Grid for the enterprise

Grid had its beginnings in the mid-1990's in scientific computing but the notion of distributed computing has been around for decades. Grid was originally conceived and designed in this community to allow access to computing resources that were geographically dispersed. The notion was that underutilized resources in places other than where the researchers were physically located could be used. Also fundamental in the formative thinking was the prospect of sharing access to data, typically in the form of files, that was being jointly produced and used by collaborators in disparate locations.

In the interim these notions have further developed. They are still driving forces behind Grid, but they have been significantly extended to encompass more of IT than just computers and data. As we detail later, Grid enables a loosely-coupled, service-based IT environment. It is the broad spectrum of IT "resources" that can be a "service" that elevates Grid beyond just scientific computing. Importantly, Grid will have applicability in a larger cross-section of the IT world, specifically the enterprise. Enterprise is the commercial, large and medium business IT space.

HP is focused on delivering Grid technology to the enterprise to enable our customers to derive benefits in the form of reduced costs and complexity and to allow for doing things that could not be efficiently achieved before.

What is Grid?

Before going further, there is a need to define some terms related to Grid. This is especially important since so many different Grid definitions exist.

We offer the HP point of view on Grid. If you accept the perspective we offer, either as a whole or in part, then you will grasp the strategic importance of what Grid, true Grid, can offer to you. Grid can offer strategic importance whether you are a CIO, IT manager, IT administrator, or an end-user wanting to, singularly or in collaboration, solve problems.

HP is focused on the management and execution of Grid services. This sounds too simple. Before you accept the seeming simplicity of this, we further dissect the statement. We first need to explain a Grid service. In the context of the Grid we render virtually any IT resource in the form of a "Grid service." Think of this service as you do for a service you receive in the consumer sense; you have a need for it, you find it, you request or purchase it and finally it is provided to you. All parts of what we think of as IT can be rendered as Grid services; computer systems, a quantity of computer cycles, storage space, a printer or some printed pages, an application, a data file, a database or set of records in a database, and so on. Once these IT entities are in the form of a Grid service, then the Grid infrastructure itself will allow them to be registered, discovered, provisioned, accessed, shared, removed, managed, monitored, metered, and even billed for. These last sentences describe, albeit briefly, a very powerful concept. There is however, yet another aspect. Grid services and the supporting Grid technology enables the secure sharing and access to these services by members of a virtual organization. A virtual organization is one or all of the following: ephemeral, geographically distributed, in separate ownership or management domains, and has specific membership. The high-level view is then that Grid enables a loosely-coupled, service-based IT world.

HP is focused on the industrial-strength management and execution of Grid services in the scientific space and importantly, also in the commercial enterprise.

Why Grid now?—Technology enablers

There are three important enablers that make Grid technologies viable.

- **The global reach of the Internet.** Geographic proximity is no longer required to conduct commerce, science or business. The ability to access IT resources from anywhere, to anywhere is today almost pervasive. This is a direct result of the growing use of IP (Internet Protocol) addressing being used on IT resources. Over the coming years it will be even more so as IP addressing will be applied to even more things in our business and personal lives.
- **Abundant bandwidth between entities connected to the Internet.** This permits the exchange of substantial amounts of information which could be in the form of data or applications. The available bandwidth is growing at significant rates further "shrinking" the distance between IT resources where ever they reside.
- **The development of open standards for "Web Services" and "Grid Services."** These terms refer to a broad set of definitions, at various stages of evolution and industry acceptance, to permit applications to communicate. The core Web Services specifications permit applications to discover each other and communicate over the internet using standard languages and protocols such as XML, WSDL, SOAP, and UDDI. The adoption of this core technology by the Grid community, and development of extensions such as OGSI (Open Grid Services Interface) designed to meet Grid requirements, has been a key development over the past two years. Grid Services extend Web Services with a set of interfaces and behaviors defined by OGSI, in areas such as service state and transient service instances.

Why Grid now ?—Business drivers

With the above enablers and the maturing state of Grid technologies we see them, in concert, being directed toward solving several emerging mega-trends in IT.

