DLM/ILM Questionnaire

Questions

Development Status and Objectives. Describe your vision of D/ILM and identify what features or functions your platform currently provides and what features or functions are still in development.

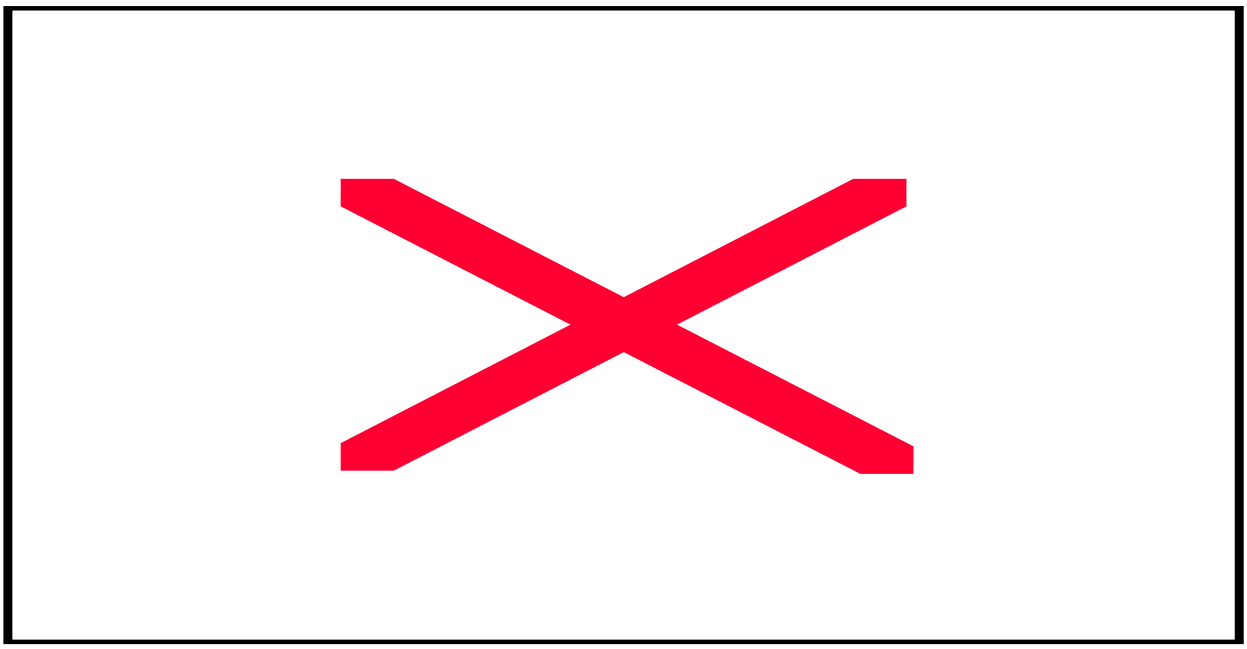
In its broadest sense VERITAS defines Data Lifecycle Management (DLM) as:

> *"The process of managing business data throughout its lifecycle from conception until disposal, across different storage solutions and within the constraints of the business process".*

Defining DLM as a process decouples the activity from the products and solutions that may be in place to support the process. This also supports the common understanding that Lifecycle Management is already part of most businesses today, though typically only parts of the process are (semi-) automated. For example, creating backup copies of data is fairly well automated in most organizations, whereas archival or extraction of a specific sub set of business data typically requires manual intervention.

While at first glance it appears simpler to keep most or all data online for as long as possible, the actual management cost and complexity, as well as the indirect real-estate/operational/management costs (i.e. the cost beyond the initial storage purchases) are prohibitive. Realistically then this means that a uniform high-end primary storage solution for all corporate data is no longer economically feasible for any reasonable sized enterprise business process. Various industry regulations and corporate compliance practices add another dimension to this complex problem whereby the regulated data needs to be managed in a controlled, immutable environment, typically as a separate repository.

Given this need for managing data in a non-uniform infrastructure, let's consider the "well-known" facts. We know that not all data is created equally; *some data is more critical than others*. Data criticality is seen in relation to the business process it supports or the business legal requirements it fulfills. We also know that the *business criticality of data will change over time*. So while maybe less optimal than uniform storage, over time data of different criticality would automatically need be stored on and moved to different classes and types of storage solutions and data would be deleted when expired and considered of zero value to the business – the DLM model.



