HP's Information Supply Chain

Optimizing Information, Data and Storage for Business Value

John Webster
August, 2011

Technology Insight Series

Evaluator Group

enabling you to make the best storage decisions

Introduction

The need for organizations and business enterprises to gain control of their burgeoning data volumes is now a strategic imperative. And it's not just the amount of stored data, increasing at an average CAGR of 50% that needs to be managed. Multiple data sources including databases, image repositories, mobile devices and the web are now being converged and leveraged as new sources of information for business value.

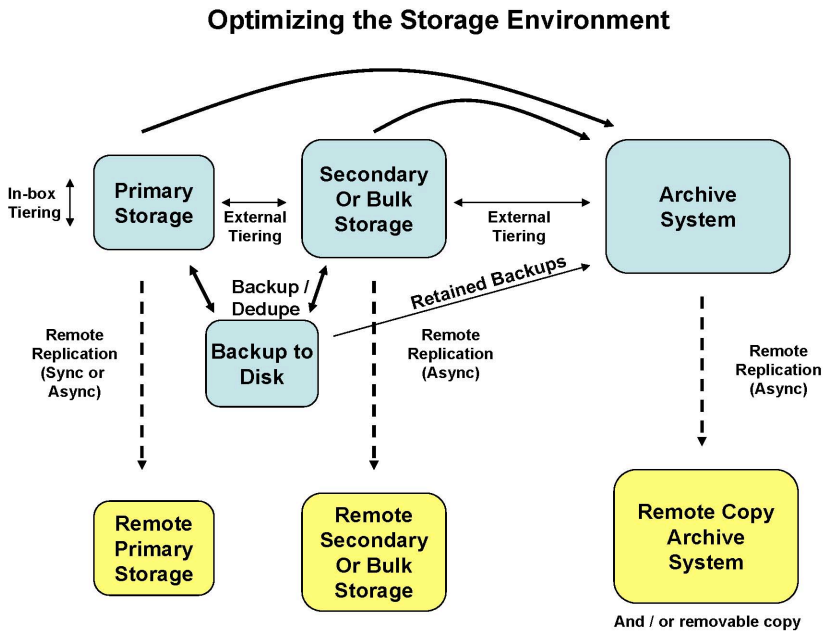
IT administrators have traditionally considered the tiering of data, data protection, and archiving as independent processes—narrowly focused and specific to a particular application or storage objective. The reality is that these topics are all related. Buying more primary storage will eventually force the need for more backup storage capacity. The need to add capacity to accommodate application growth will eventually increase the amount of data that needs to be retained for compliance and risk management.

These three processes—tiering, data protection, and archiving—are part of a bigger picture. They all relate to managing data in an enterprise data center environment and effectively utilizing storage systems. To not understand their interrelationships in that bigger picture can lead to sub-optimal operations and complexity that is difficult to make more efficient as time goes on.

Economics and the need to meet specific requirements drive enterprise-level data management. The requirements fit into two categories:

> Operational requirements that reflect to both evolving and ongoing business needs
>
> Information governance requirements that respond to internal and external business controls or regulatory requirements

Evaluator Group has found it very useful to establish a common understanding for Information Technology professionals regarding tiering, data protection, and archiving. All three processes handle the same data objects depending on their relationship to time. They are created at the primary storage level, moved to lower performance tiers and copied to the data protection systems over time, and then archived or deleted. This common understanding is needed in order to develop an overall storage strategy. We call this the "Information Supply Chain."



In this paper we examine each of the major technologies in the Information Supply Chain that should be considered as part of an integrated strategy and then show how they interrelate. Understanding the interrelationship between the three elements will allow trade-offs to be made between them such that the best economics can be achieved while meeting the business needs. We then examine HP's storage product portfolio as an example of how primary, secondary, and archival storage systems can be integrated with data and storage management software to yield an Information Supply Chain.

Needed: An Information "Supply Chain"

Data is commonly captured and managed in silos. Production data is stored in high performance, high value arrays. Multiple copies of backup data are retained in secondary storage devices and managed separately for data protection purposes. Archival repositories—a third silo—answer the need for long-term data retention and regulatory compliance.

