HP StorageWorks grid strategy white paper

Executive summary

Today, many enterprises employ network storage to satisfy diverse, dynamic application requirements. While storage area networks (SANs) and network attached storage (NAS) provide significant benefits to businesses, it is imperative that storage environments evolve dramatically to ensure solutions to ongoing—and accelerating—IT challenges. Going forward, businesses will need greater simplification in acquiring and managing storage, and greater agility in the way storage is provisioned and greater application focus for the capabilities it provides. HP is planning to deliver these requirements by extending networked storage in an innovative way, by creating an HP StorageWorks grid.

The HP StorageWorks grid will be built from intelligent building blocks called Smart Cells. These elements will incorporate commodity hardware. In addition to the expected control and storage hardware, Smart Cells will incorporate a flexible operating environment that allows storage functions to be downloaded as needed—and allows Smart Cells to be repurposed if necessary.

To manage the HP StorageWorks grid, HP will extend HP OpenView Storage Area Manager (HP OpenView SAM) by adding new automated management capabilities and increasing the integration of storage management with the enterprise—as described by the HP Adaptive Enterprise.

HP StorageWorks grid evolution will take advantage of research that has been underway at HP Labs for several years. These projects will provide technologies to strengthen the HP StorageWorks grid's module federation, configuration, and provisioning capabilities—key to creating the holistic storage environment that HP intends to deliver.

This paper describes the HP StorageWorks grid, then explains the grid, its components, and the building blocks with which HP plans to create and evolve it. In addition, this paper introduces key enabling technologies, and discusses compatibility with existing storage environments. This paper supports and extends the information originally provided in the "HP StorageWorks grid: The Storage Vision for the Adaptive Enterprise" white paper.

Introduction

HP introduced the vision of a storage utility in 1998. This utility vision, articulated by the HP Enterprise Network Storage Architecture (ENSA), pioneered SANs as the basis for consolidating storage for heterogeneous servers. Over time, this vision expanded to include comprehensive management that leveraged broad storage virtualization capabilities. Integrated storage resource management—including discovery, configuration, and monitoring for all HP subsystems in a storage network—created a management foundation upon which storage provisioning and other capabilities were layered. HP also advanced the consolidation of block and file services using management appliances to create NAS/SAN convergence. Continuing, HP evolved tiered storage with the introduction of Serial ATA (SATA) and Fibre Attached Technology Adaptive (FATA) drives into HP StorageWorks arrays. Today, HP StorageWorks SANs continue to evolve with increasingly simplified management, greater scalability, broader availability options, and advanced managed capabilities including Information Lifecycle Management (ILM).

HP recognizes the importance of continuing to evolve the storage environment both to accommodate increasingly demanding IT requirements and to anticipate future needs. Therefore, HP is evolving the historical storage utility into a far more capable collaborative storage network called an HP StorageWorks grid—HP is taking a revolutionary step along an evolutionary path. This new environment is driven by the need for storage to support a higher vision for the enterprise—the HP Adaptive Enterprise. It is achieved by incorporating new technologies as they become commercially available. This vision is fully focused on simplifying IT management, through advanced storage management and integrating storage management with other aspects of IT management.

HP StorageWorks contributes a strong foundation from which the grid will evolve—far beyond the necessary first steps taken by networked storage. In addition to protecting customer investments, building on existing capabilities makes the HP StorageWorks grid that much more attainable. The HP StorageWorks Enterprise Virtual Array (EVA) family of storage arrays offers a modular, scalable architecture, with performance and manageability features based on advanced virtualization technology. Other HP storage arrays offer business continuance solutions that will be vital moving forward. The new HP StorageWorks Reference Information Storage System (RISS)—itself an HP StorageWorks grid—offers advanced, scalable ILM capabilities for common types of documents. The HP Extended Tape Library Architecture (ETLA) provides tape library intelligence that lends itself to federation with the HP StorageWorks grid. To bind all of this together, HP will leverage existing Fibre Channel SAN technologies as well as evolving Ethernet fabric capabilities. And, of course, everything will build from industry standards and cooperation.

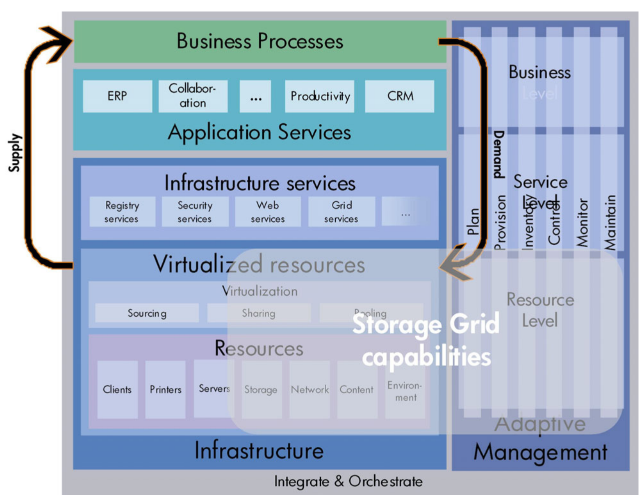
The HP approach

HP Adaptive Enterprise

In 2003, HP introduced a vision for the future of business and IT—the Adaptive Enterprise. It is described by the Darwin architecture (Figure 1), which clearly shows how business requirements drive application requirements, which in turn drive the IT infrastructure that supplies vital information to the enterprise.

HP StorageWorks provides the managed storage infrastructure that stores, delivers, and maintains information. Information is delivered on-demand. With the newly announced ILM thrust, HP has committed to automating the placement, migration, protection, recovery, and retirement of information—according to policies created by the business.

Figure 1. Grid storage delivers Adaptive Enterprise/Darwin value



Today's storage infrastructure (switches, arrays, tape, and optical libraries) provides reliable physical data repositories. Also, storage management applications simplify the management of complex heterogeneous environments. But even with the best of today's management solutions, today's storage administrators must still spend a lot of energy learning and dealing with the peculiarities of each type of storage system in their environment. Even though today's SAN environments are vastly scalable, the scaling often comes at the cost of greater management complexity—in the form of more objects (subsystems to manage, types of subsystems) and new management interfaces to learn. In other words, in spite of all the advances made in the last several years, today's storage environment is still very technical and management remains too device-centric.

