Grid Computing Technology

Georgiana MARIN
Romanian – American University Bucharest, ROMANIA
georgiana.maryn@gmail.com

This paper presents the grid computing technology, the recent developments in this field. The idea of grid computing has its origins in the early development of computer networks and technologies involved, given that CPU cycle use "free" or "unused" was seen as an optimal and cost-effective way to use all capabilities of the hardware resources which were very expensive machinery.
Keywords: grid computing, technology, virtualization, data base, file system

1 Introduction

The term grid is referring to a distributed computing infrastructure, able to provide resources based on the needs of each client. Grid technology can largely enhance productivity and efficiency of virtual organizations, which must face the challenges by optimizing processes and resources and by sharing their networking and collaboration. Grid computing technology is a set of techniques and methods applied for the coordinated use of multiple servers. These servers are specialized and works as a single, logic integrated system.

The grid has developed as a computing technology that connects machines and resources geographically dispersed in order to create s virtual supercomputer. A virtual system like this is perceived like it has all the computing resources, even if they are distributed and has a computing capacity to execute tasks that different machines cannot execute individually.

In the past few years grid computing is defined as a technology that allows strengthening, accessing and managing IT resources in a distributed computing environment. Being an advanced distributed technology, grid computing brings into a single system: servers, databases and applications, using a specialized software. In terms of partnership between organizations, grid technology may include the same enterprise organizations as well as external organizations. Therefore, grids may involve both internal and external partners, as well as only internal ones. [1]

The complexity of the environment in which the grid will be developed and the requirements that have to be applied depend on the environmental impact of trade, defining the relationship of trust, security considerations, globalization and integration period, of the company involved, in the market. Based on the different levels of complexity for the enterprise, grids can be categorized as follows:

- Infra-Grid – this type of grid architecture allows optimizing the resource sharing within a division of the organization's departments. Infra-grid forms a tightly controlled environment with well defined business policies, integration and security.
- Intra-Grid – it's a more complex implementation than the previous because it's focused on integrating various resources of several departments and divisions of an enterprise. These types of grids require a complex security policies and sharing resources and data. However, because the resources are found in the same enterprise, the

focus is on the technical implementation of the policies.

- Extra-Grid – unlike intra-grid, this type of grid is referring to resource sharing to / from a foreign partner towards certain relationships are established. This grids extend over the administrative management of local resources of an enterprise and therefore mutual convention on managing the access to resources are necessary.
- Inter-Grid – this kind of grid computing technology enables sharing and storage resources and data using the Web and enabling the collaborations between various companies and organizations. The complexity of the grid comes from the special requirements of service levels, security and integration. This type of grid involves most of the mechanism found in the three previous types of grid.

Grid computing technology is not the only distributed technology; among other distributed technologies we can include: web technology, peer-to-peer, clustering and virtualization.

The similarities between the grid computing technology and the others distributed technologies are:

- A grid hides its complexity, interconnects different resources and provides a unified perception of the environment;
- Can share files and communicate directly through a central broker;
- Clusters and grids groups the resources in order to solve a problem and can have a tool for unified management of all components [2].

The differences between the grid computing technology and the others distributed technologies are:

- A grid allows machines to work together and collaborate in a way that involves more than simple communication.
- Grids can use the relations many to many to share a wide variety of resources, not just files.
- A cluster is made of computers of the same type located in a central location requiring a single system image – SSI. Unlike the cluster, where the emphasis is on performance of parallel processing, in a grid is on the resource sharing. SSI is not anymore a requirement, computers from grid are heterogeneous and geographically distributed.
- Unlike virtualization, which refers to a single system, a grid allows virtualization of heterogeneous resources everywhere, even global, to form a pool of IT services.

2. Grid computing requirements

The requirements of grid computing technology are: security, fault tolerance, global name spaces, scalability, adapting heterogeneity, persistence and non persistence, extensibility, complexity management and autonomy.

