Grid Computing: A Brief Technology Analysis

Roger Smith
CTOnet.org

Executive Summary

Grid computing is a means of allocating the computational power of a large number of computers to a very difficult problem. The goal is to access computers only when they are needed and to scale the problem so that even small computers can make a useful contribution. From a business perspective, this is an opportunity for major vendors like IBM and Sun to sell services to clients who require computers for raw, high-compute problems. From a social perspective, it is an opportunity to harness unused computer cycles and apply them to a socially valuable project.

Grid computing is enabled by relatively high-performance computers, robust computer networks, grid management software, and the divisibility of difficult scientific problems. Together these allow a job to be subdivided and distributed to thousands or even millions of computers to calculate a solution.

The grid computing business has been pioneered by United Devices and IBM, and joined by Sun, HP, and others. It is just beginning to capture enough corporate interest to indicate that is might become a profitable venture. Businesses can generate revenues in three major ways: (1) offer computing time on large hardware grids, (2) offer their services in creating a grid computer for a customer, or (3) offer their services in operating a system once created. The two latter appear to provide the best revenue stream and avoid major capital expenditures.

Introduction

Computers were originally large, centralized assets that could tackle problems in sequence. This soon expanded to the ability to process multiple jobs semi-simultaneously. DARPA’s funding of parallel computing in the 1980’s and 1990’s led to computer architectures involving multiple CPU’s, multiple memory stores, and techniques for dividing and distributing problems (Lewis 1992; Lester 1993). Many of the companies providing these machines perished when DARPA ceased supporting the research, but the lessons learned live on in all computers and workstations. For example, the popular Intel Pentium and Xeon chips contain multi-processor architectures.

More recently, the explosion in power of low-cost, desktop computers and the proliferation of broadband internet access have led to the vision of a globally distributed multi-processor computer. In a sense, every computer that is connected to the Internet is one node in an extremely large computing machine. Grid computing seeks to leverage this machine for both socially beneficial and financially profitable applications.

Given the fact that an average computer is idle 90% of the time and that 99% of its capabilities are never tapped, as measured by the computational stress on the CPU, there is a huge opportunity to apply this power in a beneficial manner. The grid technologies originally developed for global distributed computing are also being applied in centralized computing centers to create high-performance resources that can be rented to companies that need such power very infrequently, or who do not wish to manager the computing environment necessary to maintain their own hardware.

Grid computing has the potential to reduce computation time on complex problems from a period of months to hours. This presents a significant business opportunity if there are enough customers who need such a capability. The name "grid computing" refers to the goal of providing computer services to users in the same way that water and electricity are provided. Customers need not own the means to produce their own water or electricity; instead, they outsource that capability to a utility and purchase only as much as they need (Fordahl, 2005).

Grid computing received a boost from obscurity when it was demonstrated across eleven computers at an IEEE/ACM computing conference in San Diego in 1995. This demonstration allowed participants to interact with a high fidelity simulation of the Chesapeake Bay ecosystem. Following that conference, Ian Foster, Argonne National Labs, and DARPA were just a few of those interested in creating standards for linking computers into a grid (Foster 2003; Waldrop 2002).

Applications

There are a number of existing applications for grid computing. It has primarily been used to support scientific research into large problems concerning weather, astronomy, and medicine. However, the number of potential applications seems to grow each year, which has lead to increasing corporate interest in turning it technology into a business.

There are currently two dominant models of grid computing applications. The *Social Model* views this capability as a resource to be harnessed for the good of society. Projects like SETI@Home, AIDS@Home, and the Human Proteome Folding Project (Figure 1) have created systems that divide a problem and distribute it across the Internet to the computers of people who have volunteered to apply their machines to these problems (Hoffman, 2005).

Figure 1. Client Application for Human Proteome Folding Project

Source: Hoffman 2005

The *Commercial Model* views this as an opportunity to create a large processing center and sell its capabilities to customers by the hour. This frees the customers from the need to purchase, configure, and manage their own large network of computers. Where large computing jobs are infrequent enough that renting is less expensive than purchasing computers, the vendors of grid services hope to capture that company as a customer (Fordahl, 2005). Even the computer gaming world has discovered useful applications of grid computing for maintaining massive worlds for multi-player gaming. Second Life uses the players' home computers to help build and manage the game world (Figure 2).