The new reality is that these silos are all interrelated. They are part of what can now be seen as an integrated information "supply chain." Storage technologies available today, including automated storage tiering, deduplication, replication, non-disruptive migration, and information-aware archiving can be integrated to yield a high value information delivery service. The benefits can be expressed in two ways:

- Efficiency: Optimization comes from the integration and delivery of the parts as a solution.
- Economy: The CAPEX and OPEX savings resulting from consolidation of resources—both infrastructural and operational—that are inherent in the solutions

We start by looking at data tiering across the storage infrastructure as a way to integrate the enterprise storage architecture.

Making the Connections via Storage Tiering

Storage tiering is now a well-known data management technique, but undergoing some advancement driven by based on the emergence of new automation capabilities including the well known need to manage burgeoning data volumes. Tiering can now be divided into two basic categories: internal or "within-the-box" storage system tiering, and external tiering across storage systems.

Internal Storage Tiering

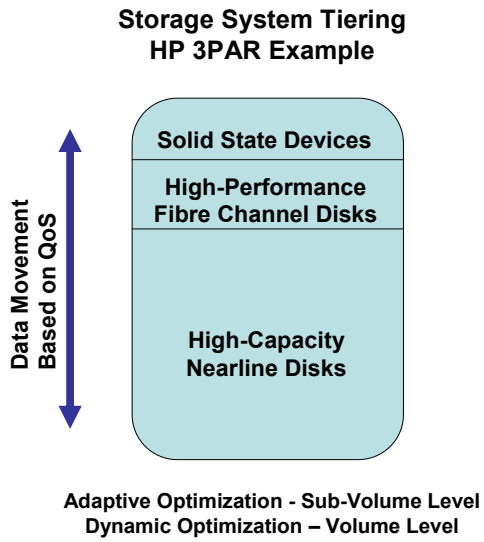
Internally tiered storage systems typically have different types of storage devices managed by an ability to move data among the devices. Data movement may be automated or manually driven. The granularity of the data objects moved may be set at the volume (or LUN) level or at a segment (sub-LUN) level.

The tiers within a storage system can be comprised of groupings of storage devices with different cost and performance characteristics such as:

- Solid State devices,
- High performance hard disks with fibre channel or SAS interfaces
- Lower performance, higher capacity disks with SATA interfaces

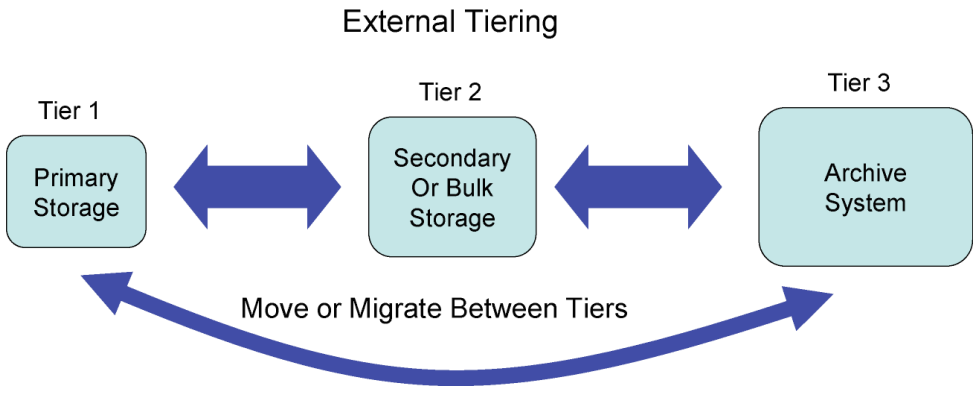
An example of internal storage tiering is HP 3PAR Utility Storage. HP 3PAR implements SLC-based, 50 GB SSDs as the high performance tier, integrated with 15K RPM, fibre channel disk and 7.2K RPM SATA disk. The performance of the FC disk is approximately 2.5 times better than the SATA disk while the performance of the SSD tier is roughly 12.5 better than the FC disk tier depending on a number of factors.

