SearchStorage.com

E-Guide

Virtual Tape Libraries and Data Deduplication in Enterprise Storage Today

A virtual tape library (VTL) today has varying definitions but most agree that an enterprise-class virtual tape library should back up at least tens of terabytes of data daily and be able to store hundreds. High-end virtual tape libraries must scale easily and support complex backup and recovery environments. This helps ease the ability of adding disk capacity in turn backing up more and more data all the time. Download this guide to learn more about the solutions surrounding virtual tape libraries and data depulication in today's enterprise storage environments.

SearchStorage.com
E-Guide

Virtual Tape Libraries and Data Deduplication in Enterprise Storage Today

Table of Contents

Fujitsu upgrades data deduplication backup hardware, launches VTL and NAS devices

By Sonia R. Lelii, Senior News Writer

Fujitsu America is expanding its data deduplication backup portfolio with upgrades to its Eternus CS High End enterprise virtual tape library (VTL) and two new configurations for its CS800 S2 data backup appliances for midrange and small- to medium-sized (SMB) customers.

The latest version of the CS High End supports cascading synchronous replication of deduped data, encryption and write once, read many (WORM). The CS High End scales to 2.4 TB with a single controller and can scale to 1.6 PB with multiple systems. It scales from four to 32 back-end ports, two to 112 physical tape drives and 10 tape libraries. It can handle 50,000 physical tape volumes.

The new members of the CS800 S2 family are the NAS Entry and VTL models, which join the NAS Basic and NAS Performance systems Fujitsu launched in April. The NAS Entry scales to 4 TB and is the smallest member of the family, while the VTL model supports 80 virtual tape drives and 9,000 virtual cartridges. Like the other NAS configurations, the NAS Entry also supports Symantec OpenStorage (OST), CIFS and NFS connectivity.

Fujitsu licenses Quantum Corp.'s data deduplication software for its backup devices. "Data deduplication is a must-have when you do disk storage these days," Fujitsu American storage product marketing manager Steve Crawford said.

Nick Sundby, consulting director for IDC's European Storage Group, said the enterprise system is geared more toward organizations that will likely keep tape for archiving and backup, while the CS800 platform is for those who want to replace tape.

"The CS800 is more about tape replacement since it can do NAS, dedupe and VTL," he said. "You can consolidate tape in the high end product. That really is designed for companies that have a long-term commitment to tape rather than a replacement."

NetApp Inc., which discontinued product development on its NearStore VTL early this year, last week said it would resell the CS800 S2 in Europe. Sundby said he could see NetApp expand that deal to other parts of the world, and add the High End product as well if the CS800 sells well. NetApp was outbid by EMC Corp. last year when it tried to acquire data deduplication backup market leader Data Domain.

FUJITSU

PUT YOUR DATA ON A DIET
ETERNUS CS800 WITH DEDUPLICATION

ETERNUS® CS800 DATA PROTECTION APPLIANCE FEATURING DEDUPLICATION

- **Intel® Xeon® Processor 5500 series**
- Turbocharges data reduction for your storage
- Reduces your capacity requirements for backup by 90%
- Lowers your expenses for data protection

More information is available at **solutions.us.fujitsu.com/eternus**

intel Xeon® inside™

Powerful.
Intelligent.

How to choose an enterprise-ready virtual tape library

Robert L. Scheier

Definitions of an enterprise virtual tape library (VTL) environment vary, but vendors and analysts generally agree that an enterprise-class virtual tape library should back up at least tens of terabytes of data daily and be able to store hundreds of terabytes of data. These high-end virtual tape libraries also need to scale easily and support complex backup and recovery environments, such as those that include multiple remote offices or multiple VTLs in the storage fabric.

The ability to easily add disk capacity is important not only because organizations are backing up more and more data all the time, said Lauren Whitehouse, an analyst at Milford, MA-based Enterprise Strategy Group (ESG), but because companies often buy their first VTL to meet the needs of a specific department and then add to it as they see the value it provides.

Because of the number of production servers involved, failed backups are considered unacceptable in an enterprise virtual tape library environment, said Andrei Shishov, VP of backup platforms engineering at EMC Corp. Reliability is often provided through features such as clustered failover among virtual tape library nodes, as well as redundant components such as power supplies within individual nodes.

Performance is another basic requirement for an enterprise virtual tape library, so it can back up data quickly enough to fit within an organization's backup window and restore files quickly when needed, said Whitehouse. Exact definitions of what is sufficient performance vary, although Peter Eicher, director of product marketing at FalconStor Software Inc., estimated the minimum at 300 MBps to 400 MBps. In addition to improving reliability, clustering can boost performance by spreading the work of reading and writing data, and/or deduplicating it, among multiple VTL servers.

