Grids – A Low Cost, High Speed Alternative to Traditional High-Performance Computing

1 Introduction

IT Resource Inefficiency Abounds

- **Average Daytime Utilization of Windows servers is less than 5%**
- **Average Daytime Utilization of UNIX servers is between 15-20%**
- **Average Desktop Utilization is less than 5%**

Source: IBM Corporation, *Taurus – Taxonomy of Actual Utilization of Real UNIX and Windows Servers*; GM13-0191-00IBM

Companies today rely on information technology more so than they ever have in the past. Trillions of dollars have been spent in the last decade by companies trying to optimize all aspects of their operations, such as financial, audit, supply chain, back office, sales and marketing, engineering, manufacturing and product development.

Yet, today enterprises stand at a crossroads – most of the investments made in technology have yet to achieve their financial objectives or provide the expected boost in corporate productivity.[1] Additionally much of the deployed infrastructure remains hopelessly under-utilized.[2]

Grid Computing, a fast-maturing technology, has been called the "silver bullet" that will address many of the financial and operational inefficiencies of today's information technology infrastructure.

After developing strong roots in the global academic and research communities over the last decade, Grid Computing has successfully entered the commercial world. It is accelerating product development, reducing infrastructure and operational costs, leveraging existing technology investments and increasing corporate productivity. Today, Grid Computing offers the lowest cost high-throughput solution, enabling companies to migrate from expensive HPC systems.

This paper outlines key factors that should be considered by enterprises as they build business cases for deploying Grid Computing solutions. A *Values Analysis* is conducted and various *value elements* are discussed in detail. Additionally, three types of risks are evaluated.

[1] Standish Group Survey, *The Chaos Report*. Project Failure rate estimated US$ 250B/year. 30% of projects failed, 52% are "challenged" and 16% succeeded.
P. Strassman, *The Squandered Computer*;
Grid Technology Partners, A. Abbas, *Computer Paradox and Grid Computing,* June 2002, GridToday

[2] IBM Corporation, *Taurus – Taxonomy of Actual Utilization of Real UNIX and Windows Servers* – GM13-0191-00; Average daytime utilization across Windows servers is less than 5%, while corresponding aggregate daytime utilization of Unix servers is usually in the range of 15-20%. Daytime desktop utilization is usually less than 10% - and drops to being negligible in the evenings.

2 Grid Computing Business Value Analysis

In the post-bubble economy, value is more important than ever. In this section we highlight some of the key *value elements* that provide credible inputs to the various valuation metrics that can be used by enterprises to build successful Grid Computing deployment business cases. Each of the value elements can be applied to one, two or all of the valuation models that a company may consider using, such as return on investment (ROI), total cost of ownership (TCO) and return on assets (ROA).

2.1 Value Element 1: *Leveraging Existing Hardware Investments & Resources*

As outlined in the sidebar on page one, there is a tremendous amount of unused capacity in the IT infrastructure at a typical enterprise. Grids can be deployed on an enterprise's existing infrastructure, including the multitude of desktops and existing servers, thereby mitigating the need for investment in new HPC systems and other hardware.

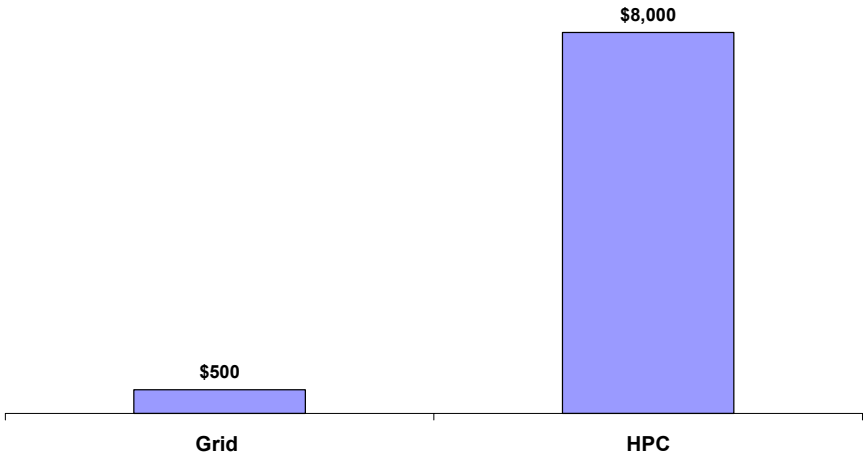

Figure 1: *Relative Cost per Gigaflop for Grid vs. Traditional HPC*

For example, grids deployed on existing desktops and servers provide over 93% in up-front hardware cost savings when compared to HPC systems.[3]

