

White Paper

# SEPATON S2100: ContentAware Architecture

There is no end in sight to the exponential growth of data that IT managers have to manage, store and protect. This problem is compounded by corporate business continuity programs, green initiatives, and increasingly stringent government regulations that require companies to have both long-term data retention, data encryption, and advanced disaster protection. As the volume of data and the complexity of data management continue to grow, companies face a variety of data protection challenges. This paper describes the ways that the ContentAware™ architecture of the SEPATON® S2100® virtual tape library (VTL) is helping enterprises meet these challenges while reducing risk and saving money.

## Table of Contents

---

Today's Data Protection Needs ..... 1

Best-in-Class VTL Technology ..... 2

SEPATON S2100-ES2 Virtual Tape Library ..... 3

Designed for Performance..... 4

High Availability Protection..... 5

Modular Scalability of Performance and Capacity ..... 6

Plug and Play Simplicity ..... 8

Software Applications for Advanced Functionality ..... 10

Addressing Today's Data Protection Needs..... 12

## Today's Data Protection Needs

---

To stay competitive in today's information economy, businesses need instantaneous, worldwide access to data generated by a wide variety of computer systems. As a result, there is no end in sight for the exponential growth of data that IT managers have to manage, store and protect. This problem is compounded by corporate business continuity programs, green initiatives, and increasingly stringent government regulations (SEC, HIPAA, SOX, etc.) that require companies to have long-term data retention, data encryption, and advanced disaster protection. As the volume of data and the complexity of data management continue to grow, companies face a variety of challenges, including:

- Backing up increasing volumes of data within ever-shortening backup windows
- Meeting business continuity objectives and regulatory requirements by ensuring data availability through failures and site-wide disasters
- Reducing floor space and power consumption to meet green initiatives
- Cutting administration cost and streamlining data center complexity through virtualization and consolidation
- Adding capacity and performance of backup/restore technology to meet current and future needs while staying within budget
- Eliminating labor-intensive, repetitive media management tasks
- Getting the maximum return on investments in data protection technology
- Managing data in mixed environments (i.e., mainframe and open systems)
- Protecting remote data in decentralized organizations

The options available to meet these challenges were limited. Physical tape libraries are labor-intensive and prone to backup failures. Restoring data from tape is also slow. While disk-to-disk technologies solve some of the challenges posed by tape, they have performance issues, increased capacity requirements, and ongoing management complexity.

As a result, virtual tape library (VTL) technologies are becoming the industry standard. According to an Enterprise Strategy Group report<sup>1</sup>, "Not only are users replacing tape libraries with new disk-based solutions, they are doing so on a significant scale: Sixty percent of users surveyed by ESG believe that more than 40% of their current tape-based capacity will reside on disk in just three years." The report also states, "Virtual tape library (VTL) is the most widely deployed tape replacement technology today."

## Best-in-Class VTL Technology

---

SEPATON has designed a best-in-class VTL technology that combines the benefits of both physical tape and disk-to-disk technology with the SEPATON ContentAware architecture to deliver the industry's fastest, most scalable, and most automated backup solution. The result is an appliance that reduces the cost and risk associated with storing, protecting and recovering data. This system was designed from the ground up to be highly efficient for backing up and restoring both large volumes of data and individual files. The SEPATON technology delivers:

- Industry-leading performance that cuts both backup and restore times dramatically
- Simple, modular scalability of both performance and capacity that saves time and money by allowing you to purchase only what you need, when you need it
- High availability, high reliability design that nearly eliminates downtime and meets business continuity objectives
- Automated management functions that eliminate the complexity of implementation, configuration and ongoing management tasks
- Complete emulation of leading physical tape libraries and tape formats as well as integration with leading backup software applications
- Support for mainframe, open systems computing, and mixed environments
- Software applications enable simple addition of new features and functionality including advanced data deduplication and remote replication

This white paper will describe the SEPATON technology in detail and highlight the powerful features that make it unique in the industry and enable it to deliver the highest level of business value.

### Financial Services Firm Cuts Backup and Restore Time

IT staff at a global financial service firm did not have time to backup, restore and perform off-site tape reclamation with their tape systems. Two tape drives running 95% of the day for reclamation could not reclaim off-site tapes efficiently. They used a SEPATON S2100-ES2 VTL to achieve dramatic results:

- Cut daily backup time in half
- Reduced restore time for 25 GB of data from one week to just 30 minutes
- Reduced off-site tapes in half

The company is moving to a high availability, multi-site data protection infrastructure with duplicate backup and restore systems including a second SEPATON S2100 for disaster recovery.

