

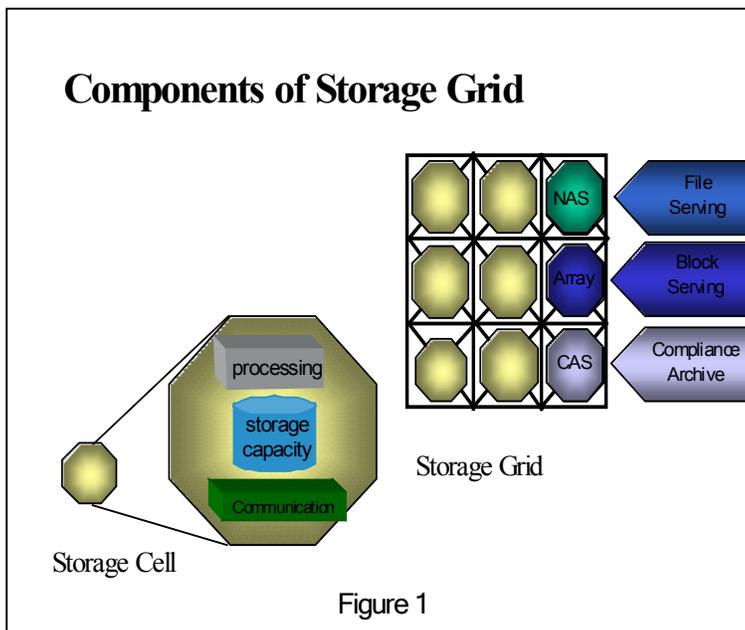
Hewlett Packard's StorageWorks Grid

Date: August, 2004
Author: Nancy Hurley
Title: Senior Analyst

Abstract: Recently, some in the storage industry began to think HP may have forgotten that old slogan, HP Invent. In reality, HP has been working to develop one of the most innovative approaches to meeting storage management challenges on the market. HP's Storage Grid Architecture is a revolutionary new approach to creating a dynamic, flexible, scalable, intelligent storage environment.

What is the StorageWorks Grid

Hewlett Packard has developed a truly unique approach to solving many of the problems that users encounter in today's networked storage environments. Organizations that have networked their storage have certainly reaped benefits from being able to share resources and better centralize management, but let's face it there are still plenty of problems. Although the cost of disk has come down significantly, the dramatic increase in the amount of data that



needs to be stored has caused storage budgets to rise. Today it is still difficult to scale without ultimately increasing the management burden. Despite the myriad of storage network management solutions available today it is extremely difficult to attempt to manage an enterprise storage environment from a central location. Effectively protecting data is still a significant challenge, and now many organizations have the burden of protecting data for compliance purposes. Today, the only way users can address their needs is with a number of disparate hardware and software solutions designed to perform very specific storage tasks.

Yet despite all of these very real issues, storage vendors continually message what seems like pie in the sky visions of automated policy based ILM and utility/adaptive/autonomic/on demand storage environments. Users are having difficulty today meeting TCO goals and effectively meeting backup

windows, so how can they move their organizations to these automated, intelligent environments? Hewlett Packard's strategy is to migrate users to an Adaptive Enterprise via the adoption of the StorageWorks Grid. The StorageWorks Grid is the underlying architecture that will enable users to scale as needed and lower costs with both commodity hardware and flexible software that enables storage "cells" in the Grid to dynamically adapt to business needs on demand (Figure 1).

Unlike other storage vendors' proposals to deliver on a Grid vision with a combination of standalone hardware systems being managed via policy based software overlays, HP has developed a completely integrated solution. Developed with HP Labs, one of the most innovative groups of people on the planet, the StorageWorks team has come up with a "smart cell" based architecture that will enable users to create an intelligent, flexible and scalable storage "Grid" network. This Grid network will allow users to scale in multiple dimensions (capacity, performance, resilience) on an as needed basis, and have the flexibility to use the smart cells for any storage function the business demands. The Grid can be managed as a single image, significantly reducing the complexities associated with managing multiple storage systems and services.