To adapt to changes in the business environment, and evolve to take advantage of new technologies, an adaptive enterprise needs a more organic, agile storage environment—a simpler way of growing and managing IT resources. IT administrators must be liberated from device-oriented management and encouraged to focus on business and application needs. They should then be able to input these needs into management applications, which reliably and automatically attend to the technical aspects of managing the ecosystem.

For storage, this is accomplished by creating a physical infrastructure that can easily scale up or out in response to application-driven requirements, and scale up without adding administrative complexity. These capabilities will be delivered through a self-managing HP StorageWorks grid. The HP StorageWorks grid will provide a single, flexible, scalable, shared pool of storage that will include provisioning and other storage management features.

HP StorageWorks grid overview

The HP StorageWorks grid is a collaborative storage network that performs all the functions of today's networked storage while scaling differently and providing fully integrated management of the whole HP StorageWorks grid as though it were a single system. The enhanced storage network scales out by adding disks and other storage devices in modular building blocks. This is similar to adding disks to conventional array controllers or tape drives to libraries. In addition, it scales up by aggregating the performance and other capabilities of its component modules. The aggregation aspect is important: unlike conventional approaches that increase management complexity when adding controllers, the HP StorageWorks grid always presents just one entity to manage, and storage capabilities are always delivered from the resulting single system image. In other words, the promise of large-scale storage pooling—virtualization—is fully realized and taken advantage of. Unlike current approaches that virtualized arrays or tape libraries or other types of storage, the HP StorageWorks grid includes all kinds of storage devices and manages them appropriately. Administrators deal with business application requirements and management software attends to intrinsic storage management tasks.

Figure 2. HP StorageWorks grid highlights

Storage Grid Architecture	Value
Grid-based architecture	•Scaling in multiple dimensions •Greater resilience
Standard Smart Cell hardware	•Reduced cost •Efficient, prompt incorporation of technology advances
Plug-in services architecture	•Simplified addition of new capabilities •Path toward application integration
Downloadable personality traits	•Rapid, dynamic deployment of functionality •Simplified upgrades
Single system image	•Vastly simplified management •Simplified access

The HP StorageWorks grid consists of a modular hardware infrastructure. The individual building block modules, called Smart Cells, are built from high-volume commodity hardware. In addition to storage media (for example, disk or tape drives), the Smart Cells contain a CPU and, optionally, cache memory. Smart Cells are interconnected to form a powerful, flexible storage network. All Smart Cells have a set of common software installed in them. In addition, their operational capabilities, or personality, are defined by extra software services that are loaded on them. Capacity allocation, access path management, data resilience, index/search/retrieval, and copy/replication are examples of these services.

Smart Cells are characterized with useful combinations of physical storage attributes and management control points. Also, all Smart Cells are based on a common operating environment (hardware and software). Networking Smart Cells together creates a peer-to-peer grid that forms a very flexible, unified, agile ecosystem. The ecosystem is capable of real-time scalability and delivering real-time information services as described in the Darwin architecture.

The HP StorageWorks grid represents an evolution of today's SANs—the grid is designed to be completely compatible and coexist with today's network storage deployments. The grid can be added to an existing SAN, like a conventional storage array, by connecting it to an available switch port. Grids can also be created from scratch. To simplify management even further, existing arrays—both from HP and other vendors—can eventually be incorporated into the grid using special Smart Cells. The grid is always managed as a single entity—analogous to a storage array.

HP plans to introduce new technologies and capabilities into the marketplace to deliver the HP StorageWorks grid, which are described in the following sections.

Computational grids and the HP StorageWorks grid

HP is focused on the management and execution of Grid services. In the context of the grid, virtually any IT resource is rendered 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 is thought 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. After these IT entities are in the form of a Grid service, then the grid infrastructure itself enables 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, another aspect. Grid services and the supporting grid technology enable 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.

The HP StorageWorks grid will be a good citizen in the grid environment, which will do the following:

- The HP StorageWorks grid is designed as a *service-oriented architecture*. As explained elsewhere in this paper, the HP StorageWorks grid's capabilities are rendered as services and exposed to its consumers through remotely accessible interfaces.
- Further, these service interfaces will be expressed as Grid/Web services. That is, interface definitions and protocols standardized by the W3C, Grid Global Forum (GGF), and other bodies will be used to ensure that the HP StorageWorks grid's services can be discovered and accessed in the same way as other grid resources.

Additionally, HP is a leading sponsor of the GGF and is participating in its working groups that are standardizing the Open Grid Services Architecture (OGSA) data services.

Key capabilities and technologies

The HP StorageWorks grid provides an architecture for delivering a number of significant advances over today's SANs. This section describes key elements including:

- Modular, extensible architecture
- Single system image—the underlying logical unification that forms the basis for much of the access and management simplicity delivered by the HP StorageWorks grid
- Security—an overall approach toward securing the evolving services for the adaptive enterprise
- Module controller software—a common operating platform upon which management software functionality is loaded
- Data services—traditional storage services that protect data and manage the storage infrastructure
- Extended data services—services that manipulate data and provide advanced protection advanced protection capabilities, and more
- Content services—services that process and present information to applications
- Other services—services for information transfer across systems

Modular, extensible architecture

While today's storage networks deliver a broad range of capabilities, HP believes that a more versatile infrastructure is needed to take the storage environment to the next level. From an application perspective, the most important storage requirement is to provide reliable access to information whenever it is needed. This implies that information is available for delivery at any access point, in conformance with service levels that are set on per-user, per-application, or some other meaningful basis.

From an administrative perspective, storage should be easily acquired, deployed, and managed. Storage capabilities should be scalable in many dimensions including capacity, performance, resilience, and geographic dispersion. Furthermore, it is desirable that storage capabilities be deployed as they are needed. For example, if additional network bandwidth is needed, more switches or fabric connections should be added. If more storage delivery performance is needed, more array controllers should be added. Incremental capacity needs are met with more disk drives or other storage media. This is not possible with today's SAN and storage system products, which tie large disk capacities with fixed amounts of delivery (controller) capacity—capacity and performance attributes are inextricably linked. Adding disks into an array does not ensure that sufficient subsystem performance will be available to deliver data at the same rates as were achieved before the capacity was added.

