Implementing Grid Computing – Optimizing Resources to Meet Business Requirements

Mrs. Reshma Desai
Lecturer, Computer Science Dept. Thakur College of Science and Commerce, Mumbai – 400 101
E-Mail: resh_d@hotmail.com

ABSTRACT
Grid computing is a form of distributed computing that involves coordinating and sharing computing, application, data, storage, or network resources across dynamic and geographically dispersed organizations, allowing companies to benefit on account of reduced costs and increased efficiencies via grid infrastructure, without being locked in to a non-growing system. Grid technology also provides the ability to store, share and analyze large volumes of data, ensuring that people have access to information at the right time, which can improve decision making, employee productivity and collaboration. In a grid environment, resources are virtualized to create a pool of assets. Workload is spread across servers and data can be seamlessly retrieved. Infrastructures can now dynamically adapt to business requirements, instead of the other way around.

By the late 1990s, a more generalized framework for accessing high-performance computing systems and distributed data (e.g. Globus3) began to emerge, and during the start of this decade, the change quickened with growing recognition of the synergies between grid and the emerging Service Oriented Architectures.

The Global Grid Forum (GGF) is involved in the definition of a number of important standards for scheduling and workload management, application deployment, resource provisioning, data movement, and data access. Most notable is a broad architectural focus called the Open Grid Services Architecture (OGSA) which defines a very rich vision of execution, management and data services to enable the creation of grids.

One of the basic uses of grid computing is to run an existing application on a different machine. The machine on which the application is normally run might be unusually busy due to a peak in activity. The job in question could be run on an idle machine elsewhere on the grid. There are at least two prerequisites for this scenario. Application must be executable remotely and without undue overhead. Remote machine must meet any special hardware, software, or resource requirements imposed by the application.

Another benefit of a grid is to better balance resource utilization. Some grid implementations can migrate partially completed jobs. In general, a grid can provide a consistent way to balance the loads on a wider federation of resources. This applies to CPU, storage, and any other types of resources that may be available on a grid along with industry-specific value propositions. This paper will try to throw light on available dynamic infrastructure for grid computing technology, which can be deployed to allow capturing and analyzing customer information to increase the speed and accuracy of business decisions.

KEYWORDS

INTRODUCTION
What is grid computing?
Grid computing can be seen as a distributed computing model that 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 others, over a network and across organizational boundaries, using a set of open standards and protocols. Grid users see these resources as a large virtual computer.

The simplest way to think of grid computing is as the virtualization and pooling of IT resources, such as compute power, storage and network capacity, into a single set of shared services that can be provisioned or distributed and re-distributed as needed. Just as an electric utility deals with wide variations in power demands without affecting customer service levels, grid computing provides a level of control and adaptability to IT resources that can respond to changing computing workloads while being transparent to end users. Grid computing operates on these basic technology principles:

1) Standardize hardware and software components to reduce incompatibility and simplify configuration and deployment.
2) Virtualize IT resources by pooling hardware and software into shared virtual resources.
3) Automate systems management, including resource provisioning and monitoring.

Grid computing can be used to optimize the existing resources by giving businesses the ability to respond to volatile business needs at high speed -businesses today operate in an unpredictable, global environment. Predicting business demands, competitive threats, supply chain risks and regulatory requirements are increasingly challenging. An underlying grid infrastructure provides IT with the agility to do so.

Grid computing models
There are also some fundamental grid models based on the type of basic services provided. Resources can basically be
computing power, provided by servers or individual computers, data storage capacity, provided by information and data repositories, or network bandwidth, provided by networked infrastructures.

- **Computational grid**
  A computational grid is an infrastructure that allows resources to donate computing power to the grid whenever the workload demands. This infrastructure is suitable for applications that demand as much processing power as possible or additional processing power during certain periods of time.

- **Data grid**
  Within a data grid infrastructure are the components used to provide grid capabilities to the data and information virtualization disciplines. It provides the ability to supply homogeneous access to heterogeneous repositories of data. It allows the data consumers to see an unified image of the respective information or data spread across different resources, potentially based on different technologies.

