

Extended Data Life Management: Protecting Data Over Long Periods of Time

NOTICE

This Technology Brief may contain proprietary information protected by copyright. Information in this Technology Brief is subject to change without notice and does not represent a commitment on the part of Quantum. Although using sources deemed to be reliable, Quantum assumes no liability for any inaccuracies that may be contained in this Technology Brief. Quantum makes no commitment to update or keep current the information in this Technology Brief, and reserves the right to make changes to or discontinue this Technology Brief and/or products without notice. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or information storage and retrieval systems, for any person other than the purchaser's personal use, without the express written permission of Quantum.

TABLE OF CONTENTS

Introduction	3
EDLM Overview	3
Library Partitions	3
EDLM Components	5
Library Managed Partition (LMP)	5
EDLM Scanning Drives	5
Policy Based Scanning	6
Integration with StorNext Storage Manager	7
Candidate Policies	7
Scanning Policies	7
Results Action Policies	8
Putting the Policies Together Into Policy Sets	8
EDLM Recommended Practices	9
Using EDLM to Support Archiving	9
Using EDLM to Support Disaster Recovery (DR)	10
Frequency of Scheduled Scans	10
RAS Ticketing	10
Sizing the EDLM Library Managed Partition	10
Summary	11

INTRODUCTION

Media errors have historically been one of the main causes of backup interruptions and failures. In 2005, the Quantum Scalar® tape libraries led the industry in addressing this problem with the introduction of the Advanced Reporting feature. Advanced Reporting offered a media integrity report that made it very easy for users to identify suspect or bad media in their environment, significantly reducing backup failures due to media errors. Within a few years most library manufacturers had the capability to monitor and report on media errors and problem media, largely resolving this issue for the industry.

Over the last few years the role of tape has changed to more of a long-term retention solution. For most organizations it makes sense to use disk to serve as the short-term backup solution, typically moving the data to tape after 30-60 days. Retention periods for most corporate data are longer than four years, with many companies keeping their data forever. As archived data ages it is accessed less frequently; and DR data is stored for years and may never be accessed. Once data is no longer accessed the media integrity statistics are no longer collected, so storage administrators don't know if their data in long-term storage can be restored. A new mechanism is needed to ensure that the data can be restored or retrieved when needed. Quantum developed the Extended Data Life Management (EDLM) feature in the Scalar i6000 tape libraries to address this need.

EDLM OVERVIEW

EDLM is an optional feature of the Scalar i6000 that provides policy based data integrity checking. In Quantum StorNext® environments, EDLM offers a policy to trigger tape scans based on StorNext's tape suspect count – enabling truly proactive, automated scanning of tapes. In addition, when suspect tapes are discovered through the EDLM scan, the Scalar i6000 works with StorNext so that StorNext can move those files to a new tape.

Library Partitions

Before EDLM can be fully explained, library partitions must be understood. The Scalar i6000 has the capability to partition the library into multiple logical library partitions. There are two types of partitions that can be created.

- **Host Managed Partitions (HMP):** These “online” partitions are presented to applications and are typically used to support backup and archive operations. The Scalar i6000 supports the creation of up to 16 host managed partitions. A HMP must contain at least one tape drive, six cartridge slots. A Partition license must be installed to use this feature.
- **Library Managed Partitions (LMP):** These “offline” partitions are not presented to applications and are typically used to support special library operations that are conducted in the background, with no impact to operations of host managed partitions. Currently there are three types of LMPs that users can create:
 - **EDLM Partitions:** At least one EDLM partition must be created to conduct EDLM scans. Typically, only one EDLM partition is needed as this partition can conduct scans on cartridges within the entire physical library. An EDLM license must be installed to use this feature.
 - **Active Vault Partitions:** These partitions are used primarily as an alternative to archive vaulting outside the library.
 - **Automated Media Pools:** This partition creates a common pool of media and slots that can quickly and non-disruptively be added to a host managed partition or application.

