



Qualities of Grid Computing that can last for Ages

ASAGBA, PRINCE OGHENEKARO; OGHENEOVO, EDWARD E. CPN, NCS.

*Department of Computer Science, University of Port Harcourt, Port Harcourt, Nigeria.
pasagba@yahoo.com, edward_ogheneovo@yahoo.com. 08056023566*

ABSTRACT: Grid computing has emerged as an important new field, distinguished from conventional distributed computing based on its abilities on large-scale resource sharing and services. And it will even become more popular because of the benefits it can offer over the traditional supercomputers, and other forms of distributed computing. This paper examines these benefits and also discusses why grid computing will continue to enjoy greater popularity and patronage in many years to come. Finally, we discussed about virtual organization (VO) as one of the key characteristics of Grid computing. @ JASEM

Grid computing has recently enjoyed an increase in popularity as a distributed computing architecture that is becoming highly suitable for corporate computing. The reasons for this increase in the past decade or so are not far fetched. A limited IT budget resulting in a lean economy has forced enterprises to more fully utilize their existing computing assets. This has also led them to become more flexible to respond to rapidly evolving markets by being able to intelligently allocate finite resources to the appropriate business applications (Akinyemi, et al, 2007). Grid computing is a special type of parallel computing, which relies on complete computer (with on-board) CPU, storage, power supply, network interface, etc. connected to a network (private, public or the internet) by a conventional network interface, such as Ethernet. This is in contrast to the traditional notion of supercomputer, which has many processors connected by a local high-speed computer bus (Berman, et al, 2003). Grid computing is useful in areas that requires enormous computing power and challenges such as financial modeling, earthquake simulation, and climate (or weather) modeling as well as drug discovery, seismic analysis, economic forecasting, e-commerce and Web services.

From the Search for Extraterrestrial Intelligent (SETI) to the Search for ways to utilize the unused computing power across the enterprise, grid computing has come of age. It promises to harness the spare clock cycle of all your computers and use this new-found power to speed up the most complex of your computational or data processing demands. It also gives access to all the storage, all the data, of all those PCs and networked systems working on any processor-intensive task (Davey, 2003), and (Foster, et al, 1999) It enable sharing, selection, and aggregation of geographically distributed autonomous resources, such as computers (PCs, servers, clusters, supercomputers), databases, and scientific instruments, for solving large-scale problems in science, engineering, and commerce. It leverage existing IT infrastructure to optimize computer resources and manage data and computing workloads (Buyya, 2007). Grid computing has provided avenue of accessing information technology resources optimally.

Grids are collections of networked computers that pool their resources together in such a way that users may utilize processing, storage, software, and data resources from any of the interconnected computers, leading to greater resource sharing and higher utilization ratio. Such grids can have many different definitions and objectives and may exhibit many different properties. One of the key characteristics of a grid is that it allows organizations to pool computing resources (processors, storage, information applications etc.) to enable users to benefit from a potentially far larger pool of resources than would otherwise have been available to them (Carpenter, et al, 2004). The European Union has been a major proponent of Grid computing and many projects have been funded through the framework programme of the European Commission (Wikipedia, 2008D).

Advances in networking technology and computational infrastructure make it possible to construct large-scale high-performance distributed computing environments, or computational grids that provide dependable, consistent, and pervasive access to high-end computational resources. These environments have the potential to change fundamentally the way we think about computing, as our ability to compute will no longer be limited to the resources we currently have on hand (Foster, et al, 2001), and (Foster, et al, 2005A).

CPU-scavenging, cycle-scavenging, cycle stealing, or shared computing creates a "grid" from the unused resources in a network of participants (whether worldwide or internal to an organization). Typically this technique uses desktop computer instruction cycles that would otherwise be wasted at night, during lunch, or even in the scattered seconds throughout the day when the computer is waiting for user input or slow devices. Volunteer computing projects use the CPU scavenging model almost exclusively. In practice, participating computers also donate some supporting amount of disk storage space, RAM, and network bandwidth, in addition to raw CPU power. Since nodes are likely to go "offline" from time to time, as their owners use their resources for their primary purpose, this model must be designed to handle such contingencies (Wikipedia, 2008D).

