Grid, Storage Grid, and the HP StorageWorks Grid

Joseph Zhou ◆ Senior Analyst, Enterprise Storage Solutions ◆ ess@dhbrown.com ◆ August 24, 2004

The concept of Grid computing promises an on-demand and highly utilized computing environment that appeals to many enterprises. Recent developments in server virtualization technologies enable server resources to be centrally managed and distributed – similar to the way a storage network partitions storage resources. The current state of technology development has inspired the concepts of "server Grids" and "storage Grids." HP, an established technology leader in the storage market, has delivered the first "storage Grid-enabled" product among the major storage vendors. The HP StorageWorks Grid, based on Smart Cell technology, is a visionary architecture for an intelligent, scalable, reliable, and agile storage platform. The Smart Cell architecture also promises reduced infrastructure, hardware, and management costs. Nevertheless, DHBA envisions the ideal Grid computing environment to be a single Grid consisting of both server and storage resources; it remains to be seen whether HP will eventually converge its server Grid and storage Grid developments toward this vision.

The Grid Paradigm

As Information Technology (IT) progresses, more and more business processes are being integrated with associated IT solutions in order to achieve operational efficiency. This is causing IT infrastructures in today's enterprise computing environments to grow to enormous sizes. As a result of this growth, organizations are facing constant increases in IT costs. To reduce costs and improve return on investment (ROI), enterprise customers are seeking solutions that enable them to cost effectively manage their infrastructures.

Improvements in the following manageability aspects can help customers realize significant cost savings:

- *Complexity Management* – Simplifying administrative tasks can reduce operational overhead. This simplification can be achieved by unifying management across the IT infrastructure and automating manual processes.
- *Utilization Management* – Increasing the utilization level of each computing device can reduce overall infrastructure costs. Resource sharing by applications and users is key to improving utilization and can be achieved through physical and logical consolidations of the IT infrastructure.

"Grid computing" represents the ultimate solution to both of these manageability concerns. In the Grid computing paradigm, computing resources across the enterprise are connected into a massive grid, which can be centrally managed in a unified and granular manner. The Grid distributes computing resources to all applications and users in an on-demand fashion to eliminate the over-provisioning of resources. An ideal Grid will be embedded with a high-level of intelligence so that it will be self-contained and self-tuned to provide around-the-clock availability and optimal performance with minimal manual intervention.

D.H. Brown Associates, Inc.
222 Grace Church Street
Port Chester, NY 10573

Phone: (914) 937-4302
Fax: (914) 937-2485
www.dhbrown.com

An ideas International Company

Although the Grid concept has existed for quite a while and has been widely promoted in the industry, it remains only a vision for the vast majority of commercial customers. Current Grid implementations are generally proof-of-concept reference systems that lack the advanced functions required by commercial customers. In enterprise environments, where heterogeneous and legacy devices are the norm, Grid computing will not be feasible until key IT solution providers support a common Grid architecture.

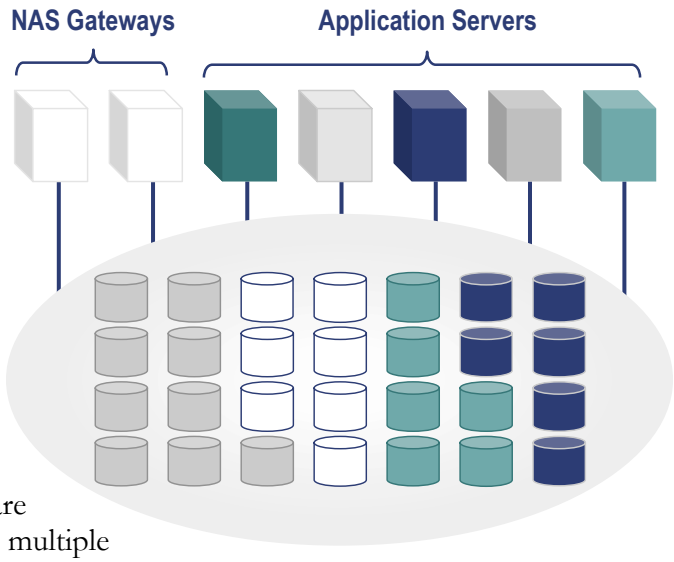
Although industry-standards for Grid computing are still being developed, emerging standards such as the Open Grid Services Architecture (OGSA) defined by the Global Grid Forum (GGF), are helping to bring Grid computing to the forefront. Since OGSA builds on web services, it is heavily influenced by other emerging specifications defined within the World Wide Web Consortium (W3C), the Internet Engineering Task Force (IETF), the Organization for the Advancement of Structured Information Standards (OASIS), and other standards organizations.