Costs savings are not limited to diminished hardware and software expenditures, but are also derived by eliminating expenditure on air conditioning, electricity and in many cases development of new data centers. It is important that these savings not be overlooked when comparing Grid Computing solutions with SMP based systems.

HPC systems, for example, are heavy pieces of equipment – a single cabinet containing just 16 nodes can weigh approximately 1200 lbs. It and its accessories will produce more than 20,000 BTUs an hour and will need to be kept between 60°F and 68°F.[4] The

[3] Compared with IBM ASCI White. The savings are even larger when compared to multi-way systems from Sun and HP.
[4] A rule of thumb is that electronics reliability is reduced by 50 percent for every increase of 18°F above 70 F. Hence, insufficient cooling even for a short period of time can be devastating to all electronics in the data center.

additional cooling requirements can easily overwhelm even the best of commercial cooling systems and may require expenditure on supplemental cooling systems. Additionally, new power sources may be required if the existing ones cannot handle the two 15 Amp circuits required for each 16 node system.[5] If the above environmental, power and spatial requirements cannot be met within the existing data center, then a new one will have to be constructed (or leased). Current Tier IV data center build cost is US $6,600.00 per 16 node HPC system and generally takes 15 to 20 months to construct.[6] These non-trivial infrastructure upgrade expenses alone are usually enough justification for leveraging existing computational resources by deploying grids.

2.2 Value Element 2: *Reducing Operational Expenses*

Grid Power

Anthrax Research @ United Devices' Global Grid

Using grid-based in-silico techniques, researchers narrowed down potential candidates for an anthrax vaccine from 3.5 billion to 100,000 in less than 18 days using grid software and services valued in the low hundreds of thousands of dollars.

Had an HPC system been deployed, the cost would have been prohibitive and taken 1.2 years of continuous computation.

Grid Computing brings a level of automation and ease previously unseen in the enterprise IT environment. Key self-healing and self-optimizing capabilities free system administrators to focus on high- value activities that are longer term and more impactful. The ability for grids to cross departmental and geographical boundaries uniformly increases the level of computational capacity across the whole enterprise and enhances the level of redundancy in the infrastructure. This is a major breakthrough for system administrators who always seem to be chasing systems outages.

The operational expenses of a Grid Computing deployment are 73% less than for comparable HPC based solutions.[7] Many of the existing cluster solutions are based on open source cluster management software that is complex and unsupported. Operational expenses associated with these deployments have been so high that many enterprises are being forced to outsource management of HPC systems to the suppliers themselves.

Additionally, both small and large enterprise grids are being deployed in as quickly as two days – with little or no disruption to operations. Cluster system deployments on the other hand are taking 60-90 days, in addition to the days to configure and deploy the applications. Deployments may take longer if the existing enterprise data center is out of capacity. Additionally, HPC installations in data centers can cause substantial disruptions and potential downtime. In fact, 54% of data center site infrastructure failures are coincident with human activities.[8]

[5] Linux NetworX, *Planning for Your New Arrival*, 2001
[6] W. Pitt Turner, *Industry Standard Tier Classifications Define Site Infrastructure Performance*, The Uptime Institute
[7] Grid Technology Partners analysis based on 600 Gflop grid (desktop and server) and equivalent HPC solution.
[8] Kenneth Brill, *We Have Met the Enemy of Site Uptime and He is US*!, Uptime Institute

2.3 Value Element 3: *Creating a Scalable & Flexible Enterprise IT Infrastructure*

Traditionally, IT managers have been forced into making large step function increases in spending to accommodate slight increases in infrastructure requirements. Even sparsely populated SMP machines can be a capacity & cost overshoot. IT managers either spend the money for a system that remains underutilized or force the users to live under the tyranny of an overloaded system until such time as the load justifies purchase of another system. Neither scenario is tenable in a fast evolving business environment.