Because enterprise customers "potentially have not only the local data center, but maybe some remote offices to be concerned about," support for a variety of backup applications is

also a must-have, said Whitehouse. Then there's the need, added Shishov, for the virtual tape library software to work with the widest possible variety of backup software, servers and server components, such as host bus adapters and disk drives.

The desire to cut the purchase and energy costs of disk drives, and to reduce the bandwidth required to replicate data among various sites (such as for disaster recovery), have made data deduplication a must-have feature for enterprise virtual tape libraries. Some vendors are also introducing massive array of idle disks (MAID) or spin-down features that power up disks only when they read or write data.

Ease of management is key to minimizing the VTL's total cost of ownership, and includes everything from the clarity of the user interface to the ability to easily view and manage groups of VTLs from a single console. Specific features to consider are how easy it is to access all VTLs in a group through a single sign-on, and the ability to view a consolidated report for all the systems and make simultaneous changes in common configuration settings.

Security wasn't a key requirement cited by users or analysts (see How virtual tape libraries compare for a look detailed look at enterprise tape library requirements) because VTLs are generally used in data center environments, which are regarded as secure. Security becomes more important when data leaves the data center stored on tape, so most users said they perform encryption at the tape library and not the VTL. However, some vendors, such as FalconStor and Sepaton Inc., do offer encryption in their VTLs.

Depending on whether a virtual tape library is used along with or as a replacement for tape, users might need it to write directly to or read directly from a physical tape. While most VTLs can import and export tape, said ESG's Whitehouse, another important consideration is whether the backup software's catalog is updated to reflect any changes in the data stored on the virtual tape library so backup administrators can more easily track the location of the backed up data.

How virtual tape libraries compare

Enterprise requirement	Metrics	Enabling technologies	Major vendors
Scalability/Performance	Back up at least tens of terabyte daily; store hundreds of terabytes overall	Clustering; grid technologies; automatic load balancing; independent scaling of VTL; dedupe hardware	EMC Corp., FalconStor Software Inc., Fujitsu Siemens Computers, Hewlett-Packard Co., Sepaton Inc., Sun Microsystems Inc.'s StorageTek Division
Reliability	Continue operating, with minimal impact on performance, after hardware or software failure	Within VTLs, RAID and redundant hardware. Among VTLs, clustering, grid architectures with automatic failover.	Copan Systems Inc., EMC, FalconStor, Fujitsu Siemens Computers
Manageability	Ability to view and manage multiple VTLs as single units; automation of routine tasks	Centralized management consoles; dynamic resizing of VTL cartridges	Data Domain Inc., Diligent Technology Corp. (an IBM company), FalconStor, Overland Storage Inc., Sepaton

Source: Analyst interviews

Second-generation virtual tape libraries

Lauren Whitehouse

When virtual tape libraries (VTLs) emerged several years ago, they improved how data backup applications wrote data to disk. Tape-centric backup applications understood how to communicate with and manage tape devices, and then write data in tape format.

Virtual tape library vendors that launched those initial solutions were focused mostly on developing the emulations for popular Fibre Channel tape libraries; proving compatibility with leading backup apps; ensuring that the virtual tape library wasn't a single point of failure in the backup process; and figuring out how to get virtual tape copies easily offsite for disaster recovery (DR)purposes. Virtual tape library vendors masked any complexity by packaging their solution as a purpose-built appliance and partnering with disk vendors to deliver turnkey solutions.

As VTLs have moved from early adopters to mainstream users, the stakes have changed. Ease of deployment and use, improved backup performance and reduced physical tape media management issues are basic requirements today. Virtual tape library vendors now need to accommodate the more advanced needs of mainstream users and adapt to the changing climate.

Changes in the virtual tape library market

There are a number of factors contributing to the changing virtual tape library climate. First, the base technology is more mature. VTL vendors have been able to exercise their products in a variety of environments and collect valuable feedback from customers. Second, today's virtual tape library buyer is better educated not only about VTL technology but their own data protection requirements. This next generation of virtual tape library buyers (that's you) is more careful about vetting vendor claims. For example, what type of scalability is required to achieve advertised throughput estimates? And are data deduplication ratios realistic?

Another angle is the advancement of backup solutions. Most, if not all, of the traditional backup vendors have embraced disk-to-disk (D2D) backup and a new crop of backup

solutions built specifically for D2D backup has emerged. Today, backup vendors are keenly aware of user needs for addressing backup window problems with better backup performance, improving recovery time objectives with rapid recovery techniques, and reducing the capacity of data transferred and stored with capacity optimization technologies.

Because many of the newer backup products address these primary concerns, how have virtual tape library vendors responded? It turns out they're introducing a crop of new features to win over the pragmatic buyer (you again), which is great news for IT organizations.