In the broader DLM context, a different class of storage does not only imply different price-performance levels, but also different levels of protection, manageability, immutability, cost & charge-back, etc. This highlights one of the key differences between HSM and DLM. HSM primarily focuses on optimizing data availability in a virtual online model (across a hierarchy of storage – typically disk and tape), whereas DLM takes all other aspects of the data's lifecycle in consideration too, including the protection levels, data retention and destruction of data.

A well-established DLM process then would provide a highly automated segregation and management of data by required performance, availability, recoverability, discoverability, cost allocation and immutability, while retaining transparency to applications and end-users.

VERITAS Data Lifecycle Manager enables companies to solve their problems of data growth, compliance, data security, data organization, and resource utilization by automating the management of data—from creation through disposal—according to defined policies. In addition, VERITAS Data Lifecycle Manager provides powerful, high-speed search and index technology that reduces the time and cost of electronic records discovery.

VERITAS Data Life Cycle Manager is designed to handle e-mail and file archiving in Microsoft Exchange and NTFS (Windows NT File System) formats. VERITAS Data Lifecycle manager extends existing backup/restore capabilities to address retention and retrieval by sharing hardware and media licenses with VERITAS NetBackup and VERITAS Backup Exec. This allows VERITAS customers to leverage their previous investments in data protection (tape libraries, SANs, and drive licenses) to address compliance.

Through this common media management layer, VERITAS saves customers money by reducing unnecessary copies of data, improving utilization of devices by leveraging them for multiple purposes, such as both backup and archive, and eliminating multivendor support issues by not requiring a new storage infrastructure or IT skill set.

VERITAS Data Life Cycle Manager is ideal for customers that need a data lifecycle management/information lifecycle management solution for:

- regulatory compliance
- space management (HSM)
- corporate governance

VERITAS Data Lifecycle Manager v5.0 will initially support Microsoft Exchange and NTFS (Windows NT File System) environments. (VERITAS intends to add support for Lotus, IM, and database archiving via third parties later in the year.) All data entering DLM can be content indexed so that both file and message data can be easily searched and manipulated.

VERITAS Data Lifecycle Manager provides media support for multiple storage tiers, including disk, nearline storage, and tape and optical media. High availability is supported through VERITAS Cluster Server and Microsoft Cluster Service.

VERITAS Data Lifecycle Manager is based on VERITAS' seven-year-old VERITAS Storage Migrator codebase/engine, which is currently shipping as v4.1. VERITAS Data Lifecycle Manager v5.0 contains new functionality such as content indexing, formal policies, and data retention.

Going forward, VERITAS plans to expand platform support and add a formal API to the product for close integration with such products as document management systems.

Development Partners. List the vendors with whom you are working to deliver your D/ILM solution functionality.

VERITAS is working with Network Appliance, Princeton Softech, EMC, Pegasus Disk Technologies, and Plasmon.

VERITAS and NetApp have a closely integrated project team to ensure seamless interoperability between VERITAS Data Lifecycle Manager and the NetApp® NearStore® and SnapLock™ storage products.

Data Naming Scheme. What is your method for identifying data storage requirements or characteristics and for using those requirements or characteristics in building policies?

VERITAS relies on formal SRM solutions, such as VERITAS Storage Central and VERITAS Storage Reporter, to gather and present statistics on data and storage resources. These can be used either by the customer or by VERITAS Consulting Services as part of a data lifecycle assessment.

Administrators create named policies that incorporate a variety of rules regarding data selection, relocation, and retention. Once created, policies can be run in a "preview-only" configuration to see what data would be selected and what actions would have been performed.

Selected data can be sent to any number of virtual storage locations called “datastores.” This allows data to be stored in logical groupings that may span a variety of media. For even greater control over data placement, data can be assigned to “affinity groups.” These groupings ensure that data is stored on the same piece of media. This is especially useful when implementing a three-tier migration model from disk to nearline disk to tape.

Access Frequency. Identify the mechanism by which your solution identifies how frequently specific files or datasets are being accessed and whether and how access frequency is used in migration policies.

VERITAS Data Lifecycle Manager uses a filter driver on NTFS volumes, an agent for Network Appliance primary and nearline storage, and/or scheduled jobs to detect and act on data that meets the criteria set forth in the policies (for example, “not accessed in 90 days” or “not modified in 1 year”). VERITAS Storage Central can also be used to run a variety of reports that can be translated into Data Lifecycle Manager policies. Runs of the policies (jobs) can be scheduled at flexible times to accommodate external hardware, networks, or existing processes.