- **The rate of accumulation of IT gear is significant.** Servers, storage, PCs, laptops, printers, switches, routers, PDAs, etc. are filling up data centers and offices. With this volume comes complexity and difficult management and utilization issues.
- **Centralization will be the bane of tomorrow.** Co-located resources managed by a single central authority will eventually, if not already, become unsustainable. Multi-national companies are not centralized. Corporate data centers are, of necessity, not all in one location.
- **The ownership of the content of science, commerce and consumers is dispersed, but not necessarily public.** It resides in geographically removed locations, within separate ownership domains and separate management domains. However, if it is not accessible to collaborators, partners, customers and colleagues then it has little value.
- **Geographic dispersion of virtual teams is essential, and it's happening today.** Very few of us only interact with colleagues in the office next door.
- **The Earth's rotation leads to a "follow the sun" model.** While it is sleep-time at one organization's data center it is prime-time at another. Spreading computing load across otherwise lightly loaded resources provides better ROI.

Somewhat glibly, we can envision all roads, driven by the business enablers we mentioned and aided by the technology enablers, leading to a distributed and dynamic resource sharing model that is loosely-coupled.

Why Grids?

If you are a CIO or IT manager you're probably being asked to do more with less. Global management of IT requires that CIOs be able to guarantee the security and reliability of the systems and support every aspect of a business. So this really places a new level of importance and focus on the IT function. It must be able to simultaneously, in three dimensions—predictably, flexibly and reliably—provide a significant return on IT.

The corporate IT function is accountable for driving down costs, creating new value, making IT as a service more agile, secure, reliable, flexible and adaptable, to any sort of change. In today's business environment that is the only constant. Virtually all CIOs are faced with one thing, and that is, they demand more from their IT infrastructure and from their IT partners.

What is needed is a platform for managing change, which links business and IT together in real time. HP's late CIO, Bob Napier, who successfully managed our own internal IT infrastructure through the largest merger in IT history, coined a phrase that we think really captures the problem: "every business decision that you make triggers an IT event or a series of IT events." One of the things that the Grid will allow you to do is tie the business architecture through service level agreements to the IT architecture. At that level it allows for the necessary operations to provide a flexible and agile architecture, on a global basis.

Proprietary architectures with vertically integrated IT stacks have become islands of automation today. They are prohibitively expensive to manage and maintain through any significant change. CIOs are specifically demanding a new enterprise architecture, one that's open, one that welcomes change and modularity so that it can be evolved in an incremental fashion. Grid technology can provide this.

Grid mythology

There is a certain amount of "mythology" surrounding Grid. This is partially a result of misunderstanding, over-loading of terms, overlapping technologies and hype. When considering technology insertion and adoption in the enterprise none of these is acceptable.

Let's just pick a few of the more common misunderstandings and debunk them here.

Myth	Reality
Grid computing is just about scientific computing and does not fit commercial applications	Grid computing has been driven by new and demanding applications from the worlds of technical and scientific computing. Just like the internet and the Web were.
Grid computing is restricted to a particular class of applications, those which are "embarrassingly parallel"	Grid computing is about running application workloads on a pool of shared resources. Scientific workloads which are easily parallelized are a particularly good fit to this model. Research and development into commercial grids is being driven by the desire to run commercial workloads across a pool of shared resources, either in an enterprise Grid or managed utility service provision model.
Grid computing will never work because of {security, resource management … insert your favorite technical issue here}.	Grids tend to be deployed today in scenarios where these issues are easy to address, typically within a single administrative domain. Innovation in {security, resource management etc}, especially for multi-domain Grids, is driving the efforts of the system vendors and a growing number of independent software vendors and start-ups.
Grid computing will never work because of {licensing, server hugging, charging … insert your favorite business or political issue here}	The drive towards the grid/utility/service-centric/adaptive model is unstoppable. Over time software, storage, server and IT service businesses will change to adopt this model – driven by the demands of their customers.
Grid computing is restricted to a particular class of applications, those which are "embarrassingly parallel"	Grid computing is about running application workloads on a pool of shared resources. Scientific workloads which are easily parallelized are a particularly good fit to this model. Research and development into commercial grids is being driven by the desire to run commercial workloads across a pool of shared resources, either in an enterprise Grid or managed utility service provision model.

