VISHVESHWARAIAH TECHNOLOGICAL UNIVERSITY

S.D.M COLLEGE OF ENGINEERING AND TECHNOLOGY

A seminar report

on

"Grid Computing"

Submitted by

Nagaraj Baddi

(2SD07CS402)

8th semester

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

2009-10

VISHVESHWARAIAH TECHNOLOGICAL UNIVERSITY

S.D.M COLLEGE OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

*Certified that the seminar work entitled **"Grid Computing"** is a bonafide work presented by **Mr. Nagaraj.M.Baddi**, bearing USN **2SD07CS402** in a partial fulfillment for the award of degree of Bachelor of Engineering in Computer Science Engineering of the Vishveshwaraiah Technological University Belgaum, during the year 2009-10. The seminar report has been approved as it satisfies the academic requirements with respect to seminar work presented for the Bachelor of Engineering Degree.*

Staff in charge
(S. L. DESHPANDE)

H.O.D CSE
(S. M. JOSHI)

Name: Nagaraj M. Baddi

USN: 2SD07CS402

INDEX

1. <u>Introduction:</u>

Grid is a type of parallel and distributed system that enables the sharing, selection, and aggregation of geographically distributed "autonomous" resources dynamically at runtime depending on their availability, capability, performance, cost, and users' quality-of-service requirements.

Figure 1. The Grid

One of the main grid computing strategies is to use software to divide and apportion pieces of a program among several computers, sometimes up to many thousands. Grid computing can also be thought of as distributed and large-scale cluster computing, as well as a form of network-distributed parallel processing. It can be small — confined to a network of computer workstations within a corporation, for example — or it can be a large, public collaboration across many companies or networks shown in Figure 1.

2. History:

Parallel computing in the 1980's and '90's:

Focused on providing powerful mechanisms for managing communication between processors, and development and execution environments for parallel Machines.

The first modern Grid is the I-WAY (SC95):

Aggregate a national distributed tested with over 17 sites networked together by the vBNS. Over 60 applications were developed for the conference a rudimentary Grid software infrastructure to provide access, enforce security, coordinate resources, and other activities.

Projects after I-WAY:

The Globus and Legion infrastructure, the Condor experimented with high-throughput scheduling. Mars and Prophet experimented with high- performance scheduling.

NWS (Network Weather Service) focused on resource monitoring and prediction and Storage Resource Broker focused on uniform access to heterogeneous data resources. NetSolve and Ninf focused on remote computation via a client- server model.

3. <u>How Grid Computing Works:</u>

A scientist studying proteins logs into a computer and uses an entire network of computers to analyze data. A businessman accesses his company's network through a PDA in order to forecast the future of a particular stock. An Army official accesses and coordinates computer resources on three different military networks to formulate a battle strategy. All of these scenarios have one thing in common: They rely on a concept called **grid computing**[3].