Figure 2. Second Life Grid System Management Map.

Source: Rosedale 2003

Table 1 provides a partial list of the applications of grid computing. The social model dominates many of the research projects by volunteering computer power from across the Internet or a campus-wide network. The commercial model is being applied by large vendors who hope that this field represents a profitable new service for customers.

Table 1. Current Grid Computing Projects or Services

Project	Purpose	Model
SETI@Home http://setiathome.ssl.berkeley.edu/	Search for Extraterrestrial Intelligence through the analysis of radio signals from outer space (Figure 3).	Social
Human Proteome Folding Project http://www.grid.org/projects/hpf/	Investigate the folding of proteomes in an attempt to understand the human genome.	Social
NSF's Network for Earthquake Engineering Simulation grid http://www.nees.org/index.php	Simulation tectonic activities leading to earthquakes and their impacts on populated areas	Social
DataGrid http://web.datagrid.cnr.it/	Analysis of data on particle physics, earth observation, and biological problems	Social
TerraGrid	Analyze Earth climate data.	Social
CoGrid http://cogrid.colostate.edu/	Grid computing resource for residents of Colorado	Not Defined
SARSgrid http://sarsgrid.nchc.org.tw/	Study the spread of SARS in East Asia	Social
GliederfusllerGRID	Global Analysis of Arthropod Evolution at the High Performance Computing Center Stuttgart	Social
Biomedical Informatics Research Network http://www.nbirn.net/	Analysis of medical data like the human genome project.	Social
EDIAMOND http://www.ediamond.ox.ac.uk/	Analysis of mammographs	Social – but potentially Commercial
Second Life http://secondlife.com/	Massively Multi-player Computer gaming	Commercial
Sun ONE Grid http://www.sun.com/service/sungrid/overview.html	10,000 Sun computers for hire. Currently applied to analysis of financial and oil exploration data.	Commercial
IBM http://www-1.ibm.com/grid/	Grid system technologies, standards, and design services.	Commercial

Multiple Sources: Fordahl 2005; Hoffman 2005; Waldrop 2002; Rosedale 2003; Ellisman 2004.

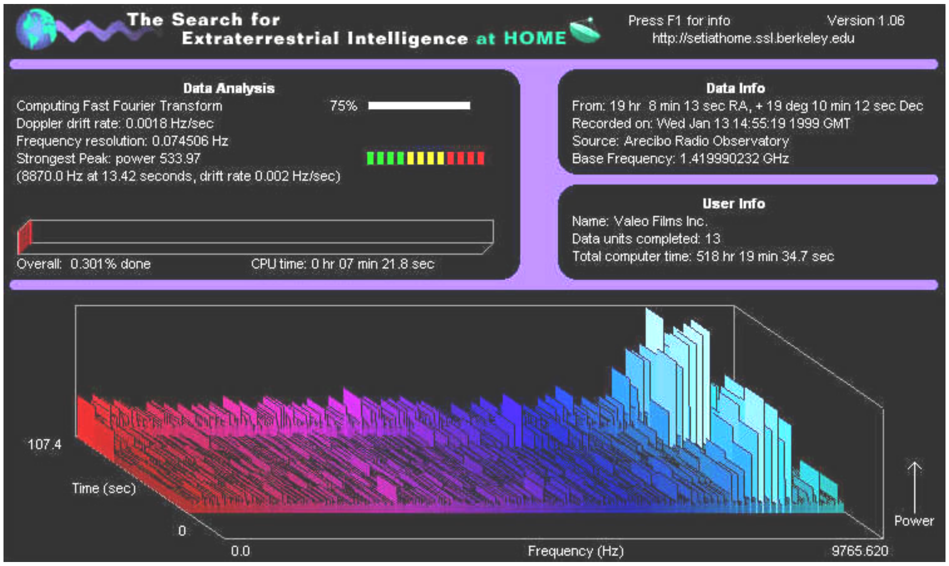

Figure 3. SETI@Home Fourier Transformation Graph.