It is misleading that some people label the Grid, by itself, as a product. The reference implementation of the Grid, the Globus toolkit, is just that; a toolkit. There needs to be infrastructure to support it; say servers, instruments, storage, etc. At the highest level an application needs to invoke the Grid middleware to get access to, or enable sharing of, the resources in the form of Grid services. This may be in the form of a portal to more easily connect a ``user'' to a Grid. However, the Grid middleware alone is not a complete solution.

When customers are ready to evaluate Grid as a technology for their environment or their problems, it is often valuable to do an evaluation of deploying Grid as a solution. This can involve an analysis of the situation and the problem to be addressed to simply determine if Grid is in fact the right answer. If Grid technology is appropriate then it still requires a scoping of the effort, the actual deployment, and the enablement of the environment. It may even extend to maintaining the Grid. Net result is that energy needs to be applied to derive value from Grid.

Management and execution of Grid services

When the commercial enterprise considers technology adoption their requirements are more stringent than in the scientific space. Knowing this, HP is focused on the management and the execution of Grid services. This is the barrier to entry for enterprise customers.

The ability to manage, including monitoring, metering and billing, the IT resources rendered as Grid services is crucial. If there were no ability to manage these services, the enterprise would be powerless to understand or control the IT environment. Without control there is risk of inoperability or failures which would in turn affect the business. Additionally, as we had mentioned previously the growing amount of deployed IT equipment necessitates a robust management solution.

In the commercial space the Grid services are expected to be more complex than the simpler compute-intensive batch processing common in the scientific arena. There will be multi-application services with SLAs and possibly dependencies. Being able to robustly and simply “execute” these types of services is necessary for enterprise customers. As a rendered IT resource is provisioned to a requestor, there still remains technologically challenging tasks to insure successful delivery of the service and its results.

For the above reasons HP is focused on the **management** and **execution** of Grid services. We have developed several technologies that address these needs.

Challenges and requirements for commercial Grids

To fulfill the promise of Grid we, as a vendor, and as an industry, have much work to do. This is because hard problems exist and solutions to them are required. Here we highlight a few of the difficult issues and what, as an example, HP is doing about them.

The hard problems that need to be addressed for commercial enterprise use are; fully open standards, heterogeneity, complexity, robustness, and trust and security. The only way that Grid will be able to provide the universality it offers is if it is built on open standards. Another part of the universal applicability is to ensure it can interoperate completely in heterogeneous IT environments. In this way all IT resources, regardless of manufacturer, can participate and be rendered in a enterprise Grid. The complexity of the deployed Grid solutions in the enterprise will require a robust management solution to reduce and mask this complexity, making the most efficient use of the enterprise CIO or IT manager's time and resources. The body of code for the Grid middleware must be robust and rigorously tested; software best practices need to be applied to significantly limit risk and downtime. The establishment of trusted and secure solutions is a baseline necessity for Grids to be deployed beyond the four-walls of a single enterprise.