New modular architecture from HP addresses these issues (Figures 3 and 4). The storage ecosystem is built with intelligent building blocks called Smart Cells. Based on commodity hardware, these modules have attributes like capacity (TB), performance (megabytes-per-second or operations-per-second), resilience (redundancy, speed of recovery), data permanence (archive, backup, read/write properties), and the type of storage provided (block-serving, file-serving, other object-serving). Smart Cells consist of storage devices—disk, tape, or optical drives—a controller module, network interfaces, and other elements needed to create the HP StorageWorks grid's physical infrastructure. The controllers are federated together into a unified grid that presents a single image to manage. Service modules can be downloaded to the modules and deployed as needed.

The HP StorageWorks grid is designed to scale granularly by attribute, rather than by storage subsystem capabilities. Capabilities are added by hardware modules (storage capacity, backup capacity, and so on) and software modules (content indexing and retrieval, replication, and more).

Importantly, the HP StorageWorks grid can be added to an existing SAN fabric (Figure 4). The two connect together through a switch or router. This represents a relatively painless evolutionary path from today's SANs: customers have the choice of adding an HP StorageWorks grid to their existing environment (analogous to the way storage arrays are added today), or migrating data from their existing arrays to the grid. Investment protection continues to be an important consideration in the HP storage strategy.

Figure 3. HP StorageWorks grid

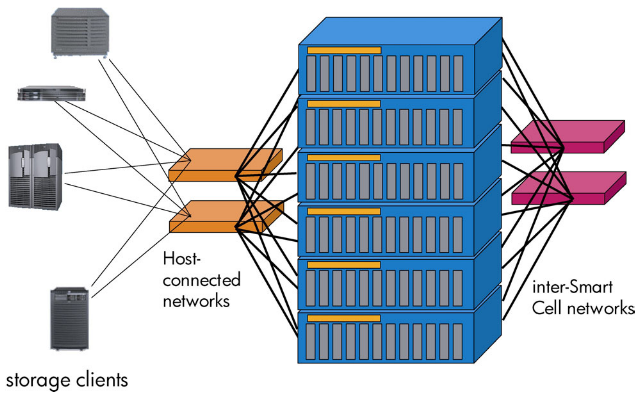


Figure 4. Conventional array controller and HP StorageWorks grid

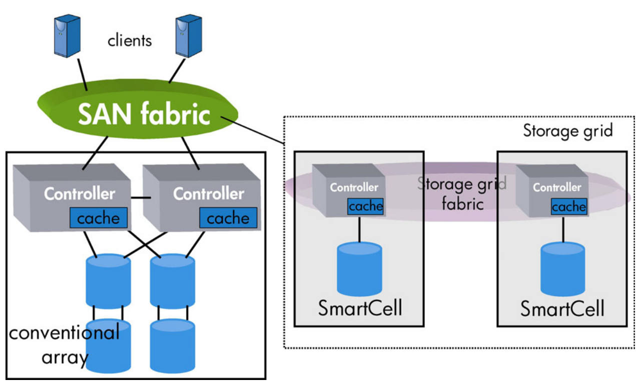
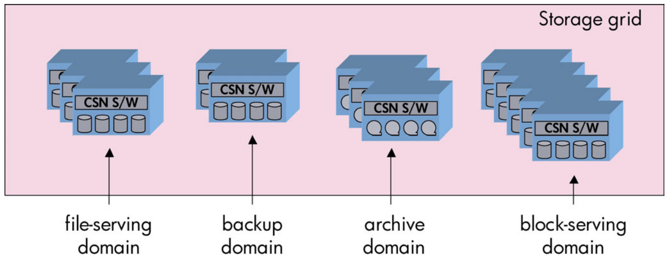
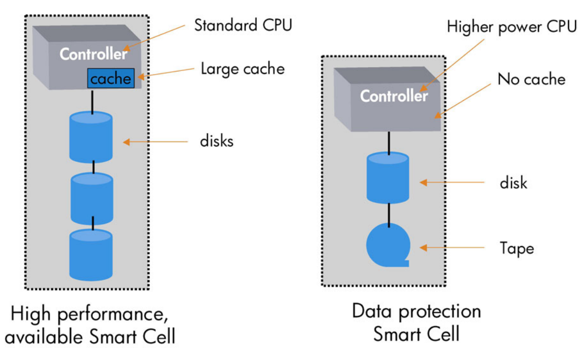


Figure 5. HP StorageWorks grid domains



In the HP StorageWorks grid, Smart Cells function as nodes, and are joined into domains with specific attributes (Figure 5). The grid is expanded by adding new modules, which are automatically detected and automatically incorporated into an appropriate domain. A storage domain is a portion of the HP StorageWorks grid that is made up of Smart Cells of the same type (having the same service modules loaded on it), and dedicated to a particular purpose such as block serving, file serving, archive, or backup. An HP StorageWorks grid may include many domains while still presenting a single system image for management purposes. The domains are provisioned automatically and dynamically to satisfy application and other requirements. For example, if an administrator determines that one domain should have more capacity or throughput, while the demands being made on another domain warrant less capacity, he or she can initiate a dynamic re-provisioning of the Smart Cells and a migration of the effected data. This action could also be automated. This can be accomplished, in part, because the component modules are designed to actively participate in storage management—they can embed more and different functionality than traditional large-scale storage systems. Figure 6 shows examples of how two types of Smart Cells might be constructed.

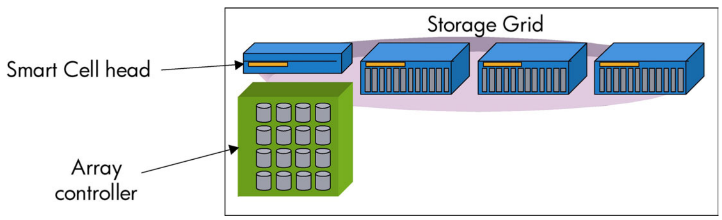
Figure 6. Smart Cell examples



Importantly, Smart Cell software combines to create a unified HP StorageWorks grid. Smart Cell controllers implement all the necessary software to coordinate and behave like a single storage system. This new system, though, is far more capable than today's single-purpose disk arrays or tape libraries. The data path is completely virtualized—any Smart Cell can manage any I/O. Smart Cell software maintains consistency between Smart Cells, as well as ensures data redundancy and reliability. These and other capabilities are all accessed through services and services interfaces, described in the following sections. The result is that the HP StorageWorks grid can incorporate—in very meaningful ways—functionality like automatic placement and movement of data on tiered storage, rapid and intelligent search and retrieval for information, and dynamic provisioning from appropriate storage domains.