The figure below illustrates the basic concepts of using EDLM. First, the user configures an EDLM library managed partition (LMP). The LMP contains all the scanning drives used by the physical library and also contains any EDLM slots that would be typically used to load DR cartridges or vaulted archive cartridges into the library without introducing them back into an online partition.

EDLM Operations

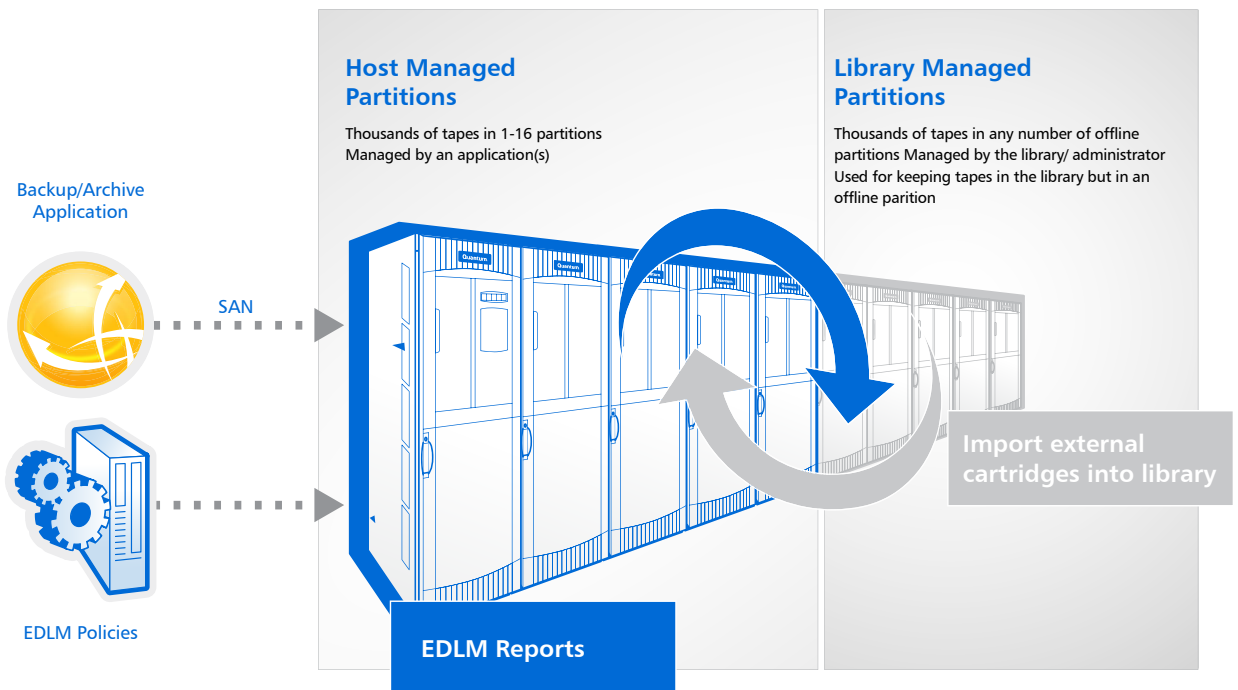


Figure 1: EDLM Operations Illustration

Next the library administrator configures the scanning policies to be used. These policies define what tapes get scanned, when they are scanned, and how they are scanned. More detail on scanning policies are provided later. Once the policies are configured, the library will conduct the scans and create results reports that can be viewed through the library GUI or exported.

Note: EDLM comes standard as part of the StorNext Archive Enabled Library appliance, but it functions the same regardless of the library used. For this reason a distinction is not made in this paper between EDLM running on the Scalar i6000 or the StorNext Archive Enabled Library appliance.