A grid uses the resources of many separate computers, loosely connected by a network (usually the Internet), to solve large-scale computation problems. Public grids may use idle time on many thousands of computers throughout the world. Such arrangements permit handling of data that would otherwise require the power of expensive supercomputers or would have been impossible to analyze. What distinguishes grid computing from conventional cluster computing systems is that grids tend to be more loosely coupled, heterogeneous, and geographically dispersed. Also, while a computing grid may be dedicated to a specialized application, it is often constructed with the aid of general-purpose grid software libraries and middleware (Wikipedia, 2008D).

Evolution of Grid Computing

The term Grid Computing originated in the early 1990s as a metaphor for making computer power as easy to access as an electric power grid. CPU Scavenging and Volunteer became popular in 1997 by distributed.net and later in 1991 by SETI@home to harness the power networked PCs worldwide, in order to solve CPU-intensive research problems (Berman, et al, 2003).

The concept of grid computing was the brainchild of Ian Foster, Carl Kesselman and Steve Tuecke, and it was made known in one of their seminal presentations, "The Grid: Blueprint for a new computing infrastructure." Their efforts created the Globus Toolkit, which contains computation management, storage management, security management, monitoring and other related services. They are widely acclaimed as the "father of the grid."

Grid computing got its name because it strives for an ideal scenario in which the CPU cycles and storage of millions of systems across a worldwide network function as a flexible, readily accessible pool that could be harnessed by anyone who needs it, similar to the way power companies and their users share the electricity grid. Grid computing can encompass desktop PCs, but more often than not its focus is on more powerful workstation, servers, and even mainframes and supercomputers working on problems involving huge datasets that can run for days. And grid computing leans more on dedicated systems, than systems primarily used for other tasks (Foster, et al, 2005B).

Distributed Computing

Distributed computing deals with hardware and software systems containing more than one processing element or storage element, concurrent processes, or multiple programs, running under a loosely or tightly controlled regime (Wikipedia, 2008D). The core objective of a distributed computing system is to interface users and resources in a manner that is transparent, open, and scalable.

In distributed computing a program is split up into parts that run simultaneously on multiple computers communicating over a network. Distributed computing is a form of parallel computing, but parallel computing is most commonly used to describe program parts running simultaneously on multiple processors in the same computer. Both types of processing require dividing a program into parts that can run simultaneously, but distributed programs often must deal with heterogeneous environments, network links of varying latencies, and unpredictable failures in the network or the computers (Wikipedia, 2008D).

Other Forms of Computing

Prior to the emergence of Grid computing, there were other methods of computing. These will be briefly discussed in this paper.

Mainframe Computer

Mainframes are computers used mainly by large organizations for critically handling bulk data processing such as census, industry and consumer Enterprise Planning, and financial transaction processing. Modern mainframe computers have abilities not so much defined by their single task computational speed (Wikipedia, 2008A).

Nearly all mainframes have the ability to run (or host) multiple operating systems, and thereby operate not as a single computer but as a number of virtual machines. Mainframes are designed to handle very high volume input/output (I/O) and emphasize throughput computing. Since the mid-1960s, mainframe designs have included several subsidiary computers (called channels or peripheral processors) which manage the I/O devices, leaving the CPU free to deal only with high-speed memory (Wikipedia, 2008A). Mainframes exhibit fault tolerant computing and one draw back has being the high costs of hardware and operating system.

Supercomputer

Supercomputers were introduced in the 1960s. They were designed primarily by Seymour Cray at Control Data Corporation (CDC). Today's supercomputer tends to become tomorrow's ordinary computer. CDC's early machines were simply very fast scalar processors, some ten times the speed of the fastest machines offered by other companies. Supercomputers are used for calculations involving intensive tasks such as quantum mechanical physics, weather forecasting, climate research, molecular modelling (computing the structures and properties of chemical compounds, biological macromolecules, polymers and crystals), physical simulations (such as simulation of airplanes in wind tunnels, simulation of the detonation of nuclear weapons, and research into nuclear fusion), cryptanalysis, and the like (Wikipedia, 2008C).