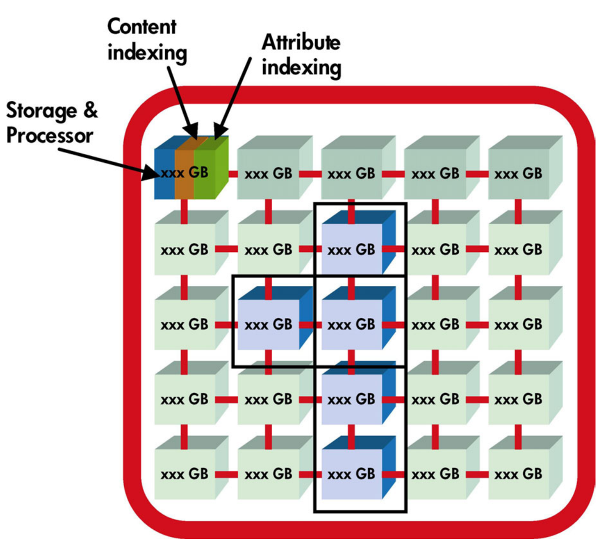
To protect existing investments, the HP StorageWorks grid will be compatible with existing network storage installations. Ultimately this will be achieved in two ways. Initially, a new HP StorageWorks grid can be deployed along side an existing SAN and bridged through a common switch port. Later, for greater integration, a special Smart Cell, which will function like an array controller aggregator and network adapter, will be provided. This special cell will attach in front of qualified conventional storage arrays and provide them with the required connectivity and intelligence to fully participate in the HP StorageWorks grid (Figure 7). Support for the SNIA SMI-S interfaces will be the primary requirement for a legacy array to participate within the HP StorageWorks grid; that is, exposing its features and functions through standard management interfaces will allow the device to become a full participant with the HP StorageWorks grid.

Figure 7. Adapting existing arrays to the HP StorageWorks grid



HP StorageWorks RISS, a powerful ILM solution, now offers a first look at an HP StorageWorks grid solution (Figure 8). In the HP RISS, Smart Cells—shown as cubes in the figure—consist of several disk drives that are controlled by an intelligent module. The module incorporates data placement, index, search, and retrieval services as well as the intelligence and connectivity to enable all modules to be federated into a single managed entity—an HP StorageWorks grid.

Figure 8. HP StorageWorks RISS



Single system image

The HP StorageWorks grid creates a utility that is much easier to manage than today's SANs. This is accomplished in part by integrating all components to present a single system image from an administrative point of view. In other words, the HP StorageWorks grid is viewed and controlled as a single entity. The system is designed from the ground up to be self-managing—tasks traditionally associated with storage resource management are accomplished by the utility itself, with no administrative involvement.

The system includes the necessary features for an administrator to monitor, configure, and control the system as a unit. The system does not require an administrator to know anything about individual elements (Smart Cells). The only exception is that when a hardware failure occurs, manual intervention is needed to repair or replace the failed unit. Importantly, the single system image software will provide fault isolation and failure identification to vastly simplify the task of trying to find a failed component in the sea of components in a grid.

Administrators are required to:

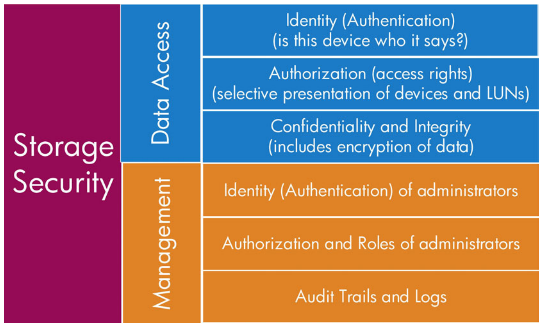
- Deploy or remove Smart Cells from the HP StorageWorks grid, either to increase the scale of the HP StorageWorks grid or to replace failed units.
- Initiate, and monitor the progress of, the installation or upgrade of Smart Cell service modules. This can be to either redeploy a Smart Cell to a different purpose (for example, to migrate Smart Cell capacity from block serving to archiving) or to deploy a new version of a service module.
- Monitor HP StorageWorks grid usage and performance, which enables the administrator to determine when Smart Cells should be added, removed, or redeployed, or failed modules replaced.
- Set policies to control certain aspects of the HP StorageWorks grid self-management. For example, policies control the number and placement of data replicas created for the purpose of protecting the system against failures. Another example might be data placement: in most cases the selection of which Smart Cells are used to house which data is determined by sophisticated algorithms with data reliability and system performance being the primary guiding principals. However, in some cases the administrator might want have some control over this (for example, to ensure that the data from one part of his or her organization does not share physical media with another).
- Set access control for the HP StorageWorks grid, that is determine which user, hosts, and applications can access which data and which HP StorageWorks grid services.

Client applications can establish a connection with any Smart Cell. The data path is virtualized more broadly than in today's storage networks, allowing requests to be transparently forwarded to the appropriate Smart Cell.

Security

Security is vital to the integrity of the HP StorageWorks grid and the data it contains. HP Adaptive Enterprise provides a framework showing how both storage and storage security relate to an organization's business processes, IT infrastructure, and overall security model. Policy and governance are driven from the business level; storage security will over time increasingly rely on security services such as centralized identity (logon). Storage security draws not just on the organization's security governance and attitude toward risk, but also on its centralized identity (authentication) and authorization services. While choices made for storage are independent of choices made for servers or networking, an attacker will seek weaknesses across all three. Logs and audit trails must collectively span all three, as well.