## SEPATON S2100-ES2 Virtual Tape Library

To users and backup applications, a SEPATON VTL looks and functions like a physical tape library—complete with virtual tape cartridges and virtual tape drives. As a result, it installs in minutes and integrates seamlessly with all leading backup software. The unique SEPATON architecture delivers industry-leading backup and restore times while reducing the complexity of storage management through advanced automation. (See Figure 1.) Unlike a physical tape library, SEPATON delivers Fibre Channel wire speed data transfer rates, linear scalability, advanced data compression, and rock solid reliability.

### Mainframe Compatibility

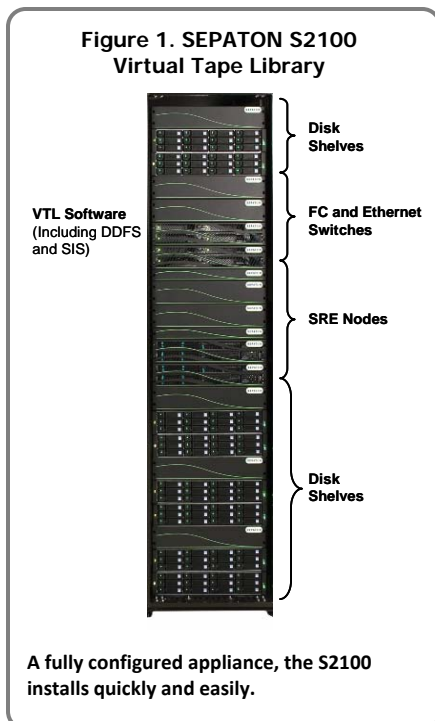
The S2100-ES2 is fully compatible with mainframes, open systems and mixed computing environments to enable you to protect your entire enterprise and to consolidate backups from a variety of server environments onto a single VTL for improved efficiency and reduced administrative complexity.

### Optimized to Backup and Restore both Large and Small Volumes

Until now, few technologies could backup *and* restore data with equal performance. While physical tape systems handle large volume backups efficiently, they have to read each tape linearly to find specific data for restores. Restoring individual files from physical tape can be particularly time consuming.

Although SATA disks perform well when executing streaming reads or writes, their performance throughput drops drastically on random seeks. The off-the-shelf file systems that are used by many disk-based technologies support general-purpose computing applications that transfer small amounts of data. They cannot handle large data I/O efficiently and are prone to slow performance caused by extensive seek activity. These file systems are designed to work well with single-application data streams, but cannot support the multiple-I/O environment needed by today's highly scalable solutions.

The unique SEPATON dynamic disk file system (DDFS) is optimized to store and restore both large and small volumes of data efficiently by getting the most performance from SATA disk technology. The SEPATON DDFS allows large I/O streams to execute efficiently by sustaining maximum throughputs and by dynamically load balancing I/O streams across all available disks without requiring tuning or defragmentation.



## Designed for Performance

---

SEPATON S2100 VTLs were designed to deliver the performance enterprises need to protect petabytes of data while staying within backup windows and restoring files fast enough to meet service level agreements. SEPATON VTLs use 4 Gb Fibre Channel technology for an aggregate maximum performance of 9600 MB/sec. It is designed to maintain this industry-leading speed during backups, restores, and deduplication processes. Four key innovations enable SEPATON VTLs to deliver this performance:

- **SEPATON I/O subsystem (SiS).** Other applications handle I/O in “user space,” an inefficient mechanism for high-volume data applications. The SiS handles I/O in “kernel space” allowing SEPATON appliances to achieve unmatched four-Gb wire-speed performance.
- **The DDFS** handles large volumes of data without fragmentation. It stores data efficiently as large 32 MB extents<sup>2</sup> instead of sector-sized (512 byte) or cluster-sized (646 byte) blocks used by other vendors. This innovative file system uses built-in heuristics that monitor the performance of disk arrays and associated controller loading to automatically choose the disk array that should receive the next I/O in the queue. It manages loading intelligently to deliver consistent maximum throughputs.
- **SEPATON Scalable Replication Engine (SRE®)** processing nodes. Each SRE node is a processing element that provides two Fibre Channel communication paths into the backup environment as well as high-end processing and memory subsystems. Each SRE hosts the SEPATON emulation application and interfaces with media servers and backup hosts. It enables you to add processing power by simply adding SRE nodes.
- **ContentAware engine** enables the VTL software to capture detailed information about every byte of stored data (e.g. type of backup, mount points, volumes, directory structure, file types, etc.) from the backup stream. This information enables SEPATON's next-generation DeltaStor® software to deliver the most efficient data deduplication technology in the industry.