The industry is also driving Grid computing into the commercial space. Major IT vendors have incorporated Grid computing into their solution road maps and are working together to develop and refine the Grid standards. The Enterprise Grid Alliance (EGA), a consortium of vendors and customers, was formed to ensure that Grid computing evolves into an open, standards-based, and interoperable environment. More specifically, EGA creates solutions, endorses and supports existing specifications for Grid computing in the enterprise, assembles component specifications, and defines new specifications where needed.

The Separation of Servers and Storage

The concept of sharing resources among systems is not new, as large amounts of underutilized storage devices have always posed an issue for IT systems staff. Many companies have moved from Direct-Attached Storage (DAS) to networked storage (i.e. Storage Area Networks [SANs] or Network-Attached Storage [NAS]) to enable applications and users to share from a common pool of storage resources. Using virtualization techniques, the storage pool can be managed from a single point and partitioned logically. Today's SAN environment consists of multiple servers that tap into a centralized storage pool, as illustrated in Figure 1.

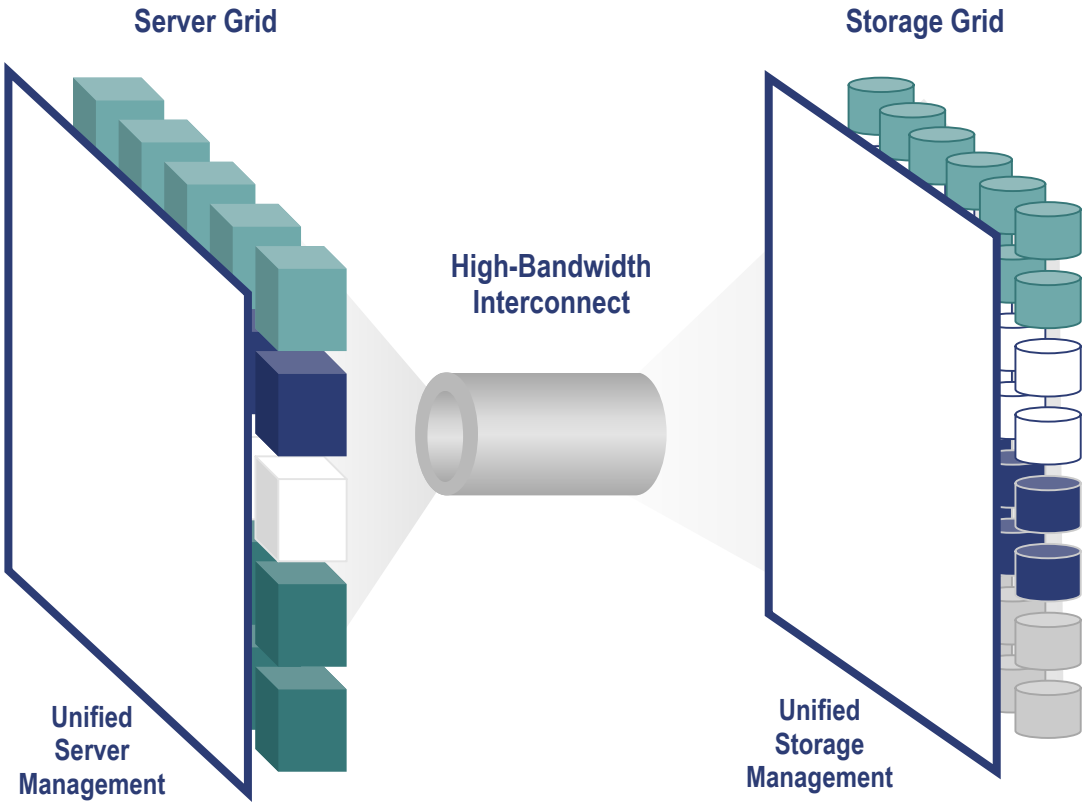
Figure 1: Today's SAN Environment – Network Servers Sharing a Storage Pool



