management System. After a security procedure, the job data and job attributes are sent to the resource.
4. The resource processes the job. The user can remotely monitor the process and possibly abort it.
5. The Grid notifies the user when the job has been finalised and sends the data back to a predefined location.

This is a simplified procedure which does not take into account the special cases where data replications have to be considered in Data Grids.

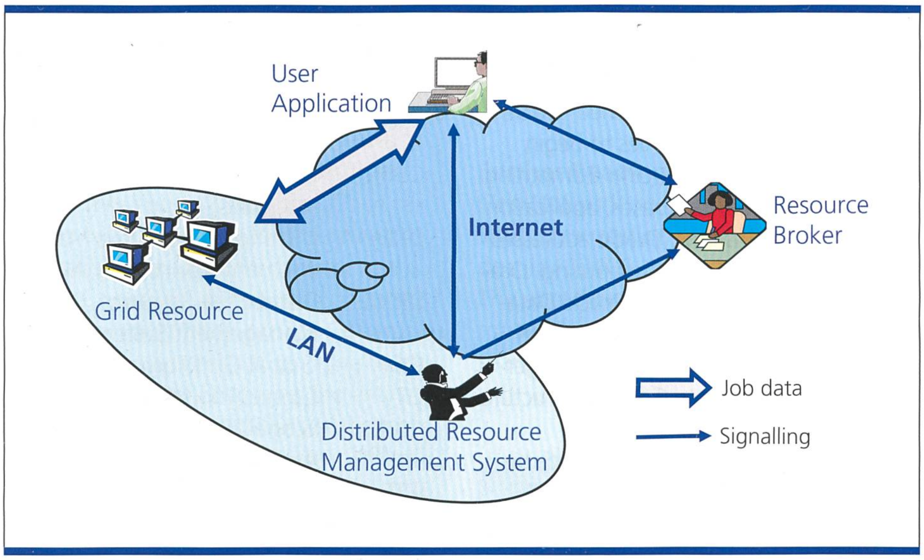


Fig. 2. Simplified Grid job procedure in a global Grid.

Grid Standardisation

The Global Grid Forum (GGF) [1] is an institution whose primary objective is to promote and support the development, deployment and implementation of Grid technologies and Grid applications. These goals are pursued through the creation and documentation of Grid standards, user experiences and implementation guidelines. The GGF recommendation process is based on the IETF procedure. Currently, several working groups are defining a Grid architecture containing all functionalities needed for Grid operation and interoperability. It is expected to have first finalised Grid standards by the beginning of 2004. Figure 3 gives an overview of the main GGF working groups and their goals.

The Open Grid Service Architecture (OGSA) [2] is a working group which aims towards a Grid system architecture based on Web Service concepts and technologies. The successful realisation of the OGSA vision of a broadly applicable and adopted framework for distributed system integration requires the early definition of a core set of interfaces, behaviours, resource models and bindings, called the OGSA Platform. This includes the specification of important services required to support Grid applications and high level functionalities required to run these services.

Building on Web Service technologies, the Open Grid Services Infrastructure (OGSI) [3] defines mechanisms for creating, managing and exchanging information among entities called Grid Services. A Grid Service is a Web Service that conforms to a set of conventions that define how a client interacts with a Grid Service. These conventions, and other OGSI mechanisms associated with Grid Service creation and discovery, provide for controlled, fault resilient, and secure management in distributed applications.

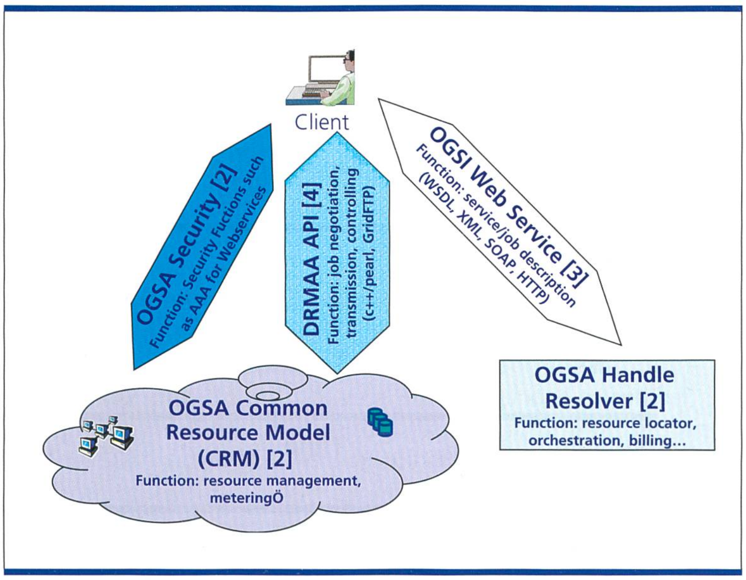
The Distributed Resource Management Application API (DRMAA) [4] is a GGF working group whose goal is to define a generalised API to distributed resource management systems (DRMS) in order to facilitate the integration of Grid applications. DRMAA provides a set of interfaces that are OS neutral, allowing to do job submission, monitoring, controlling and retrieval of finished jobs.