## High Availability Protection

---

SEPATON VTL hardware is built for reliability using SATA RAID 6-based storage. SEPATON VTL eliminates the vulnerability of tape libraries, such as mechanical tape drive failures and media management issues. In addition, SEPATON's management console is equipped with an advanced predictive monitoring feature that identifies potential issues in high-risk components before they become failures. High availability features include the following:

- RAID 6 protection against double disk faults.
- Enterprise-class high availability disk array system with built-in redundancy for all components and paths allowing any component to fail, including the controllers, without loss of availability or data access.
- Hot swappable power supplies, fans, and main chassis.
- Automatic back-end retries on failure.
- Redundant connections to disk via redundant Fibre Channel switches.
- Redundant Ethernet switches for command and control.
- SEPATON offers an optional active failover capability that provides N+1 failover protection.

## Modular Scalability of Performance and Capacity

The unique grid design of the S2100 VTL lets you add performance simply by adding SRE nodes and add capacity simply by adding disk arrays. You can start with a one- or two-node system and increase performance as you need it. SEPATON's built-in clustering software integrates additional SRE nodes into the appliance and presents them to users and to the backup software as a larger VTL with more connection points. Each S2100-ES2 can be configured with up to sixteen SRE nodes to reach performance of 34.5 TB/hour. The S2100-ES2 is currently configured for sixteen SRE processing nodes. However, the system's built-in clustering architecture can integrate up to 32 processing nodes into a single appliance. It also enables failover from one node to another if needed. This advanced clustering capability allows SEPATON appliances to deliver maximum data protection without sacrificing performance.

### Simple Capacity Scaling

The S2100-ES2 comes with advanced hardware compression that can double or triple its useable capacity and by reducing the number of spinning disks and other mechanical components needed, it can increase reliability.

SEPATON S2100-ES2 VTLs scale capacity from 7 TB to 1.2 PB (more than 60 PB with DeltaStor data deduplication software and compression enabled<sup>3</sup>) in a single appliance. Virtual tape cartridges map capacity to the backup environment. To the backup software, they are identical to physical tape cartridges. However, virtual tape cartridges have no capacity limitation (unlike their real world counterparts). As a result, they eliminate the time-consuming process of mapping capacity to available physical storage.

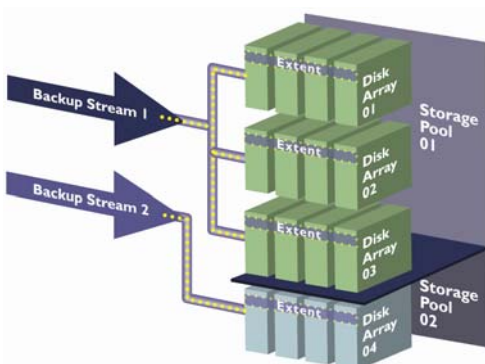
For maximum flexibility, the SEPATON architecture supports 3072 virtual devices. Each can be configured as a virtual tape drive or virtual tape library. Approximately 5.3 million cartridges can be configured per application.

### Storage Pools for Unprecedented Flexibility and Control

A storage pool is a group of disk arrays and controllers that provide capacity, performance, and data availability as one collective unit. (See figure 2.)

You can assign all disks to one pool or to several pools of different sizes (containing different numbers of arrays). For example, if you need to guarantee data segregation, you can choose a set of disk arrays and group

**Figure 2. Storage Pools**



**Extents are distributed evenly to all arrays in a given pool to eliminate fragmentation, maximize performance, and enhance security.**

them into a storage pool. You can then choose which virtual cartridges to place in that storage pool. The SEPATON system automatically handles the movement of data within the pools as well as the associated pool-level management tasks, including capacity monitoring and allocation, load balancing, and performance tuning.

### Thin Provisioning

The SEPATON architecture supports two additional capabilities that dramatically simplify scaling: thin provisioning and capacity-on-demand. Thin provisioning allows you to create as many virtual cartridges as you want without having commensurate amount of physical storage. The capacity of a tape or virtual tape that is actually used is often much lower than its maximum capacity. SEPATON software monitors the disk capacity that is actually used and when it reaches a user-defined threshold, the software automatically sends alerts to the administrator indicating the need for additional storage. Users can add capacity at any time. The SEPATON management software automatically discovers, formats, and adds the new storage to the environment without end-user involvement.