The Grid is made up of commodity "smart cells" which contain front end processors, cache (optional), disk and other storage devices such as tape. The cells can be loaded with different personalities, becoming NAS file systems, block based systems, object oriented archival systems, or perhaps backup targets. HP intends to provide up to 40 different services, and third party vendors will be able to port applications to the storage cells. The true power of this Grid architecture is that each cell is aware of the other and can collaborate in order to distribute loads, increase overall performance through parallel processing and even cache I/Os meant for another "personality". Software personalities can be loaded (and unloaded) dynamically, allowing users to reuse cells based on business requirements. Does this all sound too good to be true? No doubt, we were surprised to hear about this technology and assumed it was years away, however HP is currently shipping solutions that are based on this Grid architecture. The realization of an Adaptive Enterprise utilizing a Storage Grid is closer than you think.

The StorageWorks Grid Architecture

The StorageWorks Grid architecture is made up of "smart cells", each with software personalities loaded to provide distinct storage services. The cells are interconnected via an Ethernet fabric. By interconnecting the cells, users can scale the Grid to meet all of their storage needs. To better understand how the Grid comes together, an in depth description of each element of the Grid is in order.

Smart Cells: Each cell consists of front end processors/controllers, cache, storage devices and networking interfaces. HP will offer multiple cells, varying the number of controller processors, and amount of cache and disk capacity. This will provide low cost cell options for certain storage services, while other services requiring more processing power or more disk capacity will also be available. The cells have network interfaces, either FC or Ethernet for connection to the existing network and interconnection between cells (via an Ethernet switch)

Software Modules: Software functions are loaded onto each cell in order to create "personalities". The amount of controller processors required in each cell will be determined by the functions that will be loaded onto the cell. As an example, Figure 2 illustrates some of the key data and content services that can be loaded onto each cell. The choice and combination of functions loaded will determine the personality and storage service delivered by each cell in the Grid. The data access services support SCSI, NFS and CIFS, enabling each cell to function as a file or block based system.

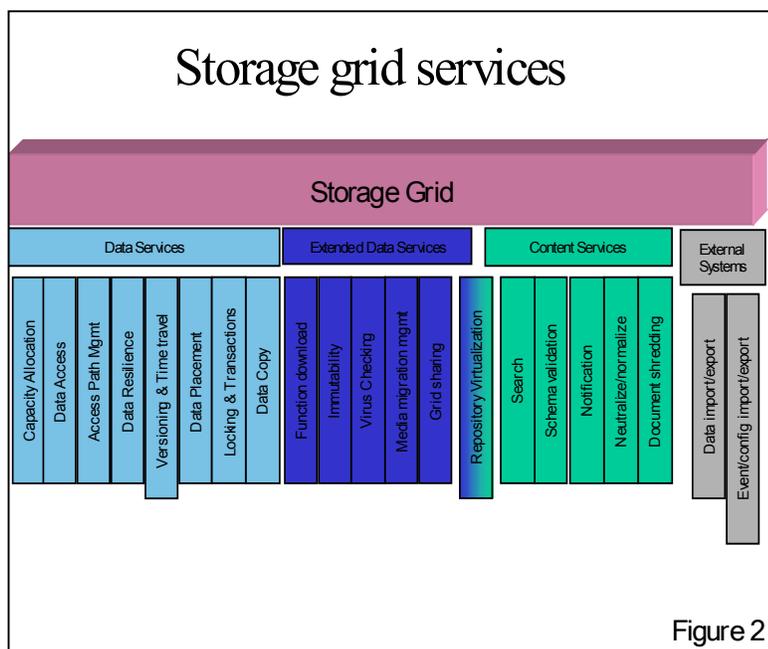


Figure 2

Common Operational Platform: This platform provides the underlying intelligence of the Grid, and it is the key differentiator between HP's Grid architecture and other file or block based scaleable solutions. The smart cells have a common intelligence and operational platform which invokes sophisticated algorithms that determine data placement, protection levels and provide for load balancing across the Grid. The operational platform also provides the means to manage all cells in the Grid as a single entity.

External Services: These services enable the Grid to import or export data to other storage systems and management systems so that the Grid can co-exist in current SAN environments.