Recent developments in server virtualization technologies, such as the HP Workload Manager, IBM Virtualization Engine, and VMware Virtual Center, enable server resources from multiple systems to be centrally managed and distributed – similar

to the way a storage network partitions storage resources. Rapidly ramping blade technology offers an even higher level of integration among servers along with improved economics by aggregating the physical server resources under a dense chassis. The current state of technology development has inspired the concepts of "server Grids" and "storage Grid." These concepts describe a computing environment that consists of two highly virtualized pools of server and storage resources, which are interconnected with a high-bandwidth I/O pipeline, as illustrated in Figure 2 below.

Figure 2: The Emerging Concepts of "Server Grids" and "Storage Grids"



As with the Grid concept itself, server Grids and storage Grids currently lack industry-standard definitions. Vendors often promote their own definitions to qualify highly virtualized environments as Grids. We believe at a minimum, a Grid must provide unified and centralized management of resources, a capability that already exists with today's networked server and storage technologies. But the qualification of Grid computing also lies in the degree of intelligence embedded into the network. For example, a server farm connected to a SAN presents two physical pools of computing resources that can be logically partitioned for various applications; however, unless their management is fully automated, the two together cannot qualify as server and storage Grids.

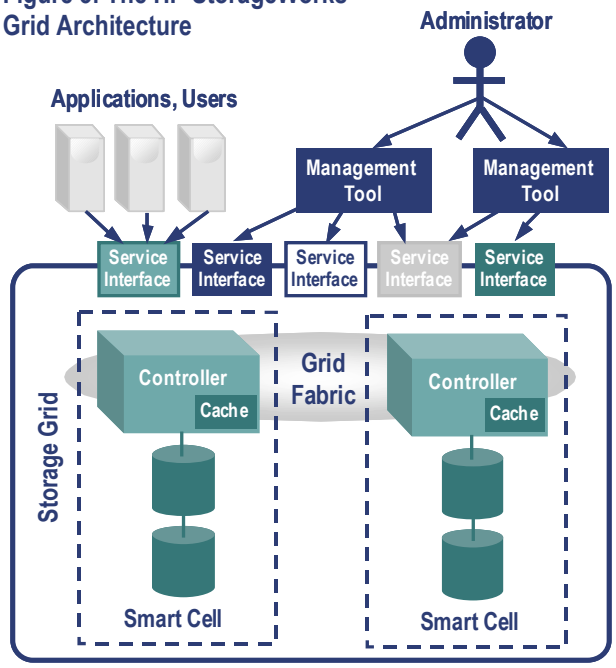
HP Embraces the "Storage Grid" Vision with StorageWorks Grid

HP recently announced its storage Grid architecture, the StorageWorks Grid, as well as its first two Grid-enabled offerings, the Reference Information Storage System (RISS) and Scalable File Share (SFS).

The essence of the HP StorageWorks Grid architecture is Smart Cell technology. Smart Cells serve as building block modules of a highly scalable storage Grid. Each Smart Cell is equipped with its own processor(s), memory, and storage (disks and/or other media). Smart Cell controllers run common software to form a unified Grid system as well as provide extra software services that define the Smart Cell's specific operational capabilities, or personalities, in the network.

The currently available Smart Cells are built with two-way rack-optimized ProLiant servers with internal storage or StorageWorks Modular Smart Arrays (MSA) and Enterprise Virtual Arrays (EVA).[1] HP plans to deliver Smart Cells that support other storage arrays, including non-HP arrays, as well as Smart Cells with other storage media, in the future.