- **Network grid**
  In a typical corporate network, computers are very often permanently connected to it while using only a portion of its bandwidth. Every machine, servers and desktops, has underused network bandwidth, which can be considered as an idle resource. And be used as resource for network grid.

- **Multipurpose grid**
  The multipurpose grid may be perhaps the more common implementation in the future of grid computing. The infrastructure of this grid should be adaptive enough to provide any of the grid models. It could be implemented as well as a meta-grid with abilities to route the requests to the grid that supports the right model to fulfill them.

- **Grid and on demand**
  On demand business is requirement of the IT industry and market. It acknowledges that the flexibility of the markets will demand flexibility in enterprises’ business processes. Flexible, adaptable, and resilient business processes require a flexible IT operating environment. The operating environment has to be virtualized; it should be based on open standards to be integrable and must have autonomic capabilities. Grid computing leverages on demand because it is a fundamental component to achieve the highest degree of virtualization, one of the key factors of the on demand operating environment.

- **Grid and virtualization**
  Virtualization is the ability to provide a unified vision of a set of resources. These resources may be geographically distributed, run on different technologies, and developed by different vendors. Virtualization is one of the key elements of the on demand operating environment, and essential on grid implementations.

### Architecture of grid computing components:[4]

- Applications
- Content Management
- Task Scheduling
- Workload Management
- Systems Management
- Job Scheduling
- Structured Data Virtualization
- File and Block Data Virtualization
- Grid Middleware

1) **Applications**
   This layer describes an environment for developed applications to take advantage of the whole set of features provided by the grid. Applications can make use of all layers in the stack, through classes, APIs, frameworks, toolkits, or Software Development Kit (SDKs). The application layer is also associated with the applications developed for a specific industry.

2) **Content management**
   Content management provides another degree of virtualization to data and information within a grid. It is related to the handling of digital media, like video, audio, images, or streams. These are not structured data, and need to be treated in a different way. It usually requires specific features, such as searching an image database for the most similar picture to the one given in input.

3) **Task scheduling**
   Task scheduling provides an environment to run small pieces of execution that, combined, create the unit of work that the requester needs to execute. The decomposition of this unit of work into smaller pieces can consist of parallel, short running tasks or of tasks that have to be run in an special order within an automated workflow.

4) **Workload management**
   Workload management is a mechanism designed to balance the workload among different resources. It is defined during the setup and used at run time. The workload management can be dynamic depending on the real time workload of the involved components and the policies applied to them.

5) **System management**
   Management is vital in every IT infrastructure. Distributed computing environments require even more sophisticated tools and policies.

6) **Job scheduling**
   Job scheduling is used to optimize the execution of jobs onto a grid. In this context, a job can be understood as the computational work needed to deliver a meaningful result within the context of an application. Its execution is governed by policies that rule the basic scheduling and prioritizing of jobs to be run.

7) **Structured data virtualization**
   Structured data virtualization provides a unified access to heterogeneous databases as though they were a large and single
Implementing Grid Computing – Optimizing Resources to Meet Business Requirements

Database. It supplies grid-based access to structured data sources of information, like databases, documental databases, XML, or flat files with certain structure and objects. This is a higher level of data virtualization. A grid can create an unified vision of the different repositories of information. The consumers of this service should see a single database that in fact is composed of aggregations and associations of other databases.

8) File and block data virtualization layer
File and block data virtualization mechanisms are related to the lowest layers of data services provided by the grid. They can live at the hardware level, associated with blocks of data managed by storage devices. File data virtualization is provided by file system implementations.

9) Grid middleware
The grid middleware provides the integration among the grid components. It is the keystone software that provides basic services to grid components. These basic services are integration mechanisms based on standards, description, and creation of services. The middleware also includes the elements that allow the grid to interact with its own components.

Case Studies and application of Grid Computing in various fields leading to Optimizing the available Resources for improved Business.
Enterprise grid computing is emerging information technology (IT) architecture that delivers more flexible, resilient and lower cost enterprise information systems. With grid computing, groups of independent, modular hardware and software components can be pooled and provisioned on demand to meet the changing needs of businesses.

