REVIEW AND ANALYSIS OF GRID COMPUTING

[1]Dr.C.Jeyabharathi, [2]C.Sivaranjani
[1]Assistant Professor & Head, [2]Assistant Professor
[1,2] PG Department of Computer Science
[1,2]Arulmigu Palaniandavar Arts College for Women, Palani

Abstract : Grid computing has emerged as an important new field, distinguished from conventional distributed computing by its focus on large-scale resource sharing, innovative applications, and, in some cases, high-performance orientation. Grid problem can be defined as flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resource what we refer to as virtual organizations. The real and specific problem that underlies the Grid concept is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations. In this paper we discuss the basics of Grid Computing, Grid Architecture, Globus toolkit, Virtual Organization, Grid Resource discovery, various Grid topologies, Grid Scheduler, Methods of Grid Computing, Simulation tools, and we compare Grid with Cloud Computing .We also mention about our Central Government's Grid project GARUDA. Finally we list out the various applications of Grid computing.

***IndexTerms* – Grid Computing, Grid Architecture, Globus Toolkit, GARUDA.**

I. INTRODUCTION

Grid computing is a type of distributed computing which allows sharing of computer resources through Internet. It not only allows us to share files but also most of the software and hardware resources. "Grid is the future Internet" is the prime mantra of the present researchers and so whole world follows Grid. Grid applications are usually scientific with large number of users and dynamic resources. Because of the dynamic nature of Grid systems, it allows participants to join or leave the system at any time. Grid computing combines computers from multiple administrative domains to reach a common goal to solve a single task. Grids can be primarily classified [1] into various types, depending on nature of their emphasis- computation, data, application service, interaction, knowledge, and utility. Accordingly, Grids are proposed as the emerging cyber infrastructure to power utility computing applications. Computational Grids aggregate computational power of globally distributed computers (e.g., TeraGrid, ChinaGrid, and APACGrid).

II. GRID ARCHITECTURE

Ian Foster et al., presents [2] the various layers of the Grid architecture by using **Hourglass model** [3]. The narrow neck of the hourglass defines a small set of core abstractions and protocols, onto which many different high-level behaviors can be mapped, and which themselves can be mapped onto many different underlying technologies. While the base of the model conveys the different underlying technologies, the top of model shows high-level behaviors that translate into services and user applications. The **fabric layer** provides the resources to which the shared access is controlled by the grid protocols. The resources normally include physical and logical entities. Physical entities are resources like storage systems, catalogs, servers, and network resources. The resource may be a logical entity like distributed file system, computer cluster or distributed computer pool, and database systems to store structured data. The Grid mechanism normally permits the capability for the resource management, which involves discovery and control. The **connectivity layer** defines core communications and authentication protocols required for Grid specific network transactions. These protocols enable the exchange of data between fabric layer resources. The **resource layer**, based on the connectivity and authentication protocols, controls the access resources.

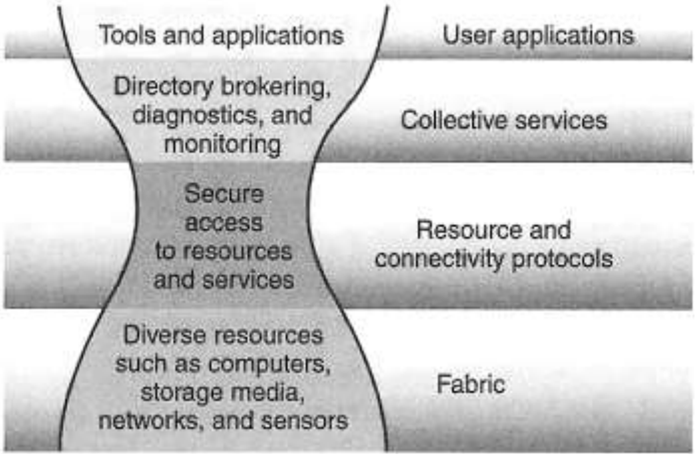

Fig 1. Model of Grid Architecture

The **collective services layer** deals with the directory brokering services, scheduling services, data replications services, and diagnostics/monitoring services. These services are not associated with any one specific resource but focus on interactions across resources. The programming models and tools define and invoke the collective layer functions. This layer is a key component in the whole grid architecture and its functioning. This is the layer that glues all the resources together in expedient exchange. The top layer, **User Applications**, comprises the user applications that operate within a virtual organization (VO) environment.

The components in each layer share common characteristics but can build on new capabilities and behaviours provided by the lower layer. This model demonstrates the flexibility with which Grid Architecture can be extended and evolved. This is the precise reason why grid architecture is taking shape and form based on the guidelines developed by the visionaries and grid standard forums.