Array-resident, Adaptive Optimization (AO) software manages the tiering process. AO software identifies active vs. inactive regions in real time within the Utility Storage subsystem using I/O rate density as the metric that determines active vs. inactive regions. Data within the active or "busy" regions are migrated up to a higher-performance tier while data within the inactive regions is migrated to a lower performance storage tier within the Utility Storage subsystem. AO supports sub-volume granularity when migrating data as well as supporting thinly-provisioned volumes. The migration process is performed as a background task and takes place in the array without intervention of software residing anywhere else. Storage administrators can designate QoS gradients for certain volumes for quick reaction time and implement other application specific controls mechanisms and migration policies.



External Storage Tiering

External tiering is also a well know IT operational practice. This type of tiering can be as simple as moving less active or less valuable data stored in a high performance array to a less costly, lower-performance storage array or to tape. Three physical storage levels are usually conceived within the external tiering structure—primary storage, secondary storage, and an online or "nearline" archive. The movement between tiers can be manual or automated based on policies usually implemented in migration software. The migration software in many cases is the same as used in archiving but with different criteria and levels. With external tiering, the movement is usually either at the volume (or LUN) level or at the file level in the case of NAS storage systems.



An example of this is HP's File Archiving software (HP FAs). HP FAs provide policy-based data migration from MS Windows file servers to the HP X9000 Network Storage System. Both are described in more detail below. The X9000 can also archive files to Windows-based CIFS archives CIFS archives as well as

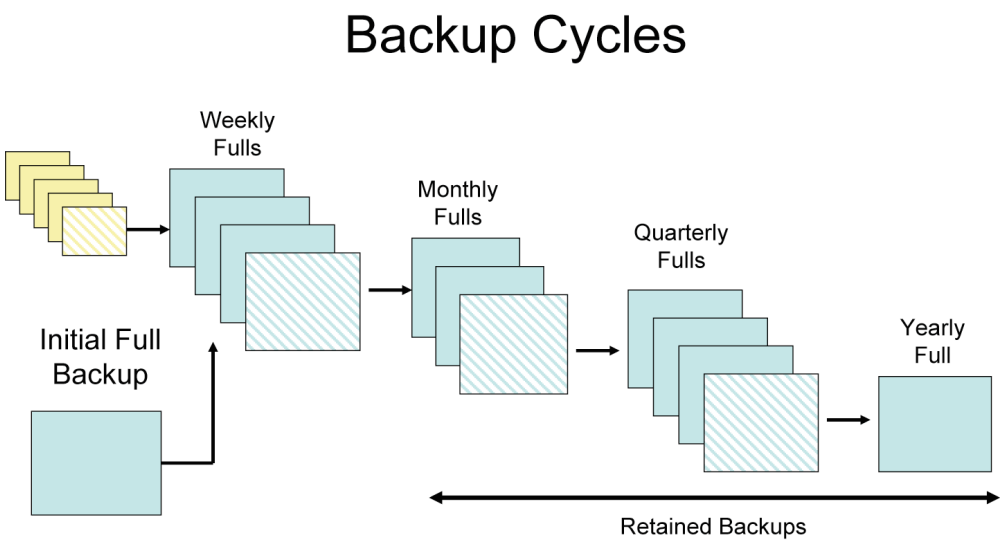
from high performance primary storage (HP 3PAR) to lower cost (HP P4000) storage. Files can be released from the source file after migration to any target archive.

When the HP X9000 is used as the archival storage target, IT administrators can implement search and access, retention management including preservation and destruction, legal holds and tamper resistant WORM for compliance and governance purposes.