1) Data Center Modernization Using Grid Technologies

Data Center Modernization Using Grid Technologies
Grid computing is an IT architecture and methodology comprised of both technology and best practices. Not every IT department will adopt every grid computing technology or technique. However, many IT departments are successfully using specific Oracle grid technologies and best practices with dramatic benefits.

Consolidation of IT resources such as servers, storage, applications and data centers can provide dramatic cost and energy savings. Forrester Research [4] estimates that average server utilization in data centers today is only about 30%. With hundreds or thousands of servers around the enterprise, the inefficiency is staggering. While application usage varies greatly by certain times of the day or year, it is also being seen as an opportunity to apply grid techniques for a combination of better management, utilization and overall efficiency.

Two key grid computing technologies, server virtualization and clustering, enable sharing of IT resources and consolidation of servers, storage and entire data centers.

Many customers are now turning to virtual machines (VMs) to consolidate multiple applications on to a smaller number of shared, centrally managed servers. A virtual machine, or virtual server, is software that simulates the operations of computer hardware, enabling an application to run on the virtual machine just as it would on a physical computer.

Oracle’s server virtualization product, Oracle VM, provides a highly efficient way to run multiple Oracle and non-Oracle databases, middleware and application environments in a single server running Oracle Enterprise Linux. Oracle VM enables IT to quickly add or release more server resources for spikes or lulls in demand, increasing utilization and reducing energy costs. Grid computing requires agile and efficient systems management. Managing a grid requires a new breed of systems management, Server virtualization and clustering are able to simplify and abstract away the underlying complexity of the software and hardware infrastructure. Grid systems management software must be able to comprehend the underlying complexity and configure and modify the infrastructure to meet dynamic business needs. Oracle Enterprise Manager provides top-down, end-to-end application management with broad coverage across Oracle databases, middleware and applications.

A good example of server provisioning has been implemented by GasNatural2, an international oil and gas utility based in Spain. They have deployed a pool of Oracle RAC cluster nodes that power their data warehouses, their business intelligence applications, electricity market applications and other internal applications. They are also able to provision RAC cluster nodes so that they can serve as production, development or test servers. The result is:

- A significant savings in hardware costs from their consolidated data warehouse;
- Tremendous performance increase in their BI applications;
- Low cost high-availability in remote locations;
- A scalable, stable environment for both transactional and business intelligence applications.

Grid computing leverages clustering and virtualization technologies at all layers—middleware, database and storage—to deliver predictable high performance and scalability for applications, especially customer-facing, cloud computing and SOA[4]. Grid computing enables continuous availability with replication, automatic failover, server failover and disaster protection.

Server failover has been available for many years from both hardware and software manufacturers. The protections that are afforded from a successful failover of a server are often critically important. Yet from a business standpoint, the drawback of setting up a standby server is the cost of hardware and software that are only used when disaster strikes. Grid computing enables standby resources to be used as active resources, resulting in higher utilization. Another consideration of server failover is that of failover time. Some applications are so critical to the business that they can not afford to be down even for a few minutes, while others can tolerate some interruption of operation.
Oracle has pioneered many techniques of server failover that provide IT departments with automatic failover capabilities for several server types. For example, Oracle Database RAC, Oracle WebLogic Server, Oracle Tuxedo and Oracle Coherence clusters can withstand failures of several servers within a cluster and still remain in operation. IT departments can simply remove failed servers from service and repair or replace them and add them back to the server grid. [4]

Load balancing, work load management and overload protection, and automatic migration and failover of services or whole servers ensure applications stay up and running.

Disaster Protection
Even clusters cannot survive a complete data center failure from natural disasters, fires, floods and so on. In these cases, failover to a remote location is required. An enterprise grid can be designed to encompass multiple locations, dynamically shifting workloads across those locations for the highest reliability.

Oracle Active Data Guard provides the ability to create up-to-date replicas of the production Oracle Database for standby and disaster recovery. In addition, these replicas can also be used for resource-intensive read-only operations such as queries, reporting or backup. The data center projects such as server consolidation, SOA development, space and power optimization and large-scale implementations of rack-mounted Linux servers can be interrelated and can be described as a grid computing approach to data center modernization.