Source: SETI@Home Web Site

How It Works

Both the Social and the Commercial Model of grid computing systems as described above follow a similar computational distribution scheme. The major difference is that in the social model, the grid computers are distributed throughout homes and businesses around the world. These computers are volunteered to work on socially important problems and there is no financial exchange between the project sponsors and the computer owners. So far, a social model has not been used for a commercial project. More details on the reasons for this are provided in a later section.

In the commercial model, computing resources usually reside within one or more central computing centers controlled and managed by the company selling the service. Vendors must control these resources in order to achieve a quality of service (QoS) level that meets their contract with the client. It is also essential to maintain the security of the data and the software provided by the client.

The specific steps for completing a grid-computing job are illustrated and described in Figure 4.

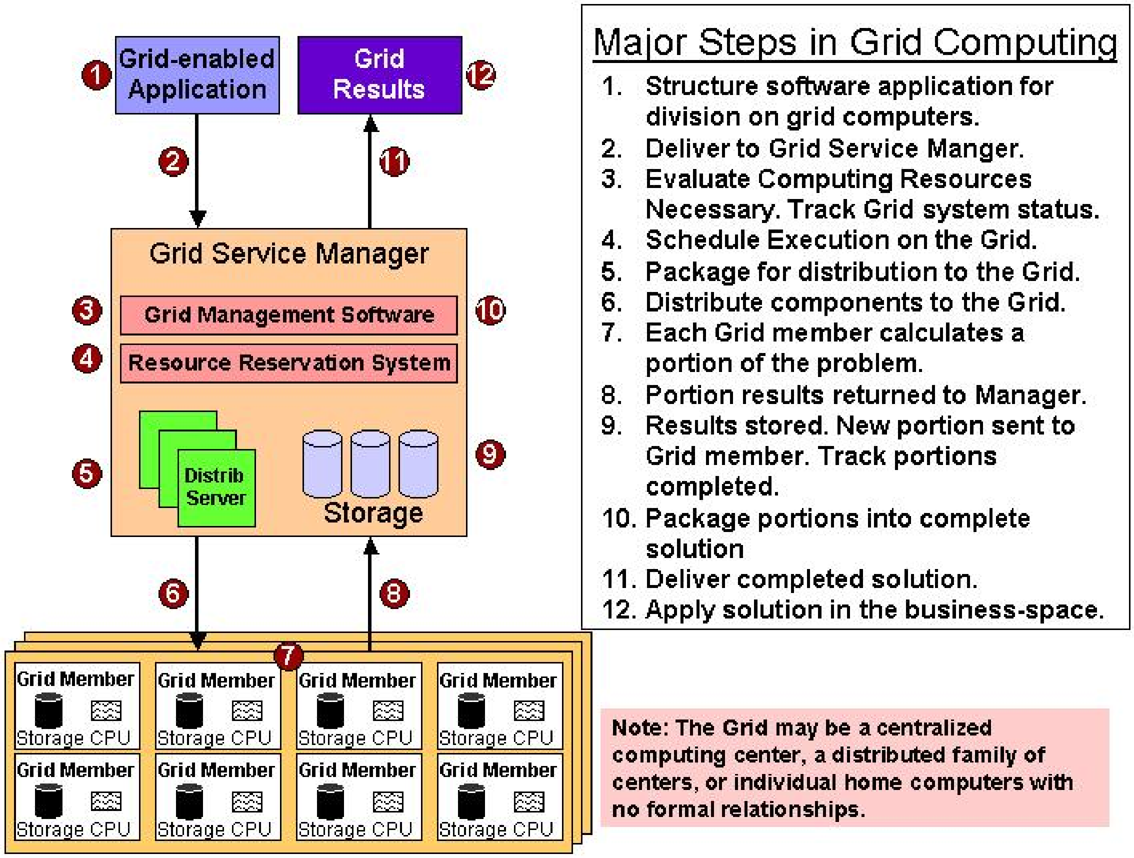

Figure 4. Major Steps in Completing a Grid Computing Job.

