The ILM journey: An HP StorageWorks File System Extender business white paper

Today's business challenges

IT departments are under constant pressure to utilize their budgets to support business with technological initiatives that provide a clear competitive advantage. IT managers themselves are measured on strict ROI terms and, for this reason, are anxious to avoid ever-increasing expenditure on the mundane and tactical tasks, which, as fundamental as they are to the business, do not offer the potential of a true strategic differentiator.

The information environment has become a focus for increased efficiency and effectiveness. The growing amounts of data that are stored, replicated, backed up, and possibly restored are absorbing large proportions of the IT budget in terms of new products, new licenses, and increased management costs. All too often, this increase in expenditure is not equally matched by an increase in competitiveness or productivity. There is more information to manage and a bigger infrastructure needed to support it.

The challenges businesses face today include:

- Explosive growth in data and storage requirements
- Insufficient time to complete critical backups
- Slow and costly single file restores from tape backups
- Increased need for 24 x 7 operations
- Budgetary pressures
- Increased cost of management

Customers are looking for innovative ways to improve the efficiency of these environments and Information Lifecycle Management (ILM) is an umbrella of approaches and solutions that attempt to provide this.

What is HP StorageWorks File System Extender?

HP StorageWorks File System Extender (FSE) is a new-generation data management software that is part of the HP ILM portfolio. FSE has data migration functionalities beyond traditional hierarchical storage management (HSM) software that makes it a cost-effective solution to managing massive amounts of file system stored data.

Unlike traditional HSM products, an FSE-managed file system can be optionally upgraded to a read only write once, read many (WORM) file system to support growing demand for legal compliance of electronic documents. In addition, FSE supports intelligent media management.

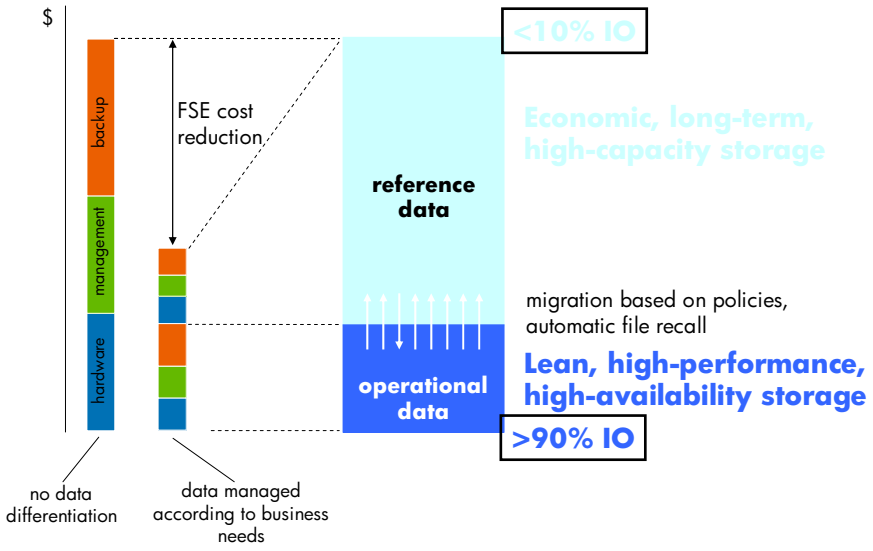
While FSE offers multiple business benefits as a standalone solution, it is an integral part of the broader ILM product portfolio from HP supporting customers as they embark on the journey of managing information stores for increased productivity and competitiveness. This white paper describes the FSE position within the ILM journey.

Key FSE business benefits

- Increases business productivity by increasing speed of access to critical data by moving and storing data on the most economic location in accordance to the value and importance of that data to the business
- Saves time by enabling faster backup and recovery by reducing the data in primary storage
- Reduces TCO by moving data from the FSE file system to the most appropriate secondary storage based on policies and reduces downtime for capacity upgrades

- Ensures business availability in that the applications and end users can continue to seamlessly access files regardless of where data resides with optimized seek times for current online data
- Streamlines operational management by integrating with archiving management software to provide a tiered storage infrastructure for all leading business applications

Attack the costs of reference data



The ILM journey

Step 1: Classifying data—not all data is equal

The HP ILM methodology builds on a logical, sequential approach or metaphoric journey.