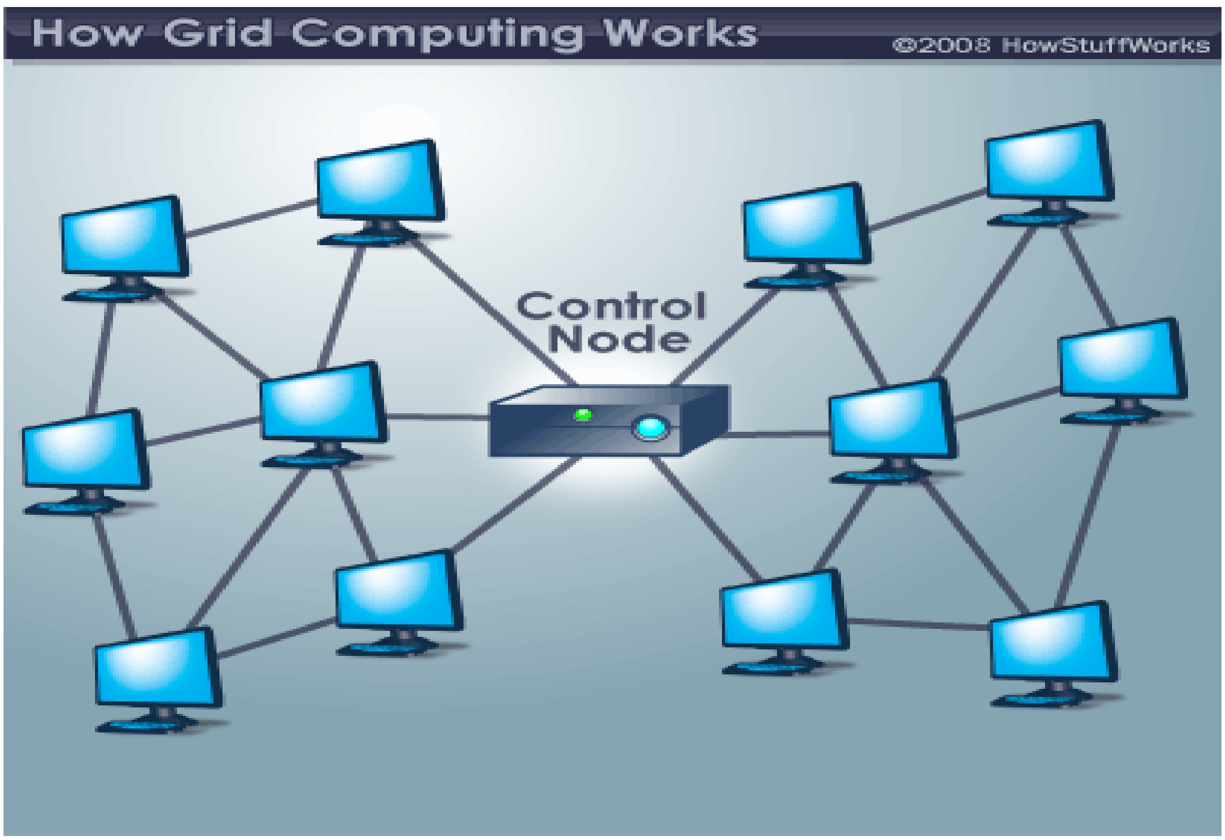

Figure 2. In a basic grid computing system, every computer can access the resources of every other computer belonging to the network.

At its most basic level, grid computing is a computer network in which each computer's resources are shared with every other computer in the system. Processing power, memory and data storage are all community resources that authorized users can tap into and leverage for specific tasks. A grid computing system can be as simple as a collection of similar computers running on the same operating system or as complex as inter-networked systems comprised of every computer platform shown in Figure 2.

The grid computing concept isn't a new one. It's a special kind of distributed computing. In distributed computing, different computers within the same network share one or more resources. In the ideal grid computing system, every resource is shared, turning a computer network into a powerful supercomputer. With the right user interface, accessing a grid computing system would look no different than accessing a local machine's resources. Every authorized computer would have access to enormous processing power and storage capacity [3].

4. **Related technologies:**

4.1 Cluster computing

4.2 Peer-to-peer computing

4.3 Internet computing

4.1 **Cluster computing**

The idea of the cluster computing is put some PCs together and get them to communicate with each other so that it is Cheaper to build than a mainframe supercomputer with Different sizes of clusters and also Scalable. It can grow a cluster by adding more PCs. The following cluster computing architecture showed [4]:

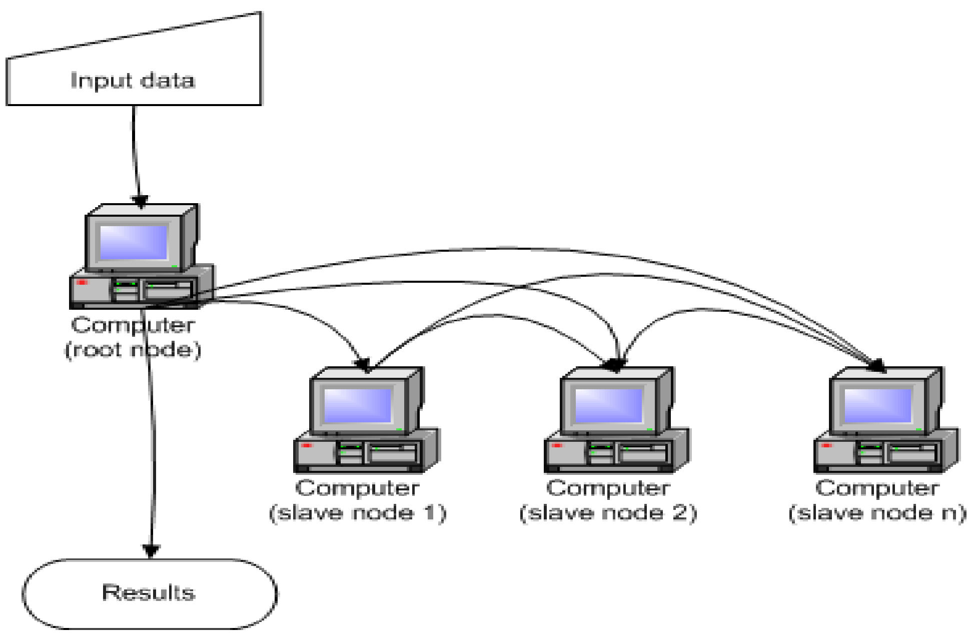

Figure 3 Cluster Architecture

4.2 **Peer-to-peer computing**

The following figure will show peer-to-peer computing the one computer connect to other computers, Can access files from any computer on the network, Allows data sharing without going through central server and Decentralized approach also useful for Grid[4].

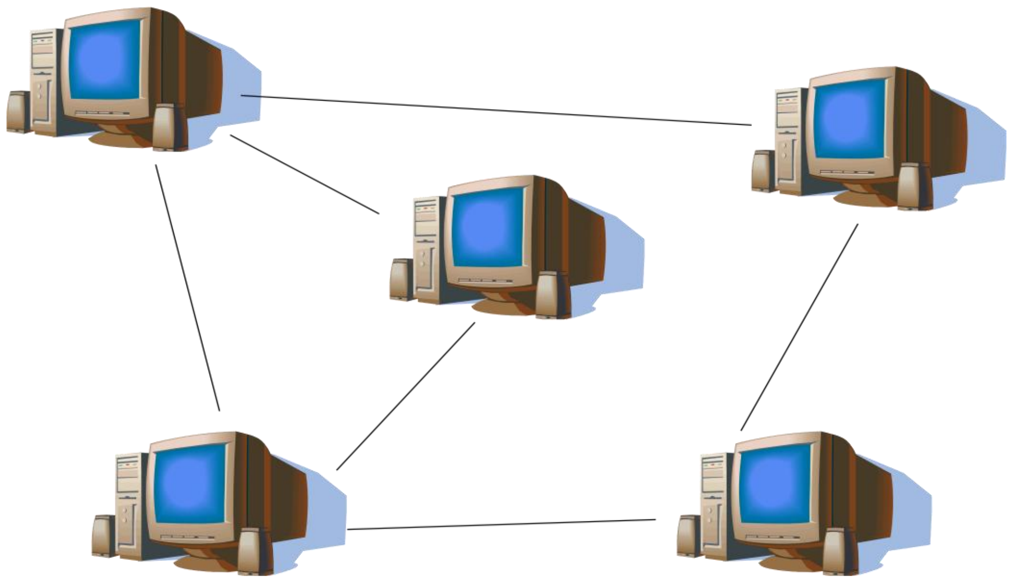

Figure 4. Peer – to – peer Architecture

4.3 **Internet computing:**

The internet computing is the idea, many idle PCs on the Internet Can perform other computations while not being used "Cycle scavenging" – rely on getting free time on other people's computers[4].