Source: Constructed by the author from concepts in all referenced papers.

Researchers at IBM have described a common software architecture to support grid computing (Figure 5). Beginning at the bottom, they emphasize that it is important to separate the functions of the compute servers, storage devices, and distribution networks. The blue and green boxes labeled "OGSA Enabled" represent standardized application programming interfaces (APIs) that are adopted from the Open Grid Services Architecture (http://www.globus.org/ogsa/). The next green layer specifies services for security, workflow, database access, file system management, directory management, and message passing. The purple layer allows all grid systems to be accessed through the World Wide Web and its servers and clients (e.g. the Internet Explorer or Fire Fox browsers). Finally, at the top are the OGSA standard services for communicating between the application source and the grid service. The application contains the software and data for the problem. These have been partitioned such that they can be distributed across a grid computer.

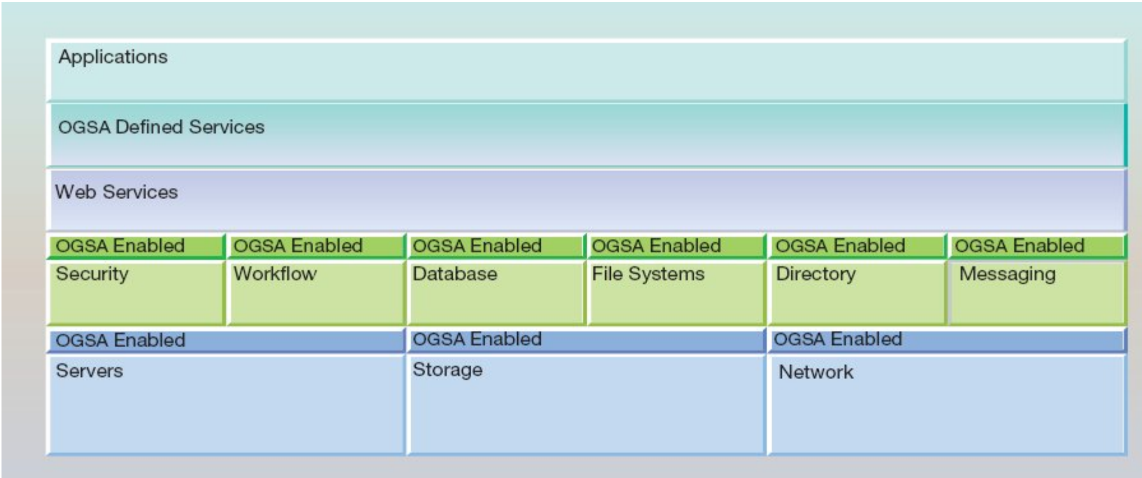

Figure 5. IBM's Architecture for Grid Computing Services

Source: Joseph 2004

Though the grid computing architecture as described above is technically sufficient for providing grid services, it does not contain the necessary functions to support a business offering. To make this technology a business service there must be additional capabilities to meter the amount of computing and storage that are being used by a specific customer's problem. It is also necessary to monitor system performance to insure that a job will be completed in the time specified in the contract with the customer. These and other services are inserted into the previous diagram and show in orange in Figure 6.