The first step in increasing the efficiency of the information environment is to clearly define what information is vitally important to the company, what is less important, and what is perhaps of no consequence whatsoever. Another variable in defining or classifying the data is the impact that aging will have on the data.

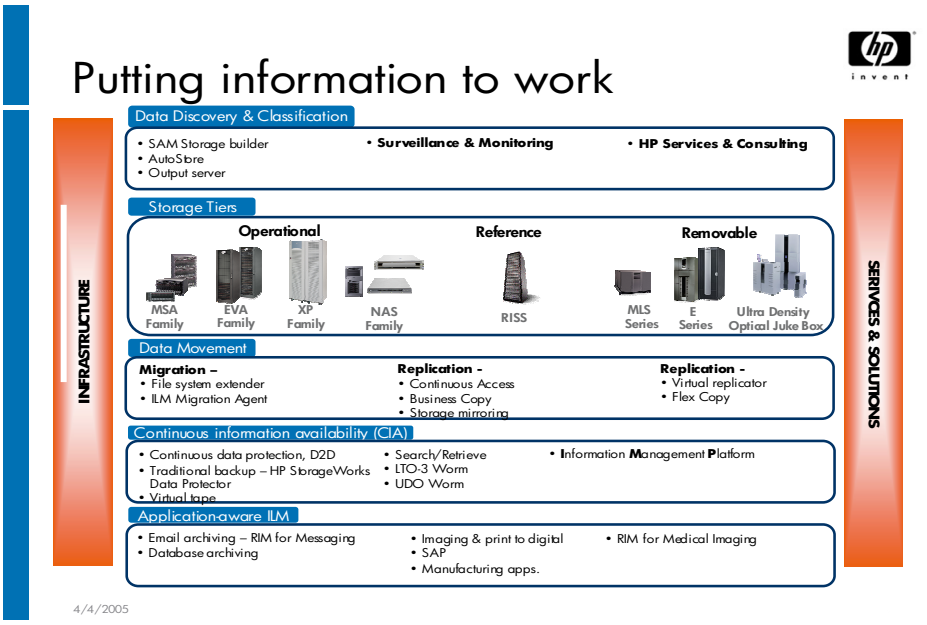
In assisting customers to manage and understand their environments, HP proposes HP Storage Essentials to assist in managing and understanding environments. HP Storage Essentials is designed as a suite of value-added plug-ins that offer advanced heterogeneous storage management functionality including storage area network (SAN) management, storage resource management, provisioning, and application infrastructure monitoring. The HP Storage Essentials File System Viewer plug-in provides information on file aging, type, user, and size.

In addition, HP Consulting and Integration (C&I) provides Data Classification Workshops to assist customers with analyzing and segmenting their data stores.

Some data will have relevance for many years, other data for a shorter duration. Some data may need to be deleted for regulatory purposes. The fact that data is classified in a logical manner leads to the idea that the underlying hardware infrastructures will also be classified accordingly.

Step 2: Storage hardware tiers

HP offers a broad portfolio of different classes and types of storage products. Performance and functionality segment more expensive from cheaper online arrays. Near-line products consist of tape as well as optical technologies and there are innovative active archiving products with integrated search and retrieval capabilities now available.



In a classic tiered storage solution, tier 1 normally provides storage for direct access by online applications such as provided by Enterprise Fibre Channel–based arrays at the highest performing level (HP StorageWorks XP arrays or HP StorageWorks Enterprise Virtual Arrays [EVA]) to Serial ATA (SATA)– or FATA–based arrays (HP EVA FATA or HP Modular Smart Array [MSA] SATA) arrays at a second tier or capacity–oriented level. Secondary tiers provide copies of online data for backup and capacity for archiving. A secondary tier may consist of lesser performing arrays, tape libraries, and also optical jukeboxes. The overarching theme is to ensure that static terabyte data stores are not needlessly using capacity on expensive Fibre Channel arrays and final state data is not needlessly backed up. Moreover, a move to tiered storage opens up the opportunity for low-cost connectivity, network-based storage services, and potential disk-based backup.