5. Grid Computing Logical Levels:

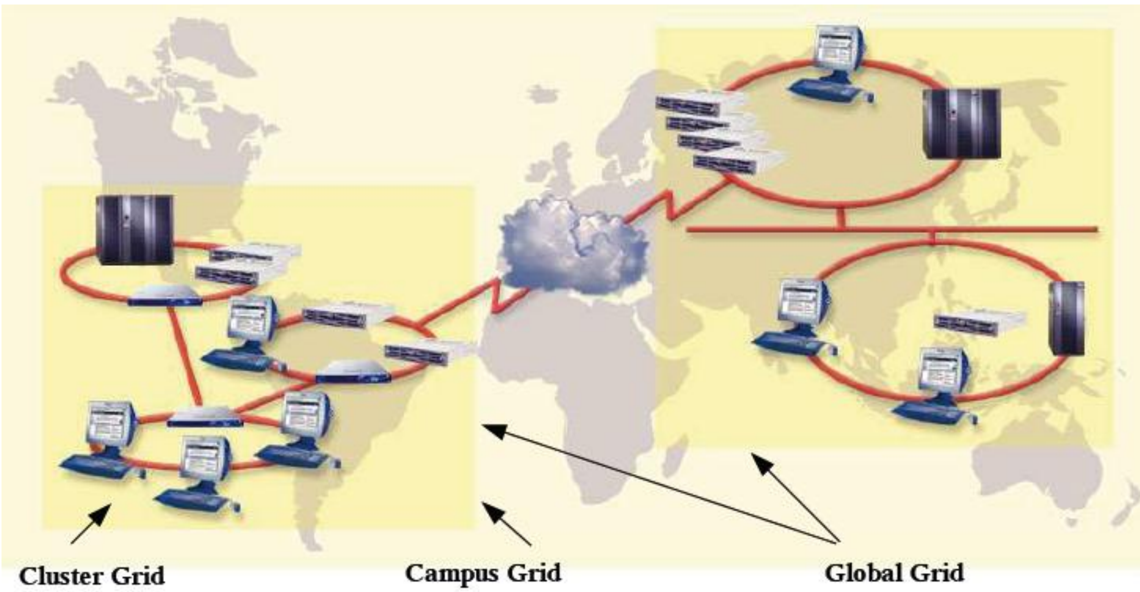

Figure 5. Grid Computing Logical Levels[5].

5.1. Cluster Grid:

- Simplest Grid deployment
- Maximum utilization of departmental resources
- Resources allocated based on priorities

5.2. Campus Grid:

- Resources shared within the enterprise/organizations
- Gives multiple groups seamless access to enterprise resources

5.3. Global Grid:

- Resources shared over the Internet
- Global view of distributed datasets
- Growth path for enterprise Grids

6. Grid Architecture:

No “official” standards Grid Architecture exist.

But:

–Globus Toolkit™ has emerged as the de facto standard for several important Grid fabric, Core Grid middleware, User-level Grid middleware and Grid applications and portals.

–Technical specifications are being developed for architecture elements: e.g., security, data, resource management, information.

The components that are necessary to form a Grid (shown in Figure 6) are as follows[6].

6.1 Grid fabric. This consists of all the globally distributed resources that are accessible from anywhere on the Internet. These resources could be computers (such as PCs or Symmetric Multi-Processors) running a variety of operating systems (such as UNIX or Windows), storage devices, databases, and special scientific instruments such as a radio telescope or particular heat sensor.

The diverse mix of resources that may be shared–Individual computers, Condor pools, file systems, archives, metadata catalogs, networks, sensors, etc., few constraints on low-level technology: connectivity and resource level protocols form the “neck in the hourglass” Defined by interfaces not physical characteristics.

6.2 Core Grid middleware:

This offers core services such as remote process management, co-allocation of resources, storage access, information registration and discovery, security, and aspects of Quality of Service (QoS) such as resource reservation and trading.

