GRID COMPUTING

The creation of a virtual-super computer!

A study article by Sandyarathi Das

ABSTRACT**:**

*This paper intends to provide an insight of an evolving network application known as “GRID COMPUTING”. Grid computing has emerged as one of the key computing paradigms enabling large-scale and enhanced scientific, mathematic and academic endeavors to be carried out via collaborations on a global scale. In short, this network application is the creation of a virtual supercomputer with a design goal to use technology resources optimally.*  *While the World Wide Web is a service of sharing information and data over the Internet, the Grid is a service of sharing computing power over a network. The ultimate aim of the Grid is to convert the global network of computers into one vast computational resource to meet the requirements of the grand challenge problems on a scientific basis. Also for the community in general it is imperative to provide a common user interface in that, application programmers and users do not have to be concerned with particulars of Web services and their underlying code computational platforms or with data file formats. Evolving technologies, as exemplified by Computational Grids and Web services give us a platform to realize this vision.*

*This paper is structured in four modules of description...i.e., introduction, working-description of the structural and functional features, emphasis on its implementation in the community in general, scientific and engineering applications.*

Table of Contents:

What is the Grid? 3

How does the Grid work? 3

Building the Grid: 4

What does the Grid means to Scientists? 7

Grid projects: 9

References: 10

# What is the Grid?

The Grid or Grid Computing simply means distributed computing. While the World Wide Web is a service of sharing information and data over the Internet, the Grid is a service of sharing computing power over a network. With sharing of computing power over the Internet, it is possible for us to create a “virtual supercomputer” with a very vast computing capability. In basic grid computing, every computer can access the resources of every other computer belonging to the network. Computers networked together can work on the same problems that traditionally were reserved for supercomputers and yet this network of computers is more powerful than the super computers built in the seventies and eighties. Modern supercomputers are built on the principles of grid computing, incorporating many smaller computers into a larger whole.

The term Grid computing originated in the year 1990, when Ian Foster and Carl Kesselmans coined it. The Grid brought to life by the ideas of Ian Foster and Carl Kesselmans and Steve Tuecke who are regarded as fathers of the Grid. However, the idea was taken forward by CERN, which initiated the LHC Computing Grid Project. . This grid has already connected CERN with the US, Canada, Europe.

# How does the Grid work?

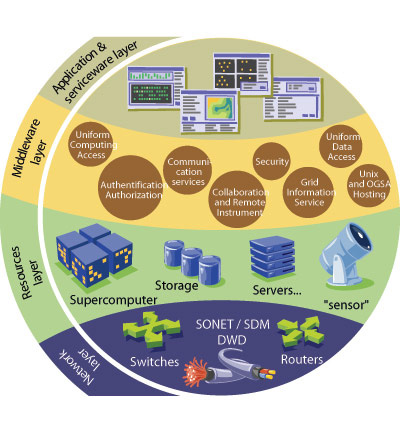
Consider for example that you want to use a program, which is currently installed on your friend’s computer. Assume that both your computers are connected via some network. Normally, in today's world, we have to install the software in our computer and then run the program. But with the realization of the Grid, we will be able to run the program on our friend's computer itself. In other words, we are using the computational power of our friend's computer to run the program. This is a simple and basic idea of what the Grid is capable of.

Grid computing consists of many computers operating together remotely and using the idle processor power of the normal computers to form a virtual super computer. Computing grids are conceptually alike the electrical grids. A grid makes use of a middleware to connect disparate resources across a network allowing them to function as a virtual whole. The goal of a computing grid is to provide users (connected in the grid) with access to the resources they need, when they need them. For instance, if a workstation user were manipulating photographs stored on a file server with Adobe Photoshop, then the workstation would be communicating with the file server to retrieve some subset of the pictures, store those pictures, run Photoshop, compute all of the user’s decisions through keystrokes and mouse movements, compute what effect those movements made within the program, and send a visual output to the monitor.