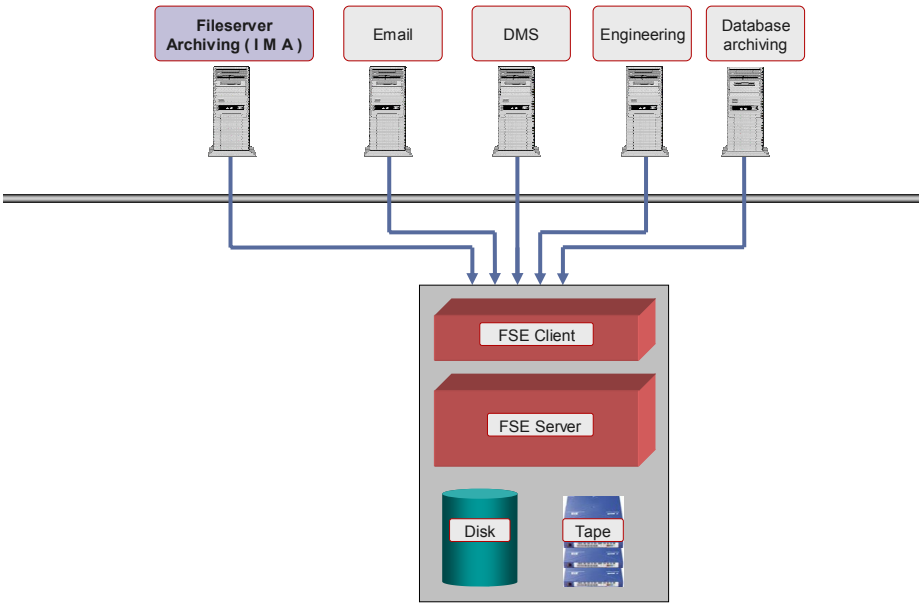
Step 3: Automated data movement by policy

The next critical ILM phase is data movement by policy. By fully understanding the data and how it must be managed, quality of service (QoS) can be improved and data management costs can be reduced. In essence, this is the migration of different classes of data according to its current perceived business value to the most appropriate storage medium. The most appropriate storage medium can be defined by a combination of the following variables, all of which affect the capital and ongoing costs:

- Performance
- Availability
- Time to recover
- Regulatory compliance
- Immutability

Policy-based data movement understands that the business value of information or data may fluctuate based on frequency of use, age, regulations and governance, and business practices. It is essential that policies allow for change and that the underlying architectures also respond with similar agility. For instance, data that may have been considered less important from a business perspective and migrated to cheaper, slower storage may be re-valued and returned to a faster performing storage tier.

FSE architecture



FSE allows for this eventuality by leaving a "stub" so that released data may be returned to its original location. This process is more efficient than a traditional "copy, delete, and restore if necessary" approach. The "stubs" point to the archive location, enabling user or programmatic requests for a released file to be automatically and transparently fulfilled. Policy-driven transference of files can be scheduled for non-peak periods, optimizing performance of the entire system.

Two and three tier storage architectures

A two tier storage architecture is one in which there are only primary and secondary tiers. The secondary tier may be either disk (usually a cheaper disk-based array, for example, MSA SATA or EVA FATA) in which case this architecture is called disk-to-disk (D2D) or disk-to-tape (D2T). D2T is the most traditional tiered storage solution where the bulk of the data is kept on tape usually in an automated tape library. The D2T architecture is suitable for low retrieval incidences as tape has longer seek times than disk. However, tape has a high density footprint and is the lowest cost media.

D2D architecture is gaining prominence as it affords higher retrieval performance and the price of capacity-oriented drives is in line with business objectives. Active archives are normally built on a D2D architecture because search and retrieval performance with sequential access technology, such as tape across massive archives, is normally unacceptable.

