GRID COMPUTING AN ADVANCEMENT OF E-SCIENCE TO COMPUTING AND BEYOND

By

Ajulo Adeshina B.
Department of Computer Science,
Federal University of Technology,
P.M.B. 704, Akure, Nigeria.
postgraduate student (m-tech. computer science)
shainaburry@yahoo.com

ABSTRACT
The Grid community has made substantial achievement in the area of Grid computing, and yet most success stories are largely limited to e-science. In part, expectations have been raised high by claims that the Grid can meet the need of distributed computing and delivery of results. Many of these claims have been demonstrated through pilot projects which unfortunately have not achieved lasting influence -- in contrast to what has been envisioned by some Grid researchers.

This presentation will go back to the roots of Grid computing: E-science. We will concentrate largely on the objective to create an invisible Grid for application users and developers.

Many existing Grid application systems provide the application developer with a non-transparent Grid. Commonly, application developers are explicitly involved in selecting software components deployed on specific sites, mapping applications onto the Grid, or selecting appropriate computers for their applications.

Moreover, many programming environments are either implementation- and technology-specific, or they force the developer to program at a low-level middleware interface.

For a case study, this talk will describe the FUTA Grid application development and computing environment, whose ultimate goal is to provide an invisible Grid to application users and developers in the area of e-research work.

Furthermore, this talk will make the case of an interesting alternative research path with the objective to automatically generate workflows based on semantic technologies.

Finally an example centered on the development of online-research work will be used to illustrate how methods originating from scientific and Grid computing can be applied to meet the needs of industry applications. This example offers renewed hope that future for the Grid may still be bright.

Keywords: Grid Computing, Grid computing system, E-science, collaborative engineering, data exploration, high-throughput computing (HTC), Meta application, and high-performance computing (HPC).FUTA Grid,
1.0 Introduction

E-science: What is it?

E-Science refers to science that is enabled by the routine use of distributed computing resources by end user scientists. E-Science is often most effective when the resources are accessed transparently and pervasively, and when the underlying infrastructure is resilient and capable of providing a high quality of service. Thus, grid computing and autonomic computing can serve as a good basis for e-science. E-Science often involves collaboration between scientists who share resources, and thus act as a virtual organization.

Grid Computing: What Is It?

The continuous progress in scientific research is itself an explanation of the insatiable demand for computational power. On the other hand, one of the results of scientific progress is the availability of more and more powerful computer platforms. This self-feeding cycle is pushing our search for knowledge towards very challenging investigations, and parallel computing nowadays plays an important role in this scenario. This is especially driven by the present-day enhancement in distributed computing, which has produced a substantial reduction in the costs of effective supercomputing facilities. Another emerging trend, due to the improvement of distributed information technologies (IT), is the acceleration of research and development processes towards concurrent and cooperative engineering. Daily workflows in academic and industrial activities are more and more based on interaction among remote entities, which in some cases are physical people and in others are agents or facilities embedding value-adding procedures.

An IT infrastructure is, most of the time, the core of such processes. In the last decade, these important evolutions have been accompanied by the so-called Internet revolution and the boom in Web applications. The extraordinary perspectives opened by the Web have reinforced the momentum towards process integration and cooperative computing. Consequently, joining together supercomputing facilities and the world of Web-based tools seems to be the key feature to opening new perspectives in industrial and scientific computational processes, and an emerging technology is being proposed as the most natural way to pursue such a goal: grid computing (GC).

The technology of GC has led to the possibility of using networks of computers as a single, unified computing tool, clustering or coupling a wide variety of facilities potentially distributed over a wide geographical region, including supercomputers, storage systems, data sources, and special classes of devices, and using them as a single unifying resource (computational grid). The concept started as a project to link supercomputing sites but has now grown far beyond its original intent, opening new scenarios for collaborative engineering, data exploration, high-throughput computing (HTC), meta application, and high-performance computing (HPC).

