Intelligent Backup with the PowerVault ML6000 Modular Tape Library

A whitepaper from Dell PowerVault Storage
CONTENTS

INTELLIGENT BACKUP WITH THE POWERVAULT ML6000 MODULAR TAPE LIBRARY 3
HIGHER LEVELS OF INTEGRATION 4
BUILT-IN LIBRARY MONITORING NETWORK AND ALERTING 5
PROACTIVE READINESS VERIFICATION 6
ADVANCED REPORTING SYSTEM 7
SUMMARY 8

FIGURES
FIGURE 1: INTEGRATING ADVANCED LIBRARY FUNCTIONS 3
FIGURE 2: PARTITIONING INTO LOGICAL LIBRARIES 4
FIGURE 3: MONITORING AND ALERTING 6
FIGURE 4: PROACTIVE PATH VERIFICATION 6

FOR FURTHER INFORMATION AND DISCUSSION
Visit www.dell.com/storage

THIS WHITEPAPER IS FOR INFORMATIONAL PURPOSES ONLY AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.
SECTION 1
INTELLIGENT BACKUP WITH THE POWERVAULT ML6000 MODULAR TAPE LIBRARY

THE POWERVAULT ML6000 ARCHITECTURE

Data protection for the corporate data center tends to go through predictable evolutionary phases—which incidentally closely match the advances seen generally in open systems, clientserver computing environments. Early data protection strategies relied on single tape drives, first located near servers, then mounted inside them. The single drives were initially replaced with small autoloaders, then with compact tape libraries. For most IT organizations today, the norm has become larger, multi-function automated storage libraries capable of consolidating backup for multiple processes in a storage network environment.

The most obvious changes required from backup systems to support this evolution include raw capacity and throughput increases, but those advances are really only the beginning. Libraries have become full storage systems which add increased value to the backup process, even providing performance advantages for the tape drives inside them, data path connectivity verification, and data moving agents. Initially, most of these functions begin externally—requiring the separate purchase, integration, and operation of add-on products or functions. But with each new generation, more of these capabilities become internal to the library. The overall direction of this evolution is clear and so is the value to the IT manager. As new generations of libraries are created, more integrated functions are provided, management and operational overhead is reduced, overall reliability is increased, and as a result, costs go down.

The intelligent architecture of the PowerVault ML6000 represents the next generation in that evolution for tape libraries, providing a new set of integrated library functions.

FIGURE 1: Integrating Advanced Library Functions

The ML6000 architecture allows a wide variety of advanced backup system functions, which normally require external servers and software, to be integrated directly into the library. This is designed to provide increased functionality, easier management, and lower costs.
The ML6000 introduces a new control and management system based on an intelligent, blade-based computer architecture built directly into the library. This intelligent engine allows the library to integrate most of the functions that earlier generation libraries provide through external servers or other components. It also expands what libraries can do, adding more complete operational reporting, proactive readiness checks, continuous system monitoring, and service-oriented action advisories. Fundamentally, it provides a unified approach to managing all the library functions from a single point, including robotics systems, drives and media, system diagnostics, service advisories, and the connectivity between the host and the library.

SECTION 2
HIGHER LEVELS OF INTEGRATION

The first major advantage to the integrated management platform is the elimination of external components needed to manage the kinds of consolidated operations typical of data center environments. These include SAN connectivity and the partitioning of a library into several, smaller logical or “virtual” libraries to support multiple backup and storage processes.

FIGURE 2: Partitioning into Logical Libraries

The PowerVault ML6000 supports native partitioning of the physical library into up to 18 separate logical libraries. Logical libraries present themselves to applications as completely distinct units, allowing a single physical library to support multiple servers, operating systems, media types, and backup software applications. Logical libraries make it easy to consolidate a number of backup operations on a single tape resource.
They even allow simultaneous operation with different kinds of media, and can connect different protocols to the library. Partitioning of a library is important in data consolidation environments because it allows a single library to support different processes. These commonly include more than one operating system, server cluster, and backup software application. Without the integrated management, these functions are supported, if at all, only by running additional library management software on separate external servers. Connectivity for SAN environments is also commonly supplied in the broader industry through external storage router products.

In the ML6000, all of these functions are native to the library, requiring no external management, no external servers, and no external connectivity products. For example, the ML6000 can be divided into up to 18 logical libraries, which can support different media types, from a simple library management interface available directly through the library control panel and at all remote monitoring locations. Connectivity is supplied through a data management system that includes embedded optional I/O blades to provide aggregation, management, and data path conditioning functions for Fibre Channel interface drives. The blade architecture includes in-band data management, including enhanced protocol management that provides significant performance advantages for the drives, as well as streamlining system management. All of these functions are supplied as native applications of the library, providing increased functionality, easier setup and maintenance, and low cost.

SECTION 3
BUILT-IN LIBRARY MONITORING NETWORK AND ALERTING

A second element of the ML6000 intelligent architecture is a separate library monitoring system designed directly into the library at different points. One of the primary tools is an internal, out-of-band Communications Area Network (CAN) that links separate systems and processors located throughout the library back to the central processor. The CAN, along with other data gathering tools, allows the collection of data from multiple elements into a central database that can be reported and analyzed, turned into active responses based on user input or policies, and used to control the activities of the library.