Figure 9. Security model



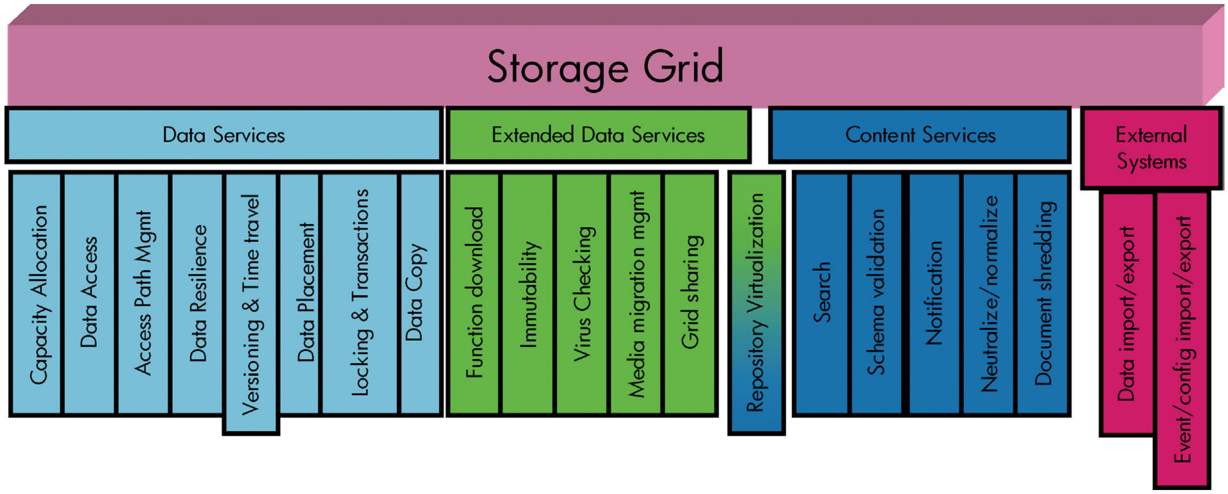
Today's toolbox includes ways to defend physical access (locked data center doors), logical disks (LUN masking, Selective LUN Presentation based on WorldWide Name [WWN]), data access paths (switch-based zoning, for example), data corruption (virus checking), permissions-based logons, and log-based auditing. The HP StorageWorks grid will take advantage of these and additional capabilities, such as those enabled by the emerging Internet Security Protocol (IPsec) and Fibre Channel Security Protocol (FC-SP) standards and central authentication directories (for example, Microsoft® Active Directory). The spectrum of attacks and mitigation methods HP is considering are described in the "Storage Security: emerging storage networking topic of interest".

Module controller software

The HP StorageWorks grid is imbued with a broad spectrum of capabilities, delivered by loadable service modules. The modules contain services—software-based functionality implemented by a Smart Cell and made available to a consumer (host, administrator, or other entity) through a well-defined remote interface. The service modules thus provide the features and management capabilities that reside in each Smart Cell's intelligent controller module. Since there will be several types of Smart Cells, differing by storage device, amount of cache, and other attributes, not all Smart Cells will be able to accept all features. This is needed, for example, to ensure the ability to create low-cost grids for cost-sensitive applications. Modular software services—including data and management services—are loaded into appropriate Smart Cells by a function download service interface. Because of the high level of infrastructure integration within the HP StorageWorks grid, any Smart Cell can service any request and forward requests as needed to any other module in a common domain.

Services fall into six broad categories, the most important four of which are illustrated in Figure 10. Services impart features—scalability, extensibility, resilience, geographic distribution, data access methods, manageability, and more—to Smart Cells. This section describes the most important services.

Figure 10. HP StorageWorks grid services



Data services

Data services enable the persistent storage of and access to data. These services together provide all the facilities normally associated with a storage system.

- The container allocation service creates containers (LUNs, file systems, and other logical devices) and allocates storage capacity to them. Invoking container services also destroy containers (obliterate information so it can no longer be read or used) when needed.
- Data access services store, update, retrieve, and delete data. They support interfaces to a broad variety of storage objects including blocks, files, archival, and backup. As well as supporting today's standard access interfaces and protocols (SCSI, NFS, CIFS, and so on), the data access services model is flexible and extensible, enabling future interfaces to be supported as the need arises.
- Access path management services establish, monitor, and manage the access path between storage clients and their data, for example, managing access permissions and multi-pathing requirements. They also are used to ensure that data is delivered with appropriate Quality of Service attributes—to defined service levels.
- Data resilience services ensure that data is protected by a mechanism (such as replication or backup) appropriate to the requirements set for the data (input through policy-setting interfaces) and the resources available. These services also ensure that the data continues to exist in spite of utility failures until expiration of the requested retention period.
- Versioning and time travel services enable tracking of versions of data objects—enabling reverting to a prior state to recover from application or human errors. Storage clients determine the frequency and criteria for creating versions or snapshots of their data. This powerful capability will effectively provide versioning capabilities to non-versioning operating systems like Microsoft Windows®.
- Data placement gives the HP StorageWorks grid functionality to handle data placement across storage containers, domains, and geographies. In most cases the selection of the appropriate locations for the data within the HP StorageWorks grid is determined automatically by algorithms within the Smart Cells. However, because business policies and application requirements may influence data placement, data placement services will allow administrators to have some control over the placement activities by setting appropriate policies.

- Copy services enable the creation of a copy of a data set for new purposes, such as data mining or application testing. Unlike resilience and other services, copy services are focused on spawning copies for uses other than data protection and recovery.

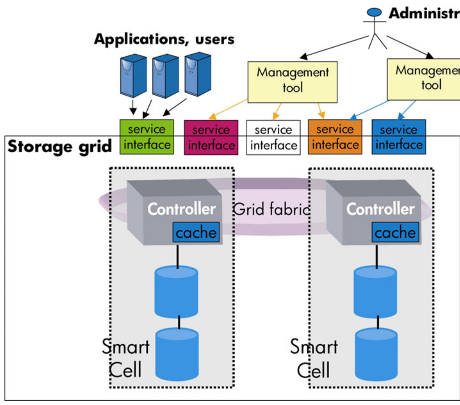
Extended data services

Extended data services are a set of facilities that extend the HP StorageWorks grid's capabilities beyond those normally associated with storage systems. These services enable the downloading of functions for data transformation or mining into the storage system.

- The function download service is the interface by which storage clients load/unload and enable/disable third-party software modules loaded into a secure, virtualized environment in Smart Cells. Each of the downloaded modules may themselves present an external interface through which consumers connect and utilize the module's capabilities. This important capability means that third parties can extend the set of services developed by HP. For example, database applications could download code that offloads search functions to Smart Cells. Other potential uses are for data format conversion routines or allowing the utility to trigger actions based on classifications based on content, attributes, time, and other criteria.
- Immutability services are used to label data as being provably untamperable. This is accomplished with digital signatures, checksums, and other devices. This is particularly important for records keeping or legal compliance applications.
- Virus checking services are used to check user data for viruses and other undesirable contaminants. The services enable virus checking and similar functionality to be loaded to the HP StorageWorks grid itself, rather than being run at a host. This facilitates data checking on the fly and other useful capabilities.
- Media migration management services are used to respond to administrative policies for controlling data placement among different media types, such as high-speed disk and tape. These are useful in ILM and other managed functions for migrating data as policies dictate. This particular service enables the administrator to set policies controlling how and when data is moved between Smart Cells supporting different media types.
- Grid sharing services are used to identify data to be shared with external entities. They also establish a virtual namespace and set data access permissions. In the future, this service will export interfaces determined and standardized by the GGF.