EDLM COMPONENTS

In order to use EDLM to scan tapes the user must have the following components installed and configured in the library:

- EDLM license: An EDLM software license must be installed on the library. Only one license is needed per physical library.
- EDLM Library Managed Partition (LMP): Once the EDLM license is installed the user will have the option of creating a library managed partition or LMP. The EDLM library managed partition is used as an offline workspace to scan tapes—ensuring that the scanning function does not interrupt normal library operations, and that there is no contention between the managing host application and the scanning process. To use EDLM, an LMP must be configured on the library.
- EDLM scanning drives: At least one EDLM scanning drive must be installed and assigned to the EDLM partition. Any number of EDLM scanning drives can be installed within the normal drive support limits of the Scalar i6000 (from 1 to 96).

Library Managed Partition (LMP)

The LMP is a special partition that is only presented to the library administrator(s)—it is not presented to any external applications or hosts. Sectioning off the LMP from all applications and hosts allows EDLM operations to run in the background with no impact to customers' backup or archive operations. In addition, it allows users to import cartridges into the LMP for scanning that they normally would not want to reintroduce into a host managed partition.

The EDLM LMP contains its own resources: scanning drives, cartridge slots, and cartridge import/export slots. A minimum of one scanning drive and six cartridge slots must be assigned to the LMP. EDLM scanning drives are connected to an I/O blade(s) to assist with managing the scans.

EDLM can scan tapes that reside anywhere within the physical library, whether they are host managed partitions for archive/backup partitions or library managed partitions. All scans occur in the background with no impact to host managed partition operations. To scan a tape from an online partition, EDLM "borrows" the cartridge from the online partition to conduct the scan. If an external application requires a tape that EDLM is scanning, the EDLM scan will be aborted and the cartridge returned to the host managed partition.

EDLM Scanning Drives

EDLM uses special purpose EDLM scanning drives to conduct the scans. Customers can install any mix of LTO-4 or LTO-5 scanning drives, allowing users to scan LTO-2 through LTO-5 cartridges. These drives can be key enabled which means that any tapes encrypted using Quantum's Scalar Key Manager (SKM) or EMC's RSA key manager can be scanned by EDLM drives.

EDLM scanning drives can only be used for EDLM scanning and cannot be used for backup or archive operations. Using dedicated scanning drives offer two distinct advantages to customers:

1. Eliminates any contention between production operations and EDLM scanning. This benefit is particularly important for customers that have numerous tapes they want to scan without having to manage complex scanning schedules. Customers can simply schedule hundreds or even thousands of scans to execute without having to work around production schedules.
2. EDLM drives are assigned to the EDLM LMP which services the entire physical library. Thus the EDLM scanning drives can be used by all partitions without any resource contention. Using a common pool of scanning drives also reduces the number of drives needed to conduct scans and conducts the scans in the least amount of time.

POLICY BASED SCANNING

Any tapes residing in the library can be scanned at any time if the user starts the scan(s) manually, but typically user definable policies are used to automate the scanning process. Three types of policies are defined:

- Candidate policies define what tapes are scanned and when those tapes are scanned
- Scanning policies define what scanning method to use on the tapes
- Results action policies determine what action to take when a suspect or bad result is returned from the scan

These three policies combine to form a policy set. A policy set always includes a candidate policy, a scanning policy, and a results action policy. Policy sets are assigned to a partition and multiple policy sets can be assigned to a partition. Figure 2 illustrates how policies are used to form policy sets which are then assigned to a partition.

