Grid Computing: A Utility Worth Looking At

Harold J. Rice

Today's challenging business environment requires organizations to adapt rapidly and be agile. They must make on-demand changes to adjust dynamically and efficiently to meet customer demands and marketplace shifts. In order to meet these unpredictable demands, a flexible and powerful information technology (IT) network and computing infrastructure must be readily available. One means of providing a cost-effective infrastructure is by the use of grid computing where an organization can gain access to computing, data, and storage resources when they need them without incurring large capital investments (i.e., data centers, etc.). This paper will provide an overview of grid computing and its benefits; it will also present an example and business model for a current commercial grid computing service.

Introduction

The theory of leveraging shared computing power and resources has been around for some time. Grid computing, while not a completely new concept, is starting to garner measurable demand, especially among larger organizations. [1] Grid computing enables organizations to optimize computing and data resources. Users can pool the grid resources for large capacity workloads and share them across networks for collaboration. Grid computing can bring tremendous productivity and efficiency to organizations facing the challenges of an on-demand world. Government labs and scientific organizations have been using grid technologies for several years, solving some of the most complex and important problems facing the world today. Now grid computing is becoming a critical component of day-to-day business. There is a profound shift afoot in how computing is used—even in basic assumptions about how it is accessed and paid for. [2] This paper will provide an introduction to grid computing, describe the basics associated with this technology, and describe its benefits; this paper will also present an example of a commercial grid computing service, available today, which can be accessed via a secure Internet connection.

Background

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. It can provide an on-demand operating environment analogous to the electric power utility where a user can turn on a power switch and instantly receive service (assuming the home or business is connected to and has an account with the electric utility). It also uses similar electric power grid concepts and techniques, such as load sharing and resource balancing.

Some people view this approach as an *information technology (IT) as a service* business model. A grid user basically sees a single, large virtual computer; this experience is similar to what an Internet user experiences when they view a unified instance of content via the Web. Consumers of a grid computing service do not want to know the specifics of how the service is delivered; they want the service available whenever they need it, and at the best available price. Grid computing is also referred to as utility computing when offered as a metered commercial service.

Definition

There are several definitions and descriptions of grid computing in use today. One such definition is provided by the Open Grid Forum (OGF) which defines a grid as "a system that is concerned with the integration, virtualization, and management of services and resources in a distributed, heterogeneous environment that supports collections of users and resources (virtual organizations) across traditional administrative and organizational domains (real organizations)." [3]

The use of open standards and protocols is a key attribute of this technology since it enables any user to access its resources. The Open Grid Services Architecture (OGSA) is an open set of standards and protocols that forms the basis for grid computing. The OGSA enables communication across heterogeneous, geographically dispersed environments.

Grid computing model

A grid can be implemented within a single organization, shared across different organizations, or offered as a commercial service. Grid computing can be used for solving considerable computational problems by exploiting under-utilized resources (central processing unit [CPU] cycles and/or disk storage) of large numbers of disparate computers treated as a virtual cluster embedded in a distributed telecommunications infrastructure. It focuses on the ability to support computation across administrative domains which sets it apart from traditional computer clusters (or traditional distributed computing arrangements). There are many scientific and business applications for this technology. A key to understanding this model is to understand the general grid architecture.

The grid architecture is multi-layered in which each layer plays a critical role. There are essentially four main layers to the architecture. In general, the higher layers are more software-centric, while the lower layers are more hardware-centric. [4] These layers are listed below, from the highest to lowest, and described in Table 1.

- Application and Serviceware layers
- Middleware layer
- Resource layer
- Network layer

Table 1. Description of grid architecture layers

Layer	Description
Application and Serviceware	These are the highest layers of the grid architecture. Applications vary by industry and the portals and development toolkits supporting these applications. Serviceware provides many management-level functions, including billing, accounting, and measurement of usage metrics, which are tracked as resources are virtualized for sharing among different users.
Middleware	This layer provides the protocols that enable multiple elements (servers, storage, networks, etc.) to participate in a unified grid environment. This layer is the intelligence that brings the various elements together through software and control. There are many different functions and protocols that this layer supports.
Resource	This layer consists of the actual resources that are part of the grid, which includes primarily servers and storage devices.
Network	This layer is the underlying connectivity for the resources in the grid. This can include high-capacity fiber optic links and layer 2/3 switch routers that provide the connectivity, performance, and intelligence required for the network infrastructure.

