International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:268
Vol.2, No.6, June 2013 DOI:10.15693/ijaist/2013.v2i6.54-60

Formation and Design Considerations of Grid Architecture

Prof. Y.K.Shah
Amrapali Institute of Technology and Sciences
Haldwani-263139 (UK) - India

Mrs. Mukta Jukaria
Amrapali Institute of Technology and Sciences
Haldwani-263139 (UK) - India

Miss Supriya Shah
Haldwani-263139 (UK) - India

ABSTRACT

Grid computing has enormous impact in the field of scientific research as well as in the realms of application world. The online presence of reliability and ease of flexible operations are the main features. The evolution of distributed computing with the virtualization of various forms of grids is the key to the understanding its utility. The utility management infrastructure consists of diverse nature of numerous virtualization in grid leading to functionality in web services core/hosting, work load and information. The virtualization layers are briefly described here. The levels of virtualization can be explained enormously owing to the never ending demand of consumer world.

INTRODUCTION

To begin with Grid computing was a concept which started to shape into implementations in universities/institutions sometime in early 90's. Now a days it has transformed into reality not only in scientific or academic world but also in commercial ventures. It has now integrated into products of both incipient and well established IT enterprises to enrich a user's experience.

Evolution of Distributed Computing-Grid Computing

In this form, the user views the independent and physically disparate computing devices as one large virtual machine.

With evolution progressing, now the Grid computing supports the concept of virtual dynamic organizations by providing secure and coordinated access and sharing of heterogeneous and geographically distributed resources such as Applications, Data, Processor Power, Network bandwidth, Storage capacity and many others, over a network and across organizational boundaries by utilizing a set of open standards, protocols and rules. Grid computing can be viewed as the function of a large Virtual Machine which works as-

(i) Distributed Computing Model - For Grid computing the infrastructure is geographically distributed or scattered. This would demand some sort of communication among them.

(ii) Network- Any network is actually a physical structure that interconnects the scattered resources. It generally provides the pipes to connect some component with components using some protocols and open standards.

(iii) Open standards and protocols- These provide mechanisms to create communications among components separately and differently developed by various vendors. Thus the vendors or developers program and concentrate entirely on the business logic.

(iv) Shared Resources- The distributed computing infrastructure provides the ability to work by resource sharing like CPU, disks and memory. And the resources can also involve other elements such as networks, databases, storage devices etc.By sharing such resources the user has a homogenous system consisting of heterogeneous elements. So a request can make use of a resource that is outside of the organization to which it belongs. This gives the concept of Virtual Dynamic Organizations as described below.

(v) Virtual Organization- A virtual organization needs to be dynamic so that if the requester needs more resources than the provider can provide, in that ease the grid should search for more resources, and deliver what the requester has demanded by expanding the organization. So for a short time these two parties would adhere to same organization (virtual organization intending to provide a resource to the requester). As a matter of fact a virtual organization has to be dynamic always expanding as per the needs of the requester. This dynamism has to be repetitive also.

(vi) Secure Access- Necessarily the implementation of virtual dynamic organizations requires secure mechanisms to ensure protection of all shared elements.

There must be a mechanism for guarantee or authentication of the service providers and the consumers. The system is also supposed to keep the updated list of users and service providers in Access Control Lists(ACLs).The security has to ensure privacy as well and that information passed on to the network is not altered or modified after delivery.

(vii) Coordinated Access- Coordination is basically related to scheduler for jobs, processes or higher level task processing. Quality of Service(QOS) while distinguishing between regular users and premium users has to be effective in fulfilling business objectives.

Coordinated Access is required to fulfill in many ways. Suppose someone submits a job to the grid which seemingly can be paralyzed. In that case grid may split the job into many sub jobs, process these sub jobs and return a unique consolidated and computed result to the requester. At the same time if the requester prefers to receive the results as those are coming, the grid should allow that. Both of these conditions require coordination effort for a single action. On the whole the grids are expected to coordinate with all the requestors, producers, consumers and donors for smooth working.

Further the grids are expected to satisfy the requirements of the consumers and to utilize the full capabilities of the donors. Thus coordination requires implementation of policies, some metering abilities and optimization of grid's use. Optimization includes turning off machines during non-working hours and also granting higher priority to the premium users.

Various Computing Models of Grid

The grid models can be identified on the basis of donors services and demands generated by requesters.