For example, you can create a single two-TB cartridge in a VTL without mapping the virtual cartridge to physical disk storage. SEPATON's DDFS allocates storage to the cartridge only when data is actually written by the backup application. If you write only 100 GB to your two-TB cartridge, only 100 GB of disk storage is allocated (not two TB). This real-time allocation function gives you tremendous flexibility by allowing you to create virtual cartridge schemas without requiring the physical storage to be in place. Other value-added applications that use storage are free to use capacity from the same disk pool without requiring users to make any storage allocation decisions.

### Capacity-on-Demand

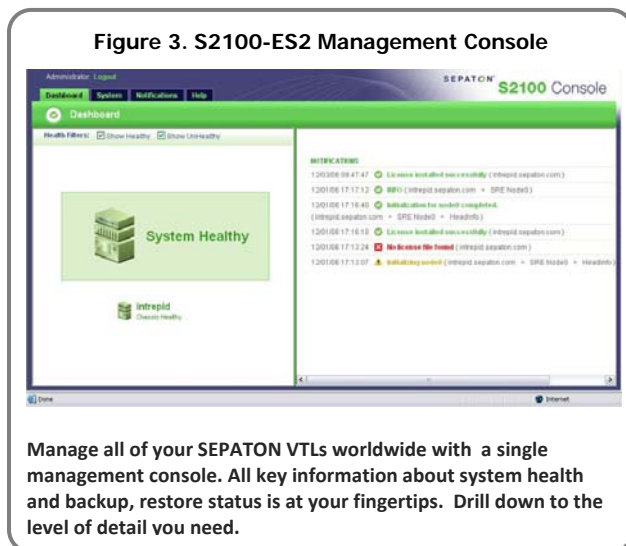
Capacity-on-demand allows you to have additional storage pre-installed but not enabled until it is needed. SEPATON software monitors the amount of licensed storage that is used. When more storage is required, you simply install a new license key to unlock the latent storage.

## Plug-and-Play Simplicity

SEPATON products simplify and streamline all aspects of data protection from configuration and installation of new a new VTL to ongoing administration and maintenance. SEPATON VTLs are turnkey appliances that include: one or more SREs that host the SEPATON VTL software, one or more RAID 6-protected disk shelves, and an integrated web-based management console. Unlike other VTL technologies that require you to procure and configure a variety of hardware and software components, you can have a SEPATON system running in minutes.

### Automated Storage Management

The nature of a backup environment is far more dynamic than a primary storage environment, where disk storage and its protection mechanism are generally allocated and then left intact for an extended period of time. As a result, any cost savings achieved with a disk-to-disk backup system can be quickly offset by the labor cost of disk management. SEPATON's DDFS automates all aspects of disk storage, including: performance management, space allocation, sparing, capacity growth, classes of service, and advanced features.



SEPATON's web-based management console gives you a single tool to monitor and manage all of your VTLs worldwide (see figure 3). The management console has a wide range of functions for active monitoring and management as well as automatic tracking of all physical components of your VTLs, including servers, switches, storage, etc.

An intuitive, wizard-driven graphical user interface guides you through all common management tasks, such as setting up a VTL, adding storage, or adding SRE nodes. You never have to worry about RAID levels, stripe size management, cache settings, or other complexities of storage management. The software automatically brings added storage on-line, configures the devices to their optimal settings and incorporates it into the VTL cluster.

SEPATON software monitors all aspects of system health (fan RPM, temperature, bad disk blocks etc.) and provides a summary view that includes troubleshooting and device diagrams. You can drill down to any aspect of the VTL hardware status including disk

drive manufacturers, firmware levels and cooling status. If status data exceeds predefined parameters such as hard errors, soft errors, warnings, or capacity thresholds, the software's email home function automatically logs detailed system status information and alerts SEPATON support and your designated administrator(s). The system also supports SNMP trapping to integrate with third-party system management software.

### Non-Disruptive Design Integrates Seamlessly

SEPATON protects customers' investments in existing infrastructure by integrating seamlessly with all leading backup/restore applications, including:

- Atempo Time Navigator™
- Backup Exec™ by Symantec
- BakBone® NetVault™
- CA BrightStor® ARCserve®
- CommVault® Galaxy
- Computer Associates BrightStor®
- EMC NetWorker™
- HP Data Protector
- IBM® Tivoli® Storage Manager
- Veritas® NetBackup™ by Symantec
- Oracle® Secure Backup
- SyncSort®

### Emulation Protects Investment

SEPATON software emulates a variety of tape library robotics and tape transports, allowing you to use SEPATON's solution without changing your driver software, backup application, backup policies or procedures. Data is stored and read back in the native format of your backup application. SEPATON's emulation functionality has undergone extensive interoperability testing and certification with every major ISV and major operating system. SEPATON also supports all leading tape libraries, including StorageTek®, Quantum®, ADIC®, HP®, and IBM®.