Second-generation VTL requirements

The Enterprise Strategy Group surveyed virtual tape library users about the types of features they would like to see in their current virtual tape library solution. Data deduplication and improvements in virtual tape library management topped the list, followed closely by improved scalability, and better recoverability and performance. Some users, struggling with methods for getting data offsite for DR purposes, prioritized disaster recovery features and support for physical tape on the VTL back end. Lastly, support for protocols—such as FICON and ESCON for mainframe support, and Ethernet for iSCSI support—rounded out the list.

These users are concerned with overall functionality. VTL vendors are beginning to factor this into their development roadmaps. For example, implementing replication between VTLs has become a standard feature for most solutions. Virtual tape library buyers are interested in how that capability maps to current processes and existing environments. Can the replication occur to more than one location? Is it possible to encrypt the data while in flight? Is scheduling (at off-peak times) or bandwidth throttling available? What techniques are offered to reduce the capacity of data transferred?

Here are some of the second-generation features virtual tape library users are interested in. Capacity and performance scalability. The ability to scale storage capacity and performance are high-priority requirements that go hand in hand. Some early VTL solutions had design bottlenecks where the VTL architectures didn't fully take into account how backup applications work or what performance limits they would hit. Often, the existing backup

processes or jobs would have to change just to take the pressure off virtual tape library solutions. Second-generation VTLs must have the ability to seamlessly increase throughput performance as backup window pressure and recovery objectives dictate. The same holds true for capacity limits, and how easy it is to deploy additional disk capacity in the environment.

Capacity optimization. As organizations back up more workloads to disk and retain data on disk for longer periods of time, capacity optimization capabilities such as compression and data deduplication become more critical. Second-generation virtual tape libraries will need to have these features packaged in a way that addresses functional requirements. That includes deduplicating data across multiple virtual tape library heads, turning deduplication off depending on the workload, and deduplicating data in real-time (inline processing) or batch mode (post-processing) to address performance concerns.

Central management. Organizations that have implemented multiple VTLs (due to scale limitations or by design) created an unexpected management issue: virtual tape library sprawl. Central management of policies and a consolidated view of multiple VTLs can alleviate the management burden, and are important features in the next phase of VTL products.

Disaster recovery. Another burden of adopting a D2D strategy is safeguarding data in the event of a local outage or disaster. Best practices call for a copy of backup media to be stored offsite. Those first VTL vendors either implemented a "tape export" command to generate a physical duplicate of a virtual tape or directly managed the physical tape library to create duplicate media outside of the backup window. Neither approach satisfied the requirement that the backup app maintain full knowledge and control of duplicates. The alternative was to initiate a virtual-to-physical tape copy through the backup app, which meant the backup application was burdened with additional processing and the data was dragged over the network unnecessarily. Second-generation VTLs are leveraging more efficient methods for creating physical media for offsite storage. Taking advantage of backup vendors' APIs, such as Symantec Veritas NetBackup's OpenStorage programming interface, allows VTLs to initiate duplication to tape or a secondary location, while ensuring that the media catalog is synchronized. DR preparedness is also being addressed via local-to-remote virtual tape library-to-virtual tape library replication. Second-generation VTLs

take this basic feature a step further by allowing users the flexibility to decide which workloads require site-to-site replication; supporting multiple replication topologies, such as 1:1, and many-to-one ratios; as well as supporting bi-directional replication.

iSCSI support. Improvements in iSCSI performance, wider deployment of 10Gb Ethernet networks and increased support for iSCSI from storage vendors may drive more demand for Ethernet interface support. Users who want to leverage an IP SAN instead of, or in addition to, FC will look for VTLs that support iSCSI connectivity.

We know that organizations of all sizes are under pressure to have secondary data available and quickly accessible in downtime situations. Not doing so can be costly—if not devastating—to organizations. Virtual tape libraries have emerged as a means to this end for a segment of the market. VTL vendors are responding to these challenges as evidenced by the excitement around their incorporation of data deduplication features. For users, it means scouting out vendors that are improving on these capabilities so you can be prepared to take advantage of the benefits being offered by the next generation of virtual tape library products.

Resources from Fujitsu America, Inc.

Best Practice Guide: RMAN From Oracle with the ETERNUS CS800

Disaster Recovery with ETERNUS DX – The Data Safe – and VMware SRM

Improving Backup Effectiveness and Cost-Efficiency with Deduplication

About Fujitsu America, Inc.

The Fujitsu Interstage Business Process Management suite helps companies Sense and Respond to fluctuating market and business dynamics through continuous process visualization and optimization. Organizations can quickly visualize the "as-is" process as it "really-is" with our innovative Automated Process Discovery (APD) offering and then iteratively optimize these processes using Interstage BPM on premise or via the Cloud. Interstage Business Process Manager brings together functionality that has been developed based on customer feedback over the past 18 years and has been deployed by over 2500 enterprise customers worldwide. It provides rich capabilities that bring business and IT professionals together to dynamically discover, design, simulate, automate, analyze, and optimize business processes.