On the basis of services provided, some grid models can be type fed. The resources can be computing power rendered by servers, the data storing capacity provided by information and data repositories and network bandwidth managed through network infrastructures. A grid is supposed to handle higher value added services from these one or more models. In fact these grids optimize the principles of the basic models, leveraging and combining them in innovative manner for more advanced services.

The Computational Grid-

This is an infrastructure which on demand of work load, provides the computing power. This is suitable for the following applications-

(i)Wherever maximum processing is required during certain periods of time and a single processing machine is unable to provide the same.

(ii) When resource scavanging in desktop machines and under used servers is required.

When users browse an internet, reads an e-mail or creates office documents presentations or word processing files then in that case the idle cycles of CPU can be utilized by Grid. Further this would also enable the client application for more than one project by reusing the infrastructure.

Server Oriented Grids-Practical Application

Consider an example of Petroleum Industry. Geologists gather lot of information in oil fields whose analysis will concentrate on composition, geology and streams of reservoirs in that locality. Here analysis would require computation of high level graphics and complex processing capabilities. Further this is on line processing as the engineers and geologists have to make important real time decisions for the oil company, more and faster processing is sometime more effective which could be had from grid computing. The server grid can provide the capability of submitting the jobs from the field which are executed in the main regional or the local data centers thus improving the activity of the professionals in the oil field.

Let us take another example of the application potential of server grid. In financial sectors managers have to use complex spreadsheets to carry out financial calculations. Many times such types of spreadsheets have to iterate as many times as possible to come in the proximity desired and achievable situations. The result would be more accurate depending upon no of iterations. This process is often carried out for profitability analysis. The profitability analysis and stock evaluations by share holders or loan approvals can be conveniently carried out on grid servers. Though presently these operations are executed on the desktop models of the customer assistants, yet if grid utilized then the result will be returned faster.

On the above lines World Community Grid's mission is to create a very large/computing grid which will benefit the humanity.

Data Grid

There are many components within a data grid infrastructure which enable grid capabilities to the

1) Data

And

2) Information virtualization disciplines

This provides the ability to supply homogenous access to heterogeneous storages of data. This also allows the data consumers to view a unified image of the respective information or data spread across different resources potentially based on various technologies.

Information Infrastructure-

The information is meant to be useful data to the user. Information grid integrates the heterogeneous sources of information in a comprehensive manner. This also allows applications and users to view data base as a singular entity hiding the complex manipulation of various databases. This information spread across them is published by the grid as if it was total centralized and accessed together.

This grid must also make connections to the final heterogeneous and disparate databases. This is desirable that replication and caching mechanism be inbuilt in the IT infrastructure to make it more efficient and smooth manageable.

File System and Block Data Infrastructure

The information infrastructure provides unified access to information but the approach of access by file system and block data structure in not unified to files or data blocks.

Data is the unit of work of operating systems and hardware devices. However blocks of data or files are often of no interest to the end users who would like to know the end results only. There are standard mechanisms (like SAN-FS) which view the storage mechanism (components from different vendors as a single disk in a storage network.

Network Grid

In most of the corporate network system the computers are often connected permanently and a small portion of bandwidth is always mutilized. As a matter of fact every machine, servers and desktops have underutilized network bandwidth and so this can be considered as idle sources, while many times a given user machine may require more resources than it has and so a bottleneck is reached.

A network grid has some more efficient mechanisms to provide services similar to peer to peer grid. An IBM download grid has been developed reducing the trans-atlantic network traffic.

Multipurpose Grid

This would be probably the most important application in the futuristic grid computing catering to the user demands. The infrastructure of such a grid has to be adaptive in nature with economy associated with it. Such applications could be termed as Meta Grid.

These meta grids will be capable of routing the requests to most suitable and efficient grid that would provide the right model to carry out the objective goals optimally.

IT components relationship with Grid

The most common components (for IT) are

(i)Technologies
(ii)Concepts
(iii)Paradigms

Any grid has to be related to them in some way or the other, leveraged conversely.

Grid and deep computing

Deep computing or High Performance computing (HPC) is a technology which uses clusters of machines for executing high demanding tasks. The implementation of deep computing cluster is done by connecting these machines with high speed network formations. The jobs which cluster runs is divided and then parallelized running parallel in the cluster nodes. The deep computer structure can be part of a grid and a grid can be the mechanism to provide deep computing resources.