- Security - addresses a wide range of issues, such as: authentication, data integrity, access control and updates. Grid system and technologies, in order to earn a place on the corporations market and their use in important IT departments, is necessary the grid to be focused on the security issues. The security mechanism must allow the integration of applications with any degree of diversity and to support various security policies, adopted in each particular case of entities that integrate and implement a system like this. To ensure flexibility, security must be integrated in the environment of the grid from the beginning. Future changes of the system may lead to amendments to security needs so that the existing necessary security mechanism elements can avoid the need to update and modify essential parts of the system and applications. The security mechanism must be sufficiently abstract to allow the definition and integration of any security policy. These conditions being given, the grid system must provide mechanisms that allow users and resources owners to use that policy that adapts to the requirements of security and performance and to local administrative requirements.

- Fault tolerance - Binding the developers to predict and deal with all the faults in a system leads to a significant increase of the application complexity. For this reason, grid computing would be provided with mechanisms for fault tolerant computing.
- Global name spaces - The absence of a global namespace that allows a convenient and consistent access to data and resources is a general issue in the fields that involves parallel and distributed calculations. Thus, one of the tendencies of grid computing is that each grid object must be able to access the grid in a transparent way without harming the data integrity and resources.
- Scalability - Grids adopt an architecture based on a review of non centralized resources. So, the architecture is based on the principle of distributed system: the service required by any component is independent from the number of system components.
- Adapting heterogeneity - A grid system must support interoperability between resources and heterogeneous hardware and software systems.
- Persistence and non persistence - One of the most important aspects of resource management is the aggregation in an advanced mechanism of detection algorithms and resource planning. This mechanism should ensure a high level of performance in order to respond to the high demands regarding service quality. In this regard, the resource management should firstly address to aspects concerning on the resources persistence and non persistence, because the environment proposed by the grid technology involves multiple administrative domains, a variety of heterogeneous resources whose behaviour in unpredictable even in the short term.
- Extensibility - means grid system flexibility so that they can meet the current and future requirements of users and also the ones that cannot be anticipated.
- Complexity management - Complexity is common in developing applications in a space with more dimensions, dimensions generated by: heterogeneity in using the resources and security policies, a wide variety of malfunctions, the requirements on the availability of resources, name spaces and the number of components. In this matter of abstraction that grid computing technology provides, must be sufficiently high to reduce the size of the units of measurements and allows them a efficient complexity administration.
- Autonomy - Grid systems may include resources from different organization involved in the distributed environment. Each resource owner wants to keep control over its own resources and therefore the grid system should give the opportunity to vary the level of access given to users, to specify when and how a resource can be used and also the possibility that each owner to organize their own resources according to needs, without relying on the grid system [3].

3. Advantages and disadvantages of Grid Computing Technology

Although grid computing has been developed in research laboratories, manufactures and companies have also started to adopt this technology on two of the most important benefits that grids brings: economy and performance. Grid computing aims to bring together operating systems and different hardware platforms into a single virtualized entity whose performances is higher on average then parts.

Grid computing saves financial resources both in capital and operating costs. This positive aspect is achieved by using all the computing resources of all components of the grid.

The second benefit of grid computing technology is performance in processing data integrity. By increasing the processing power, applications run faster the computing tasks and provide faster results.

Even if the benefits of grids are real, this is still a specialized technology. Grid computing is suitable for organizations that already use in a way or another the high performance computing or are already oriented in some form towards distributed computing. Although there isn't a direct barrier to the passage of an organization to use a grid system, there are still issues to be taken into consideration before such a step. One of the reasons this happens is determined by the applications, which must provide the possibility of breakdown into smaller computing tasks in order to take advantage of parallel computing. If the operation currently executed depends on the previous tasks, then processing cannot be done in parallel on distributed resources so that the application cannot benefit from the grid advantages in terms of high performance computing.

Another issue that emerged over the time is the one that occurs in administration and accounting resources. This problem is more obvious as a solution involving a grid system involves multiple departments of the same organizations and each department require proportionality between the grids contribution, through its own resources, and the gain obtained from it.