Figure 1 depicts the relationships between each layer. The architecture enables multiple and diverse applications to run over the same infrastructure.

An integral part of the grid is the ability to manage the distributed resources and track which users are members of the grid. A grid management system is responsible for determining status of the grid resources, identifying which jobs (a.k.a. submissions or transactions) to run, and selecting where to run them.

Other key functions are measurement components that determine the current utilization rates and capacity within the grid. This information is used to determine the health of the grid, identify issues such as congestion and outages, and log, audit, and account for usage of grid resources. Further necessary functions are security, a method for querying jobs, monitoring progress of a job, job submission, and recovery in case of a job failure.

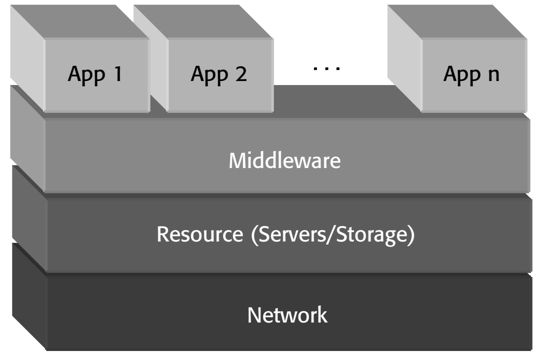

Figure 1. General grid architecture layers

Benefits

Grid computing offers substantial benefits in terms of improved productivity, efficiency, scalability, performance, and capital savings. It allows the user to access additional computing resources without needing to invest in (or over-provision) infrastructure to support temporary spikes in demand. Grid computing can be appealing and beneficial to organizations of all sizes. The major benefits are summarized below:

- Allow widely dispersed businesses and business units to create virtual organizations to share data and resources
- Provide scalability in terms of data storage—no need to over- or under-provision storage capacity
- Exploit underutilized resources; balance workloads and provide capacity for high-demand applications
- Reduce real estate requirements and energy expenses (space, power, cooling, etc.) necessary to support and manage data centers
- Reduce the need for retaining in-house IT expertise
- Support business continuity
- Provide predictable pricing; pay for only what you consume and transform IT from being a capital expense to an operational expense

A commercial service offering

Utility computing is a business model for grid computing in which resources (CPU power, storage space, etc.) are made available to the user on an as-needed basis. Users are able to dial-up or dial-down usage in real time to meet the varying demands of business.

Sun Microsystems offers an Internet-based utility computing service which is available via their Website *network.com*. It is a pre-paid computing service with no minimum volume purchase or contract required. To obtain access to the Sun utility computing grid, a customer must create an account, purchase computing time, and complete an export compliance questionnaire. The compliance questionnaire is reviewed within 48 hours and then access is either granted or denied. Customers can either run their own applications on the Sun utility computing grid, or choose from an application catalog that offers access to on-demand applications.

Running an application on the Sun utility computing grid is a general five-step process. The process consists of the following:

- Packaging the work—place all the application and data files in a single directory hierarchy
- Uploading the data to the network.com secure Website
- Describing the job—providing details about how it will be executed
- Starting a run—scheduling the job to be executed on the grid
- Getting the results—downloading the results from the job run

Sun's approach is unique compared to some other commercial offerings since it provides a platform for independent software vendors (ISVs) to publish and sell on-demand applications. The application catalog consists of two components:

- Packaging the work—placing all of the users application and data files into a single directory hierarchy and ensuring the submitted files meets certain criteria
- A high performance computing infrastructure

There is no cost to the ISVs to publish an application in the catalog, but Sun gains revenue through use of their utility computing grid to run the application. The current catalog of ready-to-use applications ranges in categories such as life sciences, computer aided engineering, computational mathematics, financial services, 3-D digital content/animation, and other general applications such as a numerical weather prediction system.