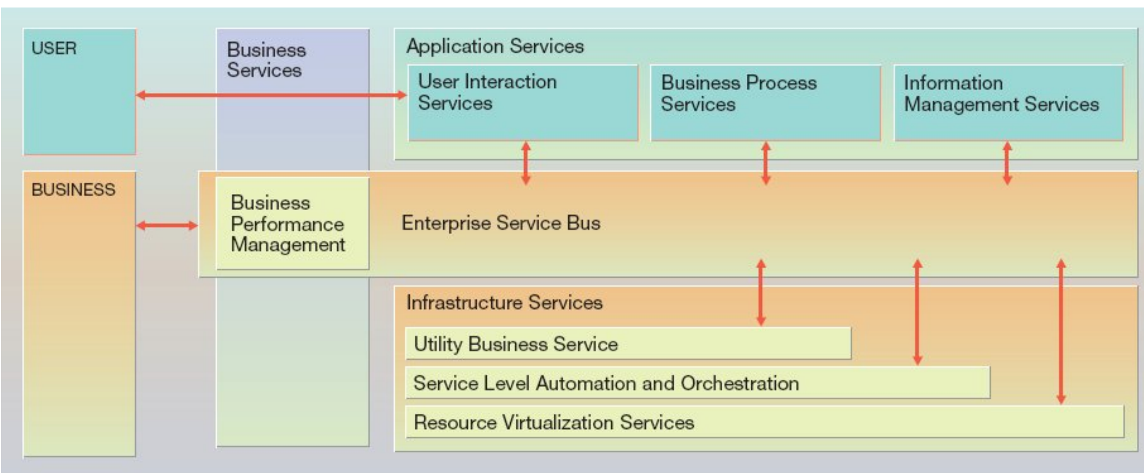

Figure 6. IBM's Modified Grid Architecture for Business Services.

Source: Joseph 2004

Technical Challenges

Grid computing, like shared time and parallel computing before it, has faced a number of technical challenges to its implementation. However, advances in computer power, memory prices, network bandwidth, and distribution services have solved or minimized many of these issues. The technology to support grid computing is affordably available and these services can be provided by a number of vendors.

Hardware

Grid computers can be constructed from any type of computer. Sun Microsystems has used their own Blade Servers and proprietary infrastructure and tools to provide their service. But, a working grid can just as well be created from a group of standard desktop computers as shown in the picture from the Italian National Institute for Nuclear Physics (INFN) in Rome (Figure 7).

Figure 7. Grid Computing Hardware at INFN in Rome.

Source: Italian National Institute for Nuclear Physics

Software

Distributing a problem across a grid of computers requires the use of a software infrastructure. The core capabilities required of this infrastructure have been provided by Ian Foster and a group of programmers working through the Globus Alliance (http://www.globus.org/). Fully 80% of all grid-computing applications reside on top of the Globus software that implements the Open Grid Service Architecture (OGSA) referenced above (Hoffman 2005).

Globus is focused on universal services for grid computing. Vendors like IBM and Sun modify these services or add to them to create capabilities that are important to their business model. It is also sometimes necessary to modify the infrastructure to operate with the proprietary corporate tools used in these vendors' configurations.

Networking

Commercially, the distribution of computational data during execution is accomplished on a network controlled by the vendor. When the computers are centralized, the network can be a very fast gigabit Ethernet with modifications to make it even faster. When the computer resources are distributed across multiple sites, the vendor can purchase networking services sufficient to meet its needs. Linden Labs has discovered that the average data exchange rates between computers on their grid is 100Kbps. This is a very nominal amount for small grids and only becomes an issue when there are 10,000 computers in the grid. However, in this case, hardware clustering can be arranged so that computers are distributing data only to a few local neighbors and do not address all of the computers in the grid (Rosedale, 2003).

Customers

The most significant issues associated with grid computing come from the transition to commercial applications. Security and quality of service are two of the primary concerns of commercial grid customers. When a customer entrusts a service provider with software and data that are important to its business, it is imperative that this information not be revealed to the competition or the world in general. Therefore, it is important for the vendor to maintain separation between the data and applications of two different customers. This has also led to the encryption of all data exchanged between computing centers and with customers (Gartner 2004).

The second important issue is that of quality of service (QoS). In a social project, it is acceptable for an application like SETI@Home to work on a problem indefinitely. SETI has tapped the resources for millions of computers for nearly eight years. Such unbounded problems are not characteristic of a commercial project. A vendor must be able to assess the computational resources required by a customer's application and determine, within some reasonable bounds, when the job will be completed. The duration and resources required for a job also play heavily into a vendor's contracted price for the service (Joseph, 2004).

Market Leaders

Grid computing applications call upon many different computing resources. As a result, there are a number of companies involved in different aspects of the business. Some of the leaders in this are shown in Table 2.