To be sure that a job fits to the Grid resources, the concept of "job categories" has been introduced. Job categories are the approach recommended by DRMAA for the encapsulation of site-specific details and completely hiding these details from applications using the DRMAA interface. In other words, the resource broker will attribute only resources with a matching job category to a specific job request. Matching criteria could be for example OS type, resource type, application type etc. Thus, the vision of a powergrid-like, seamless resource attribution is not completely true for the moment.

Grid Implementations

There exist different Grid topologies which have evolved over time. Figure 4 shows the most important ones in chronological order. Depending on the geographical extension, Grids may employ public Internet for access and reach of distributed resources. The first and simplest stage of a Grid computing network is called Cluster Grids. They consist of a cluster of servers usually located in the same network domain. A further step in evolution is represented by Intra-Grids, enterprise- wide inter-departmental Grids, usually interconnecting clusters between different network domains. Extra-Grids on the other hand are connections of Intra-Grids between geographically distributed sites (e.g. different universities) within or between enterprise organisations. Finally, the last stage of evolution is represented by Inter-Grids, a global collection of enterprise and cluster grids as well as other geographically distributed resources, all of which have agreed upon global usage policies and protocols to enable resource sharing. Inter-Grids are the most complex topology

Fig. 3. Overview of the main Grid draft standards and their relationship.



from the infrastructural and management point of view. This is the reason why Inter-Grids have not yet been realised.

Historically, the Grid idea was born in the research area, driven by the need for increased computing power and storage volume. Up to now most Grid implementations have been used for research purposes. However, as it was for the Internet in the early 1990's, Grids will likely expand to commercial applications. Several IT providers are making strong efforts to push Grid development in this direction. Many companies are offering software and hardware solutions for the implementation of enterprise-wide Grids for industrial utilisation. Here are some of the most important examples:

Among the IT companies, IBM is the most prominent believer in Grid Computing. IBM is involved in both the technology and the business issues that have led to the Grid Computing evolution. The IBMeServer™ product line offers a platform for designing and developing Grid solutions and even for managing entire Grids. In time, IBM will grid-enable many of its systems and much of its software [5].

HP plans on including Grid-enabled software such as the Globus tool-kit [7] into its servers and storage devices, but also consumer products such as handhelds, PCs and printers. HP has joined with Oracle to implement a standardised Grid platform version that will build and link enterprise Grids together [6].

SUN is one of the first companies offering Grid software solutions for enterprise-wide Intra-Grid implementations. The Sun Grid EngineTM [9] software aggregates available computing resources and delivers computing power as a network service. It is designed to harness idle computer resources, match them to individual job requirements and deliver network-wide computing power to the desktop. Big companies such as Sony, Ford and Motorola have already implemented the SUN Grid solution.

In September 2003, Oracle came out with its new product extension named Oracle 10g [10]. It allows companies to begin evolving their IT toward a Grid Computing model. For storage, databases, application servers and applications, Oracle addresses the requirements of Grid Computing. Oracle claims that together, the Oracle Database 10g, the Oracle Application Server 10g, and the Oracle Enterprise Manager 10g provide a complete Grid infrastructure software.

Microsoft is most active in Data Grids because it is congruent to Microsoft's .NET agenda. For this reason several collaboration projects with research and government institutions are ongoing. Microsoft has participated in the Global Grid Forum and is sponsoring a port of the Grid-toolkit (based on Globus) to Windows. The GXA vision and products [11] correspond to the needs of the Open Grid Services Architecture (OGSA) middleware. Microsoft is actively building and marketing GXA.

An important initiative is the Globus Project [7], a research and development project focused on enabling the application of Grid concepts. This project targets technical challenges that arise from scientific and engineering computing activities. Typical research areas include resource management, data management and access, application development environments, information services, and security. The Globus Project software development has resulted in the Globus Toolkit (GT), a set of services and software libraries which groups around the world are using to build Grids and to develop Grid applications. The toolkit includes software for security, information infrastructure, resource management, data management, communication, fault detection, and portability. With the GT3 release, the Globus Project now offers an open source implementation of the OGSI draft specifications. Future releases of the Globus Toolkit will continue to track state-of-the-art OGSA concepts.

Abbreviations

API	Application Programming Interface
CERN	Centre Européen Recherche Nucléaire
DRMAA	Distributed Resource Management Application API
DRMS	Distributed Resource Management System
GGF	Global Grid Forum
GT	Globus Toolkit
GXA	Global XML Webservices Architecture (Microsoft)
IETF	Internet Engineering Task Force
OGSA	Open Grid Services Architecture
OGSI	Open Grid Services Infrastructure
SDK	Software Development Kit
WSDL	Web Service Definition Language

Fig. 4. Grid topologies evolution.

Key Factors for a Successful Takeoff

Three key factors have been identified for the successful commercialisation of Grid Computing: meeting customer needs, no connectivity bottlenecks, and revenue generating business cases:

The customer needs which can be satisfied with Grid Computing are the need for computing power and for (reliable) storage volume. The question is: could these needs be satisfied in an easier and cheaper way than with Grids, for example with super computers and local high capacity storage drives? It seems that it only makes sense for applications requiring petaflops (10^{15} FLOPS, Floating Point Operations Per Second) of computing capacity and petabytes of storage volume, to be processed by a Grid.