In order to create a storage "personality", software functions are loaded onto a cell. Cells with similar personalities are interconnected together to form a domain. (Figure 3) Each domain would be dedicated to a particular service, such as block serving, file serving, archive or backup. The domains will be able to be reconfigured dynamically based on the needs of the business. For example, if a database application being served by a block based domain needs additional capacity or greater throughput, cells in an underutilized NAS domain could be dynamically reconfigured and added to the block level domain. This would involve unloading certain services from the NAS cell and adding services to meet the specification of the block level domain. Any load balancing and file system reconfiguration on the NAS side would occur automatically after the NAS resources were removed. Again, all cells are aware of the other, and the data path is virtualized such that any cell can manage any I/O. Data redundancy and reliability is constantly managed via the underlying operational platform.

One key feature of the Storage Grid is the ability to perform chunking- or fine grain duplicate detection of data. Instead of saving every copy of data and taking up valuable storage capacity, a special mathematical algorithm is applied to streams of data to identify recurring patterns. The system only stores a single copy of each pattern (chunk), and stores pointers to each piece of stored information that contains that pattern/chunk. The result is extremely efficient use of capacity.

HP is currently shipping their StorageWorks Reference Information Storage System (RISS) and the StorageWorks Scalable File Share (SFS), the first offerings based on the Grid architecture. RISS is an intelligent content indexing and archival system; SFS is a scaleable file system for Linux clusters based on Lustre technologies. Both solutions take advantage of the Grid architecture, which provides the ability to scale capacity and performance on an as needed basis. Although these current incarnations of the Grid architecture do not share I/O's between each cell, they will be able to be incorporated into the Grid just as existing monolithic storage systems will in the future.

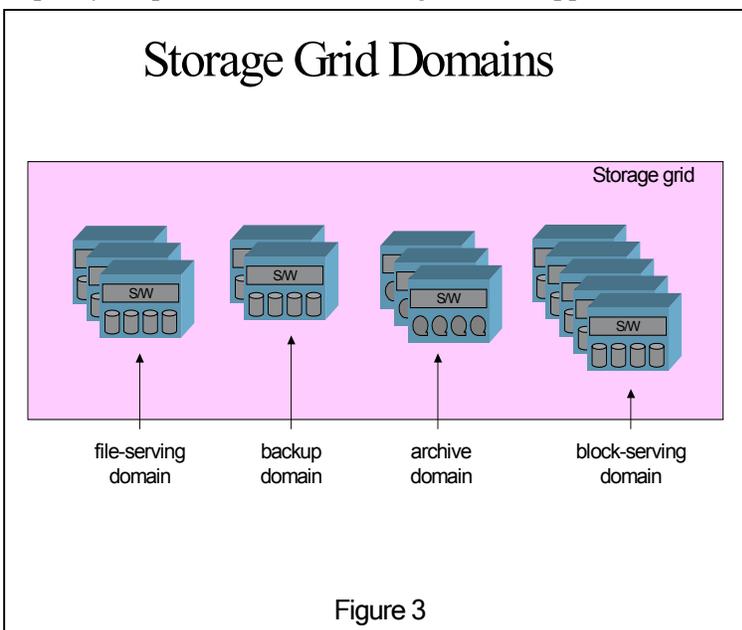


Figure 3

Migration: Implementing StorageWorks Grid

Obviously, HP does not expect users to rip out their existing storage systems and replace them with the smart cells on day one. Actually, the intent is for users to be able to incorporate existing storage resources into the Grid. Initially, users can deploy Grid based solutions such as RISS into their environment to meet current storage requirements. The cells in that Grid implementation could be added to the SAN just as an array is today. Over time, the Grid would expand as additional cells with various personalities (file, block, archive, backup etc) are added based on business demands. All of the cells in the Grid would be managed as a single entity.

In order to reduce management complexities, HP will also introduce a smart cell which will act as a controller “incorporator” for existing/legacy storage systems, enabling those systems to fully participate in the Storage Grid. Thus as the Grid expands, users can take advantage of their existing resources, with the added benefit of layering the intelligence of the Grid on top of those systems. The only requisite is that the storage arrays support SMIS, exposing their storage functions in a standard manner. The vision is that all resources in the environment are fully integrated into the Storage Grid.