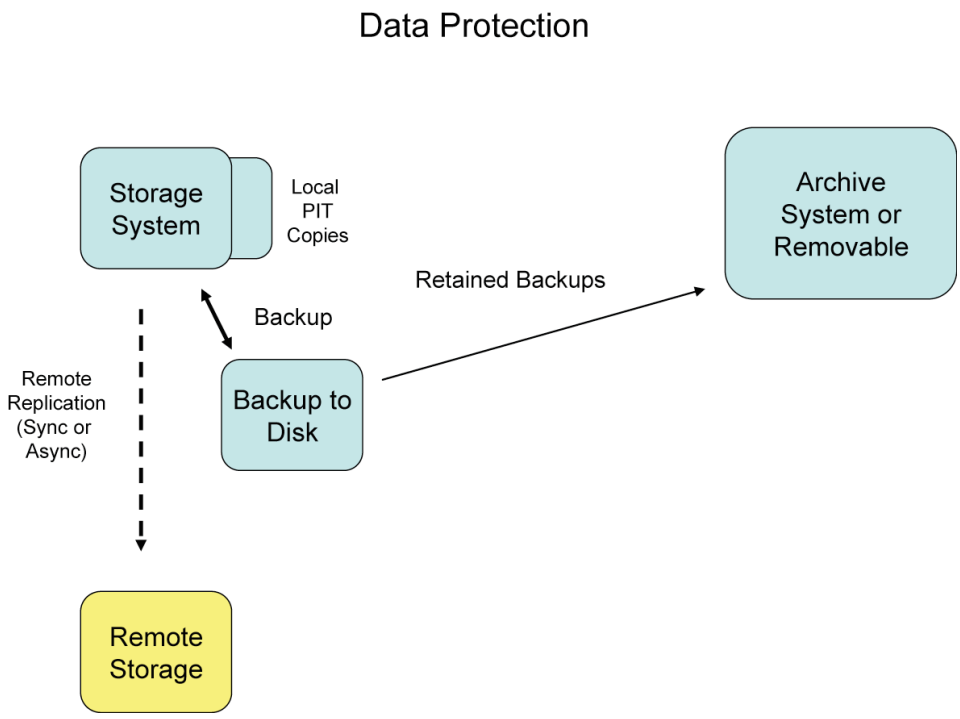
Data Protection

Data protection is commonly synonymous with backup. However, Evaluator Group believes that the two need to be identified separately because there are now a number of approaches to data protection—only one of which is traditional backup.

The backup process is closely associated with specific backup software which can be integrated with the enterprise storage architecture. To understand how this can be done, one should first look at the typical enterprise backup cycle from an operational standpoint. Incremental changes are backed up on a daily basis. Full backups are taken at longer intervals, usually weekly or monthly.



Retained backups (backups at a known point such yearly, monthly, close of the quarter, etc.) are saved. These retained backups may be sent to an archiving system to be managed as any other archived data object, demonstrating one of the integration points between data protection and archive.



Disk Array-based Data Protection

The need to quickly process backups and meet recovery times has led to usage disk as a backup target. Disk systems specifically configured as backup targets have added advanced capabilities such as data deduplication to reduce cost (less actual data to store) and transfer time in the case of remote protection (less data to transfer). Cost, performance, and operational changes have made backup to disk a prominent part of data protection deployments.

HP StoreOnce

HP's StoreOnce software is the first instantiation of HP's enterprise-wide Deduplication 2.0 strategy. It leverages unique software developed by HP Labs that can deduplicate data once then propagate it from local or remote backup clients to back-end data centers regardless of application type or platform. Therefore HP StoreOnce deduplication software can be implemented in multiple places within the data center and at remote sites.

StoreOnce will be applied first to the data protection environment. HP StoreOnce back up systems consolidate the backup of multiple servers, both local and remote, onto a single rack-mountable disk storage device. StoreOnce deduplication software reduces the backup data footprint and greatly speeds recovery when compared to retrieving backups from tape.

HP will extend this innovation into virtual machine appliances, client side backup deployments, and larger multi-node scale-out solutions. A common deduplication infrastructure will be critical in the future for efficiency and performance in large data centers.

HP 3PAR and Data Protection

HP 3PAR Utility Storage is an example of how HP implements snapshots and other data replication technologies for disk-based backup and business recovery. Snapshots allow point-in-time copies to be used – either for data restoration at the source site or for business recovery purposes from the remote site. Special features usually implemented in storage systems allow for these copies to be made periodically and non-intrusively and to minimize the amount of data transferred to the remote site.