Grid and on demand computing

This is based on the concept that market will demand flexibility in enterprises business and applications. For this the flexible IT operating environment would be needed for adaptable and resilient processes. This means that virtualized operating environment based on open standards have to be incorporated. On demand computing is the fundamental component of grid to achieve the highest degree of virtualization and the grid is also leveraged when open standards are also met fully. The autonomic abilities of the component is also a significant feature of grid computing.

Grid and Utility Management Infrastructure (UMI)-

Since the business and IT infrastructure are of varying nature in an on demand enterprise there should be new ways to make these profitable and feasible.UMI is a model or specification of how the IT services can be acquired and paid as these were a service given by the traditional utility. Utilities imply the “pay per usage” payment method.UMI conceptually decides which component a utility should have to provide IT services and its payment per usage. This would actually result in variable costs for the consumers. This would also imply for the IT facility provider the reuse and sharing of their infrastructure facilities.

Grid computing provides reasonable option to use an IT utility but it is not the only choice. The grid computing makes available provisioning, metering, billing and also provides mechanism to utilize those resources which are currently not in use. Grid also manages sharing of portions of resources and security. These are all essentially required qualities to implement an infrastructure along with UMI in general.

Virtualization in Grid

Here the virtualization is considered to be the ability to provide a unified vision of a set of resources disparately distributed. These geographically distributed resources developed by various vendors may run based on different technologies .Actually virtualization is a very important element on the demand operating environment and is essentially required for grid computation.

Provisioning in Grid

The provisioning mechanism in grid provides (on demand) additional resources to the grid. Alternately grid can provide additional resources to the grid. It is also possible that both systems can coexist in the same infrastructure and can also be leveraged.

Categories and Functions-

The grid computing has to be analyzed according to cross criteria, categories and functions. These categories are related to the layers in a grid stack as shown in the fig:-
These categories help the user relate products with their role and scope within a grid.

Categorization of Components-

The layer structure of the grid stack is as per IT industry traditions

1.Application
2.Content Management
3.Billing & Metering
4.Orchestration and Provisioning
5.Task Scheduling
6.Workload Management
7.Systems Management
8.Software Licensing
9.Job Scheduling
10.Structured Data Virtualization
11.File And Block Data Virtualization
12.Grid Middleware

Fig 1: - Grid Stack Products Portfolio

Functionality on Grid Computing-

These are the attributes an IT architect will look for to understand the main feature a product brings it to the grid. This classification presents four elements-

(1) Web Service Core/Hosting

(2) Work Load Virtualization

(3) Information Virtualization

(4) Other Disciplines

There may be other disciplines like Security, Provisioning, Billing and Metering

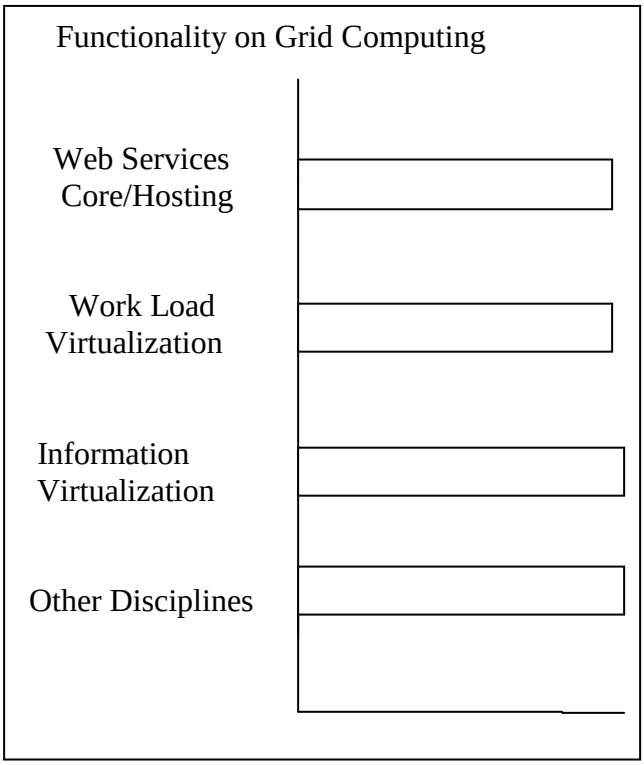

Fig 2:-Product Evaluation Criteria

The different functionality on Grid Computing can be elaborated as below:-

Web Services Core/Hosting-

All applications provide services to requesters and consume services from the grid. The environment is intended to facilitate the life cycle of the grid enabled applications, in other way, how those are developed, run and maintained etc. The examples are monitoring tools, security features and APIs for accessing lower layers of toolkits(grid stack).These toolkits must provide a common set of utilities to improve the development time experience. For instance a toolkit could provide set of tools to build applications that can be parallelized in an easy way, either to be further deployed on desktop grid or on a server grid, or high level interfaces to handle grid resources and submit jobs to them. There are other tools which include components to split the unit of works and results consolidation.