Figure 3: The HP StorageWorks Grid Architecture



Smart Cells are joined into domains – collections of Smart Cells dedicated to a particular purpose within the HP StorageWorks Grid. When new Smart Cells are added into the Grid, they are automatically detected and incorporated into the appropriate domain. With policy-based management, the domains are provisioned automatically and dynamically to satisfy application and other requirements.

High-availability of data is ensured through the cellular structure of the Smart Cells. Redundancy is maintained at the Grid-fabric level instead of the array level. This enables customers to achieve continuous access to data without deploying expensive, high-end storage arrays.

The HP StorageWorks Grid fundamentally differs from a highly virtualized conventional SAN because it is designed to simplify and automate tasks rather than mask the complexity (of a SAN) with management software. The StorageWorks Grid offers a single system image with a service-oriented architecture. The entire Grid is presented to the administrator as a single unit and managed through various service interfaces, as shown in Figure 3. The intelligence embedded into the Grid automatically provisions the resources according to goal-based policies to meet the desired service level objectives – no manual intervention is required except to physically replace failed components. Although a virtualized SAN provides a central point of management, each storage array needs to be managed

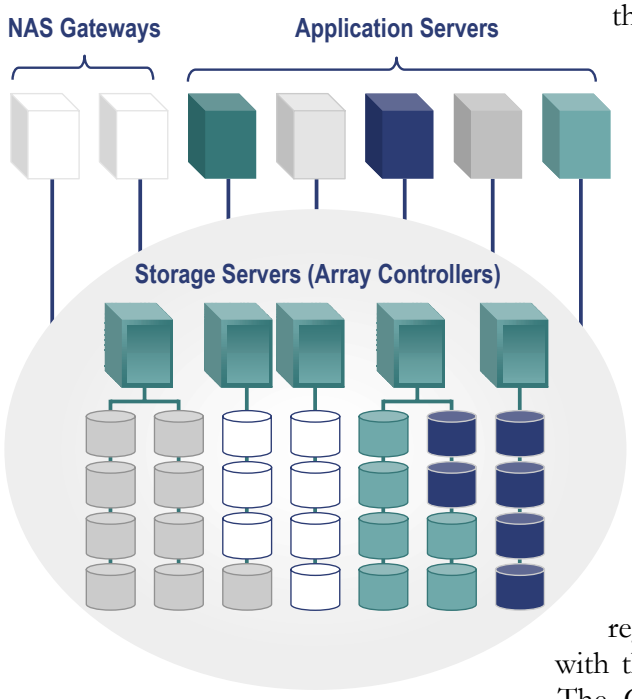
individually and service levels must be maintained manually. For example, a virtualized SAN offers a single pane of glass as a "management portal" that allows simplified access to numerous objects that must be managed, whereas a single system image presents a single object to manage. (The single system image may carry significant underlying complexities, however, these are managed internally by the storage itself.) This is the new centralized management paradigm of the Grid. In addition, the Smart Cells can communicate with each other, load balance, share I/O, and change personalities to better serve the customers' environment/needs. This responsiveness to change, which drives asset utilization and flexibility, is another key benefit of the Grid.

HP currently ships two solutions, RISS and SFS, based on the StorageWorks Grid architecture. RISS is an integrated information lifecycle management (ILM) product that offers high-performance archiving and retrieval, as well as indexing and searching of both structured and unstructured data. SFS is a high-bandwidth storage solution for High-Performance Computing (HPC) Linux clusters. While both products are built with Smart Cells, RISS employs the EXT3 file system with an HP-developed user interface, and SFS implements the Lustre file system with value-added HP features such as automation and resilience capabilities. Today, RISS and SFS are offered as separate products (to be managed as separate Grids). However, HP plans to integrate RISS and SFS, as well as other future StorageWorks Grid-enabled products into a unified Grid environment, at which point RISS and SFS Grids will become RISS and SFS domains in a single StorageWorks Grid.