- Collaborative (or cooperative) engineering means providing engineers and researchers with tools for cooperating online. These tools allow them to share remote resources, modify them, and design, implement, and launch applications in cooperation. In such a context, a grid can be seen as a global production environment, where distributed systems can be prototyped and tested. The flexibility of grids makes this system dynamic and configurable. Via the grid, researchers are able to rapidly modify their products in order to adapt them to the changes of underlying environments, infrastructure, and resources.
- Data exploration is particularly critical when dealing with huge amounts of data and their access from remote sites is needed. Several research fields, such as climate analysis and biological studies, require the storage and accessing of data up to the terabyte or petabyte range. In these cases, data are distributed on a number of remote sites and then accessed uniformly. Taking this further, redundancy can help in improving access performance and reliability. Redundancy is obtained by creating replicas of data sets, increasing performance by accessing the nearest data set. Services to manage distributed data sets and replicas are central in grid computing.
- HTC applications require large amounts of computational power over a long period of time. Examples of HTC applications are large-scale simulations and parametric studies. HTC environments try to optimize the number of jobs they can complete over a long period of time. A grid allows exploitation of the idle central processing unit (CPU) cycles of connected machines and use of them for HTC applications.
- Meta applications are applications made up of components owned and developed by different organizations and resident on remote nodes. A meta application is usually a multidisciplinary application that combines contributions from differently skilled scientific groups. A meta application
Grid Computing: Who Is Who?

GC is mature and is attracting large companies, boards, and research centers. For instance, IBM is building the Distributed Terascale Facility (DTF), with $53 million funding from the National Science Foundation (NSF) [1]. Examples of working grid applications can be found in different scientific disciplines. In the field of distributed super computing, one of the primary areas in which GC has sparked interest is in large-scale sequence similarity searches. An individual sequence similarity search requires little computation time, but it is common for researchers to perform such searches for thousands or even tens of thousands of sequences at a time. These searches, each dependent on the others, can be spread across as many hosts as are available. The storage and exploitation of genomes and of the huge amount of data coming from post genomics puts a growing pressure on computing tools—such as databases and code management—for storing data and data mining. Genomic research needs, together with requirements coming from other disciplines, gave place to the Data Grid project [2]. This European initiative joins researchers coming from European Organization for Nuclear Research (CERN), European Space Agency (ESA), and other outstanding European scientific centers and is actively fostered by the European Union. It is focused on building up an international grid to store large volumes of data and to provide a uniform platform for data mining. This helps researchers from biological science, Earth observation and high-energy physics, where large scale, data-intensive computing is essential. Other interesting grid applications are those oriented towards the promotion of synchronous cooperation between persons. The Access Grid [3] is focused on online collaboration through audio/video conferencing. The Astrophysics Simulation Collaboratory (ASC) portal [4] allows users to form virtual organizations over the Internet. People belonging to a virtual organization access the grid to collaboratively assemble code, start-stop simulations, and update and access a repository of simulation components shared with their remote colleagues. The reference point for grid communities is the Global Grid Forum (GGF) [6]. The GGF coordinates a growing number of research groups cooperating to ratify community standards for grid software and services and to develop vendor- and architecture-independent protocols and interfaces.

Grid computing is (GC) is a model of distributed computing that uses geographical and disparate resources (a logical entity such as a distributed file or a computing cluster) that are found on the network (grid network). These resources may include processing power, storage systems, catalogs, data, network resources, sensors and other hardware such as input and output devices. In grid computing, individuals’ users can access computers and data transparently without having to consider location, operating systems and other details.

Scientific computing involves the use of computers to analyse and solve scientific and engineering problems as well as construction of numerical solutions and mathematical model techniques. In practical terms, computational science is the application of computer simulations and other forms of computations to problem in various science and engineering research disciplines like the FUTA community that is an institution where research is taken to be the root to all development of the entire entire department.
2.0 Method of Study

Grid computing arises from the emerging need to transform the web into a giant repository where user can pick up resources as needed. As earlier defined a computational grid consists of distributed software and hardware facilities such as storage systems and so on. The resource-sharing architecture of grid computing is conceived as a set of universally recognized standards, protocols and web complaint technologies open to majority of existing distributed visions and methodologies. Distributed software could be located on different network that are interconnected together on different hardware component as been on internet connection. A major goal of information technology designers is to provide an environment within which it is possible to present any form of information on any device at any location. A major goal of information technology designers is to provide an environment within which it is possible to present any form of information on any device at any location.

In Grid terminology each department on the FUTA community grid that forms a collaboration corresponds to a virtual organisation and almost “perfectly” fulfils the basic Grid concepts of resource sharing, virtualisation and transparent access. Grid middleware and services are used to orchestrate the job submission to the remote sites as well as the data exchange and transfer involved to analyse data.

The main building blocks or services are categorized as thus.