These services, through appropriate interfaces, enable the utility to be managed by administrators, and provide higher-level functions to the data stored in the utility (Figure 11).

Figure 11. Storage network services interfaces



Content services

Content services are those that deliver features that provide awareness of the content of data objects. These significantly extend the capabilities of the storage beyond those normally associated with storage systems.

- Search services enable fast, intelligent searching of the utility for data meeting specified criteria. These services are invoked by applications or users through application plug-ins. They form the basis of the utility's fast, scalable search and retrieval capabilities.
- Schema validation services enable validation that a document submitted to the repository matches schema requirements. This highly intelligent capability includes schema/type assignment, discovery, inference, and validation (active or passive) mechanisms. It also enables automatic discovery, updating, and deleting of stored schemas.
- Data normalization services enable document comparisons against a schema. In addition, they enable the data to be modified or augmented as appropriate to conform to schemas. This is useful for creating the "application neutrality" that ensures that information is usable in the future, regardless of whether the application that created it still exists. Through data normalization, a document may be stored in several formats.
- Document shredding services, based on a schema, enable the segmentation of an incoming document into component parts for storage in the utility. For example, a multi-media news article may be segmented on deposition into the repository into its component text, graphics, and audio portions.
- Notification services generate events in response to changes in the repository. These services are used to alert clients to changes in the content of the data deposited in the HP StorageWorks grid.

External system services

These services are used to import or export data to other storage systems or can provide configuration or event information to or from an external management system.

Other services

The services described to this point have provided functional capabilities and interfaces to enable application-awareness, content-awareness, service level delivery, data integrity, and security. In addition to these, a number of generic services enable the utility to exchange data or configuration information with external storage or management systems. They provide the key interfaces for managing the utility from enterprise frameworks and other systems.

These services include:

- Scheduling
- Reporting
- Logging
- Policy
- Scheduling services that enable defined activities to be automatically executed at programmed intervals or times
- Access services that form the basis for auditing, authentication, and other broad security-related capabilities
- Rights enforcement

Still other services provide interfaces to storage resource management applications such as configuration management, asset management, health monitoring and reporting, and planning.

Applying services to Smart Cells

It is clear that the vision of HP for the future of storage utilities is comprehensive and ambitious. The vision—a services-rich HP StorageWorks grid that presents a single administrative image—can satisfy all of the requirements of an adaptive enterprise. Smart Cells, the network infrastructure, and many foundational services have been described. Foundational to the HP StorageWorks grid is the concept of a common operational platform. Putting these important concepts together reveals some of the power of the HP approach.

In today's storage environments, arrays and other storage systems are typically deployed on a relatively ad hoc basis—systems are acquired as applications demand capacity or other attributes. While consideration may be given to the overall storage network architecture, systems are generally acquired for specific purposes, and their attributes are weighed at the time of purchase. This results in multiple implementations of similar functionality, with attendant burdens of learning how to manage each of the implementations. For example, different arrays from either the same or different vendors may implement basic configuration management, local and remote replication, and other functions in very different ways. This introduces undesirable management complexity into the infrastructure.

With the HP Smart Cell-based grid, the situation is very different. Smart Cells have a common intelligence and operational platform, with specific storage functionality being deployed either by way of specific types of embedded storage devices or through modularly downloadable services. This means that the personality—storage- and application-relevant attributes—can be easily changed. For example, Figure 12a shows a high performance Smart Cell, while Figure 12b shows an archiving Smart Cell. Each has a specific purpose, and so each is loaded with the appropriate services required to implement the requisite functionality.

Figure 12a. High performance Smart Cell

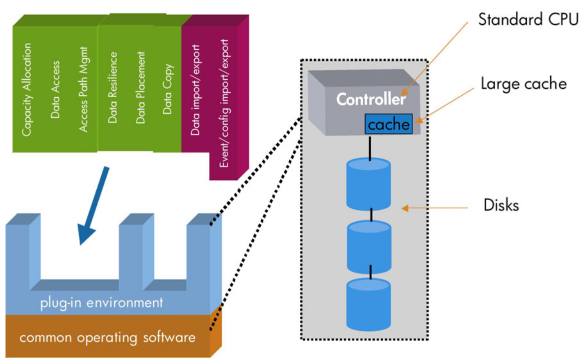
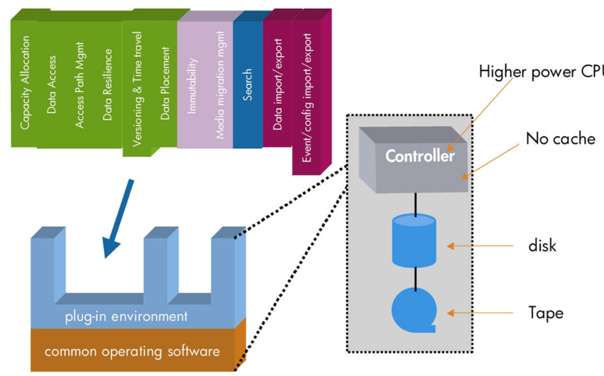


Figure 12b. Archive Smart Cell

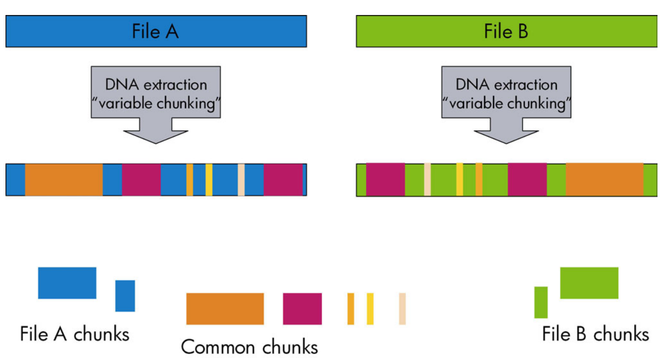


Smart, efficient information placement

With today's disk arrays and data placement methods, data is stored exactly as the operating system or application sends it to the storage system. If 20 users each saves his or her own copy of the same file, then the storage system will obediently save all 20 copies. The copies may all be saved on the same array, or they may be placed across many arrays—or NAS appliances—depending on how the users are connected. This results in needlessly inefficient storage utilization.