Work load Virtualization

The main factor in workload virtualization is the ability to execute a unit of work at the appropriate moment under some specified schemes. The meta scheduling is defined at a single machine level (within a grid) and at grid level consisting of many computation pools. Normally the unit of work and terminology are defined according to the environment of application.

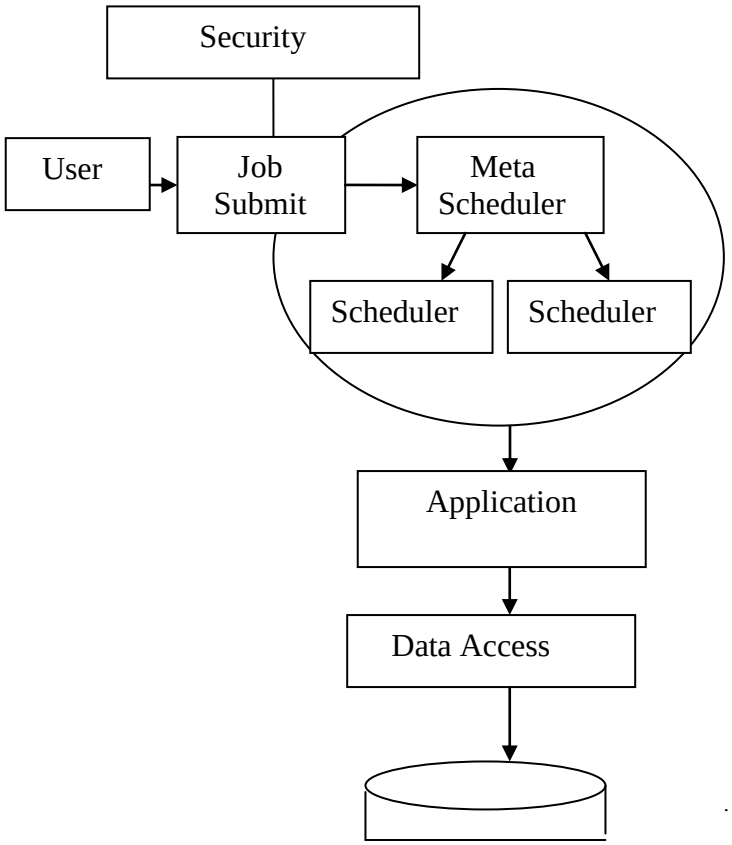

Fig 3:- Focus of scheduling environment

The main function of the scheduler is to ascertain that activity will be carried out over a certain period of time. The scheduler would consider the workload of the resources which are supposed to run some specified activities. Accordingly only those machines would be loaded which are already not occupied. And if the machine accepts that workload then it would simultaneously delay the execution of the current and new activities having lower priority assigned by the scheduler.

Information Virtualization

This covers very large and different types of data infrastructure (for grid) and the wide variety of products available to implement them. This would enable the grid to have the unified version of vast repositories of data. The virtualization is used here in the sense of aggregating distinct entities to become one virtual entity. Other subdivisions of layers can be

defined depending on the architectural blueprint been used. The virtualization can be implemented at the-

(1)Block Data
(2)File
Or
(3) Information Levels

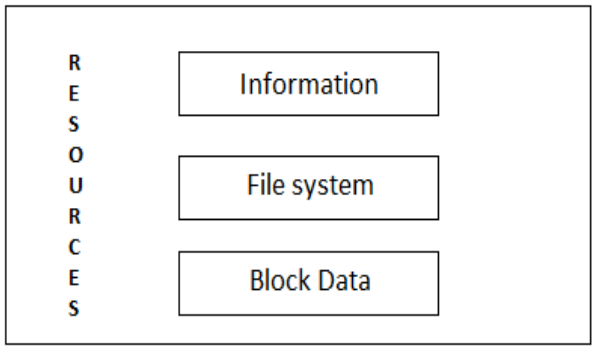

Fig 4:- Virtualization Layers

The block data virtualization provides the unified access directly either from hardware or from a driver of a given hardware that would run on a given operating system.