Cluster Computing

A cluster consists of multiple stand-alone machines acting in parallel across a local high speed network. Distributed computing differs from cluster computing in that computers in a distributed computing environment are typically not exclusively running “group” tasks, whereas clustered computers are usually much more tightly coupled. Distributed computing also often consists of machines which are widely separated geographically Wikipedia (2008D).

One of the main ideas of cluster computing is that, to the outside world, the cluster appears to be single system. Often, clusters are used for primarily computational purposes, rather than handling I/O-oriented operations such as web service or databases. The relatively low cost of clusters makes them excellent power plants for grid (Foster, et al, 2001).

The Concept of Virtual Organization

The Grid community often refers to the notion of a “virtual organization” (VO). A virtual organization exists as a corporate, not-for-profit, educational or otherwise productive entity that does not have a central geographical location and exists solely through telecommunication tools (Meliksetian, et al, 2004). In the context of this paper, the notion of a VO corresponds to the set of resources that are pooled and the set of users who can harness these pooled resources.

A Grid system is a virtual organization comprising several independent autonomous domains (Akinyemi, et al, 2007). Grid enables people to be members of many VOs and each VO gives one access to various computational, instrument-based data and other types of resources. It is very natural for these VOs to produce a Grid portal which provides an end-user view of the collected resources available to the members of the VO. By producing a portal with “one-stop shopping” for users who participate in a VO, the VO makes its resource much more useful and accessible for their users (Joseph, et al, 2004).

Grid technologies and infrastructures support the sharing and coordinated use of diverse resources in dynamic, distributed virtual organizations. Enterprise computing systems must increasingly operate within virtual organizations (Vos) with similarities to the scientific collaborations that originally motivated Grid computing. Depending on the context, the dynamic ensembles of resources, services, and people that comprise a scientific or business VO can be small or large, single or multi-institutional, and homogeneous or heterogeneous. Individual ensembles can be structured hierarchically from smaller systems and may overlap in membership. Virtualization enables consistent resources access across multiple heterogeneous platforms. Virtualization also enables mapping of multiple logical resource instances onto the same physical resource and

facilities management of resources within a VO based on composition from lower-level resources. Furthermore, virtualization lets us compose basic services to form more sophisticated services – without regard for how these services are implemented (Meliksetian, et al, 2004).

Finally, virtualizing Grid services also underpins the ability to map common service semantic behaviour seamlessly onto native platform facilities. This virtualization is easier if we can express service functions in a standard form, so that any implementation of service is invoked in the same manner (Foster, et al, 2002).

Qualities of Grid Computing Over the Conventional Distributed Computing Methods

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 (Foster, et al, 2001). Grid computing employ middleware to harness IT resources across a network, thereby, making them to function as a virtual system. The goal of a computing grid, like that of electrical grid, is to ensure that users has access to the resources they required and in due time.

Grids address two distinct but related goals: providing remote access to IT assets, and aggregating processing power. The most obvious resource included in a grid is a processor, but grids also encompass sensors, data-storage systems, applications, and other resources. One of the first commonly known grid initiatives was the SETI@home project, which solicited several million volunteers to download a screensaver that used idle processor capacity to analyze data in the search for extraterrestrial life. In a more recent example, the Telescience Project provides remote access to an extremely power electron microscope at the National Centre for Microscopy and Imaging Research in San Diego. Users of the grid can remotely operate the microscope, allowing new levels of access to the instrument and its capabilities (Educause, 2008).

Driven by increasingly complex problems and propelled by increasingly powerful technology, today’s science is as much based on computation, data analysis, and collaboration as on the efforts of individual experimentalists and theorists. But even as computer power, data storage, and communication continue to improve exponentially, computational resources are failing to keep up with what scientists’ demand of them. The Grid offers a potential means of surmounting these obstacles (Foster, et al, 1999). Built on the Internet and the World Wide Web, the Grid is a new class of infrastructure. By providing scalable, secure, high-performance mechanisms for discovering and negotiating access to remote resources, the Grid promises to make it possible for scientific collaborations to share resources in an

unprecedented scale, and for geographically distributed groups to work together in ways that were previously impossible (Berman, et al, 2003), (Dejan, 2003), and (Foster, et al, 1999).