The Ideal Grid of the Future

The industry visionaries who promoted the concept of separate server and storage Grids generally overlooked the fact that storage networks are enabled by storage servers. Although storage servers are commonly referred to as storage controllers by storage vendors, they are in fact single-purpose servers specialized for serving storage to other network servers (see Figure 4).

Figure 4: Networked Storage Resources Are Served by Storage Servers



In the HP StorageWorks Grid, the Smart Cell controllers are essentially a new breed of storage servers. Smart Cell functionality is enabled by the StorageWorks Grid software loaded on those servers. Consequently, from a hardware perspective, Smart Cells are just servers with storage resources directly attached to them.

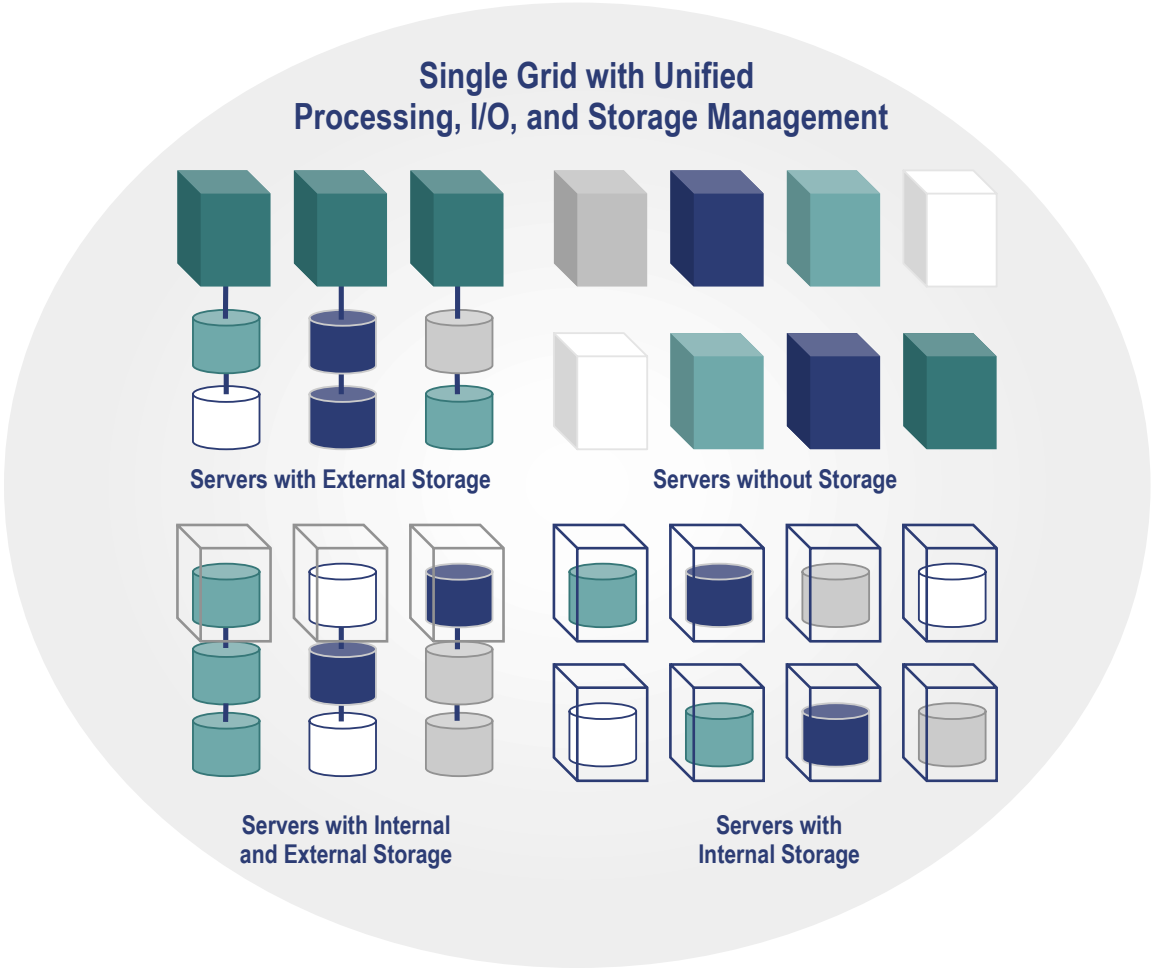
In an ideal Grid environment, as envisioned by DHBA and shown in Figure 5 below, servers will register their processing, I/O, and storage resources with the Grid when they are connected to the Grid fabric. The Grid will then provision the resources according to application and user requirements. Some of the processing