- Computing Element. The term Computing Element is an abstraction for any kind of computing resource such as a cluster, a supercomputer or just a single machine. Typically, a Computing Element provides a standard Grid interface to hide the complexities of the underlying computing hardware and software.
- Storage Element. A Storage Element is the equivalent concept for a storage system. It provides an abstraction for different several storage systems such as file systems, mass storage systems etc. Currently, Storage Elements often provide a file based interface in contrast to data stored in relational databases or object databases.
- Replica Catalogue. Data is distributed and often replicated to several sites for fault tolerance and availability reasons. A replica catalogue is a special purpose metadata service that is used to locate (file) replicas in the Grid.
- Information System. The Information System keeps track of all services in the Grid. It is used for discovery of services such as Computing Elements, Storage Elements, and other higher level services and applications. Additionally, the Information System is sometimes used to monitor the status of services as well as of other higher level Grid applications.
- Resource Broker. End users or high-level services typically do not directly interact with all Grid services. It is therefore the task of a Resource Broker to discover computing resources via the Information System and dispatch jobs to suitable locations. Resource Brokers can either be services on their own or directly available as part of a client tool for job submission. Note that the term Resource Broker is an over simplifications since the implementation of such a component typically consists of the following sub-components: resource discovery and match making, workflow management and job execution.

The origins of grid computing

Just like with the Internet, academic institutions were at the forefront when it came to developing the first generation technologies and architectures that formed the basis of grid computing. Institutions such as the Globus Alliance, the China Grid, and the UK e-Science Grid core program were some of the first to incubate and grow grid solutions to maturity, preparing them for commercial adoption.

Grids were borne out of the research and academic communities’ very real need to collaborate. A crucial component of research is the ability to disseminate knowledge -- the more efficiently you can share not only vast amounts of information but also the computational resources that help you create this data -- the more refined and informative a level of quality in collaboration you can achieve.

A counterpart to this need to disseminate knowledge is in the commercial world. Grid computing can also address this need, because the integration of business processes and transactions, facilitated by Web services standards, continues to grow in importance. As the adoption of commercial grid computing continues, standards (such as those proposed by organizations like the Global Grid Forum, or GGF) will benefit from the real-world, hardened, and practical requirements to which commercial applications will subject them. Currently, grid computing benefits from the early identification and development of standards-based technologies in the academic world that is matched with more practical and robust implementations that commercial businesses require. There is no reason to imagine that this synergy will not continue as grid computing matures.
3.0 RESULT
The complexity of the FUTA intra grid system is being taken away by the grid middle ware layer called the digipede which is a middleware ware software which gives a complete solution to the operation of the FUTA intra grid computing solution.

Grid as a new Information Technology (IT) concept of "super Internet" for high-performance computing: worldwide collections of high-end resources – such as supercomputers, storage, advanced instruments and immersive environments. These resources and their users are often separated by great distances and connected by high-speed networks. The Grid is expected to bring together geographically and organizationally dispersed computational resources, such as CPUs, storage systems, communication systems, real-time data sources and instruments, human collaborators, as the user looking at a Grid sees essentially one, large virtual computer built on open protocols with everything shared — applications, data, processing power, storage e.t.c.

The on line research work is resident on the middle ware of the FUTA intra grid which is a basis for research into the past research work in different department of the institution such that reference could be made to it for further research this includes different thesis both for the under graduate and graduate such that it could easily be seen on the data storage of the intra grid which I connected on a WAN network within the institution.

Many grid projects has been developed round the globe but the usefulness of the FUTA intra grid has been in the area of research which has made it possible for both the lectures and student and other visiting researcher to have the work resident on the data storage of the FUTA intra grid which is tagged “Research@FUTA project.

4.0 CONCLUSION
Grid technologies are the convergence of distributed and parallel computing coupled with several other areas such as web and security. The exploitation of computational grids using desktop grid computing for research in developing countries become necessary due to high cost of parallel supercomputers and commodity clusters, also is the need to generate results from large and complex scientific data. I can conclude that FUTA intra grid has really contributed to fundamental of research work on FUTA community in that it gives the basis t all research work that are interrelated and gives the department a monitoring function such that research are not double worked on by individual students. With I have to say the scientific environment, the industry, business, engineering and other areas can benefit from grids and desktop grids and also say that grid are more than just a technology that will come and go, but rather a way forward through which our resources must evolve if it is to stand our current scientific research throughput and demand.

5.0 REFERENCES