Chunking (Figure 13) applies special mathematical algorithms to streams of data to identify recurring patterns. The system then stores only a single copy—or redundant copy if instructed to do so—of each pattern (chunk) and pointers (metadata) to each piece of stored information that includes the chunk. The information can then be stored very efficiently, and quickly and reliably reassembled from its component chunks when an application requests a piece of data. The result is a virtualized information repository that offers fine-grain duplicate detection and extremely effective storage asset utilization.

Figure 13. Chunking: fine-grain duplicate detection



HP recognizes that significant amounts of enterprise data may reside in other repositories—storage arrays residing on an existing SAN, for example. The HP repository can federate these legacy repositories by creating an abstract interface and acting as a proxy for those repositories. For example, search queries can be mapped and forwarded as needed. Similarly, data access requests can be forwarded.

Evolving the vision

The HP vision extends from the history and tradition that has been delivered by the HP ENSA strategy since 1998. HP is now extending the vision to fully support the HP Adaptive Enterprise.

Technological basis and proof points—shipping technologies

The HP StorageWorks grid has its roots in the SANs that have been the heart of the HP strategy since 1998. Over the years, Fibre Channel SANs have evolved to deliver broad, flexible storage services to the majority of server operating systems. SANs, because of their support of server and storage consolidation, greater data availability, and simplified management, now dominate many corporate IT environments. Beginning in 2004, Ethernet fabrics will augment Fibre Channel, extending storage networking even further. HP is working with the industry to create the standards and products that will make this possible. Thus, iSCSI (SCSI protocol delivered over TCP/IP), RDMA (by the iSER standard later in 2004), and other advances will enable this evolution.

The new HP StorageWorks RISS, the first advanced, integrated ILM product from HP, is a shipping example of an HP StorageWorks grid. Smart Cells—intelligent modules containing controllers and storage devices—are the basic building blocks. They are federated together to create an organic HP StorageWorks grid. The grid has a single system image, so administrators see it as a single-managed entity rather than a conglomeration of component cells. Also, increasing the number of modules has the simultaneous effects of scaling both capacity and performance. This means that no matter how large the HP RISS grows, information retrieval performance remains nearly constant—something not possible with today's conventional storage arrays.

Storage arrays and other NAS systems are also evolving, forming a strong foundation for the HP utility vision. The HP StorageWorks EVA family, (for example, HP StorageWorks Enterprise Virtual Array 3000 [EVA3000] and HP StorageWorks Enterprise Virtual Array 5000 [EVA5000]), with their modularity, advanced virtualization, and management capabilities, will provide important technologies for the HP StorageWorks grid, as well as participating in it.

In parallel, storage management is advancing at an increasing rate.

Emerging technologies

The preceding examples will be augmented with technology being developed by a variety of sources. For example, HP has long been working with Lustre.org (http://lustre.org/) to develop a scalable clustered file system for use in high performance technical computing applications. Unlike more traditional file systems, Lustre is an object-oriented file system; it serves a variety of file types, including NFS. This capability provides a basis for creating scale-out architectures—built on commodity hardware—that can flexibly accommodate a variety of file types. Lustre also provides vast scaling of capacity, performance, and client connectivity—and it is designed to run on commodity hardware.

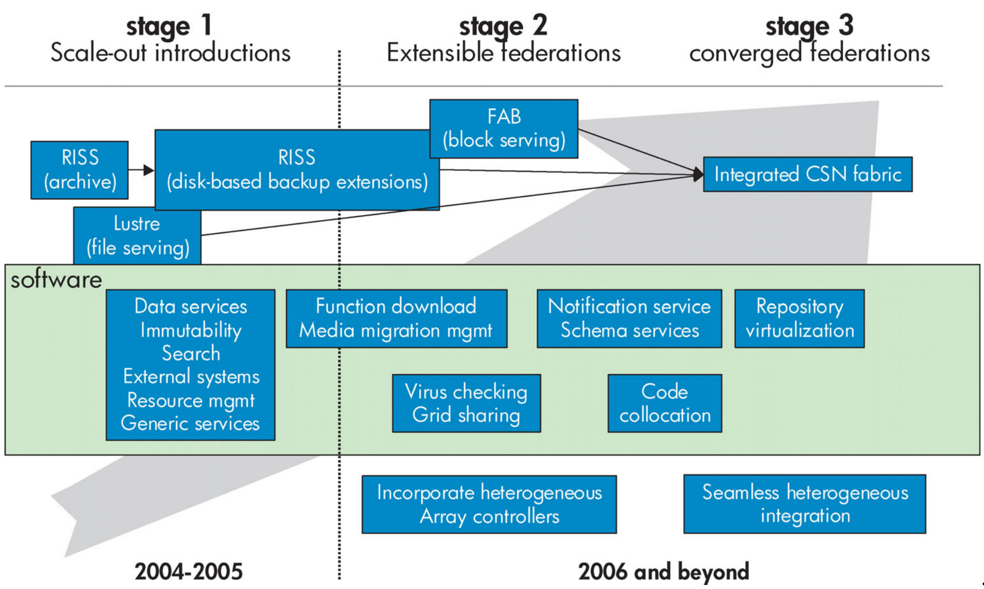
HP Labs

HP invests heavily in advanced research that can be leveraged in new products and has several projects that are expected to contribute to realizing the HP StorageWorks grid vision. One project is exploring new ways of federating modular components into scalable, economical storage subsystems. Other projects are studying ways to analyze and automatically configure and provision complex storage systems. These and other technologies will be instrumental in realizing this new vision.

Delivering the vision

The HP StorageWorks grid represents an ambitious extension of the storage utility and as such will take several years to develop. The grid represents a more capable, application-focused environment than is possible today. HP will introduce new technologies to the marketplace, and these leading-edge products must work together more integrally than today's storage networks, and they will be managed far more simply and easily. Also, the HP StorageWorks grid must evolve in concert with industry standards, mentioned earlier, and take advantage of third-party technologies as they become commercially viable. Figure 14 describes plans for achieving this.