The benefits derived from these techniques are compounded. This in turn enables IT departments to save energy, reduce systems management costs, and get a better return on their hardware investments. The widespread adoption of open standards, IT resource virtualization, on-demand provisioning, highly automated systems management and real-time monitoring has created a new generation of data center best practices. Oracle Database, Oracle Fusion Middleware and Oracle Enterprise Manager are designed with this next generation data center in mind.

Inference
By using these grid computing techniques and these Oracle products, IT professionals will find that the data centers they are building for the next decade of business challenges can also provide immediate benefits in terms of cost savings, sustainability and operational agility.

Case Study of Implementation of e-LEARNING Framework using Grid Technologies
Implementation of a network grid supporting an e-learning infrastructure that embraces many of the requirements for exchanging information in the educational and research fields.

This study presents a grid environment to support many educational and research requirements for exchanging information. Knowing the main ways that education can benefit from a grid, we can draw up the basic technological needs associated with the development of e-learning. The case study of the e-learning infrastructure is based on Access Grid. [5]

Access Grid (AG) is an ensemble of resources including multimedia large-format displays, presentation and interactive environments, and interfaces to grid middleware and to visualization environments. The AG technology was developed by the Futures Laboratory at Argonne National Laboratory and is deployed by the NCSA PACI Alliance.

Formal education is undeniably one of the most important aspects of social development: But the growing need, newer technologies can leverage educational activities are as follows:
- Creating virtual classrooms by interconnecting lecturers to geographically scattered students
- Making educational material, such as tutorials and recorded lectures, available worldwide through high-storage infrastructures
- Digitalizing and making books available through high-storage infrastructures
- Integrating library search engines and digital content

How grid technologies can be deployed to develop a worldwide learning framework or e-learning framework.

Case analysis
Though satellite broadcasting and database technologies can handle virtually unlimited amounts of data efficiently, Even then the advantage of grid technologies is that it relies on non-specialized existing computing resources, can provide a cost-effective solution for above mentioned framework. Certain functional and non-functional technical requirements for designing the e-learning framework-
- Simplex video broadcasting: When broadcasting a lecture, transmitting a high-quality image of the lecturer along with a blackboard is generally a desired requirement
- Duplex audio broadcasting: In contrast to the video broadcasting philosophy, audio information has to be transmitted in both directions so that an efficient contact can be established between the parties (lecturer and audience).
- Scheduling of live lectures: This system must offer users a comprehensive framework for scheduling live lectures. Interface for submitting new lectures, and a scheduler that best fits the lectures submission requirements with the availability of resources (broadcasting rooms).
- Storage capacity: “Off-line” content also plays a major role in e-learning; this includes pre-recorded video and/or audio lectures, tutorials, articles, books, and so on.
- Unified access to stored content: The system must provide users with a comprehensive and unified listing of all the content stored in its data-bases.
- Security (authentication): The system must provide facilities for authorizing users to access only the contents that they are eligible to.
Implementing Grid Computing – Optimizing Resources to Meet Business Requirements

- **Performance**: The system must provide good audio and video quality for the live lectures, and good network throughput for downloading off-line content.
- **Scalability**: The system has to be scalable to a potentially unlimited number of parties.
- **User-friendly operation interface**: The system must be easily operated so that non-specialized users, from lecturers to students, can access it without extra help.

**How will the process operate?** [5]

1) People required to run this system:
   - Professor (one who would give lectures), Student (they can be anyone interested in getting education from the system), Operator (this will be the person responsible for assisting professors and students during the presentation of a lecture as they prepare the environment for the broadcast (sending and receiving). Administrator: This is the person responsible for performing the basic management tasks in the system. This role is typically personified by a system analyst or a network analyst.

   The over all architecture [5]

2) The software component architecture used in this framework
   - Physical Infrastructure
   - Operating System
   - Basic Communication and Storage Grid Services
   - Collaborative Applications
   - OGSA

   **The Physical Interconnection Architecture**
   - Classroom: A lecture “virtual room” with equipment which includes a computer to which a projector and a microphone (or a video-camera) are connected, and these are managed by a specialized operator.
   - Individual site: This denotes the places from which the students may join the conference using personal computing resources.
   - Broadcasting room: This is the place from which the professor broadcasts the lecture to all the students sitting in classrooms or individual sites.