Three tier architecture contains an additional tier normally to cost-effectively achieve larger scale of storage and is often referred to as disk-to-disk-to-tape (D2D2T) or a virtual tape library. The first tier is for operational data and most commonly the second tier is a disk-based platform used for "staging"

or intermediate data access. If there are relatively high requests for recently archived data, a three tier architecture makes business sense. The duration of data on the secondary tier is determined by business need but is usually no longer than three months. The final repository is tape that can be kept available for near-line requests or transported to vaults as part of a disaster recovery plan.

Step 4: Continuous information availability

The business requirement of maintaining accessibility to records as and when required is related to many technologies—backup, archiving, and active archiving.

Backup and recovery continues to be an important part of a data management and protection plan. The challenges of this function are: growing data loads, nonexistent backup windows, high network traffic, and media management problems. As data stores grow dramatically, it is becoming more essential to perform a thorough assessment of the storage environment and the nature of the data itself to create an appropriate and cost-effective method of "continuous information availability." It is at this stage that businesses may divide data management into the following subsections:

- Backup and recovery
- Record management

Backup and recovery are about having timely access to current business data in case of disaster or file corruption. It is a process that ensures businesses do not lose important information or operational time after a data loss. Backup processes are designed to be automated and include incremental and differential backups, which allow multiple versions of files to be recovered. In essence, backup is used to protect current business data rather than older archive data. Therefore, backups are designed for disaster recovery purposes.

Are backups taking too long because unchanged, obsolete data is being backed up too often? Do the contents of a backup file meet regulatory requirements? These questions may often be the cues for a business to explore an HSM/archiving solution and focus on the record management aspect of the business.

Backup and archiving: The differences

Archiving is about keeping a long-term record of a business application or information state. Archives are typically kept for auditing, regulatory, or reference purposes rather than for application or data recovery. With archiving it is important that files are not over-written with more recent versions of files as in backup solutions. Archives must be exact copies of the data at a certain point in time. Naturally, this implies large volume growth as complete copies of data are required.

Additionally, because backups utilize compression and proprietary file formats to speed their execution, it is virtually impossible to search backup files for specific pieces of information. Instead, restoration of the backup file to a disaster recovery server must be performed, and the search can then be executed. To satisfy a discovery motion, for example, all of the backup files that a company has must be restored and searched. This is a very costly and time-consuming project. Even then, there are no guarantees that a specific piece of information will be found, for example, if a user deletes a file before scheduled backup takes place.

Archiving systems, on the other hand, should retain complete copies of data saved for a specified period of time that cannot be deleted until the end of that predetermined period.

HSM and archiving: The differences

HSM began in the mainframe sector where it was primarily used to reduce storage costs and is so called because it has the intelligence to move files along a hierarchy of storage devices that are ranked in terms of cost, speed, retrieval times, and overall capacity. Files are transferred along the hierarchy to less expensive forms of storage based on a set of rules or policies, all of which are

transparent to the user. Users gain access to this data by clicking a placeholder file or "stub," which returns files to primary storage as needed. The HSM solution does not provide an archive solution in that the solution does not guarantee retention of the data and because the application or its users can modify or delete the data free from any controls.

An archive, in the strictest sense, implies that there are intractable policies and rules that govern the retention of the data independent of the application. This requires an archiving metadata layer that guarantees the existence and integrity of the archive. The application will always be required to access the data by using the archiving management function. Archive management that includes retention, compliance, and security must be maintained for the archive across all tiers.

Compliance: Archiving business driver

Increased legal scrutiny of business information has given rise to around 10,000 different electronic data management regulations in different regions of the world such as SEC17a-4, Sarbanes-Oxley, Basel II, which specify compliance requirements. All these global regulations have many base features in common. They require the following features:

- An enforced retention period.
- Deletion of the archive—If management overview is required, this may have a negative impact on the choice of tape as an archive media.
- Data integrity—Assurance that archived records have not been changed; quality and accuracy is assured.
- Security—Controlled access to the archive and audit trails of access.
- Data protection—Information must always be accessible and be able to sustain any eventuality, either natural or human.