Within the newest incarnation of Grid Computing, the workstation is only required to retrieve visual data and report the actions taken by the user. All of the processing and computing is done by the supercomputer. Instead of having Photoshop installed and running on the workstation computer, the actions of the workstation user are sent to the supercomputer that analyzes the information, computes what effect that data has within Photoshop, and sends back data in the form of pixels. [1] These pixels are shown on the workstation monitor in a way that is similar to the way Web pages are displayed. In this format, Photoshop is now being operated within a virtual environment. The users’ experience with Photoshop is unchanged, but their workstations’ computing power has exponentially increased because of the access to the virtual processing power of the supercomputer. Imagine the result if a cell phone had access to a ten-terahertz supercomputer. Grid technology has the ability to make that a virtual reality.

# Building the Grid:

The design of the grid is called Grid architecture. The Grid Architecture defines the role of all the components and used to build the Grid. Usually, the basic computer Architecture is defined in terms of “layers”. The same type of “layer” definition can be applied to the Grid also. Actually even home computers also can create a computer grid with relative ease, using designs such as Beowulf clusters.

****

Layers of the Grid:

The layers of the Grid will progress from the most fundamental layers to the most advanced layers. Let us analyze these layers one by one:

1.Network Layer: As explained above, the Network Layer consists of the network, which connects all the computers together. The Internet is such a network, which connects computers across the globe. Organizations will also have their own separate network called local area network (LAN). The Network Layer will consist of all the hardware components that enable us to form the network. These components include- switches, hubs, network cables, routers etc. The Grid will use the concept of an ultra-fast, high-performance network to ensure as little delay as possible in transmission. At present, there are a few existing high-performance networks which have become the testing ground for the Grid. Some of these high-speed networks are- the intra-European GEANT network and the UK Super Janet network.

2.Resource Layer: The resource layer consists of the computing resources that the Grid will utilize. These resources are in the form of supercomputers, storage networks, and observatory devices. It is important for the Resource layer to have a high performance. This is because the Grid is more powerful if the computing resources have a high performance. The Grid power is measured based on the performance of computing resources. Today, even cell phones can be considered as part of the Resource Layer, since they are also capable of connecting to a network.

3. Middleware Layer**:** The Middleware Layer is also known as the Brain of the Grid. It integrates all the resources of the Grid and enables them to participate in the Grid. This integration of resources is enabled by a vast number of software programs present in the Middleware Layer. These software tools are collectively known as “Middleware”. Middleware Layer consists of Metadata. Metadata can be defined as data that contains the attributes of other data. This means that Metadata consists of information as to how and when a particular data was collected, in what format the data is stored and in which location it is stored.

4.Applications Layer: The Applications Layer is probably the larges layer of the Grid as it contains such a huge number of Application soft wares. As explained earlier, the Grid enables us to run application software on someone else's computer, which is connected to our computer through a network without having to install that software on our computer. It enables us to run the software on the other computer using the computational resources of the other computer. The Application Layer is the layer with which people will have maximum interaction. Functionally, one can classify Grids into several types:

1.**Computational Grids** (including CPU scavenging Grids) that focuses primarily on computationally intensive operations.

2.**Data Grids** or the controlled sharing and management of large amounts of distributed data.

3. **Equipment Grids** which have a primary piece of equipment e.g. a telescope, and where the surrounding Grid is used to control the equipment remotely and to analyze the data produced. The grid will be most useful for scientists, as they have to deal with calculations and data from all over the globe. For example, consider a scientist analyzing some satellite images of the earth. These satellite images might be stored in some computers around the globe. Instead of transferring all this data into one central computer to analyze it, which is very time consuming, the Grid will enable the scientist to do the analysis and computation wherever the data is located. The Grid also enables teams of scientists from around the world to get together and analyze and perform calculations on some given data together while video conferencing. The Grid, as described above, is a dream that engineers and scientists hope to achieve in the next few years. But let us get an overview of what the reality is today. It is to be noted that the reality is catching up fast with this dream.

Some of the prototypes of the model of the Grid can be seen in the following concepts:

1**.** Distributed Computing: Distributed Computing is now an important concept in use nowadays. Consider some calculations for which the computing resource of a single computer is not adequate. Then some organizations or academic institutions link the computer resources of the computers linked to the network of the organization and perform the calculations on several computers. This is called Distributed Computing.