Table 2. Market Leaders in Grid Computing

Category	Leading Companies
Grid Services	United Devices, IBM, Sun, HP,
Grid Systems	DataSynapse, Platform Computing, Entropia
Disk Storage	EMC, Veritas
Networking	Cisco, Juniper Networks
Computer Architectures	Nauticus, Netteza, Azul
Social Projects	Grid.org, SETI@Home
Standards	Globus Alliance, Global Grid Forum
Software	Globus Alliance, Microsoft, Wolfram Research

Financial, Social, Legal, & Regulatory Issues

The Social and the Commercial models of grid computing are separated because of issues regarding the release of software and data from a commercial customer to a horde of unknown grid client machines across the Internet. Using the SETI@Home model to distribute the client application to anyone who volunteers to help provides no control over where the information goes. For example, an oil exploration contract from Exxon would require analytical software and data from their surveys. This information could be distributed to client computers at Shell Oil, the government of Saudi Arabia, and an employee at the US Department of Commerce. Even in small packets, this represents a security risk that a customer could not accept.

Distributing a commercial project across these machines also raises the issue of ownership of the results. If the computer was volunteered for the job, does the owner have any claim to the results produced? That person may be legally allowed to post the results on the Internet. A grassroots effort to collect the results of all of these packets of data could emerge in the form of a web site to which volunteers could post their data. The goal would be to accumulate enough of it to make it valuable to other parties. Given the aggressiveness with which hackers have created e-mail viruses and cheats for computer games, such a scenario is very likely.

These types of issues have prevented grid service providers from creating a business using the social model of the grid. Early in grid development there were proposals to pay volunteers for the computing resources used. However, this addressed only one small part of the problem.

Researchers like Andrew Grimshaw at the University of Virginia have created a viral form of grid computing. Grimshaw's Legion system is used on the campus network to seek out computers that are not being used and to push a grid application onto those computers to take advantage of the resource. When someone begins using one of these computers, the Legion client detects the activity and places the grid application in dormant mode so that it does not interfere with the human usage (Waldrop 2002). The concern with this type of model is that the grid application and data are difficult to control. The applications tend to spread widely and use resources that perhaps should not be used. Grimshaw and others have recognized that, once released within the corporate domain, the grid application travels among computers very much like a virus.

Market Size

It has taken some time for the concept of grid computing pioneered by SETI@Home in 1997 to transform into a business. Each of the major vendors listed above has a set of customers, but it is not clear how much money each is making in this field. In their 2002 "Grid Engine White Paper", Sun estimates the 2003 grid market at over $4 billion (Figure 8). However, this number seems unlikely based on more current information regarding the grid business.

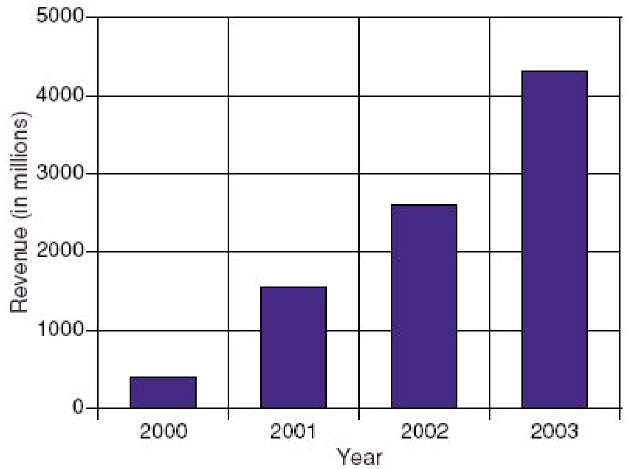

Figure 8. Sun estimate of grid computing market.

Source: Sun 2002

As an example, Sun currently prices its grid services at one dollar per hour for each processor used and one dollar per month for each gigabyte of storage used by a customer. They have established a grid-computing infrastructure of 10,000 processors spread across Texas, New Jersey, Virginia, and Scotland (Fordahl, 2005). If a system like this were running at 100% capacity, 24 hours per day, it would generate $240,000 per day from computing services, and some smaller amount in data storage fees. At this rate, this system would require full utilization for 1,667 days to generate $4 billion in revenue. That is over 4.5 years of continuous, uninterrupted business – and that from a grid system that was created three years after the above market projections were made in 2002.