FSE provides a compliant environment by assuring data integrity (by way of optional WORM technology). FSE does not rely only on WORM tape or WORM optical media. This functionality further enhances the use of FSE in an archiving solution.

In a tiered storage solution, leveraging the different cost structures of storage media is not necessarily an archive and an archiving solution does not necessarily have to use a multi-tiered approach but may exist on a single tier storage medium. But common practice is that archiving solutions tend to reside on more cost-effective storage platforms.

In summary, the goal of an archive is to retain data with integrity independent of the applications that created it while HSM solutions will migrate data to the most cost-effective storage platform based on business rules. Consequently, there are synergies between data movement solutions and archiving solutions.

Step 5. Application-aware active archiving: Putting information to work

Active archiving as opposed to "deep archiving" implies a need for frequent access to the data and is consistent with a knowledge management view of the data. Applications requiring active archived data insist on quick and reliable access times. Examples of these include industries that have frequent legal discovery obligations, reference library data, scientific simulation environments, and multi-media content. The HP StorageWorks Reference Information Storage System (RISS) is an example of an active archiving solution that provides the scalability, performance, and functionality required for the serious challenge of maintaining a long-term active archive.

Application-aware archive solutions ensure that the archiving management software is deeply integrated into the original application. In this manner, the archiving management software becomes an extension of the application itself and is completely transparent to the users. Users can scan the archived data and locate vital information based on keyword searches.

Tiered storage solutions and application-aware active archiving are linked in that archived data is migrated to a separate medium to that of operational online data. When the referential links to the moved data are not deleted, the original application can access the data transparently whether it is on the primary storage or secondary storage system. The choice of the archiving platform relates to archive management, which takes the following important factors into consideration:

- Compliance
- Legal discovery
- Long-term security and accessibility
- Data authenticity
- Keyword search

Synergies of application-aware archiving solutions and FSE

FSE or File System Extender as its name implies manages a file system that is extended across multiple tiers of media so files are located on the most appropriate media transparently to the user and application. The "client" portion of FSE is responsible for managing migration policies, release policies, file recall, and also the ability to upgrade the file system to WORM in specific legally regulated environments. The FSE server logically creates media pools and serves these pools to the FSE client. Synergy between archiving middleware and FSE is created as archiving middleware has the "hooks" into specific applications while FSE has the ability to automatically place data on appropriate storage media and manage that media. Archived data (single instanced, compressed, indexed) is readily written to FSE-controlled file systems because they appear as standard file systems and subsequently to target media pools, which may be a selection of tape, disk, and optical technology (in FSE v.4.0), all of which are controlled by the FSE server, which allows policy-based movement of data between these hardware layers or tiers.

The archiving middleware vendors such as IXOS, OuterBay, and Princeton Softech provide archiving middleware with specific application intelligence and features such as single instancing, indexing, data normalization, compression, and so on while FSE will control the location of the archived data, ensuring that it is always placed on the most appropriate storage medium.

FSE is application unaware and does not have features such as content indexing, keyword retrieval, compression, and application awareness. Archiving middleware cannot transparently span all the storage tiers and be able to integrate data on a policy basis between these tiers. Consequently, FSE used in conjunction with application-aware archiving solutions offers all the benefits of a tiered storage solution together with the application-specific benefits.

FSE is an integral part of the broader HP ILM product portfolio supporting customers as they embark on the journey of managing information stores for increased productivity and competitiveness.

Conclusion

HP offers a comprehensive range of products, services, and solutions to help customers throughout fully leverage the power of information assets throughout the ILM journey. Wherever customers may be in the ILM journey, HP can assist in part or full implementation, whether discretely in response to a particular challenge or in logical sequence as part of a holistic data management initiative. FSE is an important new addition to the HP ILM portfolio.

For more information

www.hp.com/go/ilm

http://h18006.www1.hp.com/products/storageworks/fse/index.html

5983-1885EN, 04/2005