**2**. Meta Computing: Meta computing is a specific type of Distributed Computing where many supercomputers are linked to perform complex calculations.

**3.** Cluster Computing A cluster of computers can be defined as a group of computers that work closely together in so many respects that in many situations, they can be considered as a single computer. The first cluster that came into existence was called Beowulf. It was named Beowulf after the hero who killed a dragon. In this case, the dragon is a metaphor for expensive mainframe and supercomputers. One of the biggest advantages of a cluster is that simply adding new computers to the already existing cluster can expand the cluster. However, adding too many computers may complicate the cluster and it may become difficult to handle. Nowadays, computers of hundred or two hundred computers are very commonly used in big organizations. The main difference between the Grids and Cluster Computing is that the Grid connects computers, which do not fully trust each other or are geographically widespread, whereas a Cluster is made up of computers that fully trust each other and are physically close to one another.

**4.** Internet Computing and cycle scavenging: The greatest example that defines Internet Computing and cycle scavenging in today’s world is the virtual supercomputer SETI@home. This project is based at the University of California. It analyzes data of the Arecibo radio telescope in Puerto Rico, searching for signs of extraterrestrial intelligence. SETI brings together the computing power of more than 3 million computers via the Internet. The important thing about SETI is that it is a background program as it works without impacting the normal working of the computer. All that the user has to do is download the software from the Internet. The software takes care of the rest. SETI also uses the concept of cycle scavenging. This involves scavenging for free time on a computer that is not under our control. SETI@home has inspired other @home applications such as

**5.** Peer to Peer Computing: Peer-to-Peer computing is a popular computing technique popularly used for file sharing in almost any format. A famous example of P2P computing is the website Napster which enabled people around the globe to share their music online. In P2P computing, every user has common software that will interact with other users who have also downloaded the same software. The users specify which files the software will make available for sharing with other users, and thus there is no need for a common server to store the files to be shared. Other example is the torrents. Here, there is a main “seeder”, who shares a file, for example maybe a movie, or a song, or software, etc. Then, there are many “peers” who download the file from the seeder's computer. Now, if we want to download the same file, different parts of the file are downloaded simultaneously from the seeders as well as the peers who have already downloaded or are downloading the same file.

# What does the Grid means to Scientists?

Some of the present Grand Challenge Problems that the grid has been applied to are:

1**.** High Energy Physics: High Energy Physics deals with exploring the properties of fundamental particles in the atom and to analyze the forces between them. High Energy Physics will produce around 10 Peta bytes of data per year, which is equal to 1000 bytes. This data will pertain to information regarding results of high-energy collisions of fundamental particles, and it will be accessed by thousands of physicists around the globe. The Grid is the perfect service for them to conduct their research effectively and speedily, without losing time having to transfer the data onto some central computer.

**2.**Ecosystem simulations: With Global Warming becoming very important and with efforts being made to prevent global warming and to maintain the ozone layer above the Earth's atmosphere, Earth scientists will use satellite images and satellite observations to keep track of the level of the ozone layer and the extent of it's depletion. This task may involve transfer of data of the magnitude of 150 Gigabytes from satellite to Earth. Again, the Grid provides a ready solution for such a problem.

**3.** Biology and DNA decoding**:** The biologists of today are in the hunt to decode the DNA and unlock the secrets of the human body. The number of genes in the human body is enormous. And each gene consists of thousands of DNA. Thus analyzing the DNA and attempting to break the DNA code will obviously involve enormous amounts of observations, results, and data that will be accessed by millions of biologists worldwide. Again, the Grid is a ready solution for such a situation.

**4.** Strong Artificial Intelligence (AI): Artificial Intelligence is one of the hottest research topics of Computer Science Engineers today. Artificial Intelligence is the ability of the computer to behave like the human brain. Computer Science Engineers today have implemented basic models of Artificial Intelligence using concepts such as Fuzzy logics and Neural Networking. Now Computer Science Engineers are researching on Strong Artificial Intelligence, which is the artificial intelligence that matches or even exceeds human intelligence. The Grid is a very useful tool for such a groundbreaking research. Other important Grand Challenge Problems for which the Grid will be of immense use are:

a) Protein Folding

b) Financial modeling

c) Earthquake simulation

d) Climate and weather modeling

e) Molecular Engineering

f) Nuclear power simulation

# Grid projects:

**1.** Test beds Projects: These projects use the existing Grid technology to create working test beds. A test bed is a platform for experimentation large development projects. Some of the major test bed projects at present are

a**) Astro grid:** It is a United Kingdom based test bed project. Its area of development is database management and data mining.