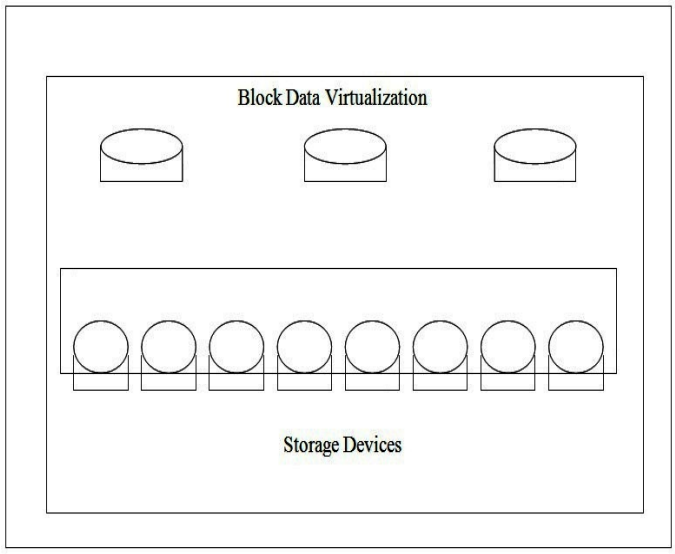

Fig 5:- Block Data Virtualization Layer

The utilization of this data is actually the operating system itself. In this case the unit of work is raw block of data of a device storing this. This enables the operating system to view physical resources that are not in the same device as a single logical volume or file system on top of the services provided by the lower layers. These devices create a logical low level structure on top of the physical structure of the storage and make is available for publishing to the operating for further use.

File virtualization level is concerned with file systems. Here the unit of work is file just like a spread sheet or a word processing document, a file document. However the practical applications demand a combination of files for the mapping of data into meaningful and useful information purposes.

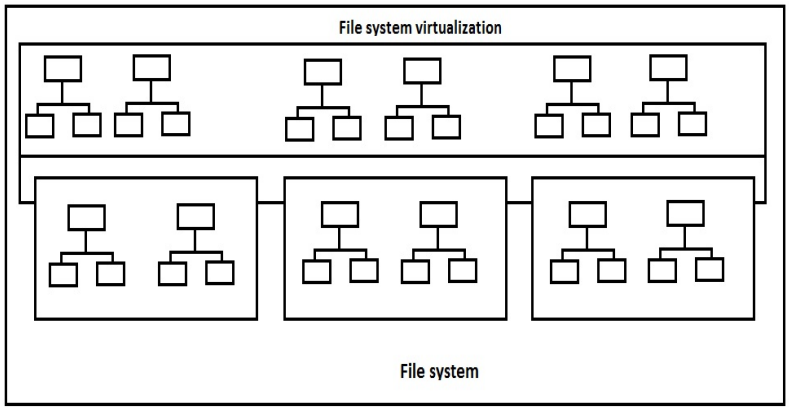

Fig 6:-File Layer Virtualization

The highest level of the information virtualization discipline is achieved in the Information Layer. This layer also known as structured data layer provides the integration
From heterogeneous resources with heterogeneous formats by accessing them using heterogeneous protocols and publishing them to the applications in a homogenous manner is done for enabling the applications or consumers of this information to have a unified vision of the information storages.

Information Virtualization Layer

Universally the information term is referred to the meaningful data significant for the end users. This layer provides an abstraction of disparate and distributed information sources, such as database management system(DBMS), flat files (e.g. comma separated files),structured files(e.g. XML documents),or a Content Management System.

This virtualization layer also refers to the ability to federate or integrate data and information from distributed heterogeneous resources into a unified repository. The whole concept behind this approach is to get a single view of the information.

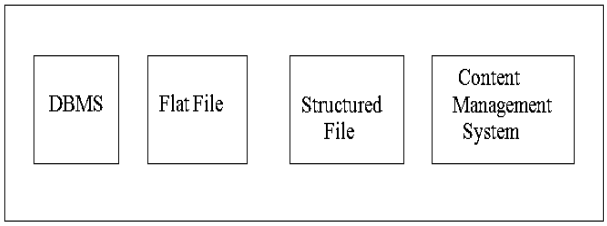

Fig 7:- Information Visualization Layer

Here it should be noted that some parallel databases are usually regarded as information virtualization tools, but generally, these products only support their own branded engines and as such cannot be considered grid enabled products.