III. VIRTUAL ORGANIZATION

In grid computing, a virtual organization (VO) [4] refers to a dynamic set of individuals or institutions defined around a set of resource-sharing rules and conditions. All these virtual organizations share some commonality among them, including common concerns and requirements, but may vary in size, scope, duration, sociology, and structure.

Researchers must join a VO in order to use grid computing resources provided by the corresponding Grid services. Each virtual organization manages its own membership list, according to the VO's requirements and goals. VO is created to run specific (group of) tasks and may include multiple specially created Grid Services. VO is created on the base of the business agreement between participating organizations and individuals each of which contribute their specific resources (computers, services, people, etc.). The agreement defines all resources and services available to VO members and conditions on which these resources and services are provided and used. VO like real organization may contain all basic services required to run typical organization but these services "physically" and administratively may be run by member organizations on behalf of the VO. The examples of VO are: members of a large international long-term collaboration in high energy physics; or group of organizations participating in severe weather simulation and prediction; virtual laboratory involving group of specialists using remotely located unique analytical equipment (e.g., electronic microscope, or mass-spectrometer) for analysis of some samples.

IV. GLOBUS TOOLKIT- GRID MIDDLEWARE

Grids are typically managed by gridware a special type of middleware that enable sharing and manage grid components based on user requirements and resource attributes (e.g., capacity, performance). Middleware connects other software components or applications to provide the following functions:

- Run applications on suitable available resources – Brokering, Scheduling
- Provide uniform, high-level access to resources – Semantic interfaces, Web Services, Service Oriented Architectures
- Address inter-domain issues of security, policy, etc. – Federated Identities
- Provide application-level status
- Monitoring and control

One of the main strategies of grid computing is to use middleware to divide and apportion pieces of a program among several computers, sometimes up to many thousands. Globus is open source grid software that addresses the most challenging problems in distributed resources sharing. The Globus Toolkit includes software services and libraries for distributed security, resource management, monitoring and discovery, and data management.

V. A TYPICAL VIEW OF GRID ENVIRONMENT

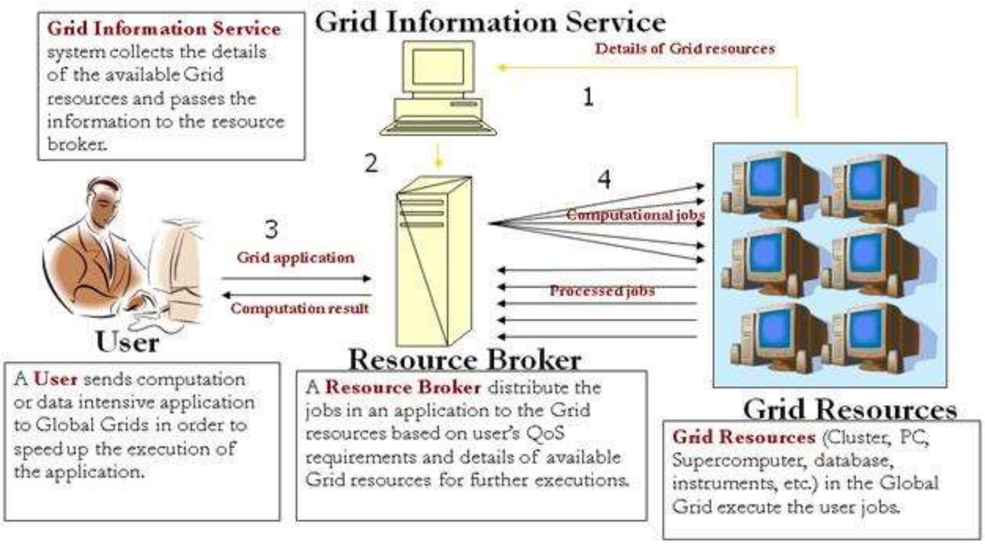

Fig 2. Typical Overview of Grid Environment

VI. RESOURCE DISCOVERY IN GRID

Resource discovery in Grid is a process of locating proper resource candidates which are suitable for executing jobs within a reasonable time. Efficient usage of the right resources is the key component of success of the Grid systems. The characteristics of the Grid systems make the resource discovery a time consuming process which can decrease the performance of the whole system. Various methods have been proposed to solve the resource discovery problem in Grid systems. They are classified into three main categories [5].