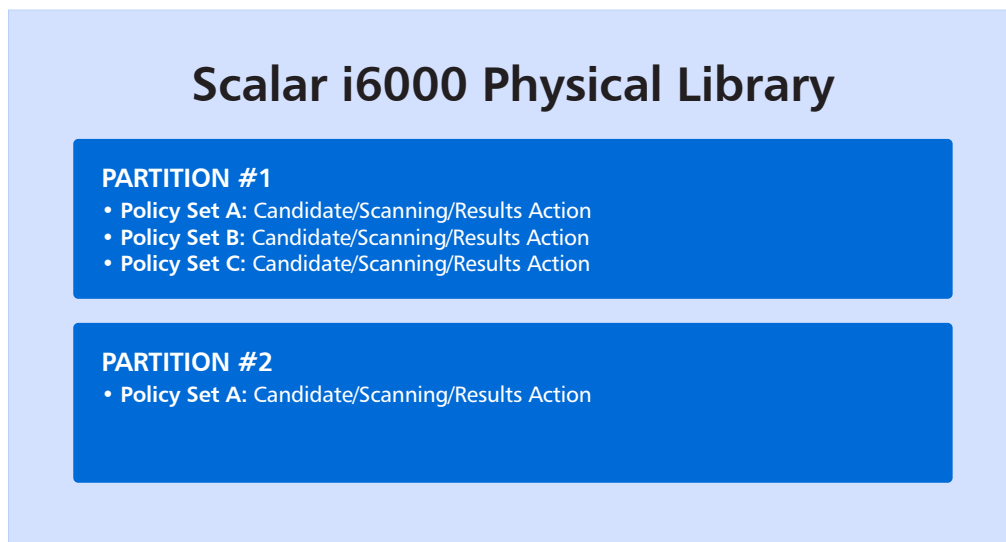


Figure 2 – Example of How EDLM Policy Sets are Assigned

In the above examples, partition #1 contains three policy sets. Each of these policy sets must contain a candidate policy, scanning policy, and results action policy. Partition #2 only has one policy set, but the policy set still contains a candidate policy, scanning policy, and results action policy. Examples of specific policies with policy values are illustrated later in this section.

Integration with StorNext Storage Manager

StorNext Storage Manager users can take advantage of two EDLM scanning policies that are integrated with StorNext. These StorNext specific policies are described in the next sections. To take advantage of these policies, customers must also have the StorNext API (SNAPI) installed.

Candidate Policies

There are two types of candidate policies that users can set.

- **Event Triggered:** Users can select specific events that will trigger the library to scan the tape associated with the event. There are three events that users can select:
 1. **Cartridge Import:** Users can set this policy so that any time a cartridge is imported into the partition, EDLM will scan that cartridge or set of cartridges.
 2. **Tape Alerts:** If selected, tapes will be scanned if the threshold for the number of tape alerts on a cartridge is met. The default number of tape alerts is three but users can adjust this number. The tape alerts for cartridges are checked after each dismount.
 3. **StorNext Suspect Count:** StorNext keeps a “suspect count” on each tape cartridge. If this option is selected the library will check with StorNext after each tape dismount to determine if the suspect count threshold has been reached. If so, EDLM will scan the tape.
- **Scheduled:** If this option is selected tapes will be scanned at regular intervals. The user selects the number of days between scans.

Scanning Policies

Once the candidate policies are defined, the end user must select a type of scan to be used with the candidate tapes. There are three levels of scanning the user can select:

1. **Quick Scan:** Does not scan the tape. Evaluates data from the cartridge memory (CM) only. A quick scan takes less than one minute per tape.
2. **Normal Scan:** Evaluates the cartridge memory (CM) and scans selected portions of the tape, focusing on areas most likely to indicate problems. A normal scan can take 20 minutes per tape. This time should remain the same regardless of the LTO media generation being scanned.
3. **Full Scan:** Evaluates the cartridge memory (CM) and scans the entire tape. The time it takes to complete a full scan varies based on the generation of the LTO media. The table below provides the time it takes to complete a Full Scan on LTO media.

LTO Media Generation	Native Throughput (MB/s)	Native Capacity (GB)	Native Max Scan Time (hrs)
LTO-5	140	1,500	3.0
LTO-4	120	800	1.9
LTO-3	80	400	1.4
LTO-2	40	200	1.4

Table 1 – EDLM Scan Times by LTO Media Generation

The scanning times listed in the table assume that the tapes are full of data. EDLM scans the portion of tape with data. If the tapes are less than full, the time it takes to complete a Full Scan will be similarly reduced.