Besides resources accounting there's also a problem with software licensing. This is particularly important because licensing costs may eventually cancel the savings obtained from the use of grids.

Securing the servers and grid administration resource is an issue that should not be neglected. In many cases servers are public addressable because of data delivering and receiving from agents that are anywhere, geographically speaking. So, these servers are exposed to attacks and unauthorized access, even DoS – Deny of Service. Experts recommends that, in this case, the companies must eliminate all unnecessary services and to carefully monitor these machines. However, this brings an extra complexity problem.

One of the benefits of grid environments is that it allows the development of specific and on demand environments for various commercial environments. An important feature of on demand environments is the ability to respond to rapid changes in the market while reducing operational costs.

Grid computing helps distributing and sharing data, which enables a collaborative improvement at both inside and outside the enterprise. Thus, companies can reduce time spent placing products on the market, can quickly solve specific problems and can address immediately to the customers' requests. One of the troubles encountered when distributing the workload over several machines is the difficulty to trace the distribution process, which can lead to bandwidth problems. Is not only the bandwidth performance criterion but also the performances of the grid, of the way tasks are distributed and how they handle the distribution server. Bandwidth management is a serious problem that must be taken into account. A grid network bandwidth covers two aspects of any grid:

- CPU utilization. CPU bandwidth in a grid represents the maximum rate at which the grid can operate.
- Using the network. Available bandwidth of the grid and the one used internally by the grid.

These two elements affects a number of different systems and it's important to balance these elements in order to avoid the risk of having an inefficient grid, a grid that is unable to manage the computing tasks, or customers to start overburden the resources that they have at their disposal.

Although there are several technologies that can be integrated into the grid or pattern which can form the basis of the mechanism of resource management and workload distribution and planning, however, they must be adapted to the needs of grids. We cannot avoid the artificial intelligence solutions to solve these type of problems, but must consider the complexity of the system and the additional resources that they are bringing.

Categories of grid computing fits into several areas of research communities. These may include areas oriented in intensive computational calculus, peer to peer, utility, data and applications and collaborations. Each of these communities can use a different model on the adoption of grid technology.

Grid computing is a new concept for commercial industry and a large area of interest has developed around infrastructure virtualization through the manipulation of resources as utilities.

Adopting a grid solution in a commercial activity depends on the ability of technology to meet the needs of improving turnover. This involves the adoption of the grid models key factors, such as adaptation of existing resources, reducing operational costs, creating a flexible and scalable infrastructure, while accelerating development, reducing the period of development and marketing of products and increase customer satisfaction and business productivity.

An important issue in grid adoption models is the complexity of IT infrastructure needed to implement a grid system. The integration complexity of heterogeneous environments is a challenge and it must be taken into account factors such as activation of grid resources in homogeneous and heterogeneous environments, enabling resources in the form of services to external participants and porting applications to grid applications. These features allow a classification of grids into the following categories, with varying degrees of complexity at the level of integration:

- Grids developed to optimize computing infrastructure;
- Computing grids, with the virtualization of processing resources.
- Data grids with virtualization and data storage resources.
- Service grids with virtualized services for easy integration.
- Virtualized applications through the composition of resources, through service interfaces, applications from various partners.

Architecture and technology standards developed for grid computing presents a crucial role in adopting grids commercially. Because these standards are constantly evolving and are not mature enough to support subsequent stages, which can evolve into after the adoption of grids, the speed at which these stages will be achieved is reduced. Grid computing solutions have been adopted in several key areas (finance, education, telecommunication, research) and answer the implementation and evolution of environmental performance requirements that are integrated. From this we can conclude that the success of grid computing depends on the integration, service orientation and ability to break down skills and applications and then expose them through the services.