The first kind of action that is provided by this integrated data gathering and reporting is increased operational information about what is happening inside the library. For example, the ML6000 gathers ongoing throughput history that shows the total amount of data transferred by the library hour by hour. This information is available for the entire library and for each logical library, and is captured in various data logs. This throughput history will allow a local or remote operator to see whether different backup jobs transferred the amount of data expected directly from the library without having to log onto different backup software applications. Other operational data collected includes information such as number of tape mounts and the total library utilization. All of this data provides an enduser a single location to see useful information about how the library itself is operating.

The same reporting tools also monitor other kinds of data that is normally used only when there is some exception or anomaly. In the ML6000, these include the ongoing history of each drive and piece of media, as well as data about voltage, current and temperature changes within each of the library subsystems. The library even captures the rotation of cooling fans and status of primary and secondary power supplies. All this information is not normally used in day-to-day operations, but changes in it may indicate the need to perform service on the library. Capturing the history of soft write errors for media and drives, for example, can provide early information about degraded performance or worn or damaged media. Voltage, current, or temperature changes may show similar trends.
SECTION 4
PROACTIVE READINESS VERIFICATION

The third kind of data gathered by the ML6000 library includes several different kinds of proactive readiness checks. Rather than waiting for conditions in the library to change, the proactive readiness checks automatically execute routines within the library and its storage network environment to verify that components and connections are ready to transmit data. Several levels of readiness checks are integrated into the ML6000. At user-defined intervals, signals are sent by the primary controller through the library’s data paths to each of the system’s tape drives. They either confirm readiness to read and write data, or they report conditions that need further diagnosis. Even more comprehensive checks are also available. Since the data paths from the host backup servers through the switched fabric to the library are often not exercised between backup jobs, the ML6000 can also check the paths outside the library. This is new with the optional I/O blade, and is accomplished by using an agent to send a signal from the host through the paths in the fabric that backup data will use to the library. Successful transfer confirms that paths—including hosts, wiring, switch ports, and zoning configurations—are ready to transfer data.

FIGURE 4: Proactive Path Verification

PowerVault ML6000 libraries with the fibre channel I/O blade proactively verify that the data paths between all hosts and all drives are ready to carry data. Because it checks connectivity while the system is inactive, path verification can discover system faults and let users take corrective action before their backup begins.
Readiness checks are especially important in data center backup environments where networks tend to be complex with many connectivity options. Because checking both the drives and the connections in the fabric that will support the backup allows users to discover problems before they can negatively affect the backup process, it is a process that can significantly increase the overall reliability of the data protection system.

SECTION 5
ADVANCED REPORTING SYSTEM

Turning the data gathered into useful information for end users and delivering those messages effectively is another area in which the ML6000 represents a next-generation advance. The ML6000 libraries communicate information in several ways. For operational data, a real-time user-readable display, available locally and remotely, provides key data points—including parameters such as immediate throughput history. For event-based reporting, the library includes a fully integrated policy engine that allows different event thresholds to trigger alerts. These alerts are issued in several ways: e-mail messages (which can link to pagers and SMS messages) to specified contacts, local and remote system displays, and SNMP alerts which allow the events to be automatically registered by standard network management framework tools. It is important, however, that the library does not rely on the network management tools for contact: network management tools, which tend to be complex to set up and manage, are not reporting backup operations in most data centers.

The alerts are set by event thresholds identified with yellow or red indicators on the control panel and in the other automated contact mechanisms. Red indicators normally mean a service call, and usually have preset factory thresholds. Yellow indicators have a range of threshold points and can have several action advisories. In some cases, the action may indicate that the end user needs to clear a piece of media or replace one that the system identifies as worn or damaged. In others, it may outline additional quick diagnostic steps that an end-user can easily take to bring the unit back into full normal operation mode, or to confirm that a service call will be necessary. If a service call is required, the library’s distributed Control Area Network and system-wide reporting and monitoring functions also provide a highly granular fault isolation capability. By tracing potential problems with a high degree of accuracy to single components, it lets service personnel show up with the right components to bring the unit back to operational status as rapidly as possible.

One of the key advantages offered by the ML6000 architecture is a roadmap that can add functions in future product releases. As the library gathers, stores, and analyzes operational data, that information can form the basis for a system with higher levels of predictive value. Ultimately, the library will allow the engineering, service, and support teams to analyze patterns and dependencies that will lead initially to more refined predictive analysis and later to autonomic system self-correction.
SECTION 6
SUMMARY

The ML6000 architecture represent a new generation of system functionality. It can save time and money for IT departments by embedding functions within the library that normally require external servers and software in earlier generation products. It can make the IT manager’s job easier by increasing the levels of data monitoring and reporting. And it can increase the overall reliability of the entire data protection process by performing proactive system readiness checks and automatically alerting users about changes that could jeopardize successful backup. Looking to the future, the ML6000 architecture is also providing a foundation and a knowledge base which will be used to increase the sophistication of future management functions.