Results Action Policies

Once the candidate tapes and the test type is defined, the end user must determine what actions they want EDLM to take when a scan results in a suspect or bad result.

1. RAS Ticket Generation: When a scan returns a suspect or bad result the iLayer management software can generate a RAS ticket and notify the end user through e-mail of the suspect/bad scan result.
2. StorNext Copy Tape Request: This policy automatically requests StorNext Storage Manager to copy all data from a suspect and/or failed tape to another tape. Once you enable this policy, you can select whether to copy failed tapes, suspect tapes, or both. Also, once this policy is selected a RAS ticket will be generated for each request to copy data indicating whether the request succeeds or fails. Note the StorNext API (SNAPI) must be installed before this policy can be used.

Putting the Policies Together Into Policy Sets

The figure below gives examples of how the specific EDLM policies are used to form policy sets and how these policy sets get applied to a partition.

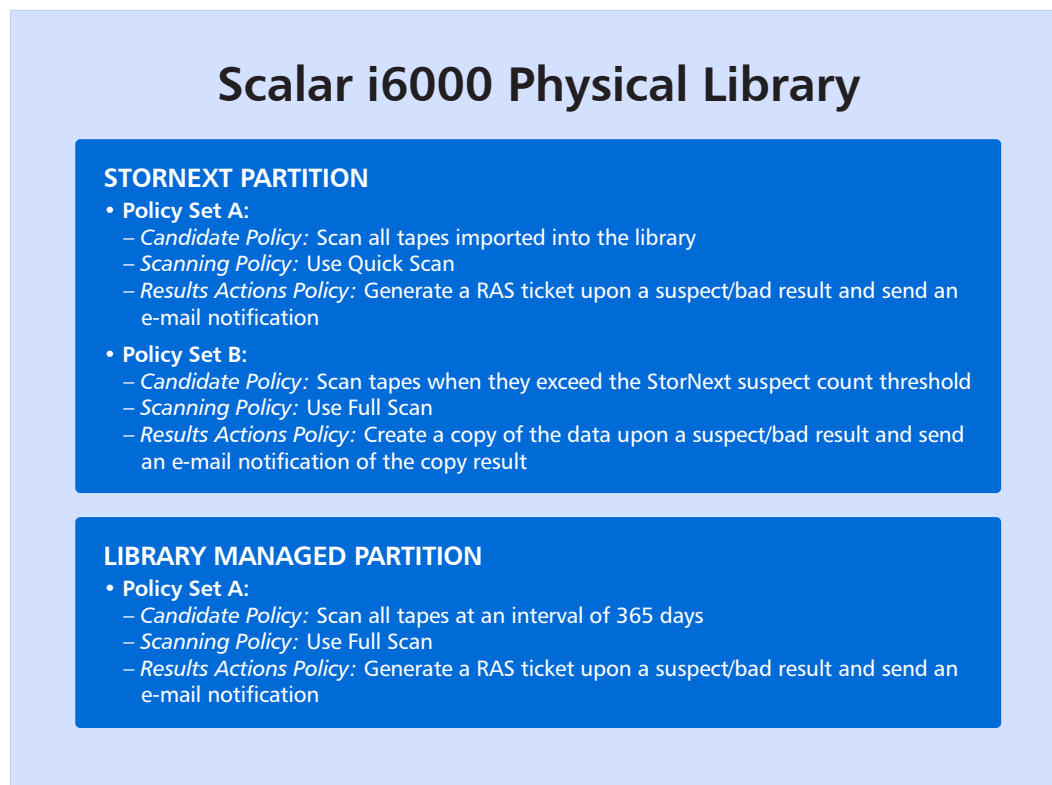


Figure 3 – EDLM Policy Sets Examples

EDLM RECOMMENDED PRACTICES

Using EDLM to Support Archiving

Customers using Scalar tape libraries to archive their data can store their tapes in three different locations. Each of these locations are described below with recommendations on how to use EDLM to protect that data.