Conclusion-

The foregone discussion presents the lucid coverage of the most significant aspects of grid computing. Essentially the various components utilized for practical applications are covered. Various aspects of grid functionality have been touched briefly. From the above it is natural to assume that information virtualization software provides virtualization in the two lower levels as well that file system virtualization software provides data block virtualization. Nevertheless, this is not valid, as it can be seen in most distributed database implementations.

REFERENCES

[1] Davies, Antony (June 2004). "Computational Intermediation and the Evolution of computation as a Commodity" "(http://www.business.duq.educ/faculty/davies/research/Economics of Computation/ pdf)" (pdf). Applied Economics.
[2] Foster, Ian; Carl Kesselman. "The Grid: Blueprint for a New Computing Infrastructure" (http://www.mkp.com/grids/). Morgan Kaufmann Publishers. ISBN 1-55860475-8.
[3] Plaszczak, Pawel; Rich Wellner, Jr. Grid Computing "The Savvy Manager's Guide"(http://savvygrid.com/). Morgan Kaufmann Publishers. ISBN 0-12-742503-9.
[4] Berman, Fran; Anthony J. G. Hey, Geoffrey C. Fox. "Grid Computing: Making The Global Infrastructure a Reality" (http:/www.grid2002.org). Viley. ISBN 0-470-85319-0.
[5] Li, Maozhen; Mark A. Baker. "The Grid: Core Technologies." (http://www.acm.org/pubs/cacm/)". Wiley. ISBN 0-470-09417-6.
[6] Catlett, Charlie; Larry Smarr (June 1992). "Metacomputing
[7] Smith, Roger (2205). "Grid Computing: A Brief Technology" Analysis (http://www.ctonet.org/documents/GridComputing_analysis.pdf). CTO Network Library.
[8] Buyya, Rajkumar (July 2005). "Grid Computing: Making the Global Cyberinfra structure for eScience a Reality" ((http://www.gridbus.org/~raj/papers/CSI Communications July 2005. pdf)". CSI Communication 29 (1). ISSN 0970-647X.
[9] Berstis, Viktors. "Fundamentals of Grid Computing" (http://www.redbooks.ibm.com/abstracts/redp3613.html). IBM.
[10] Ferreira, Luis; et.al.."Grid Computing Products and Services" (http://www.redbooks.ibm.com/abstracts/sg246650.html). IBM.
[11] Father of the Grid (http://magazine.uchicago.edu/0404/features/index.shtml.(http://magazine.uchicago.edu/0404/features/index.shtml.
[12]What is the Grid? A Three Point Checklist (http://www.fp.mcs.anl.gov/~foster/Articles/WhatIs The Grid.pdf) (pdf).
[13]IBM Solutions Grid for Business partners: Helping IBM Business Partners to Grid-enalbe applications for the next phase of e-business on demand (http://www304.ibm.com/jct09002c/isv/marketing/emerging/grid_wp.pdf)
[14]A Gentle Introduction to Grid Computing and Technologies (http://www.buyya.com/papers/Gridintro-CSI2005.pdf) (pdf). Retrieved on 2005-05-06.
[15]The Grid Café'-What is Grid?\ (http://gridcafe.web.cern.ch/gridcafe/whatisgrid/whatis.html). CERN. Retrieved on 2005-02-04.

AUTHORS PROFILE

(1) Mr. Y.K. Shah is Professor of ECE in AITS Haldwani, presently. He has more than 43 published articles on Computer & Microprocessor based applications. He has served on various national committees on computers including CDAC, DOE & SAIL. He has vast experience of working on Main frame computers at U.S.A. and Europe.

(2) Ms. Mukta Jukaria was HOD of Dept. of ECE at AITS, Haldwani. She's is pursuing her Ph.D in ECE.

(3) Ms. Supriya Shahis a computer engineer and is presently pursuing higher studies in the same field after doing M.A. (Public Administration) and 'B' Level course of DOEACC.

Davies, Antony (June 2004). Computational Intermediat

Plaszczak, Pawel; Rich Wellner, Jr. Grid Computing "The

Berman, Fran; Anthony J. G. Hey, Geoffrey C. Fox. Grid

Li, Maozhen; Mark A. Baker. The Grid: Core Technologi

Smith, Roger (2205). Grid Computing: A Brief Technolog

Buyya, Rajkumar (July 2005). "Grid Computing: Making

Berstis, Viktors. Fundamentals of Grid Computing (http:/

Ferreira, Luis; et.al..Grid Computing Products and Servic

Father of the Grid