Currently, HP 3PAR offers several products that provide data protection. There is a point-in-time snapshot product known as "Virtual Copy," along with a product for the creation of clone copies dubbed "Full Copy." In addition, both synchronous and asynchronous remote replication is provided in the "Remote Copy" product.

Full Copy

Full Copy creates a complete physical copy or a LUN. The copy may have a different RAID level than the base volume, and may be of equal or greater size than the source. Full Copy includes a re-synch option that allows the physical copy to be updated only with changed blocks. The Full Copy feature is able to make clones from both traditional and thinly provisioned LUNs.

Virtual Copy

Virtual Copy is a copy-on-write snapshot technology where the only blocks changed since the first snapshot are copied. A reserve area of space for snapshots is allocated when snapshots are enabled. Virtual Copy supports up to 500 snapshots per base volume. Snapshots may be either read-only or read-write. A Virtual Copy may be rolled back so that a volume can be synchronized with a prior snapshot.

When a snapshot is first created, a pointer to all its data is recorded. When a new data block is written, the original block is copied to a separate snapshot data space and the snapshot points to this data space instead. Snapshots are also writeable by allowing new blocks to be written in the snapshot data space. Snapshots of snapshots are also supported.

Remote Copy

HP 3PAR Remote Copy is a storage system-based set of data protection products available on the HP 3PAR T- and F-Class arrays. Remote copy is also known as remote replication and remote mirroring. HP 3PAR's remote copies may be configured with different RAID levels from their source volumes in order to provide configuration options. Remote Copy supports both synchronous and periodic asynchronous modes of operation along with support for multi-volume consistency groups. Volumes can be grouped together to maintain write ordering across the set of volumes.

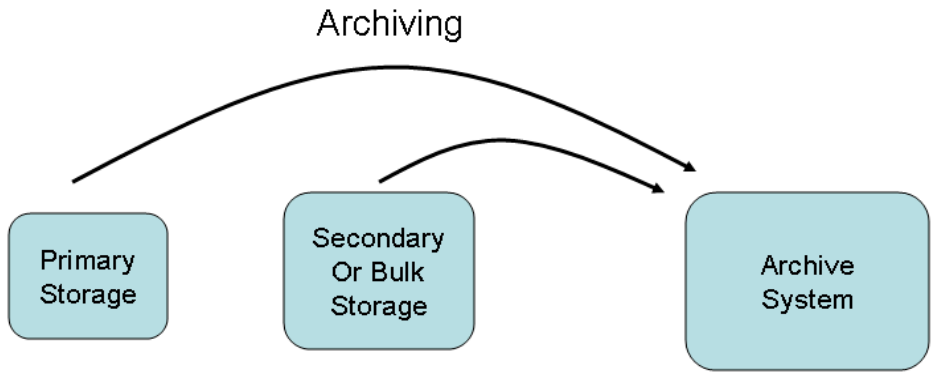
HP's 3PAR Remote Copy services provide several data replication mechanisms to provide business recovery protection. Typically, two HP 3PAR arrays are utilized: a primary for production data, and a secondary system used at a business recovery site.

Archive

The archiving element of the Information Supply Chain is about moving data that is not expected to be needed "relatively soon" to a more economical and archive-specific storage system while protecting it according to a set of rules at the time it is moved. "Relatively soon" is a time frame that varies in length based on policy or a particular business requirement. Records that may be required for insurance reasons, regulatory compliance, or for other potential re-use are candidates for an archival system that retrieves these records "relatively soon." As mentioned, these are typically considered to be online or "nearline" archival storage systems.

The key to efficient archiving using an online device is recognizing the cost for storing archival data for years and in some cases decades. The advancement of archival storage technologies is proceeding rapidly and can add cost to the archival storage environment. For example, while the media used to store data for can last for long time periods, the devices used to read the media may become obsolete sooner than the media, necessitating a migration of data to new storage devices and new media—a potentially expensive process depending on retention time and type of media used. Another consideration when evaluating for long-term storage is the cost and availability of the software needed to read the data. If this becomes obsolete during the retention cycle, IT administrators can face an even more difficult migration.