and I/O resources will be dedicated to storage serving, but an entire server will not necessarily be provisioned as solely a storage server. Such Grid implementation will enable greater utilization of resources across the IT infrastructure and further simplify the management of these resources.

Today, the HP Smart Cell technology relies on homogeneous[2] storage controllers – servers running StorageWorks Grid software. All the processing and I/O resources of these servers are dedicated to storage serving. To achieve the ideal state of Grid computing, as envisioned by DHBA, HP needs to:

- Enable the StorageWorks Grid software to run on any server hardware
- Incorporate partitioning technology into the StorageWorks Grid software so that computing resources can be assigned in a more granular fashion

Figure 5: The Ideal Grid – A Single Grid without the Separation of Servers and Storage



Note that an ideal Grid will utilize not only server resources but also computing resources on any system that is connected to the Grid, including PCs and workstations. Local storage resources on PCs and workstations should not be aggregated into the storage resource pool in the Grid because they are dedicated to particular users. However, users can take advantage of the storage resource pool provided by the Grid, which reduces the need for local storage.

Customer Impact

- For customers who are looking to deploy a storage Grid solution and are willing to invest in new infrastructures, the HP StorageWorks Grid should address many of their requirements. However, cost-conscious customers should deploy Smart Cells with internal storage in order to achieve optimal economics. (Note that the HP SFS does not support Smart Cells with internal storage today.)
- Customers who plan to transform their existing storage infrastructure into storage Grids need to work with HP to integrate Smart Cell controllers with their existing storage arrays. However, they may do so at their own risk as the other vendors servicing their existing arrays may not support such implementations. This issue should become a negotiation point with HP.
- Customers who are adamant about avoiding vendor lock-in should either wait until Grid standards are finalized (and multiple vendors are delivering Grid solutions) or negotiate exit clauses into their contracts with HP. Although the StorageWorks Grid promises to be an open platform that embraces heterogeneous environments, customers may decide to deploy other Grid architectures in the future and incorporate the existing StorageWorks Grid components as generic hardware resources rather than as another Grid.

The DHBA Bottom Line

HP has been the industry leader in developing advanced storage solutions, having pioneered SAN technology in the late 1990's. Today with StorageWorks Grid it is the first major vendor to deliver storage Grid-enabled products. HP's StorageWorks Grid is a visionary architecture for an intelligent, scalable, reliable, and agile storage platform. It leverages HP's storage and server technology expertise, as well as research done at HP Labs,[3] to create a new storage environment that delivers long-promised capabilities in novel and very useful ways. Although the initial StorageWorks Grid offerings take the form of point-solutions (archive and file serving), HP has announced a promising road map that will unleash the full benefit of the storage Grid in the near future.[4]

Nevertheless, while StorageWorks Grid delivers a cutting-edge technology solution, the architecture is based on the idea of separate server and storage Grids. DHBA believes the full potential of the Grid concept will not be achieved until a converged Grid consisting of both server and storage resources is created.

[1] RISS supports Smart Cells with internal storage today, and MSA/EVA arrays in the future. SFS supports Smart Cells with MSA/EVA arrays today, and internal storage in the future.

[2] Homogeneous in terms of software functionality only – although the current Smart Cell implementation uses ProLiant servers only, the architecture does not rely on this configuration.

[3] There are proof-of-concept prototypes available in HP Labs today that can be leveraged into market offerings in the near future.

[4] See "HP StorageWorks Grid Strategy White Paper" for details. ftp://ftp.compaq.com/pub/products/storageworks/whitepapers/5982-6007.pdf.