Storage Platform Characterization. Explain the mechanism that your solution provides for characterizing the performance capabilities and costs of specific hardware platforms for data storage so that this information can be used to target the appropriate storage platforms as destinations for automatically migrated data.

VERITAS Data Lifecycle Manager does not use performance-profiling features to characterize hardware. VERITAS Storage Central and VERITAS Storage Reporter can be used to gather data and view SRM reports. VERITAS Data Lifecycle Manager can then be used to set policies and automatically target this data to the appropriate storage platforms.

End of Useful Life. Explain how your platform facilitates the automated removal and clean-up of data that has outlived its useful life and restoration of freed capacity for use by applications.

Retention policies can be configured and named which contain retention properties. These retention policies can be referenced/used within a VERITAS Data Lifecycle Manger policy to provide a consistent way of managing data throughout its lifecycle within Data Lifecycle Manager. The policies can be configured to archive, migrate, or delete the data on the primary storage based on a variety of metrics (age, size, name, etc.).

Once a data object exists in a datastore, any online copies of that data can be truncated to immediately free space. This can be event-driven, as in the case of exceeding a free-space threshold, or initiated as necessary by administrators.

VERITAS Data Lifecycle Manager alone cannot ensure the complete annihilation of data. It can overwrite expired media, but a simple overwrite or “zeroing” of the media won’t prevent its recovery. VERITAS Data Lifecycle Manager, in conjunction with VERITAS NetBackup or VERITAS Backup Exec, can eject the appropriate media from a library so that it can be physically destroyed to ensure permanent deletion. VERITAS Data Lifecycle Manager, in conjunction with Network Appliance™ NearStore and SnapLock, provides the capability to store data permanently based on specified retention dates in WORM (Write-Once, Read-Many) volumes. When a retention date expires the data is deleted and is no longer accessible.

Policy Articulation. Describe how policies are created and how they are applied to existing data.

Policy creation begins with selecting a location to look for the desired data. The location could be a whole machine, a set of volumes, or a list of directories and subdirectories. The next step is to define the selection criteria by using attributes such as:

- File name
- File type
- File size
- Accessed, created, and modified dates
- File system attribute (read-only, hidden, archive)
- User ID or group
- Sender/Recipient
- Importance/Sensitivity

These policies can be assigned to one particular machine or they can be “shared policies” that are available to the entire environment. Shared policies allow for changes to be made once and propagated to all affiliated devices.

Device Support. What storage devices does your product support? Are there any proprietary devices (controllers, arrays, HBAs, switches, SAN topology, virtualization products, etc.) that are required for your solution to work?

VERITAS Data Lifecycle Manager supports migration/archival of data from Windows servers, Exchange servers, and Network Appliance storage. Target devices supported include local and network disk storage, optical drives and libraries, and Network Appliance NearStore with optional SnapLock for WORM data permanence, and tape, as well as all devices/libraries that VERITAS NetBackup or VERITAS Backup Exec support, including SAN attached drives.

Resource Consumption. How much bandwidth and server CPU "overhead" is introduced by your solution (e.g., to support polling processes, migration processes, agent processing, access frequency counting, etc.)?

This varies by the amount of data being pulled into or recalled from VERITAS Data Lifecycle Manager. The product uses Microsoft SQL and a content indexing subsystem (Autonomy IDOL), which can also consume varying degrees of resources depending on the actions they are performing.

Due to these large resource variables, VERITAS recommends that VERITAS Data Lifecycle Manager be installed on a separate server from the servers it manages. (VERITAS is in the process of profiling Data Lifecycle Manager for resource utilization as this questionnaire is being completed.)

Flexibility. Does your solution have multi-vendor support? Can data, once integrated into your D/ILM scheme, be migrated readily between your solution and other solutions in this space? Is your solution interoperable with other solutions in this space?

Yes, data can be moved from one storage device to another, and VERITAS Data Lifecycle Manager also supports copying data in its archives to multiple pieces of media for safekeeping/compliance. VERITAS Data Lifecycle manager supports the devices listed in the response to the "Device Support" question above.

One of the great strengths of VERITAS Data Lifecycle Manager is its flexibility and ability to work with third-party applications. For example, a number of industry-specific applications manage content or workflow but not underlying storage resources. VERITAS Data Lifecycle Management makes the storage of an entire enterprise available to all applications with minimal disruption.

Speeds and Feeds. How should a prospective customer compare the relative performance of competitive solutions in this space? What are the appropriate performance measures or metrics to use in evaluating competitive solutions?