Fixed and limited processing capacity	**Nearly unlimited processing power**
Limitations to project scope	**New levels of speed and accuracy**
Large up-front capital expense required	**No capital expense necessary**
Hardware prone to obsolescence	**Power of grid grows with asset upgrades**
Demand often outstrips supply & vice-versa	**Power demand / supply always matched**

Figure 2: *Benefits of Grid vs. Traditional HPC*

Grid Computing allows companies to add resources linearly based on real-time business requirements. These resources can be derived from within the enterprise or from utility computing services.[9] Never again do projects have to be put on hold from lack of computational capacity, data center space or system priority. The entire compute infrastructure of the enterprise is available for harnessing. Grid Computing can help bring about the end of departmental silos by exposing computational assets curtained by server huggers and bureaucracy. Yet, while departments will be making their resources accessible to the whole enterprise, the right Grid Computing solution can allows them to maintain local control.

2.4 Value Element 4: *Accelerating Product Development, Improving Time to Market and Raising Customer Satisfaction*

Earlier we focused on the key cost savings derived from grid deployments. However, in addition to providing cost savings, Grid Computing has a direct impact on the top line by accelerating product development at enterprises and helping bring product to market quicker. The dramatic reduction in, for example, simulation times can get products completed quickly. This also provides the ability to perform much more detailed and exhaustive product design – since the computational resources brought to bear by the grid can quickly churn through the complex models and scenarios to detect design flaws.

[9] Gateway Processing on Demand service offers CPU capacity at 15 cents a CPU/Hour.

For example, in the life sciences industry, companies large and small have turned to Grid Computing to shorten the drug discovery and development process. Grid Computing is being used both in the drug discovery phase to screen suitable, drug-like molecules against disease targets and also for clinical simulation, healthcare ecosystem modeling, and pharmacokinetic simulations. In short, Grid Computing is allowing drug companies to get the most out of their R&D expenditures by developing the *right* product and getting it to market in the shortest possible time. Companies can save almost US $5M per month in R&D expenses for each month shaved off the drug development process. [10]

In addition, Grid Computing has been playing a major role in helping the oil and gas industry process this data efficiently and pinpoint suitable areas for drilling. In each of the phases, companies have had to choose between the resolution of data collected and the time to process it. Grid Computing can save companies millions of dollars by allowing them to not only collect and analyze high resolution data to pinpoint drilling sites, but also to do it with great speed. As indicated earlier, the cost of prospecting at the wrong location is substantial. Many oil and gas companies today are either replacing SMP-based solutions or complementing them with grid deployments.

As a final example, in the government sector, agencies and contractors are turning to Grid Computing to support security and intelligence related simulation and analysis projects in such diverse areas as biodefense, weapons & machinery design, warfare modeling, pattern recognition & matching, network load testing, information extraction, and decryption/encryption. Grid Computing provides these organizations with a way to meet the federal E-Government mandate of finding more efficient, cost-effective ways to contribute more effectively to the nation's security.

Virtually all others sectors including financial, insurance, automotive, entertainment and defense are also gaining competitive advantages by deploying grid solutions.

2.5 Value Element 5: *Increasing Productivity*

Enterprises that have deployed Grid Computing are seeing tremendous productivity gains. Consider for example the productivity gains of an electronics design and automation company – run times of jobs submitted by its engineers were reduced by 58% by deploying a grid. Corporate-wide productivity gains by this reduction have been assessed at US $9M annually.[11] Similarly drastic reductions in run times and associated employee productivity gains are being seen in grid deployments in a variety of enterprises.

Productivity gains serve as a crucial measure in any business analysis because they have a direct impact on corporate bottom line. Yet in many instances of technology deployments, calculating productivity gains is more voodoo than science.[12] However, in the case of Grid Computing, enterprises have found it relatively easy to determine the reduction in processing time due to increased computational capacity offered by the grid and the resulting emancipation of employees' time. Thus, productivity gains remain a strong driver for grid deployments.