The foregoing discussion has led to the conclusion that the most relevant approach for IT administrators is to store the data in the context of the application that accesses the data. The context may be in a file format or an object format. Managing data that is stored for long-term across generational changes of storage devices can be expensive both from capital expenditures for the generational change and for operational expenditures to transition the data to the newer technology.



Moving the data to the archiving system can be done a number of ways. It must be understood that data is ultimately to be moved to the archive and protected there—freeing the capacity consumed from the source (typically primary storage) and from the operational expenses of repeated protection cycles. The operational sequence can be been defined as one of three actions:

- Archive and Delete – move the data and when successfully archived, delete the copy on the source.

- Archive and Stub – move the data and when successfully archived, leave a stub in place of the data moved so that if accessed, it can be recalled from the archive system.
- Archive and Leave – move the data and make the determination to delete the data or alter the data to a new version at some later point.

Protecting the data at the time of archive is also critical part of the process. This may be a remotely replicated copy of the information to another archiving system and/or a removable copy of the data that is physically stored in a protected location.

HP X9000

HP most recent archival platform is the X9000 Scale-Out NAS array running content archive software. This solution can scale to over 1000 nodes and manage up to 16PB of content consisting of over 1B objects. It also features advanced snapshot capability with support for over 1M snapshots.

The X9000 Content Archive is a true, policy-managed, multipurpose archive. Content sources include:

- File
- Database
- Email
- Microsoft SharePoint
- PACS systems

A federated query engine that functions across X9000 nodes is capable of real-time content search.

As mentioned, HP's File Archiving software can use X9000 as a target for Windows-based file stores. This combination supports policy-based migration of files from Windows servers to the X9000 via all three of the operational archive sequences mentioned above. Stub files on the Widows server are used for file retrieval. X9000 also supports integration with HP's Email Archiving software for Exchange and Domino.

HP Data Protector

HP Data Protector is backup and disaster recovery software that can be used to automate data protection for physical servers and virtual machines. There is a wealth of features available with Data Protector to augment the primary purpose of making protected copies of information. Integration with applications and HP array storage systems enable protection with minimal impact on operating environments. HP StoreOnce backup systems and Virtual Library Systems can be integrated with Data Protector to incorporate the latest data reduction technologies and improve the performance of the protection cycle. Advanced features such as encryption, single point of management including virtual machines, recovery to points in time, synthetic and virtual full backups, and reporting capabilities allow integration into many different environments and usage requirements.

HP Services

To support its Converged Storage initiatives and help customers integrate primary, secondary and archival storage platforms into their environments, HP has also developed an extensive storage services portfolio. It includes general storage technology and infrastructure support, targeted engagements like

support for large and complex data migrations, and fully outsourced management of storage infrastructure.