Though Sun is certainly not the only vendor in this business area, it is difficult to imagine that so many problems have been solved using grid computers. Business and science periodicals typically use social applications of medical problems to illustrate the power of grid systems. Very little has been published with details from a commercial application. However, the fact that Sun has invested in a 10,000-processor grid and the infrastructure to support it, indicates that there are potential customers for the service.

Platform Computing Inc. has conducted a survey of business executives regarding their purchases of grid computing and their plans for such purchases in the future. That survey indicates that grid computing has become an area of interest to many companies and that interest may be growing (Figure 9). Though this is encouraging for the technology, it does not assess the value of the market and the survey was conducted by a grid computing vendor.

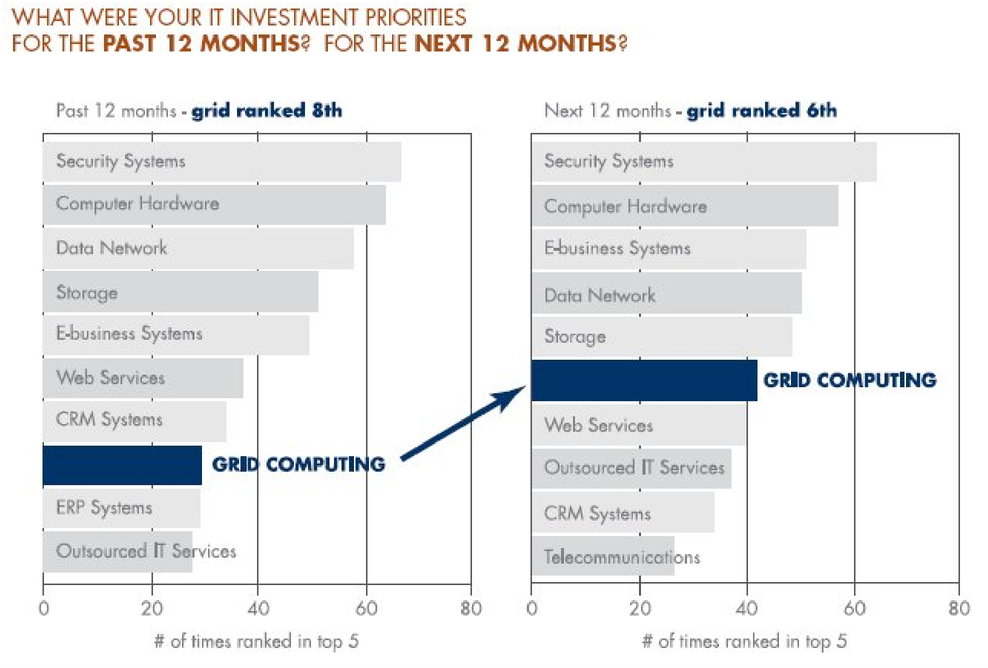

Figure 9. Platform Computing Survey of Executives

Source: Platform Computing 2004

Strategic Positions

United Devices, IBM, Sun, and HP have discovered a grid computing services model that will support a business. It precludes the use of the social model pioneered by SETI@Home and establishes computing centers with thousands of processors for hire. It is not clear to what degree these will be profitable. However, United Devices, DataSynapse, and Platform Computing have built their business on constructing grids for use by customers in the customer's facilities. This model generates cash flow from each contracted job.

If there are sufficient customers to purchase computing time from a central grid, then the first model will generate considerable revenue by renting computers by the hour until the hardware is paid for and the later cash streams cover only operating costs before generating a profit. The second model generates profit from every job completed. However, it is limited to the number of customers who want their own grid system. A third approach is to provide grid management services for these computer systems on customer sites. There is sufficient unique knowledge required for these systems, that in-house IT departments will not be able to operate them. Many customers may also require assistance in dividing their software such that it can be hosted across a grid of thousands of computers.