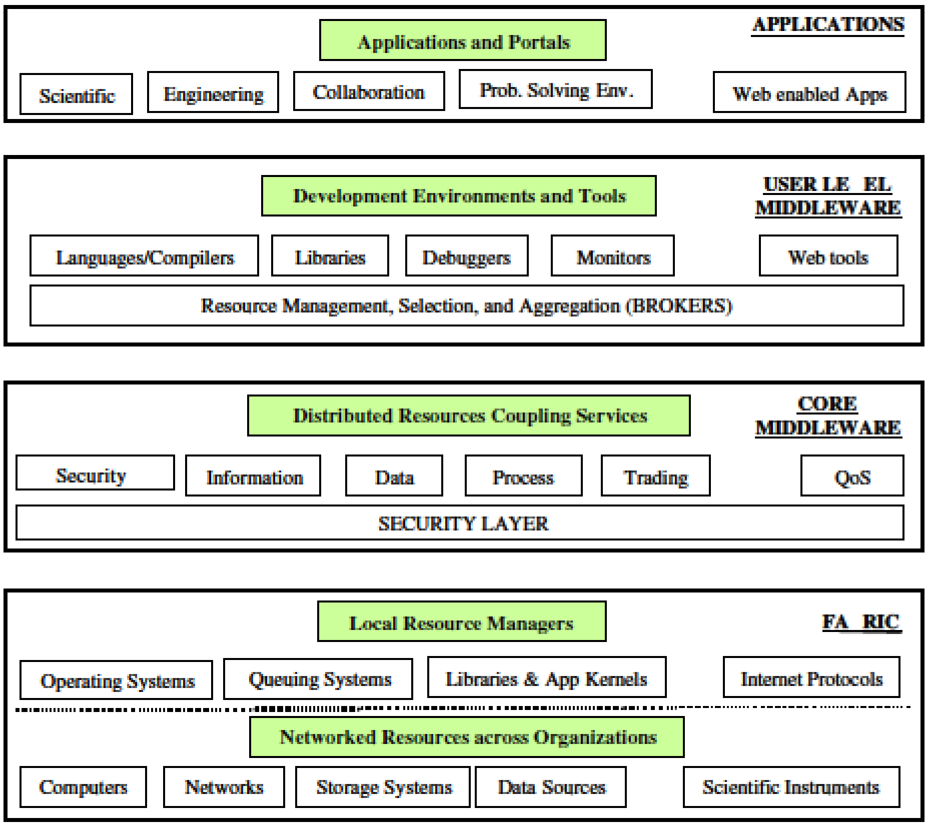

Figure 6. A layered Grid architecture and components.

6.3 <u>User-level Grid middleware</u>. This includes application development environments, programming tools and resource brokers for managing resources and scheduling application tasks for execution on global resources [6].

6.4 Grid applications and portals: Grid applications are typically developed using Grid-enabled languages and utilities such as HPC++ or MPI. An example application, such as parameter simulation or a grand-challenge problem, would require computational power, access to remote data sets, and may need to interact with scientific instruments. Grid portals offer Web-enabled application services, where users can submit and collect results for their jobs on remote resources through the Web.

7. Grid Applications

7.1 Distributed supercomputing
7.2 High-throughput computing
7.3 On-demand computing
7.4 Data-intensive computing
7.5 Collaborative computing

7.1 Distributed Supercomputing

Aggregate computational resources to tackle problems that cannot be solved by a single system[6].

Examples: climate modeling, computational chemistry

Challenges include:

- Scheduling scarce and expensive resources
- Scalability of protocols and algorithms
- Maintaining high levels of performance across heterogeneous systems

7.2 <u>High-throughput computing:</u>

Schedule large numbers of independent tasks.

- Goal: exploit unused CPU cycles (e.g., from idle workstations)
- Unlike distributed computing, tasks loosely coupled
- Examples: parameter studies, cryptographic problems

7.3 <u>On-demand computing:</u>

- Use Grid capabilities to meet short-term requirements for resources that cannot conveniently be located locally.
- Unlike distributed computing, driven by cost-performance concerns rather than absolute performance.
- Dispatch expensive or specialized computations to remote servers.

