THE CLIPPER GROUP
Navigator™
Navigating Information Technology Horizons℠
Published Since 1993 Report #TCG2004019 March 8, 2004

VERITAS Data Lifecycle Manager Actively Exploits Data from Cradle to Grave

Analyst: Michael Fisch

Management Summary

All enterprises should be thinking about data (or information) lifecycle management, because it may be the best – or only – way to meet particular storage requirements of the enterprise. Data lifecycle management is a concept for moving data through different storage quality-of-service tiers as its value changes over time. It seeks to actively and intelligently exploit the lifecycle of data from cradle to grave, with the end goal of an optimized balance between meeting storage requirements and minimizing total cost of ownership.

There are several specific reasons to consider such a solution, any or all of which may apply to your enterprise:

- Meeting regulatory, legal, and/or corporate governance requirements for data retention,
- Improving application performance by trimming a too-large database,
- Lowering storage hardware acquisition costs, and
- Speeding up data management operations like backup, restore, upgrades, or replication.

VERITAS' latest solution is called *Data Lifecycle Manager v5.0*. It automatically migrates file and message data based on policy from primary storage to one or more secondary storage tiers. Key features of the product include:

- Support for Microsoft *Windows*, *Exchange*, and (in the future) Network Appliance fabric-attached storage (FAS) systems,
- Extension to user desktops via the VERITAS *Desktop and Laptop Option* for *NetBackup* and *Backup Exec*,
- Advanced search capability using mathematical algorithms,
- Common media management with NetBackup and Backup Exec,
- Policy-based management,
- Transparent access to migrated data by users and applications,
- Role-based security, and
- Integration with VERITAS *Command-Central Service* for delivering storage as a service.

Read on for details.

IN THIS ISSUE

The Data Lifecycle

Data has a lifecycle, similar to plants and animals. A data object initially springs to life as the creation of an application or user. The object could be a record of a financial transaction, an e-mail, a document, or a thousand other kinds of digital information. The data is placed on a persistent storage medium like disk or tape for safekeeping. Users and applications may access and alter it frequently at first, such as a document undergoing an editing process. This is the *active* state. After the initial flurry of activity, it becomes *fixed* or *inactive*. An enterprise is much less likely to access or change inactive data, though it still needs to keep it for reference purposes – just in case. An unexpected event, like a financial audit or a long-dormant customer deciding to repurchase, could revert inactive data back to active. After a period of time, however, the value of data eventually falls to zero and it can be deleted. **In a broad brushstroke, this is the cradle-to-grave lifecycle of data.**

The question that follows is: How can (or should) an enterprise take advantage of the lifecycle of data? This is focus of *data lifecycle management* (DLM)[1], a concept for actively and intelligently exploiting this lifecycle. **DLM seeks to apply the optimal storage quality of service (i.e., price/ performance tier) to data as its value changes over time.** It is a process that entails classifying data based on policy and then moving it among different tiers over its useful life. The primary reason is to strike an optimal balance between meeting the storage requirements of the business and minimizing total cost of ownership (TCO).[2] DLM avoids two common extremes: (1) Assuming the value of all data is high and spending too much on storing and protecting it or (2) Spending inadequately on storage and suffering performance bottlenecks, downtime, data loss, or penalties for failing regulatory compliance.

Reasons to Consider

So why bother? Why not just go on with "storage as usual?" There are several reasons you may need to consider DLM, depending on the particular pains and issues that your enterprise is facing:

- **Data retention is necessary to meet regulatory, legal, and/or corporate governance requirements.** The recent trumpeting of compliance concerns has pushed the issue of data retention into the spotlight, but it has always been an important business issue. Best practice suggests that all enterprises should define and enforce data retention policies, if only to meet the legal, tax, and operational requirements of the business. In reality, though, it may be the requirement to comply (and fear of penalties for non-compliance) with regulations like Sarbanes-Oxley, HIPAA, SEC 17a-4, etc., that cause many enterprises to take action. Data retention means not only keeping data but also having the ability to readily search and access it.
- **Application performance is slow because of a large, fast-growing database.** Databases are like cargo ships in that they become slow and unstable when overloaded (i.e., too much data). User response times increase, which can negatively impact worker productivity. Instability can cause downtime, also impacting productivity. Pruning inactive records or messages from the database and moving them to a secondary storage tier offloads the extra "cargo", streamlines the database, and makes it faster and more nimble. It also saves on the higher cost of primary storage. The alternative is to continuously tune the database and periodically throw more hardware at it in the form of server processors, memory, and primary storage capacity.
- **Storage hardware acquisition costs are too high.** Archiving inactive data saves the cost differential between the primary

[1] This is also known as *information lifecycle management*.

[2] For details, see *Tiered Storage Classes Save Money – Getting The Most Out Of Your Storage Infrastructure* in **The Clipper Group Explorer** dated August 29, 2002, at http://www.clipper.com/research/TCG2002030.pdf.

and secondary tiers of storage. This is especially attractive if the amount of data is large. Storing data on ATA disk and/or tape can be a fraction of the price of high-end storage arrays with Fibre Channel or SCSI. The savings are magnified, if an enterprise uses multiple replicas of primary data for purposes like disaster recovery, testing, and development – as many do. The total savings would be a multiple of the number of replicas. Smart DLM policies also help promote high utilization of storage resources, especially more costly primary storage.

- **A too-large file system or database inhibits timely backup, restore, upgrades, or replication.** The larger a database or file system is, the longer it takes to backup, restore, replicate, or perform a software upgrade. If any of these processes take an inordinate amount of time, it can cause unacceptable downtime, interfere with production operations, and/or delay application testing and development cycles. There may be more than one way to solve the problem, but using DLM to keep the primary repository at a reasonable size may be the most cost-effective and durable solution. If you also need to address one of the other issues above, faster administrative processes could be icing on the cake.