Although grid computing is an important step in information technology it does not

involve the disappearance of existing resources and advanced equipment such as supercomputers, mainframes, clusters etc. The grid has no purpose in replacing the current technology resources, performance management, but the efficiency in extracting optimum benefits from investments made in such resources. There are still many areas (particle physics, simulations, computational intensity processing) in which execution of tasks involves only supercomputer (or dedicated resources in general performance), only areas where the grid cannot cope, or rather does not provide the necessary desired performance in such cases.
Growth of technology and applications on this area, in the last period, may lead to a future in areas indispensability of enterprise grid computing. Such a critique must always be approached in the case of adoption of this technology to enable a more coherent adaptation to the needs for which a solution is desired grid.

4. Technologies used in computational grids

Performance criteria in a grid environment are dictated by several of the most important aspects are planning mechanisms and the mechanisms of load balanced and integrating resources and data necessary for their calculations on the same node. The choice in favour of another technology to support the role of a computational grid environment is a problem of decision that can be solved only after a careful analysis on a number of factors who are determining the particular environment in which to apply the solution and the specific requirements of grid.
Heterogeneity. Any facility planning software that is used in the grid should be able to support heterogeneous computing nodes. Even if the current implementation does not need them, yet this method is a useful strategy for the future expansion of the system. Thus the nodes of a service reservoir have to bear the same service or set of services irrespective of the choice or implementation of their type.
Security. Mechanisms planning software must support the required levels of security specified in the particular needs of any environment that will be integrated grid system. However, many implementations of grid are located in the same administrative domain where security requirements are less stringent than in distributed environments that cover more organizations. Under these conditions many grids shows a low outside the additional security provided by the operating system of nodes, since all the nodes are behind the same firewall as well protected as any system in the network company. Thus a compromise must be made because the existence of an additional security level without a decrease in performance (speed of the executable is directly affected), but leads to a secure data protection.
Reliability. Since the planning provides some of the most important performance, the mechanisms must ensure the reliability levels needed to resolve defects. And in this case the solutions on this issue are not unique and their usefulness varies depending on the environment in which a solution will be integrated grid. Such technology should be appropriate and adapted to the user.
Thus, in some environments the emphasis is on completion of the total load of calculation, so that some form of recovery of computing tasks is sufficient. If a sub-task of calculation fails due to a defect node, network or software, then that burden is being passed back to another appropriate node. This mechanism provides some fault tolerance but without any warranty and / or improve the response time of the task, is known that it will grow with each retransmission of sub-tasks.
On the other hand, in other environment the response time is the most important criterion and the planner should be able to ensure the return of preserving the response but decreased during execution.

Such solutions usually involve a launch of multiple copies of sub-tasks to different nodes in grid computing. This technology is suitable for planning grid environments are those that present a high flexibility towards upgrades and additional features that ensure the fundamental planning.

Data Virtualization is an architectural approach that hides the location and data format for the user. Level virtualization architecture that addresses the data shows a direct proportion to the level of complexity with the complexity of the level of computing resources. Data handled in a grid environment can be structured or unstructured, ordinary files and other objects stored in file systems, also in case of a grid covering a development involving several organizations, external data sources can cover a wide large format and can be publicly accessible, available on a contract basis, can be shared between companies working together in a consortium or assumptions to be shared in an academic or business environment based on collaboration in a joint project, some data may support active interrogation while others do not. In these conditions of an environment that involves a high degree of dynamism so the question of criteria for a decision on the choice of technologies to fill the grid environment. The answer is not unique and the selection is based on a number of important factors and their relationship shows a coefficient related to the particular specifications of each environment that integrates a grid system. Thus Cluster file systems or distributed network can integrate successfully into the grid system. Also aggregation and data replication may be needed in certain grids.

Cluster file systems is particularly suitable for constraints within a single administrative domain because they are usually sensitive to the degree of separation and differentiation between the cluster nodes, therefore cluster systems are less applicable in distributed environments that cover more administrative areas. In the same way file systems are network-oriented environments such as logos or single-site campus.