Figure 14. Roadmap



With the announcement of the HP StorageWorks RISS, HP has released the first HP StorageWorks grid product and will soon offer a Lustre-based product. These products represent steppingstones toward a flexible, scalable ecosystem that is based on a single system image. Over time, HP will extend the grid to support more kinds of storage arrays, including third-party arrays. Virus checking and other services will be added to the grid over time, building upon the services model described earlier. Also, additional data serving capabilities, in the form of support for more kinds of data objects, will be introduced.

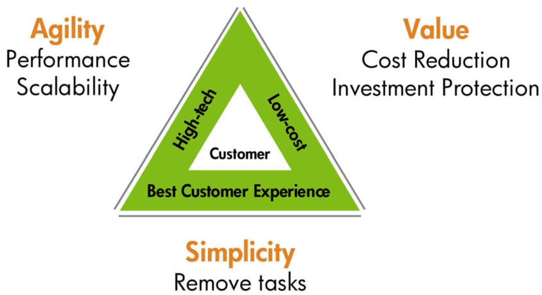
In the 2008–2009 timeframe, the vision will be completed as the fabric becomes more unified and broader virtualization is introduced. This will enable broadened information management and migration capabilities and greater application context.

All of this, of course, happens in the broader Adaptive Enterprise context. Therefore, the HP StorageWorks grid evolves alongside other HP delivered IT capabilities. The HP StorageWorks grid grows to deliver the entire set of capabilities described by the Darwin architecture, and it does so transparently to applications. Storage management converges with the rest of IT management to deliver truly integrated management and provisioning. Today's single pane of glass approach will evolve greatly in terms of simplicity and business focus.

Benefits of the HP StorageWorks grid

The HP strategy strongly supports the basic value proposition of delivering high technology, low cost, and the best total customer experience.

Figure 15. How HP StorageWorks delivers the HP value proposition



High tech

The HP StorageWorks grid delivers scalability, high performance, and availability features in ways that transcend traditional storage implementations. Smart Cells enable the grid to scale simultaneously for capacity, performance, and availability. Downloadable software personalities provide a versatile, powerful complement that enables the HP StorageWorks grid to precisely and dynamically satisfy application requirements with new accuracy and responsiveness. The content-aware HP RISS with its federated Smart Cell architecture is an excellent example. This new storage/archive/information retrieval solution is architected to vastly scale in the dimensions of performance and capacity. Unlike traditional storage products, the HP RISS is always aware of the information it stores—the information (data along with application context and content) is stored and indexed. Retrieval happens with virtually the same speed regardless of how large the HP RISS is. In other words, as you add HP RISS Smart Cells, HP ensures that performance and availability are maintained. Traditional storage solutions scale for capacity, but [array controller] performance scaling is often severely limited.

Low cost

HP continuously strives to help you manage storage acquisition, deployment, and operation costs. The HP StorageWorks grid continues the tradition. Smart Cells are built from commodity hardware, enabling the rapid incorporation of new technologies as well as continuously keeping costs down. Also, by enabling Smart Cell capabilities to be changed with downloadable software personalities, the HP StorageWorks grid preserves hardware investments as storage needs evolve.

Best total customer experience

Simplification is the key to excellent customer experience. The HP StorageWorks grid does this by adhering to standards and delivering powerful, integrated, automated asset management capabilities. Downloadable software personalities combine with policy-driven automation to eliminate many operational tasks. Those tasks that remain are executed largely by management applications, ensuring high levels of consistency and accuracy with minimum administrative intervention.

To learn more about HP StorageWorks grid benefits, refer to the HP StorageWorks grid business paper ("HP StorageWorks grid: The Storage Vision for the Adaptive Enterprise") referenced in For more information section.

Summary

This paper has described the HP strategy for taking the storage vision to the next level. The HP StorageWorks grid is the next generation architecture for delivering the services for the adaptive enterprise. Based on the HP ENSA heritage and its deep foundation in networked storage, the vision aims to strengthen the relationship between storage and the applications it serves and to increase the application focus of information management. This begins with today's ILM products, including the HP StorageWorks RISS, and evolves to broader capabilities delivered by a managed HP StorageWorks grid.

The vision is about creating a much more capable utility, while at the same time making storage management far simpler and more automatic. The HP vision is to create an HP StorageWorks grid—an entity that can scale vastly in many dimensions, and scale nimbly as required by changing business requirements. Today's SANs will first be augmented by these HP StorageWorks grids, and will ultimately combine into the grid.

The HP StorageWorks grid is an ambitious undertaking. It leverages technologies that HP has offered for many years, including storage networking, virtualization, and replication. It also relies on new and evolving standards like the SNIA's SMI-S (for storage management), IPsec (for IP/Ethernet security), and FC-SP (for Fibre Channel SAN security). Beyond these, the HP StorageWorks grid will also introduce new technologies to the marketplace. These will be used to deliver new dimensions of scalability, stronger integration of storage subsystems, and vastly simplified management.

The HP strategy is ongoing and as such leverages historical philosophies, including investment protection, modular extensibility, and leveraging of industry standards. Added to these is a strong new HP context: the Adaptive Enterprise and the tight partnership it creates between IT and the business it serves.

The HP vision is clear. This paper has described it in detail, and the path for achieving the vision has been staked out. Throughout the rollout, the HP StorageWorks grid and its component products are being designed to work alongside today's storage systems and networked storage infrastructures. Compatibility is assured. The HP StorageWorks grid truly represents a revolutionary step along an evolutionary path.

For more information

ENSAextended technical white paper (2003)

"FAB: enterprise storage systems on a shoestring," (PDF), Svend Frolund, Arif Merchant, Yasushi Saito, Susan Spence, and Alistair Veitch, HOTOS, May, 2003, Kauai, Hawaii

"Pangaea: a symbiotic wide-area file system," (PDF), Yasushi Saito and Christos Karamanolis, to appear in ACM SIGOPS European Workshop, September, 2002, Saint Emilion (near Bordeax), France

HP Labs: Automatic Storage Area Network Fabric Design and Racking and Wiring (http://www.hpl.hp.com/research/ssrc/services/architecture/SANdesign)

ILM white paper, January, 2004 (ftp://ftp.compaq.com/pub/products/storageworks/whitepapers/5982-3398EN.pdf)

Storage Security: emerging storage networking topic of interest, May, 2004 (HP publication 5982-5975EN)

Storage and Adaptive Enterprise white paper (May, 2003)

HP StorageWorks grid: The Storage Vision for the Adaptive Enterprise (http://www.hp.com/storage/storagegrid)

5982-6007EN, 06/2004

invent