VERITAS Data Lifecycle Manager

Data Lifecycle Manager v5.0 is VERITAS' latest and most advanced solution for DLM. It migrates data based on policies from primary storage to one or more secondary storage tiers, including online, nearline, or offline storage. It preserves user access to the data and includes advanced capabilities for search and retrieval.

Features

The key features of Data Lifecycle Manager are:

- **File system and messaging application support** – It archives data on servers running the NTFS file system, including *Windows Server 2000* and *2003*, *Windows Storage Server* and *Server Appliance Kit*). It also supports the *Microsoft Exchange Server 5.5*, *2000*, and *2003* for messaging archival. Furthermore, it will be integrated with Network Appliance's *DataONTAP* operating system to support its line of fabric-attached storage (FAS) platforms.
- **Extension to end-user desktops** – When used in conjunction with the VERITAS *Desktop and Laptop Option* for *NetBackup* and *Backup Exec*, Data Lifecycle Manager can also archive data from end-user desktops.
- **Advanced search capability** – It indexes data using mathematical algorithms, not simple keyword lists, so searches are faster and more comprehensive. It can also index and archive backup data retroactively (from VERITAS NetBackup or Backup Exec), providing enterprise-wide and historical search capabilities. Good search and retrieval are essential for a DLM solution.
- **Common media management with backup/restore** – Data Lifecycle Manager shares a media management facility with VERITAS NetBackup and Backup Exec, allowing them to use the same storage resources. These can include disk, optical, and tape in SAN- or direct-attached configurations. WORM[3] devices are supported, as well. Common media management with backup promotes storage consolidation, easier administration, and higher resource utilization.
- **Policy-based management** – Administrators can set up policies that automatically trigger data migration based on criteria like age, frequency of access, type, etc. It can also include capacity utilization of a storage resource, such as a maximum utilization threshold that triggers a migration. *See sidebar below for details about policies.*

[3] *Write Once Read Many* media provides non-erasability and non-rewritability for meeting certain regulatory requirements.

- **Access transparency** – Users and applications can access archived data through normal interfaces. After migrating data, Data Lifecycle Manager puts a placeholder in the primary storage resource. If access to the migrated data is requested, it restores the data from the secondary storage tier to the primary.
- **Role-based security** – Administrators and users can be assigned and limited to specific functions to ensure security and balances for the DLM process.
- **Integration with VERITAS *CommandCentral Service*** – Along with NetBackup, Backup Exec, *SANPoint Control*[4], and *Storage Reporter*, Data Lifecycle Manager interfaces with CommandCentral Service, which is VERITAS' "manager of managers" for delivering storage as a service. As a result, Data Lifecycle Manager is already an active, integrated component of VERITAS' vision for utility computing.

Architecture

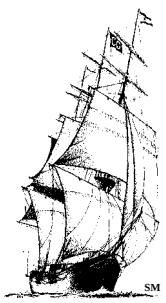
Data Lifecycle Manager is designed to support an enterprise environment and consists of the host server as well as application agents. The host server runs the primary DLM application and enforces policies, moves data, and performs searches and retrievals. It can share the same server as NetBackup and Backup Exec to promote server consolidation. It can also be clustered for highly-available data access. The application agents run on the Windows or Exchange servers that are participating in the DLM process. If archiving is extended to desktops, the Desktop and Laptop Option for NetBackup or Backup Exec must run on end-user systems. There is a centralized administrative console for the entire DLM environment.

Defining Data Lifecycle Management Policies

The key to successful policy definition is to allow the broader context of the business to shape them. Business stakeholders and IT personnel need to be involved in the process because no single group has a full sense of what can or ought to be done. Functional personnel have the best sense for their particular business process and application requirements. IT personnel have the best understanding of technical issues and ramifications. Where regulatory and legal requirements are concerned, executives and even legal counsel will want to provide input and direction. How data is handled is critically important to an enterprise, so communication and a consensus-oriented approach are recommended to ensure policies reflect the needs of the whole business.

Conclusion

An enterprise might consider DLM solution for a variety of reasons: data retention/compliance, application performance, storage cost reduction, and streamlining data management operations. **If you are looking for a DLM solution that covers Windows servers, Microsoft Exchange, and/or – soon – Network Appliance storage platforms, then consider VERITAS Data Lifecycle Manager.** The highlights include its comprehensiveness, transparency to applications and users, and sophisticated search and retrieval. It is especially attractive for existing NetBackup and Backup Exec customers because of the integration and synergies between the solutions. Data Lifecycle Manager can simply ride on top of the existing backup/restore infrastructure – now that's leverage!



[4] See *VERITAS SANPoint Control – Untangling the Web of Networked Storage* in **The Clipper Group Navigator** dated May 23, 2003, at http://www.clipper.com/research/TCG2003023.pdf.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Michael Fisch **is Director of Storage and Networking for The Clipper Group.** He brings over eight years of experience in the computer industry working in sales, market analysis and positioning, and engineering. Mr. Fisch worked at EMC Corporation as a marketing program manager focused on service providers and as a competitive market analyst. Before that, he worked in international channel development, manufacturing, and technical support at Extended Systems, Inc. Mr. Fisch earned an MBA from Babson College and a Bachelor's degree in electrical engineering from the University of Idaho.

- ***Reach Michael Fisch via e-mail at Mike.Fisch@clipper.com or at 781-235-0085 Ext. 25.***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.