The real and specific problem that underlies the Grid concept is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organization. The sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources, as is required by a range of collaborative problem-solving and resource-brokering strategies emerging in industry, science, and engineering (Farago-Walker, 2001).

Grid computing is capable of handling most of the world's most challenging problems due to its computational power and ability. This wonderful IT infrastructure depend on computer power as well as data and often large amount of heterogeneous, distributed data from various groups are stored in diverse systems. Distributed systems are scientific tools for solving large-scale problems in science and engineering, and capable of managing data and computing tasks.

Benefits of Grid Computing

The concept of grid computing is aimed towards addressing the demands to leverage and reallocate existing IT resources. Some of the benefits of grid computing are summarized below:

- i) **Exploitation of Under-Utilized Resources:** Exploitation of under-utilized resources are carried out by:
 - running an existing application on different machines;
 - Exploiting idle times on other machines;
 - Aggregating unused disk drive capacity into much larger virtual storage, to enhance performance;
 - Creating better balance of resource allocation; and
 - Improving users view of usage patterns of resources in an organization.
- ii) **Reduces Computational Time:** Computational time is reduced for complex numerical and data analysis problems.
- iii) **Provide Information Access:** In the life sciences sector, an information accessibility option is added to maximize the exploitation of existing data assets by providing unified data access during the querying process of non-standard data formats (Farago-Walker, 2001).
- iv) **Reduces cost by optimising existing IT infrastructure:** The grid facilitate reduction of costs in companies by optimising the use of existing IT infrastructure investments and by enabling data sharing and distributed workflow across partners, and therefore enabling

faster design processes. (Foster, et al, 2005B).

- v) **Providing access to parallel CPU capacity:** Grid computing offers potential access for large-scale parallel computation to enhance performance in computationally intensive applications.
- vi) **Offers improved reliability:** Grid technology offers alternate approach to achieving improved reliability through software other than hardware. Parallelization can boost other reliability by having multiple copies of important jobs run concurrently on separate machines on the grid. Their results can be checked for any kind of inconsistency, such as failures, data corruption and tempering. Automotive computing can be utilized such that when there are problems in the grid can be healed automatically even before the operator or manager is aware of them (Foster, et al, 1999).
- vii) **Provision of resource balancing:** The grid offers good resource balancing measures that can handle occasional peak loads, job prioritization, and job scheduling.
- viii) **Effective management of resources:** With grid technology, management of organization can easily visualize resource capacity and utilization to effectively control expenditures for computing resources over a larger organization. This task is made possible by aggregating utilization of data over a large set of projects, which can help an organization with the ability to plan for the future.
- ix) **Interoperability of virtual organizations:** The grid offers collaboration facilities and interoperability of different virtual organization by allowing the sharing and interoperation of the heterogeneous resources available.
- x) **Access to additional resources:** The grid offers access to other specialized devices such as the cameras and embedded systems.
- xi) **Harnessing heterogeneous systems together:** Grid computing can be used to harness heterogeneous systems together into a mega computer by applying greater computational power to a task.
- xii) **Grid virtualization:** Grid computing offers grid virtualization, thereby making a single, local computer to function as though, it greatly influenced or simplified the development needed for such powerful applications.

Conclusion

We have examined grid computing and distributed computing. We also, discussed the origin and

benefits Grid computing can offer in terms of qualities and capabilities and why we believe it will continue to be on high demand for many years to come. Attempting to harness the under-utilization of processing power across the enterprise should be a priority of every organization and must appeal to businesses as they face increased pressures to maximize returns on IT investments. Grid computing, indeed, has emerged as a new methodology of solving problems that requires very high computational capabilities which in some recent past, either was not feasible to solve or too expensive to purchase perhaps due to the high cost of financial acquisition of the conventional distributed computing systems.

One major characteristics of a grid is that it allows organizations to pool computing resources (processors, storage, information applications, etc.) to enable users to benefit from a potentially far larger pool of resources than would otherwise have been available to them. However, the European Union has been a major proponent of Grid computing because many projects have been funded through the framework programme of the European Commission. Therefore, managers, board of directors and academia communities should tap into this existing IT infrastructure to develop and increase computational productivity at a relatively low cost rather than wasting all that unused processing capacity.