Customers should measure any data lifecycle management/information lifecycle management product on:

- Cost and complexity. Data lifecycle management should not have to mean a complete overhaul of your environment, should not require a lot of new hardware, and should function across a variety of data types.
- Overall ease of use: The product should provide transparent setup, monitoring, and administration tools, using wizard-based features where possible.
- Integration with existing infrastructure to leverage existing investments in data protection, media management, and IT skills.
- Flexibility: The solution should include integrated file and message support in one product, reusable policies, scheduling, broad media support, etc.
- Scalability of product: It should be possible to add additional servers/capability to the existing solution, shared policies, device support, etc.
- Reliability/support of product: Data lifecycle management is a long-term investment. Will the vendor still be around in 10 to 20 years? Is high-quality support available for the product?)
- How fast does the product ingest data, recall data, or move data within itself? (Can it keep up?)
- Is the data it keeps on disk/tape in an open format?
- How well can data in the archive be searched? How many file formats are indexed?

Performance of the storage and tape library systems can also affect performance. For example, migrating data from NetApp fast primary storage to fast nearline storage (NearStore) while providing data permanence capabilities

through SnapLock is extremely fast and robust. This may not always be the case if data is being sent to a tape library or alternative storage system.

Interconnect Support. List the network or fabric interconnects that you support. Is there an optimal interconnect for data movement for your solution? What about for management?

VERITAS Data Lifecycle Manager supports SAN-based primary storage if it can be mounted locally to the Windows server and can use SAN-based tape storage via VERITAS NetBackup or VERITAS Backup Exec.

Protocol Support. List the storage networking technologies supported, including FCP, Ethernet, iSCSI, FCIP, iFCP, Parallel SCSI, SAS, and network file system protocols (NFS, CIFS, HTTP, DAFS).

VERITAS Data Lifecycle Manager is at the application layer and therefore utilizes whatever networking technology is available on the Windows server.

Data Type Support. List the data types supported by your solution, including file types and file systems, databases, and hybrids such as email. Comment on the granularity of your D/ILM solution: will it support the migration of database components or subsets, subsets of email files, etc.?

VERITAS Data Lifecycle Manager currently includes Windows NTFS and NetApp CIFS file support, as well as Microsoft Exchange for messaging.

VERITAS intends to add support for UNIX platforms and applications that includes Lotus, IM, and database archiving via third parties later in the year.

Cost. List the cost of your solution or provide some means for calculating cost for a specific environment.

Scenario: Single-processor VERITAS Data Lifecycle Manager server plus two managed File Servers and one managed Exchange Server = $8,900:

- $5,000 for DLM Server (including one FS or Exchange agent)
- $1,200–$1,500 for a File System Agent
- $1,200 for an Exchange Agent
- $1,500 for the QoSS option (to move data from one tier to another)

Standards. List any relevant open standards upon which your product is built.

Microsoft Tape Format (MTF) is used for disk and tape.

Other features. List other pertinent aspects of your solution.

Miscellany: Please note any additional information that you think would be worthwhile for prospective customers to consider about your solution or other solutions in this space.

VERITAS Data Lifecycle Manager extends the strengths of VERITAS's existing data protection software stack to help companies cope with data growth, compliance, data security, data organization, and resource utilization. The software automates the placement and management of data in virtual archives that can span online, nearline, and offline storage media, including unalterable media—all according to user-defined policies.

In addition, VERITAS Data Lifecycle Manager delivers powerful, automated, high-speed search and index technology that reduces the time needed to retrieve electronic records. Furthermore, VERITAS Data Lifecycle Manager provides a unique feature to sweep historical information into these archives by automatically indexing and archiving existing VERITAS NetBackup software tapes. VERITAS Data Lifecycle Manager extends existing backup and restore capabilities to address retention and retrieval by sharing hardware and media management functions with market-leading VERITAS NetBackup and VERITAS Backup Exec software. This allows VERITAS customers to leverage their existing investments in data protection (for example, tape libraries, storage area networks, and licenses) to address compliance without adding new hardware.

VERITAS Data Lifecycle Manager complements VERITAS NetBackup and VERITAS Backup Exec software's backup and restore capabilities with index, search, retrieval, reporting, migration, and archiving to address several key compliance requirements:

- Record Retention and Retrieval: Provides a safe place to store data for as long as required, with the ability to reduce costs by placing data on the

appropriate storage media according to the retention and retrieval service needed.

- Auditable Process: Provides audit trails of changes made to data in the archive and access logs.
- Timely Reporting: Offers access to detailed information, including the ability to produce reports on demand that reflect the origin and activity of data.