   - Grid portal: This is the portal by which users perform activities such as submitting lectures, subscribing to lectures, and uploading and downloading off-line content.

   **Physical layer / Operating System**: This is the layer that comprises all the computers used for building up the e-learning framework as well as the basic operating system they run. All the grid layer should be multi-platform so that computing resources of every sort can join the grid. In this sense, virtually every type of computer – running any OS is supported. In this specific implementation, Intel®-based computers running RedHat Linux are preferred.

   **OGSA**: This is the layer where the basic grid platform sits. This software is responsible for providing the grid infrastructure upon which the basic grid services, for high performance storage and communication, are implemented. Thus, it offers the basic tools for open standard communication and storage throughout the grid.

   **Basic Communication and Storage Grid Services**: These are the grid services that implement the high-level storage and communications functionality. Thus, they offer the collaborative applications a standard interface for storing content and receiving and/or sending streaming audio and video signals. In most implementations where they are required, they have to be developed for fulfilling specific needs. In this case, custom code was integrated with products like IBM DB2 and IBM DB2 Content Manager to deliver high-quality storage and communication services.

   **Collaborative Applications**: These are the applications that manage the audio and video signals and all the other collaborative tools that are used in a lecture, like virtual blackboards, remote presentation engines, and so on. Such applications are connected to the grid by the software layers underneath. Applications IBM Lotus Learning Management System (LMS), Distributed Power Point and Remote Power Point, Peer-to-Group Media, Broadcast or Kontiki. There are a number of collaborative applications that might be set in place depending of the requirements of the specific lectures to be presented. The products listed are for managing off-line content, live presentations & collaborative virtual conferences.

   **Grid portal**
   This is the portal by which users perform activities such as submitting lectures, subscribing to lectures and uploading and downloading of off-line content.

   Grid Portal IBM WebSphere Application Server ,BEA WebLogic Server Grid System Gateway ,JBoss ,Tomcat Grid System Gateway was chosen due to the reliability and technical support availability requirements.

   The case study has used the Open Grid Services Architecture - Data Access Integration (OGSA-DAI). OGSA-DAI is a project developed by UK Database Task Force, whose objective is to provide a standard interface for a distributed query processing
system to access data in different data sources for the integration between the independent tools. The above framework has been already implemented over 4 educational institutions with over 100 users and 2 mbps of average bandwidth for broadcasting lectures as well as 5 Tera bytes of storage space for offline educational content [5]

**INFERENC E**
The above case study gives presented an example of how grid technologies can be used to build an e-learning framework able to connect a potentially unbounded number of professors and students. In building such framework grid technologies offers a much more cost-effective solution, as it employs existing low-cost and non-specialized computing resources for the job.

**CONCLUSION (overall)**
Grid computing technologies are being applied to address IT challenges faced by modern enterprises. Enterprises are reviewing their existing IT strategies based on the needs of their business. Business drivers are those key initiatives that help shape a company’s strategy. Three key business drivers pervasive across industries and geographies are: Accelerate business processes, Increase productivity and assist Collaboration. Grid computing can find solution which can be designed and implemented to provide affordable and flexible IT solutions that can meet the requirements imposed by a constantly changing environment, for Enterprise resources optimization solutions.

**FUTURE SCOPE**
In summary, we can say that there are not yet a wide range of products specifically designed for security in grids. However, most traditional IT products may be integrated into a grid system. That is because a grid can be seen as a new way of using traditional IT components. Therefore, the ability to integrate traditional IT components with specifically designed components for grid is associated with those products capabilities. A grid surely will require sophisticated security mechanisms due to its intrinsic complexity and distribution of resources. At this point, a security framework and other mechanisms are vital for the future of the grid.

**REFERENCES**
[13] CSI Communications July 2005
[16] Glenn Deen, Senior Software Engineer, IBM, James Kaufman, Intermediate Research Staff Member IBM Tobin Lehman, Research Staff Member, IBM, John Thomas Research Staff Member IBM – A practical introduction to Optimal Grid Skill Level.