1. Archive tapes stored in a host managed partition: tapes residing in a host managed partition offer the fastest access to the data and the best level of security and data protection. For these host managed partitions, the following policies should be considered:
 - a. Conduct a quick scan every time a new cartridge is imported into the partition. This action is useful to get a baseline on the cartridge. But after the baseline is taken this scan option is generally not used on its own.
 - b. Conduct a scan whenever the tape alert exceeds a threshold. The default value is three tape alerts. Using this policy will conduct a scan whenever a potential problem arises. If time allows it is recommended to use the Full Scan option to make sure all the data on the tape is scanned.
 - c. In StorNext Storage Manager partitions, the user could also configure scans to occur if the StorNext suspect count reaches its threshold. This policy could be used in addition to the tape alert candidate policy. If time allows it is recommended to use the Full Scan option to make sure all the data on the tape is scanned.
 - d. If tapes that are not accessed frequently are kept in an online partition, a scheduled scan is recommended. Conducting scans at an interval of time, somewhere between 6 months and 18 months, using the normal scan method would keep appropriate health checks on the media.
2. Archive tapes stored in a library managed partition: tapes residing in a library managed partition are typically not accessed very often—to access the data on these tapes the customer would need to move/import the tapes back into a host managed partition. Since these tapes remain in the library they have a much lower probability of being damaged, etc. For these library managed partitions, a couple of policies make sense:
 - a. Conduct scans at an interval of time, somewhere between 6 months and 18 months, using the Full Scan method.
 - b. Most of these tapes will be moved/imported directly from an online partition. Thus, these tapes would have been scanned if any issues were noticed during normal operations. While it is probably not necessary, scanning any tapes that are imported may be useful to provided an added level of protection. Setup a policy to scan tapes upon import, using the normal scan method.
3. Tapes stored outside the library: tapes stored on a shelf outside the library are typically not accessed very often—to access the data on these tapes the customer would need to import the tapes back into the library, probably into an library managed partition (the library managed partition). For these tapes, a couple of options make sense:
 - a. No policy needed, select all tapes in the partition and start a scan manually.
 - b. Since these tapes will be imported into the partition, setup a policy to scan tapes upon import, using the full scan method.

Using EDLM to Support Disaster Recovery (DR)

The key difference in using EDLM in a DR versus an archive environment is that all the DR tapes are physically exported from the library and sent offsite. This second site could be owned by the customer's organization, e.g., a second data center, or sent to a third party site for long-term storage.

1. For cartridges that are stored at a secondary customer site, it is recommended to have a Scalar i6000 at this secondary site to conduct EDLM scans for the primary site tapes. If the secondary site has a Scalar i6000 to support operations at this site, this library can be used. Extra storage capacity may be needed to size the LMP to the size needed to support the EDLM scans. Having a library on site prevents the tapes from having to be moved offsite, protecting the tapes from potential damage, reducing the security risk, and avoiding transport and handling costs. In addition to scanning tapes at the secondary site, the library at the primary site should be setup to conduct scans for the secondary site tapes. For these tapes, it is likely that the customer does not want to load them into a host managed partition, so the LMP should be used. Either of the following policy options could be used for these tapes:
 - a. No policy needed, select all tapes in the partition and start a scan manually.
 - b. Since these tapes will be imported into the partition, setup a policy to scan tapes upon import, using the full scan method.
2. Cartridges that are stored at a third site are very similar to cartridges stored at a secondary customer site. The main difference is that these tapes will need to be transported to the data center for scanning. For this reason, it is recommended that the customer conduct scans on a sample of the cartridges in long-term storage. If these scans return suspect or bad results, the customer can select more tapes for scanning. These scans would be conducted using the same policy options as described above.