In terms of technical capabilities, *network.com* provides users with access to the following:

- Secure Internet data transfer at no additional cost
- Storage allocation of up to 10 GB per active user account (to enable use of the service)
- A grid network infrastructure of 1 GB switched data network and 300 Mb dedicated bandwidth
- A Web-based access portal
- Sun Fire dual processor Opteron-based servers with 4 GB RAM per CPU
- Solaris 10 OS and Solaris 10 (x64)
- Sun N1Grid Engine 6 software

The service also offers a feature called *Network.com Internet Access* so that customers can interact, through Sun's utility data center to the Internet, with other companies that have resources the customer might want to use for a particular project.

There are some basic minimum requirements for using the service, such as:

- The application must run on Solaris 10 (x64)
- The user must own the application or have legitimate legal licenses to run applications on the service
- Applications must be scripted to work with N1Grid Engine 6 software
- The application must be self-contained (no dependencies on external libraries or data sets)
- The total size of the application and the data sets must be under 10 GB
- The user must upload application and data to the Sun Grid over the Internet via the Internet Portal at http://www.network.com

The service is priced at $1 per CPU, per hour, with no minimum purchase, no advanced reservation, and no annual contract required to access the utility network. The flat rate remains the same for running applications in the catalog. Billing is calculated for each job the user submits and runs; the user's grid CPU usage is aggregated and then rounded up to the nearest whole hour. For example, if a job used 1,000 CPUs for one minute, it would be aggregated as 1,000 CPU minutes or 16.67 CPU hours. The software rounds this up to 17 hours and the job would be billed as $17. [5]

As of May 2008, the service currently is available in 25 countries worldwide—Australia, Austria, Belgium, Canada, China, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, India, Ireland, Italy, Japan, Netherlands, New Zealand, Poland, Portugal, Singapore, Spain, Sweden, U.K., and the U.S.A.

Other service offerings

There are many other companies and service providers offering commercial grid utility computing services with different options and levels of service (e.g., Hewlett Packard, IBM, Oracle, etc.). Also, enterprises with a large computing and Internet infrastructure are offering on-demand computing and storage services to external entities. For example, *Amazon.com* currently offers a beta utility computing service called Elastic Compute Cloud (EC2) and an Internet-based data storage service called Simple Storage Service (S3). EC2 offers customers use of the Amazon server infrastructure, on a pay for computing capacity and consumed bandwidth basis, ranging from $.10 to $.80 per CPU per hour plus a data transfer bandwidth charge. The EC2 rates are tiered, dependent on how much computing capacity and data transfer is required by the user. [6]

Summary

Grid computing provides a viable solution for businesses, scientific and educational organizations, and governments to effectively use computing resources on demand, using open standards and protocols, without provisioning a large private infrastructure. There are many different private and commercial grid computing networks in use today and they can be implemented quite differently. A basic understanding of the operation and benefits is necessary in order to determine if grid computing is an option for an organization. When the use of grid computing is offered as a commercial service, available on a metered usage basis, it is commonly referred to as utility computing. How quickly or successfully a commercial computing service will be adopted remains to be seen. Grid computing offers significant benefits, but issues such as technological improvements in CPU and memory storage capacity, security, and potential licensing issues (across many grid servers) may hinder its acceptance. ■

Notes and references

1. Jernigan, J., In-Stat Whitepaper, “Utility Computing Garners Demand Up-Market,” August 2007.
2. Statement from IBM Grid Computing; www-03.ibm.com/grid/about_grid/index.shtml.
3. *Open Grid Services Architecture—Glossary of Terms*, Version 1.5, July 24, 2006.
4. Byte and Switch Magazine, “Grid Networking,” May 2003.
5. Sun Frequently Asked Questions (FAQ); http://www.sun.com/service/sungrid/faq.jsp.
6. “Amazon Elastic Compute Cloud (Amazon EC2)—Beta,” Amazon.com; http://www.amazon.com/gp/ browse.html?node=201590011.

About the author

Harold J. Rice is a manager at Noblis. His experience includes planning, requirements analysis, project management, and implementation for telecommunications systems and services for large commercial companies and federal government clients. He is currently supporting the General Services Administration (GSA) Networx program which is one of the largest telecommunications contracts for the federal government. He is a senior member of the Institute of Electrical and Electronics Engineers (IEEE). He received his bachelor’s degree in electrical engineering from the State University of New York at Buffalo. Contact him at harold.rice@noblis.org.