Grid Computing networks are composed of a high number of resources. Interconnection of these resources is crucial for the functioning of the Grid. In computing-intensive applications, or in data storage applications, a considerable amount of data has to be moved between the resources. For this reason enough transmission bandwidth has to be available in order not to create a connectivity bottleneck. The latest Grid implementations (e.g. TeraGrid [8]) focus much attention

on this point. For the first time a 10GbE (10 Gigabit Ethernet) backbone network has been implemented to connect the various Grid sites.
A potential, solid business case should indicate how to generate revenues from Grid Computing. It is clear that Grid customers will have to pay for the resources they utilise in the Grid. In contrast to the Internet, no revenues can be generated from advertising, because Grid Computing applications are based on machine-to-machine interactions. Therefore, the entire Grid revenues have to be generated from customer utilisation fees. There is still no clear Grid business case yet.

Conclusions

In the past few years, Grid Computing has become a hype topic driven by the vision of a "magic" black box delivering computing power and storage volume independently of the technical hardware and software diversities. Today, the definition of technical standards shows that these diversities have to be bridged with interfaces which cannot guarantee a real seamless cross-platform Grid operation. It is not yet clear whether or not commercial mass market introduction of Grid Computing will be successful.

Outlook

Finalised Grid standards are expected by the beginning of 2004. This should enable large scale deployment of Grid Computing networks based on multi-vendor solutions. The future of Grid Computing lies in commercial applications. Although it is rather early to assess the size of the Grid Computing market, several research firms have made forecasts of the expected market size. One of the more optimistic views comes from Grid Technology Partners, an analyst firm specialising in Grid Computing. By 2005, Grid Technology Partners expect the overall Grid Computing market to grow as high as $4.1 billion. Probably, this number includes hardware and software as well as commercial Grid applications.
It is still an open question whether the success factors mentioned in the previous chapter will be fulfilled. If so, Grid technology will likely form the foundation of a fourth wave in IT. 10

Acknowledgement

The research work and report on Grid Computing was been commissioned by Swisscom IT Services.

Further Reading on Grid Computing

Grid Networking, Light Reading article, www.lightreading.com/document.asp?doc_id=33405
The anatomy of the Grid, Foster, Kesselmann, Tuecke, www.globus.org/research/papers/anatomy.pdf

Giuseppe Mazza *holds an engineering degree in electrical engineering from the Swiss Federal Institute of Technology in Zurich (ETHZ). In 1998, he joined Swisscom Innovations where he worked on the development of optical networks and transport platforms such as ASTN, ASON and GMPLS. Today, he works in the broadband technology and broadband services domain.*

References

[1] Global Grid Forum Homepage, www.Gridforum.org/L_About/about.htm
[2] Open Grid Services Architecture (OGSA) Platform, draft recommendation, www.gridforum.org/Meetings/ggf7/drafts/draft-ggf-ogsa-platform-2.pdf
[3] Open Grid Services Infrastructure (OGSI), draft recommendation, www.gridforum.org/ogsi-wg/drafts/draft-ggf-ogsi-gridservice-29_2003-04-05.pdf
[4] Distributed Resource Management Application API Specification, draft recommendation, forge.gridforum.org/projects/ggf-editor/document/draft-rec-drmaa-1/en/1/draft-rec-drmaa-1.pdf
[5] IBM Grid computing homepage, www.ibm.com/grid
[6] Siliconvalley.com, article of 11th September 2003, siliconvalley.internet.com/news/article.php/3076181
[7] The Globus Project homepage, www.globus.org
[8] TeraGrid, the world's largest distributed infrastructure for open scientific research, www.teraGrid.org/
[9] SUN ONE grid engine software, wwws.sun.com/software/gridware/
[10] Oracle 10g Grid enabled solution, www.oracle.com/appserver/index.html?content.html
[11] Microsoft Global XML Web Services Architecture,msdn.microsoft.com/library/default.asp?url=/library/en-us/dngxa/html/gloxmlws500.asp

Zusammenfassung

Grid Computing

Dieser Artikel befasst sich mit Grid Computing, einer neuartigen Computernetzarchitektur, die es erlaubt ungenutzte Rechenleistung und ungenutzten Speicherplatz externen Benutzern zugänglich zu machen. Ein wichtiger Schritt auf dem Weg zur Implementierung eines globalen Grid Computing Netzes ist die Erstellung von Standards, welche die Interaktion unterschiedlicher Hard- und Software-Komponenten erlauben. Die ersten definitiven Grid Standards werden anfangs 2004 erwartet. Bei Betrachtung der provisorischen Standards wird jedoch klar, dass die technische Implementierung von Grid Computing Netzen gewisse Einschränkungen gegenüber der ursprünglichen Grid Vision aufweist. Andererseits glauben grosse IT Firmen, wie beispielsweise IBM oder SUN, stark an das Zukunftspotential von Grid Computing und treiben die Entwicklung mit einer Reihe gridkompatibler Produkte voran.