Grid resource discovery process uses different classes of systems like centralized and hierarchical systems and agent based systems. Even though these methods have the advantage of Open Grid Service Architecture (OGSA), they suffer from scalability, reliability and false-positive problems respectively. On the other hand, agent based systems [6] are attractive in Grid systems because of their autonomy properties. They have capabilities to determine new migration sites according to their migration policies for the distribution of resource discovery queries, so that researchers adopted Peer-to-Peer (P2P) technology in Grid environment to solve these problems.

In recent years P2P systems have been the hottest research topic in a large distributed system. Since P2P based network approach may overcome the limitations of hierarchical and centralized methods, P2P techniques are especially used in resource discovery process. The self-organization, scalability and dynamicity are the inspiring features of P2P systems. As P2P network is a kind of distributed network, the nodes of P2P network share their own part of the hardware resources like processing power, storage capacity, network connectivity, RAM, virtual memory etc.,

P2P systems are mainly divided into Structured P2P networks, Unstructured P2P networks and Super-Peer systems [7,8].Unstructured P2P resource discovery approaches handle the dynamicity of resources. The routing mechanism of unstructured approaches presents the Grid to scale. In super-peer based methods, flooding [9] mechanism is used which leads to single point of failure.

VII. GRID TOPOLOGIES

7.1. Intragrid

Machines participating in the grid may include systems from multiple departments but within the same organization. Such a grid is also referred to as an intragrid. Here trust is based on personal contracts. The intragrid offers the prospect for trading or Brokering resources over a much wider audience. Resources may be purchased as a utility from trusted suppliers.

7.2. Extragrid

Grids external to an organization is called extragrid. Resources of a consortium of organizations connected through a (Virtual) Private Network. As the grid expands to many departments, policies may be required for how the grid should be used. For example, there may be policies for what kinds of work is allowed on the grid and at what times. There may be a prioritization by department or by kinds of applications that should have access to grid resources.

Also, security becomes more important as more organizations are involved. Sensitive data in one department may need to be protected from access by jobs running for other departments. Dedicated grid machines may be added to increase the quality of service for grid computing, rather than depending entirely on scavenged resources. Trust is based on Business to Business contracts

Security increases in importance once the bounds of any given facility are traversed. The grid may grow to be hierarchically organized to reduce the contention implied by central control, increasing scalability.

7.3 Intergrid

Interlinked grids within multiple organizations are called intergrid. A structure that allows Grids to grow in a scalable manner by allowing peering between Grids. It is nothing but Global sharing of resources through the internet. Here trust is based on certification. A grid may grow to cross organization boundaries, and may be used to collaborate on projects of common interest. This is known as an intergrid. The highest levels of security are usually required in this configuration. Interlinking of islands of Grids is needed to provide a global Grid-based cyberinfrastructure. Nevertheless, much beyond the need for interoperability at middleware level, interlinking of Grids requires peering arrangements, advanced and automated mechanisms for inter-Grid resource allocation, reservation, accounting, and scheduling. In addition, Grids need to adopt mechanisms that enable administrative separation, by allowing networks of networks, similar to many network-based systems such as the Internet, the Web, and numerous social and biological systems.

VIII. GRID SCHEDULER

Most grid systems include some sort of job scheduling software. This software locates a machine on which to run a grid job that has been submitted by a user. In the simplest cases, it may just blindly assign jobs in a round-robin fashion to the next machine matching the resource requirements. However, there are

Reserving resources on the grid in advance is accomplished with a reservation system. It is more than a scheduler. It is first a calendar-based system for reserving resources for specific time periods and preventing any others from reserving the same resource at the same time. It also must be able to remove or suspend jobs that may be running on any machine or resource when the reservation period is reached.

IX. METHODS OF GRID COMPUTING

There are various computing technologies in which Grid jobs can be executed and various virtual organizations can be connected.

- Distributed Supercomputing
- High-Throughput Computing
- On-Demand Computing
- Data-Intensive Computing

- Collaborative Computing
- Logistical Networking

High-Throughput Computing

- Uses the grid to schedule large numbers of loosely coupled or independent tasks, with the goal of putting unused processor cycles to work.

On-Demand Computing

- Uses grid capabilities to meet short-term requirements for resources that are not locally accessible.
- Models real-time computing demands.

Collaborative Computing

- Concerned primarily with enabling and enhancing human-to-human interactions.
- Applications are often structured in terms of a virtual shared space.

Data-Intensive Computing

- The focus is on synthesizing new information from data that is maintained in geographically distributed repositories, digital libraries, and databases.
- Particularly useful for distributed data mining.