7.4 <u>Data-intensive computing:</u>

- Synthesize data in geographically distributed repositories[6]
- Synthesis may be computationally and communication intensive.
- Examples:
  - High energy physics generate terabytes of distributed data, need complex queries to detect "interesting" events.
  - Distributed analysis of Sloan Digital Sky Survey data

7.5 <u>Collaborative computing:</u>

- Enable shared use of data archives and simulations
- Examples:
  - Collaborative exploration of large geophysical data sets
- Challenges:
  - Real-time demands of interactive applications
  - Rich variety of interactions

8. Difference: Grid Computing vs Cloud Computing

Cloud computing:

Cloud computing is the use of a 3rd party service (**Web Services**) to perform computing needs. Here **Cloud** depicts **Internet.** With cloud computing, companies can scale up to massive capacities in an instant without having to invest in new infrastructure. Cloud computing is benefit to small and medium-sized businesses. Basically consumers use what they need on the Internet and pay only for what they use[7].

Cloud computing incorporates infrastructure as a service (IaaS), platform as a service (PaaS) and software as a service (SaaS) as well as Web 2.0 Cloud computing eliminates the costs and complexity of buying, configuring, and managing. The hardware and software needed to build and deploy applications, these applications are delivered as a service over the Internet (the cloud).

Example:

1. **Salesforce.com** – Delivers businesses over the internet using the software as a service model.
2. **Google Apps** - Software-as-a-service for business email, information sharing and security

Grid computing:

Grid computing is a form of distributed computing whereby resources of many computers in a network are used at the same time, to solve a single problem. Grid systems are designed for collaborative sharing of resources. It can also be thought of as distributed and large-scale cluster computing[7].

Grid computing is making big contributions to scientific research, helping scientists around the world to analyze and store massive amounts of data by sharing computing resources. Here are some real world examples of Grid Computing.

Grids tend to be more loosely coupled, heterogeneous, and geographically dispersed compared to conventional cluster computing systems.

9. Advantages and Disadvantages of Grid Computing:

Advantages:

- Exploit Underutilized resources. CPU Scavenging, Hotspot leveling.
- Resource Balancing.
- Virtualized resources across an enterprise. (Like Data Grids, Compute Grids.)[8].
- Enable collaboration for virtual organizations.
- Flexible, Secure, Coordinated resource sharing.
- Give worldwide access to a network of distributed resources.

Disadvantages:

- Need for interoperability when different groups want to share resources.
  –Diverse components, policies, mechanisms
  –E.g., standard notions of identity, means of communication, resource descriptions

- Need for shared infrastructure services to avoid repeated development, installation.
  –E.g., one port/service/protocol for remote access to computing, not one per tool/application
  –E.g., Certificate Authorities: expensive to run

- But how do I develop robust, secure, long-lived, well-performing applications for dynamic, heterogeneous Grids?

- I need, presumably:
  –Abstractions and models to add to speed/robustness/etc. of development.
  –Tools to ease application development and diagnose common problems.
  –Code/tool sharing to allow reuse of code components developed by others.

10. Conclusion:

- Grid Computing is becoming the platform for next generation e-science experiments

- By Intranet Grid it is very easy to download multiple files.

- Grid is a type of parallel and distributed system that enables the sharing, selection, and aggregation of geographically distributed "autonomous" resources.

- Resource sharing and resource use in dynamic, scalable virtual organizations. We have also presented both requirements and a framework for Grid architecture.

- Grids tend to be more loosely coupled, heterogeneous, and geographically dispersed compared to conventional cluster computing systems.

11. References:

1. *"What is the Grid? A Three Point Checklist" (PDF).*

http://wwwfp.mcs.anl.gov/~foster/Articles WhatIsTheGrid.pdf.

2. **In http://www.google.com Articles search grid1.pdf**

3. http://communication.howstuffworks.com/grid-computing.htm

4. http://www.globus.org/

5. Architectures of Grid Computing, Dr. Marco Quaranta inwww.sun.com.

6. **http://communication.howstuffworks.com/grid-computing5.htm**

7. **http://www.thepicky.com/tag/Cloud Computing vs Grid Computing.htm**

8. Author and Principal Researcher Dr. Robert Cohen Rural Internet Access Authority Research Report