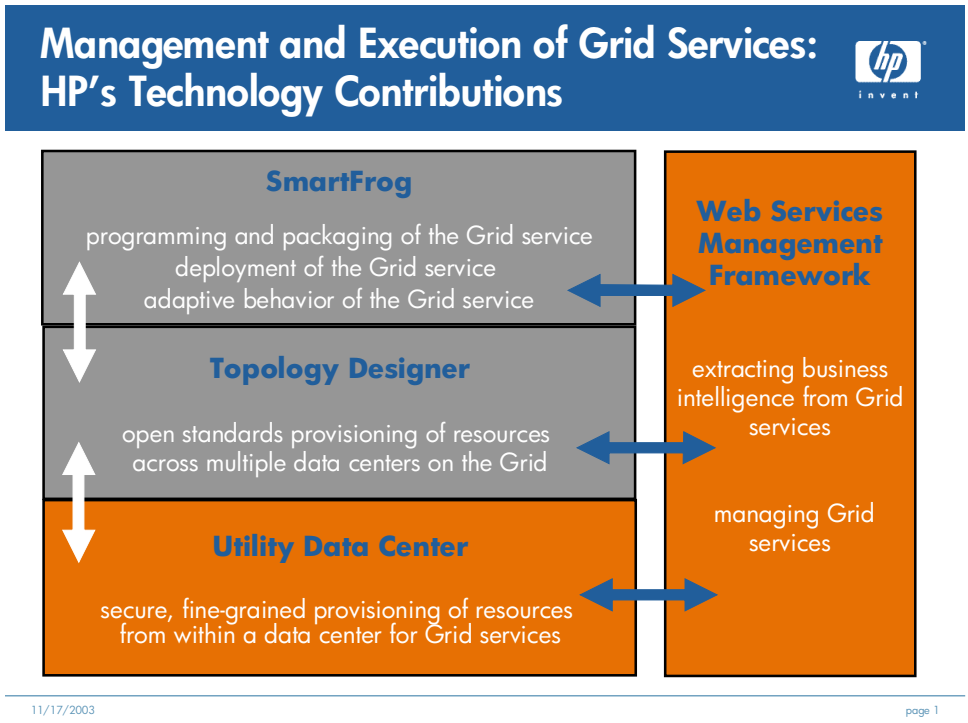
In the figure below we illustrate four technologies—UDC, Topology Designer, SmartFrog and WSMF—that are both inter-related and address some of the hard problems in making Grid suitable for the enterprise.

UDC: The Utility Data Center is a wire-once, dynamic, programmatically configurable, data center with fine-grained resource allocation and hardware security from resources in a shared, virtualized resource pool. The UDC lets organizations allocate and reallocate resources on the fly, transforming the data centre from a static repository of applications and data into a dynamic computer-power generating facility. This HP product has value to the enterprise but in this paper we point out that the marriage of UDC with Grid is particularly powerful. If a UDC is presented with a request for a Grid service that it can not currently meet, it could programmatically reconfigure to meet the request. For UDC and Grid, the whole is greater than the sum of the parts.

Topology designer: HP Labs has developed a software interface to enable the design of Grids of UDCs, or data centers, via an open, easy to use GUI. This reduces the complexity of defining and managing Grids.

SmartFrog: HP Labs has developed a software technology to enable the execution of complex service "packages" when they are invoked as a Grid service. For example, consider a rendering application that has SLAs, version requirements, input data coherency requirements, etc. SmartFrog has a descriptive language and agents to reliably deploy and execute such Grid services.

WSMF: HP's Web Service Management Framework was designed to mange Web services *using* Web services. It is being refactored to do likewise for Grid services. Importantly WSMF allows for management, monitoring, control, metering and event handling for infrastructure, applications, Grid services and business practices. As an open, heterogeneous standard, WSMF has been submitted to the WSDM working group of the OASIS body. WSMF has been used to allow operational (IT) and business (sales and revenue) views of services.



The Future

Hewlett-Packard believes that Grid technology is a significant technology that will permeate the IT industry and affect how science, business and e-commerce are performed. We are at the beginning of the path to this, but it is not going to occur as rapidly as some predict. This is the normal evolutionary route new, and complex, technologies follow. Technologies that are as broad as Grid always evolve in an organic way.

HP, as the World's largest consumer IT company, also has an interest and intent to push Grid technologies into the arena of printers, appliances, PDAs, etc. Reflection on the points we made about what Grid can provide to the enterprise, in fact, map to the consumer and appliance space. For example, think of a printer as a Grid service that can be discovered and provisioned by a Grid user. The fulfillment of this is further in the future but it yields a unifying structure to all of IT.

The call to action is to focus on open standards and industrial strength implementations of Grid technologies such that the commercial enterprise is enabled to utilize the technology. This is not a statement that the scientific space is not important. To the contrary it is in the scientific arena that the ideas are formed and the early deployments are tested.

History may well show that Grid is the next wave of IT technology.

07/2004