[10] J.A. DiMasi, *The Price of Innovation: New Estimates of Drug Development Costs*, Journal of Health Economics 22 (2003) 151-185

[11] Grid Technology Partners customer interview, March 2003.

[12] A case in point to illustrate this is email. It is estimated that an employee may spend up to 4 hours a day just managing and processing their email.

3 Risk Analysis

In section 2 we discussed some of the areas to which Grid Computing adds immediate value for companies. In this section we evaluate the key risk factors that usually plague technology deployments and analyze the vulnerabilities of Grid Computing deployments.

3.1 Lock-In

Like most software (and hardware) vendors, Grid Computing vendors would probably prefer it if their software locks-in a customer for a recurring or a future revenue stream. However, in the case of Grid Computing deployments, IT managers will not be making any significant investments in *durable complementary assets, which* promote lock-in. While cluster and SMP solutions will require a significant investment in hardware, software and supporting infrastructure may contribute to customer lock-in. Customers should pay keen attention to which vendors are supporting the Grid Computing standards activities at the Global Grid Forum.

3.2 Switching Costs

Once a grid has been deployed, the primary switching cost will be driven by the effort required to integrate and enable enterprise applications to work on whatever replacement grid infrastructure has been selected. This is generally not too significant, but nonetheless should be carefully reviewed. Another way to mitigate switching costs is to introduce new grid software in the enterprise to support new grid-enabled applications – while letting the existing software deployment and its integration with legacy grid software remain unchanged.

3.3 Project Implementation Failure

The final risk factor is of project failure, either due to bad project management or incorrect needs assessment. One way to mitigate the risk of project failure is to take advantage of hosted pilot and professional services offered by grid software vendors. This will allow the IT manager to accurately pre-assess the suitability of the grid software, level of integration required and feasibility (application speedup times, productivity gains etc.). Hosted pilots are conducted solely on the vendors' data centers and do not impact the operations of the company.

4 Summary

Grid Computing is making a tremendous impact at enterprises in all business sectors. Companies are realizing that Grid Computing offers tremendous return on investment with fast payback and a drastically lower total cost of ownership as compared to other technology solutions such as HPC systems.

Contact Grid Technology Partners to learn how Grid Computing can help you reduce your IT infrastructure cost, increase top line revenue and increase your department and company's productivity.

Appendix: Good-Fit Industries for Grid Computing

Life Sciences — Companies like Novartis and Sanofi~Synthelabo are putting grid solutions to use in areas like sequence alignment, gene identification, homology, computer-aided drug design, molecular modeling, and combinatorial chemistry.

Geosciences — Companies performing reservoir modeling, 2D & 3D seismic processing, and horizontal drilling benefit from Grid's ability to increase scope and improve accuracy of data analysis.

Financial Services — Portfolio modeling, stochastic valuation reporting, Monte Carlo simulations, and asset liability management are among the areas where a grid platform can speed analysis and improve critical decision-making.

Government — From health-related groups to classified organizations, government entities apply grid solutions to their work in areas like advanced simulations (including weapons simulation or atmospheric / life science modeling), pattern recognition, and decryption.

Academia — Academic research organizations like TACC at University of Texas at Austin and Purdue University are finding lower cost alternatives for their processing in a wide variety of areas -- including biochemistry, engineering, acoustic modeling, geosciences, and computational finance.

Industrial Manufacturing — Companies doing work in computational fluid dynamics (CFD), finite element analysis (FEA), crash simulation, process modeling & optimization, and airfoil simulation can realize significant benefits with secure grid solutions.

Electronic Design Automation — The challenge of verifying and simulating significantly complicated circuitry in microprocessor design can be met by grid platforms, which can speed analysis and run completion dramatically.

Entertainment and Media — From graphics processing & rendering to digital content compression & encoding, the work of both large and small studios can be accelerated via grid technology.

Chemical and Material Sciences — Materials scientists, chemists & biochemists, physicists, and chemical engineers performing simulation and modeling can benefit from a grid's increased compute power for faster processing.