There are advantages to this implementation that may not be obvious at first. For instance, the Grid scales as cells are interconnected via an inter-Smart Cell network, which eliminates the need for extensive expansion of the FC fabric. While there will be FC interconnects into the Grid, the architecture does not require that additional FC ports be added as cells are added. This could significantly cut down on the cost of the network infrastructure. As iSCSI becomes more prevalent, and the smart controller aggregator cell becomes available, one could conceive that someday there would be no need for a FC fabric at all.

The implied simplification of management processes is obvious; instead of managing multiple elements an administrator only need manage a single image. Of course considering that single image has multiple personalities and potentially a number of policies orchestrating the interaction between the cells, management will still be complicated, but it will not compare to the complexities of attempting to manage the coordination of actions between multiple disparate systems. Yet what may not be immediately obvious is that this architecture is the incarnation of “ILM in a can”.

ILM processes focus on data placement, retention and migration. Policies determine which storage resource data should be placed on over the course of its lifetime. Often, data will be migrated from a primary storage resource, either file or block, to a secondary or archival system. The data will have various levels of protection over its lifecycle, at times requiring incremental D2D backups and at other times requiring daily updates to tape. HP’s Storage Grid will provide both the hardware and software required to implement ILM, all managed as a single image. Cells in the Grid can act as primary, secondary, backup and archival systems. While sophisticated algorithms select which cells will house certain data based on reliability and performance guidelines, administrators can set policies to further direct data placement for ILM purposes.

Other vendors that are promoting ILM intend to utilize multiple software solutions tied together via a workflow manager to perform ILM processes. The software will have various levels of integration with storage hardware, for the most part the software will need to have access to API’s or SMI-S CIMs in order to manage the placement and migration of data on the arrays or file systems. Few vendors have a complete software portfolio that offers all of the data management solutions required for end to end ILM implementation. HP’s vision of a Grid based ILM implementation is definitely more streamlined and in theory much more scaleable and easier to implement.

Is this Truly Unique?

There are a number of startup vendors that have developed scaleable NAS and block solutions, that on the surface may look to be similar to the HP StorageWorks Grid architecture. However there are a number of differences between what could be considered single personality systems and a collaborative Grid of storage solutions. Scaleable NAS solutions do enable users to easily add capacity by connecting additional “cells”, creating a single file system when adding more capacity. Yet, these cells cannot also be block based, and definitely cannot take on different “personalities” by simply changing which storage services are loaded on the cell. In essence they are in fact purpose built solutions. This is the same for scaleable block based solutions. In contrast, the HP cells are not built for a specific storage function, but rather are flexible enough to take on any of 40 functions (see Figure 3 for a partial list of functions).

Again, the key differentiator and truly unique aspect of the HP solution is the operational platform which enables collaboration between the cells. No other vendor has proposed such a flexible and dynamic architecture to meet all of users’ storage requirements.

The Bottom Line

The StorageWorks Grid is an incredibly innovative architecture that should eliminate a number of storage complexities while reducing TCO. Of course, the full realization of an adaptive Grid environment is still a long term vision. HP has developed “standalone” Grid architecture implementations in RISS and SFS, which prove the architecture can scale in multiple dimensions, increasing performance and growing capacity on an as needed basis. However, there is still some development work to be done before the collaboration between multiple “cell personalities” becomes a reality. HP expects that this capability, which is currently working in the labs, will be available in 2006. Lest anyone thinks that’s too late, the bottom line is no other vendor in the storage market has announced a strategy or roadmap that would bring this level of flexibility, functionality, collaboration and on demand capability to the storage market, and definitely not for commodity pricing. The technological effort that is required for HP to truly deliver on this comes from the power of HP Labs, and only one other vendor in the market has a lab like HP’s dedicated to this type of innovation.

No doubt, the other storage vendors will also develop new solutions to meet the needs of the storage customer; we are not suggesting they will not respond to market requirements and user needs. However, integrating hardware and software to the level HP has in their Grid architecture is no small feat, and we expect other vendors will have to take different approaches simply because their system architecture will demand that- unless they start from scratch to develop a solution similar to HP’s. HP has an incredible opportunity to change the face of the storage market, the technology is there, so this comes down to execution. HP has been a storage powerhouse in the past, if they play their cards right they will turn the concept of an Adaptive Enterprise into a reality.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. and is intended only for use by Subscribers or by persons who have purchased it directly from ESG. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482-0188.