Both solutions use the advantages of technological advances achieved in computer networks. Distributed file systems have been developed following several administrative areas covering a wide geographic dispersion. However only single file systems do not provide sufficient data management mechanisms, especially where large volumes of data are stored in databases or other structured storage resources. Thus we have an important role of federal databases and aggregates.

Aggregated databases are trying to achieve consolidation of all data formats from disparate systems into a single parallel database. Although this model provides scalability it is also oriented in the implementation, covering a secure administrative area.

Federative data bases are trying to keep track of data distribution in different areas but to provide unified interfaces that hide implementation and data source to user. A federative database provides a greater level of flexibility as opposed to aggregated databases and a higher potential performance problems when characteristics of remote data to the user. So, another mechanism may be added: data replication.

Another aspect to be taken into account in choosing the virtualization of data to support the heterogeneity of data sources both in terms of platform (providing access and data) and in terms of technologies or techniques used for data storage.

Linking data and calculations is a fundamental problem and refers to their alignment on the same node in the grid to allow for optimal performance. Myriad of available technologies and products that aim to solve this problem shows a touchstone: the lack of collaborative approach based on open standards (this methodology to be approached from a considerable number of organizations in the business environment in recent years only). Thus, each technology presents a remarkable intelligence, but within the limits of their own views on the environment and the grid components that involve, or it can integrate them in time. In this way each technology manifests its decisions regarding the settlement calculations with data alignment in a manner isolated from the others. In conditions of low complexity and where planning decisions are obvious mechanism for calculating the load (the entity charged with moving the appropriate entities for calculating the node) and the mechanism for moving data (the entity charged with moving the data entities) may relate so that both data and calculation takes place on the same node. In terms of complexity of a grid and especially the expansion in complexity over time, especially regarding the integration of geographically distributed resources with grades of spreading increasingly larger planning mechanisms linking computing tasks with mechanisms data handling is becoming more difficult because the decision conflicts are becoming more frequent. Alignment of data and calculations on the same node is a strategy that further studies should be conducted because it allows obtaining optimal performance in grid environments and systems that exceed the current average.

5. Principles of development of a Grid System Grid, applications and requirements

Grid systems must ensure the transparency of the following ways: access, location, heterogeneity, failure, replication, scalability, concurrency and behaviour. Users and developers should not know exactly where an entity is located to interact with it and should not be forced to acknowledge the failure of components / resources of the grid system. This grid system must be equipped with self-recovery mechanisms in order to accomplish tasks. This represents the mask (transparency) on the vision system.

Grid architecture must be adaptable to as many solutions on the grid user needs and goals. A rigid system in which the known policies is limited, decision are pre-selected transaction cannot meet these demands. This grid system should allow greater flexibility for both semantic applications without imposing predetermined solutions to unique methods of implementation. Grid systems must also allow users to choose the desired type and level of functionality and appeal to their own compromise between functionality, cost and performance. This philosophy should be integrated into the grid by specifying the architecture and functionality but not implementation of core aspects of the system. The core should therefore be made of extensible components, which can be updated and replaced whenever possible / necessary while implementing the system should provide an implicit development of each functional component useful to normal users.

In general one can speak in a grid environment for the existence of four categories of users: application users, application developers, system administrators and managers. Grid system architecture and implementation must be realized that it allows users to applications and their developers is focused on their task and not involve their direct work on issues related to installation and implementation of grid infrastructure, but can also be provided and ensuring access to it the grid infrastructure and related details if required.

Another aspect of flexibility to be ensured to grids is to keep the machine's operating system will be integrated into the system. Such grid systems must be developed so as to work with and above the machine's operating system and involve such changes, on configuration, but fewer in number and complexity.