**b) BIRN**: BIRN or Biomedical Informatics Research Network is a test bed project based in USA. It deals with establishing a good Information Technology (IT) based infrastructure to boost Biomedical Research.

**2.** Grid-Tech Projects: Unlike the Test bed projects, Grid-Tech Projects are involved with the development of the Grid itself, whereas the Test bed projects use the Grid as a tool for development in other fields. Grid-Tech Projects are concentrated upon developing the middleware and existing hardware.

**a) BIOGRID**: Biogrid is a Japan based Grid-Tech Project that aims at construction of a supercomputer network.

**b) CROSSGRID**: Cross grid is a large scale Grid-Tech Project that is spread throughout Europe. Research takes place in several institutions across Europe. Cross grid enables to develop programming tools for real-time simulations and visualization in fields such as Earth Sciences, Physics, and Biomedicines.

**3**. @HOME Projects: We have already discussed a @HOME project earlier known as SETI@HOME. @HOME projects involving distributing the vast computing work across all computers distributed over the Internet. @HOME projects are very popular because of the widespread nature. Observations, data and computing can be done across the Internet without any special requirements or superior computing resources. Some @HOME projects other than SETI@ HOME are:

**a) Climateprediction**.net: This Project uses the concept of @HOME computing to research on Global Climate.

**b) Genome@home**: This project strives to use the information gathered by the Human Genome Project to provide significant improvements in the field of medicine. In order to do so, scientists need to develop a level of understanding of the available information.

**4.** Field-specific applications: These projects aim at using the Grid technology to pursue research in specific fields of scientific application A few major Field- specific projects are:**A) Earth System Grid II**: The Earth System Grid or ESG is a climate research project. It uses Grid-Technology in combination with Supercomputers and data analysis servers.

**B) FUSIONGRID**: This project is dedicated to research on the concept of Magnetic Fusion. The above are the major project areas where the Grid technology is extensively used. However, there are other smaller areas, which also use Grid technology.

**1. Grid Portals**: The Grid is mainly used by two types of people – those who develop the Grid and those who use the Grid. Some of the main Grid software developers today are Globus, UNICORE and Condor. But the Grid itself is used by Grid users like scientists, engineers etc. via a Grid portal. Instead of creating a new environment for the Grid, which will make the Grid look like a new application, the Grid Portal allows scientists and engineers to concentrate on their area of work by just making the Grid an extension of their already existing desktop environment.

**2. Grid Outreach initiatives**: Grid developers have tried to widen the range of users of the grid. At present scientists and engineers form the majority of the users. Developers have tried to include educational institutions and students into this user list by including educational websites on the Grid.

Grid Projects in India:Initiatives have also been taken in India to popularize the concept of Grid computing. The biggest of them all is GARUDA, which are India's Grid initiative connecting 17 cities across the country. Educational Institutions are playing a major part in this project with the Indian Institutes of Technology (IIT's) and all C-DAC centers participating in it.

# References:

* [www.gridcomputing.com](http://www.gridcomputing.com)
* Electronics For You magazine
* [www.gridcomputingplanet.com](http://www.gridcomputingplanet.com)
* [http://support.csis.pace.edu/CSISWeb/docs/MSThesis/ZimmererAndreas.pdf](http://support.csis.pace.edu/CSISWeb/docs/MSThesis/ZimmererAndreas.pdf" \t "blank)
* [http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/lhc-interesting-facts.htm](http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/lhc-interesting-facts.htm" \t "blank)
* [http://ewh.ieee.org/r10/madras/newsletter/06/nl-jan06.htm](http://ewh.ieee.org/r10/madras/newsletter/06/nl-jan06.htm" \t "blank)