Frequency of Scheduled Scans

For candidate policies that use the schedule option, users must define the number of days between scans. Days are used to define the interval between scans because it provides the user with flexibility in defining the schedule. However, it is more appropriate to think of the interval in terms of years. Each user application is different but generally Quantum recommends setting the frequency of scans from one to three years.

RAS Ticketing

For all EDLM uses, it is recommended to enable the RAS ticketing option as part of the Results Action Policy. The number of tapes that fail at any one point in time will be small so it shouldn't cause a high number of RAS tickets and notifications. It is important for administrators to know when their data is at risk and when they potentially have a cartridge that could cause disruptions to operations.

Sizing the EDLM Library Managed Partition

EDLM contains two LMP components that must be sized appropriately to efficiently scan tapes: the LMP slots and the scanning drives.

The EDLM LMP is a partition that must be assigned slots like any other partition. The EDLM LMP should be sized for the maximum number of tapes needed to reside in the EDLM LMP at one time. For archive applications this may be all the tapes that are vaulted outside the host managed partition or it could be a subset of those cartridges. For DR applications it will typically be a subset of the DR cartridges.

To size the number of EDLM scanning drives define the number of tapes to be scanned, the method of scanning (see the Scanning Policies section), and amount of time allotted for the scans to complete. For example, if a customer had 500 LTO-5 tapes to be scanned every month and the full scan method was selected, the customer should purchase at least three LTO-5 EDLM scanning drives, as defined below.

- $500 \text{ tapes} \times 3 \text{ hours/tape} = 1,500 \text{ hours}$ or 62.5 days with one LTO-5 drive
- $62.5 \text{ days} / 30 \text{ days per month} = \text{at least three LTO-5 EDLM scanning drives}$

Many calculations will become more complex as multiple LTO generations need to be scanned. The following list provides the number of LTO tapes that can be scanned each day using a single drive

- LTO-5 media: 8 tapes/day
- LTO-4 media: 12 tapes/day
- LTO-3 media: 17 tapes/day
- LTO-2 media: 17 tapes/day

The capabilities listed above are maximums, it is good practice to size the LMP with 1-2 extra drives to allow for additional scans based on triggered candidate policies and allow for some extra time to complete the scans.

SUMMARY

Protecting active data on tape over a relatively short period of time is very different than protecting inactive data on tape over long periods of time. To efficiently protect inactive data over long periods of time, a new process and capability is needed: Extended Data Life Management meets this need. EDLM runs in the background, allowing users to add this process without impacting their normal backup and archive operations. Policies automate the entire process so minimal administrative time is needed to implement EDLM.

All organizations retaining archive or DR data for more than three years should implement EDLM to protect their data over the long-term retention period. Existing Scalar i2000/i6000 libraries can easily be upgraded to accommodate EDLM scanning without impact to existing operations. Scalar i6000 libraries can also be placed at secondary sites specifically for EDLM scanning.

For contact and product information,
visit quantum.com or call 800-677-6268

Quantum®

Preserving the World's Most Important Data. Yours.™

©2011 Quantum Corporation. All rights reserved. Quantum, the Quantum logo, DXi, Scalar and StorNext are either registered trademarks or trademarks of Quantum Corporation and its affiliates in the United States and/or other countries. All other trademarks are the property of their respective owners.

About Quantum

Quantum Corp. (NYSE:QTM) is the leading global specialist in backup, recovery, and archive. From small businesses to multinational enterprises, more than 50,000 customers trust Quantum to solve their data protection, retention and management challenges. Quantum's best-of-breed, open systems solutions provide significant storage efficiencies and cost savings while minimizing risk and protecting prior investments. They include three market-leading, highly scalable platforms: DXi®-Series disk-based deduplication and replication systems for fast backup and restore, Scalar® tape automation products for disaster recovery and long-term data retention, and StorNext® data management software for high-performance file sharing and archiving. Quantum Corp., 1650 Technology Drive, Suite 800, San Jose, CA 95110, (408) 944-4000, www.quantum.com.