Logistical Networking

- Logistical networks focus on exposing storage resources inside networks by optimizing the global scheduling of data transport, and data storage.
- Contrasts with traditional networking, which does not explicitly model storage resources in the network.
- high-level services for Grid applications
- Called "logistical" because of the analogy it bears with the systems of warehouses, depots, and distribution channels.

X. SIMULATION TOOLS

To experiment various Grid jobs in Grid environment simulator are recommended. The following are few examples of Grid simulator which are currently used.

- GridSim – job scheduling
- SimGrid – single client multiserver scheduling
- Bricks – scheduling
- GangSim- Ganglia VO
- OptoSim – Data Grid Simulations
- G3S – Grid Security services Simulator – security services

GridSim is a Java-based toolkit for modeling, and simulation of distributed resource management and scheduling for conventional Grid environment. GridSim is based on SimJava, a general purpose discrete-event simulation package implemented in Java. All components in GridSim communicate with each other through message passing operations defined by SimJava.

XI. GRID AND CLOUD COMPUTING:

With cloud computing, companies can scale up to massive capacities in an instant without having to invest in new infrastructure, train new personnel, or license new software. Cloud computing is of particular benefit to small and medium-sized businesses who wish to completely outsource their data-center infrastructure, or large companies who wish to get peak load capacity without incurring the higher cost of building larger data centers internally. In both instances, service consumers use what they need on the Internet and pay only for what they use.

The service consumer no longer has to be at a PC, use an application from the PC, or purchase a specific version that's configured for Smartphone, PDAs, and other devices. The consumer does not own the infrastructure, software, or platform in the cloud. He has lower upfront costs, capital expenses, and operating expenses. He does not care about how servers and networks are maintained in the cloud. The consumer can access multiple servers anywhere on the globe without knowing which ones and where they are located.

Cloud computing evolves from grid computing and provides on-demand resource provisioning. Grid computing may or may not be in the cloud depending on what type of users are using it. If the users are systems administrators and integrators, they care how things are maintained in the cloud. They upgrade, install, and virtualize servers and applications. If the users are consumers, they do not care how things are run in the system.

Grid computing requires the use of software that can divide and farm out pieces of a program as one large system image to several thousand computers. One concern about grid is that if one piece of the software on a node fails, other pieces of the software on other nodes may fail. This is alleviated if that component has a failover component on another node, but problems can still arise if components rely on other pieces of software to accomplish one or more grid computing tasks. Large system images and associated hardware to operate and maintain them can contribute to large capital and operating expenses.

TABLE I. Difference between Cloud And Grid

PARAMETER	CLOUD	GRID
Goal	Collaborative sharing of resources	Use of service (eliminates the detail)
Level of abstraction	Low (more details)	High (eliminate details)
Degree of scalability	Normal	High
Ownership	Multiple	Single

Interconnection network	Mostly internet with latency and low bandwidth	Dedicated, high-end with low latency and high bandwidth
Discovery	Centralized indexing and decentralized info services	Membership services
Resource management	Distributed	Centralized/Distributed
Number of users	Few	More
Configuration	Difficult as users haven't administrator privilege	Very easy to configure

Cloud computing and grid computing are scalable. Scalability is accomplished through load balancing of application instances running separately on a variety of operating systems and connected through Web services. CPU and network bandwidth is allocated and de-allocated on demand. The system's storage capacity goes up and down depending on the number of users, instances, and the amount of data transferred at a given time.

Both computing types involve multitenancy and multitask, meaning that many customers can perform different tasks, accessing a single or multiple application instances. Sharing resources among a large pool of users assists in reducing infrastructure costs and peak load capacity. Cloud and grid computing provide service-level agreements (SLAs) for guaranteed uptime availability of, say, 99 percent. If the service slides below the level of the guaranteed uptime service, the consumer will get service credit for receiving data late.

The Amazon S3 provides a Web services interface for the storage and retrieval of data in the cloud. Setting a maximum limits the number of objects you can store in S3. You can store an object as small as 1 byte and as large as 5 GB or even several terabytes. S3 uses the concept of buckets as containers for each storage location of your objects. The data is stored securely using the same data storage infrastructure that Amazon uses for its e-commerce Web sites.

While the storage computing in the grid is well suited for data-intensive storage, it is not economically suited for storing objects as small as 1 byte. In a data grid, the amounts of distributed data must be large for maximum benefit. A computational grid focuses on computationally intensive operations.

XII. GARUDA GRID

GARUDA (Global Access to Resource Using Distributed Architecture) is India's Grid Computing initiative connecting 17 cities across the country. The 45 participating institutes in this nationwide project include all the IITs and C-DAC centers and other major institutes in India.