It appears that the market for providing knowledgeable grid construction, maintenance, and programming services is a more profitable approach to this business at this time. This approach also avoids large investments in computer hardware that will depreciate in value, eventually become outdated, and that may go unused.

The business of providing grid services from a central computing center also runs counter to the industry trend of dropping hardware prices and the emergence of new solutions from intelligent entrepreneurs. Companies like Nauticus, Netteza, and Azul may be able to create a "grid in a box" solution that will eliminate some of the complexities of creating such a system and significantly reduce the associated hardware costs and operating complexities. Such a system would present serious competition for a grid services business such as he one offered by Sun.

References:

Burrows, P. (November 9, 2004). Putting computing power on tap. Business Week Online. http://www.businessweek.com/technology/content/nov2004/tc2004119_4600_tc162.htm

Clabby Analytics. (2004). The Grid Report – 2004 Edition. Clabby Analytics: Yarmouth, ME. Downloaded from IBM grid web site. http://www-1.ibm.com/grid/grid_literature.shtml?&ca=Grid&me=W&met=rightnavliteraturelink

Coomer, J. (August 2002). Introduction to the Cluster Grid – Part 1. Sun Microsystems White Paper. http://www.sun.com/blueprints/0802/816-7444-10.pdf

Coomer, J. (September 2002). Introduction to the Cluster Grid – Part 2. Sun Microsystems White Paper. http://www.sun.com/blueprints/0902/816-7765-10.pdf

Ellisman, M. et.al. (November 2004). The emerging role of BIOGRIDS. Communications of the ACM. Vol 47, Iss 11.

Fordahl, M. (February 2, 2005). Sun Microsystems unveils grid computing initiative. *Technology Review*. http://www.technologyreview.com/articles/05/02/ap/ap_2020205.asp?p=1

Foster, I. (April 2003). The GRID: Computing without bounds. *Scientific American*. Vol 228, Iss 4.

Gartner, J. (January 14, 2004). Grids unleash the power of many. *Technology Review*. http://www.technologyreview.com/articles/05/01/wo/wo_gartner011405.asp?p=0

Hoffer, K.E. (February 1, 2005). The protein grid. *Technology Review*. http://www.technologyreview.com/articles/05/02/wo/wo_hoffman020105.asp?p=0

Joseph, J and Ernest, M. (2004). Evolution of a grid computing architecture and grid adoption models. *IBM Systems Journal*. Vol 43, Iss 4.

Lester, B.P. (1993). *The Art of parallel programming*. Prentice Hall: Englewood Cliffs, NJ.

Lewis, T.G. & El-Rewini, H. (1992). *Introduction to parallel computing*. Prentice Hall: Englewood Cliffs, NJ.

Platform Computing. (2004). The Economics of the grid: Grid computing as a strategic IT investment priority. Platform Computing web site. http://www.platform.com/pdfs/services/professional/Platform_Economics_of_Grid.pdf

Oram, A. (editor). (2001). *Peer-to-Peer: Harnessing the power of disruptive technologies*. O'Reilly & Associates: Sebastopol, CA.
Rosedale, P. and Ondrejka, C. (September 18, 2003). Enabling player-created online worlds with grid computing and streaming. Gamasutra Online. http://www.gamasutra.com/resource_guide/20030916/rosedale_pfv.htm
Sun. (2002). Grid Engine White Paper. Sun Microsystems web site. http://www.sun.com/aboutsun/media/presskits/workstations/sc2001/sun-powers.pdf
Smith, H. and Konsynski, B. (Fall 2004). Grid computing. *MIT Sloan Management Review*. Vol 46, Iss 1.
Theotokis, S and Spinellis, D. (December 2004). A survey of peer-to-peer content distribution technologies. *ACM Computing Surveys*. Vol 36 Iss 4.
Thilmany, J. (March 2003). Getting on the grid. *Mechanical Engineering*. Vol 125, Iss 3.
Waldrop, M. M. (May 2002). Grid Computing. *Technology Review*. http://www.technologyreview.com/articles/02/05/waldrop0502.asp?p=0