In developing an application grid security issues require consideration of a comprehensive approach. Below are treated briefly these considerations:

- Single sign-on. ID Mapping over the system. GSI (Grid Infrastructure Security) provides authentication, authorization and secure communication. Therefore we must fully understand the implications and security management.
- Multiplatform. Although GSI is based on open standard software running on multiple platforms though different platforms security mechanism will not always consistent. For example, a security mechanism for reading, writing and execution on a UNIX system is different from the traditional system of Microsoft Windows. This should be taken into account the different platforms will run the application.
- Use of GSI. For any specific function of the application that may require authentication or a special permit application indicated that the default mechanism to use GSI as it simplifies the development process.
- Encryption. Although GSI together with data management mechanisms ensure secure communications and data encryption over the network still need to consider what happens once the data arrived at the destination. An example in this regard is the special case where data are transmitted to a resource to be processed and are stored on a local disk in non-encrypted format, and then there is the risk that other track users to have access to that data.

Influence of resource management issues on application development:

- Choice of appropriate resources. Choosing appropriate resources an application is made by the resource management mechanism that works in conjunction with a broker. This implies that the application to specify accurately the working environment required (OS, processor speed, memory, etc.). The application shows how dependent fewer specific platform the greater the chances that an appropriate and available resource to be found and resolved quickly workload.
- Multiple computing sub-tasks. For applications involving multiple computing tasks to analyze the interdependencies between them should be made to avoid additional logic: the communication intercrosses, data sharing and tasks competing administration.
- Managing computing tasks. If the application has to provide a reactive response to the user or to release resources when the application should be designed so as to use the mechanisms of grid resource management to ensure consistency and integrity of an environment.

The mechanisms for data management in a grid to maximize the best use of limited storage space, the network bandwidth and computing resources. The following aspects of data management must be considered in the development of grid applications:

- Dataset size. If the application works with data sets of very large ineffective, if not impossible, for data to be moved on the system that will run and load calculation. In this case one possible solution is data replication (copying a subset of the data set) on the system to execute computing tasks.
- Geographical distribution of users, computing resources, data and storage. If the grid is geographically distributed environment and shows limited network speeds then the application must take into account aspects of design data access limited or low speed.
- WAN data transfer level. As grid environments and distributed networks involve extensive grid any application must take into account

issues of security, reliability and performance for handling data across the Internet or another WAN. Such applications require a logic performance for dealing with changing situations in which access to data may be slow or restricted.

- Planning data transfer. Planning involves data following two main issues: ensuring the transfer of data to the appropriate location and time at which this transfer is required, taking into account the number and size of competing data to / from any resource.

6. Conclusions

Spurred the development of both high-speed networks and the increasing computing power of microprocessors, processing grid has a remarkable impact not only at academic level, but increasingly more and enterprise in all fields. Despite technological advances, however, less than 5% of Windows servers processing power and desktops, respectively 15-20% for UNIX servers is used. Companies make profits and offers exceptional rates, with a return period of low and low total cost compared to other technological solutions.

References

[1] M. Velicanu, I. Lungu, I. Botha, A. Bara, A. Velicanu, E. Rednic – *Sisteme de baze de date evaluate,* Ed. Ase, Bucuresti 2009

[2] A. Abbas - Grid Computing: A Practical Guide to Technology and Applications, Editura Charles River Media, 2005

[3] Ian Foster, Carl Kesselman, Steven Tuecke - *The Anatomy of the Grid:Enabling Scalable Virtual Organizations*

[4] W. Jia, W. Zhou - *Distributed Network Systems.From Concepts to Implementations*, Editura Springer, 2005

[5] Virginia Center for Grid Research - http://www.cs.virginia.edu/~jfk3w/grid_group/grid_overview.html

[6] Andrew Conry-Murray – *Grid Computing's Promises and Perils* http://www.itutilitypipeline.com/showArticle.jhtml?articleID=17603526

[7] http://www1.ibm.com/grid/about_grid/ibm_grid/get_started.html http://www.cs.virginia.edu/~jfk3w/grid_group/grid_overview.html

Georgiana MARIN graduated the Faculty of Computer Science for Business Management of the Romanian American University in 2003. In 2009 she graduated the Economic IT Masters Program of the Romanian American University. She is curently a PhD Candidate in Economic Informatics of the Academy of Economic Studies.