For the dreams of the fathers of grid computing to be fulfilled, efforts should not only be made to harness and virtualize multiple computing resources, but also to abstract and hide the diversity and distribution of these various information sources to provide applications with a single, powerful virtual information store for their virtual computer. The benefits Grid computing can offer in terms of qualities and capabilities are enormous and we believe it will continue to be on high demands and will continue to enjoy greater popularity and patronage in many years to come.

REFERENCES

- Akinyemi, I.O., Daramola, J.O. and Adebisi, A. (2007), Grid-Enabled e-learning Framework for Nigerian Educational Institutions, Nigeria Computer Society, 21st National Conference Proceedings, Vol. 18, pp. 91-98.
- Berman, F., Anthony J. G. H. and Geoffrey C. F. (2003). Grid Computing: Making the global infrastructure a reality, ISBN 0-12-742503-9.
- Bourbonnais, S., Gogate, V. M., Haas, L. M., Horman, R. W., Malaika, S., Narang, I., and Roman, V. (2004), Towards an Information Infrastructure for the Grid, IBM Systems Journal, Vol. 43, No. 4, pp. 665 - 688.
- Buyya, R. (2007), Market Based grid Computing and the Gridbus Middleware, e-Science, The university of Melbourne, Australia.
- Carpenter, B. E. and Janson, P. A. (2004), Abstract Interdomain Security Assertions: A Basis for Extra-Grid Virtual Organization, IBM Systems Journal, Vol. 43, No. 4.
- Davey, W. (2003). Grid computing - spreading the word, PC PRO: Computing in the real world, pp.191-194.
- Dejan, S. M. (2003), Peer-to-Peer Computing, Hewlett-Packard Company, pp. 1- 5
- Educause, (2008), 7 Things You Should Know About Grid Computing, Educause Learning Initiative, www.educause.edu/eli, last accessed in Oct., 2008.
- Farago-Walker, S. (2001), Peer-to-Peer Computing: Overview, Significance and Impact, E-learning, and Future Trends, <http://www.elearningmag.com/>
- Foster, I. and Carl K. (1999), The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann Publishers. ISBN 1-55860-475-8
- Foster, I., Kesselman, C. and Tuecke, S. (2001), The Anatomy of the Grid: Enabling Scalable Virtual Organizations, International Journal of Supercomputer Applications, Vol. 15, NO. 3.
- Foster, I., Kesselmen, C., Nick, J. M., and Tuecke, S. (2002), Grid Services for Distributed System Integration, IEEE Journal of Computer Science.
- Foster, I. and Tuecke, S. (2005A), Describing the Elephant: The Different Faces of IT as Service, QUEUE, pp. 26 – 34.
- Foster, I., and Tuecke, S. (2005B), The Globus Project: A Status Report.
- Joseph, J., Ernest, M. and Fellenstein, C. (2004), Evolution of Grid Computing Architecture and Grid Adoption Models, IBM Systems Journal, Vol. 43, No. 4, pp. 624 – 645.
- Meliksetian, D. S., Prost, J. P., Bahl, A. S., Boutboul, I., Currier, D. P., Fibra, S., Girard, J. Y., Kassab, K. M., Lepesant, J. L., Malone, C., and Manesco, P. (2004), Design and Implementation of an Enterprise Grid, IBM Systems Journal, Vol. 43, No. 4. pp. 646 – 664
- Wikipedia (2008A), the Free Encyclopedia, http://en.wikipedia.org/wiki/Grid_computing, last accessed in Oct., 2008.

Wikipedia (2008B), the Free Encyclopedia, <http://en.wikipedia.org/wiki/mainframe> computer, last accessed in Oct., 2008.

Wikipedia, (2008C), the Free Encyclopedia, <http://en.wikipedia.org/wiki/Supercomputer>, last accessed in Oct., 2008.

Wikipedia, (2008D), the Free Encyclopedia, [http://en.wikipedia.org/wiki/Distributed computing](http://en.wikipedia.org/wiki/Distributed_computing), last accessed in Oct., 2008.