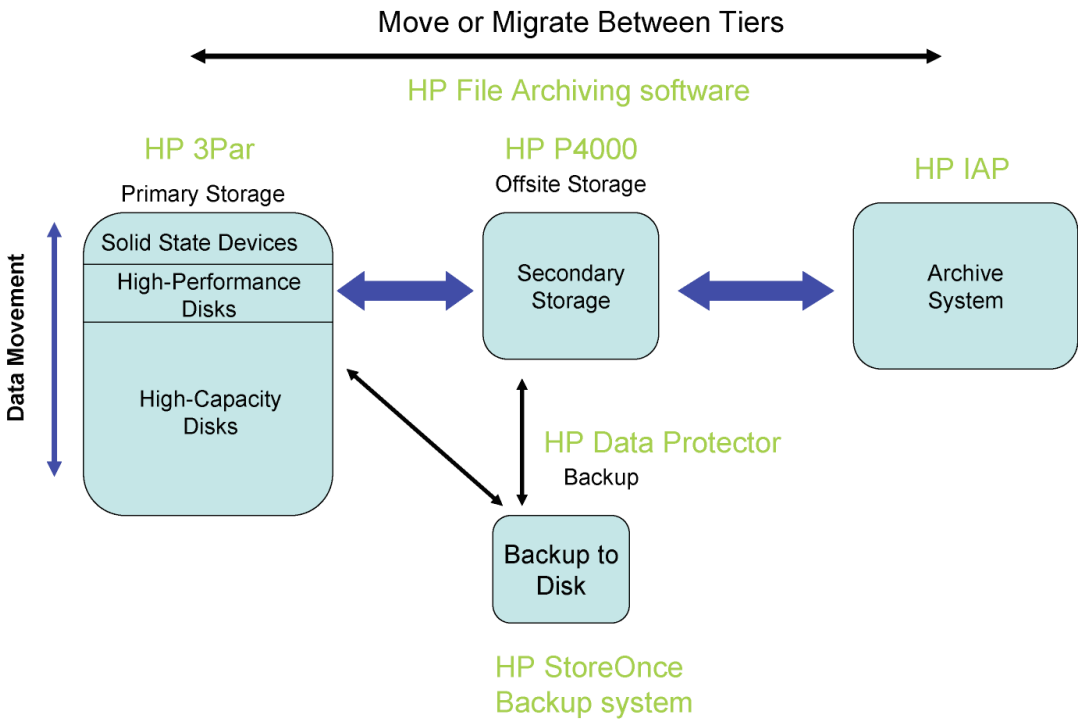
Specifically identified storage services include:

- HP Always-On support services—one-stop access to the people and technology needed to deliver personalized storage support
- HP Converged Storage Services—fine tune, protect, and enhance storage infrastructure to optimize performance and business alignment.
- Converged Storage workshops and assessments—analyze gaps in legacy infrastructure and co-develop tactical plans to evolve into virtualized and utility storage designs.
- Storage Infrastructure Transformation—for customers moving to a consolidated and virtualized server and storage infrastructure
- Data Transformation—focuses on the NAS environments with an eye to eliminating filer sprawl, architecting tiered storage, and deploying a NAS storage foundation to effectively manage the explosive growth of unstructured file data.
- Complementary backup and recovery services—protect against data loss and provide fast data recovery to meet business objectives.
- Storage Service Transformation—helps customers develop a storage architecture that supports service-oriented IT and cloud computing.

Services can be contracted separately, or wrapped into large project related acquisitions that include infrastructure and software.

HP's Information Supply Chain Model

Managing of data and protecting it involves the interaction of storing information, placing it on the right storage system based on the probability of access and performance requirements, and protecting the data based on established business requirements. An architectural depiction of this interaction for the standpoint of HP's current storage product portfolio appears below.



Tiering of data storage, data protection and archiving need to be considered as an integrated whole to enable an enterprise data storage environment that is cost-efficient from an infrastructure standpoint, and is operationally efficient even as data volumes continue to grow. Focusing on only the individual storage elements rather than the big picture view of data management will blind IT administrators to the value offered by the kind of integrated storage and data management offered by HP.

The economic gain for considering tiering, data protection, and archiving as an integrated set of processes can be much greater than implementing separately as unrelated data silos. This does not mean that each of the three has to be implemented—just that they need to be considered as part of the big picture. The planning involved when considering all three will allow trade-offs to be made between them such that the best economics can be achieved while meeting the enterprises business needs.

HP offers a unique, end-to-end, primary to secondary to archive, information supply chain. In addition HP offers comprehensive storage integration services and a single sales/delivery model. Combining these products with a comprehensive vision for future development gives the CIO a strategic view of enterprise information and supports an expanded, high-value information delivery service for information users.

About Evaluator Group

Evaluator Group Inc. is dedicated to helping IT professionals and vendors create and implement strategies that make the most of the value of their storage and digital information. Evaluator Group services deliver in-depth, unbiased analysis on storage architectures, infrastructures and management for IT professionals. Since 1997 Evaluator Group has provided services for thousands of end users and vendor professionals through product and market evaluations, competitive analysis and education. www.evaluatorgroup.com Follow us on Twitter @evaluator_group

4AA3-6019ENW