GARUDA is a collaboration of science researchers and experimenters on a nationwide grid of computational nodes, mass storage and scientific instruments that aims to provide the technological advances required to enable data and compute intensive science for the 21st century. One of GARUDA's most important challenges is to strike the right balance between research and the daunting task of deploying that innovation into some of the most complex scientific and engineering endeavors being undertaken today. The Department of Information Technology (DIT), Government of India has funded the Centre for Development of Advanced Computing (C-DAC) to deploy the nationwide computational grid GARUDA. In Proof of Concept (PoC) phase which ended on March 2008, 17 cities across the country were connected with an aim to bring "Grid" networked computing to research labs and industry.

XIII. APPLICATIONS OF GRID COMPUTING

Grid computing plays vital role in many real life applications. The followings are the few examples.

- Scientific Research
- Financial organization
- Natural Disaster Management
- Weather Forecasting
- Online Multiplayer Game

Apart from the above application areas Grid computing is also applied in the following real-life application areas.

Governments and International Organizations: Problems like disaster response, urban planning and economic modeling are traditionally assigned to national governments or coordinated by International Organizations like the United Nations or the World Bank. Imagine if these groups could apply the collective power of the nation's fastest computers and share their data archives more simply and effectively.

The military: It's a pretty safe bet that the military in many countries is already developing grid technology. The United States have traditionally used their most powerful computers for military applications. But this Virtual Organization is unlikely to let other users access its grid!

Teachers and educators: Education involves students, teachers, parents and administrators and so is a very natural application of grid technologies. E-libraries and e-learning centers are already benefiting from grid-based tools for accessing distributed data and creating virtual classrooms with distributed students, resources and tutors.

Businesses: Global enterprises and large corporations have sites, data, people and resources distributed all over the world. Grids will allow such organizations to carry out large-scale modeling or computing by simultaneously using the resources at their many sites.

Medicine: One of the most obvious applications is in medicine. Doctor had access to a grid that could handle administrative databases, medical image archives and specialized instruments such as MRI machines, CAT scanners and cardioangiography devices. This could enhance diagnosis procedures, speed analysis of complex medical images, and enable life-critical applications such as telerobotic surgery and remote cardiac monitoring.

XIV. Conclusion

In Grid systems, the number of resources and users are normally large and the attributes of resources are often dynamic. As complex and large-scale problems in science, engineering and business need more powerful computing machines, Grid is the suitable technique to solve these problems. In this paper we discussed the basic of Grid computing, Grid architecture and its various features. We also mentioned the few application areas in which Grid computing is mainly used. We conclude that as per the researchers view Grid is going to become the future Internet in short future.

References

[1] C. S. Yeo, R. Buyya, M. D. Assuncao, J. Yu, A. Sulistio, S. Venugopal, and M. Placek. Utility Computing on Global Grids. In Hossein Bidgoli, editor, *Handbook of Computer Networks*. John Wiley and Sons.

[2] S.Ratnasamy, P.Francis, M.Handley, R.Karp,and S.Shenker, "A scalable content-addressable network," in proceeding of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, California, pp 161-72, Aug 2001.

[3] Y.Tang, S.Zhou and J.Xu, "LIGHT: A Query-Efficient Yet Low-Maintenance Indexing Scheme over DHTS,"IEEE Transactions on Knowledge and Data Engineering, vol 22, issue 1, Jan 2010.

[4] Demchenko and Yu "Using WS-Federation and WS-Security for Identity Management in Virtual Organisations"

[5] Hameurlain, A., F. Morvan, and M.E. Samad, (2008) "Large Scale Data management in Grid Systems: a Survey", in IEEE International Conference on Information and Communication Technologies: from Theory to Applications (ICTTA), IEEE.

[6] Cao, J., et al. (2002) "Agent-Based Resource Management for Grid Computing", in CCGRID'02: Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid, p.350.

[7] Mohammed Bakri Bashir, Muhammad Shafie Bin Abd Latiff, Aboamama Atahar Ahmed, Yahaya Coulibaly, Abdul Hanan Abdullah and Adil Yousif, (2011) "A Hybrid Resource Discovery Model For Grid Computing", International Journal of Grid Computing & Applications (IJGCA) Vol.2, No.3.

[8] Ramya M and Nagarathna N, "Survey Paper on Resource Discovery Model in Grid Computing", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 www.ijert.org Vol. 2 Issue 6, June – 2013.

[9] C. Mastroianni, *e t al.*, (2005) "A super-peer model for resource discovery services in large-scale grids," Future Generation Computer Systems, vol. 21, pp. 1235-1248.