### Maximum Data Security

The fundamental design of SEPATON appliances make them intrinsically more secure than both physical tape and disk-based systems. Unlike a tape library, VTLs have no physical cartridge that can be stolen, lost or damaged. Because data stored in a SEPATON VTL is not mounted as it is in a disk-based system, it cannot be accessed by end-users, viruses, hackers, and other security threats.

SEPATON adds a further level of security by "striping" stored data across all available disks in a storage pool so that data sets cannot be reconstructed from individual disk drives. SEPATON also works with leading third-party encryption systems, such as Decru® DataFort™ to encrypt data that is moved to physical tape for off-site vaulting or that is replicated across a WAN.

## Software Applications for Advanced Functionality

---

SEPATON's unique architecture allows you to extend the functionality of the S2100-ES2 appliance simply by adding our breakthrough data protection applications. These applications provide powerful data protection capabilities that deliver breakthrough time-savings and efficiency.

### DeltaStor Deduplication Software

DeltaStor software is next-generation data deduplication technology that delivers the benefits of disk storage at a cost-per-GB-stored that is comparable to physical tape. Adding DeltaStor software is as easy as checking a box in the S2100-ES2 management console. As deduplication reduces data volume, capacity is automatically made available and managed through its built-in self-management functions.

DeltaStor delivers an industry-leading deduplication ratio of up to 50:1 (with compression) for a typical combination of business application data. DeltaStor software also increases the accessibility of backup data by letting you store more on disk for longer periods of time. DeltaStor software performs data deduplication outside of the primary data path, enabling it to maintain wire speed performance. Next-generation design features include built-in intelligence about file content and the backup data relationships of leading backup applications to deliver unparalleled speed, simplicity, scalability, and data integrity. Unlike other technologies that perform integrity checking by comparing file addresses or "hash" data, DeltaStor software compares full files and data at the byte level for maximum protection.

By changing the economics of data storage, DeltaStor software allows you to handle exponential data growth and significantly lengthen your online retention times for faster restores and simple scalability.

### Site<sup>2</sup> Remote Replication Software

Powerful Site<sup>2</sup>™ remote replication software lets you replicate virtual cartridges to remote disaster recovery locations over your existing WAN and manage them all from one location with SEPATON's management console. Site<sup>2</sup> eliminates the risks of physical tape loss, theft, and failure while delivering fast, secure data transfer and storage. It also leverages the media management capabilities of existing backup software applications to track both physical and virtual tapes for long-term archiving, compliance, and disaster recovery. The robust management system lets you replicate data from one VTL to another based on policies (e.g. by schedule) or by amount of new data. You can replicate data

from one VTL to multiple VTLs at the same time or centralize your data protection by copying all of your VTLs to a single backup VTL system.

As an added advantage, Site<sup>2</sup> lets you save money and reduce the complexity of managing data on multiple data protection technologies. You can back up data stored on EMC, HDS, NetApp storage and other systems, and use Site<sup>2</sup> to perform a single, consolidated remote copying of the data in these systems. The Site<sup>2</sup> bandwidth optimization feature allows up to 64 concurrent data streams between the VTL systems. A bandwidth-limiting feature lets you control bandwidth usage on your network.

Site<sup>2</sup> was designed to maintain data integrity and security. The system automatically checks data integrity to ensure virtual cartridges are copied accurately. It also performs an advanced authentication of connections between systems before a transfer starts to ensure data security. Only authenticated VTL systems can transfer and accept data. For optimal availability, Site<sup>2</sup> transport stack does automatic retries on network error conditions, and the automatic checkpoint feature lets you restart a job from a checkpoint manually or automatically after a network outage.

## Addressing Today's Data Protection Needs

---

SEPATON is addressing today's increasingly complex data protection and management issues by offering an innovative VTL appliance that delivers industry-leading performance and scalability in a high availability environment.

From its high-performance data handling to its fully automated data management functionality, SEPATON was designed to be easy to implement, manage, and maintain (see table 1). SEPATON appliances deliver the following benefits:

- Modular scalability lets you add capacity as you need it, while staying within budget
- Industry leading performance cuts backup and recovery times to seconds
- High availability and remote-site replication help you meet business continuity objectives
- Highly automated management console eliminates labor-intensive, repetitive data management tasks
- Extensible design allows you to add functionality as you need it
- Advanced emulation capability helps you get the maximum return on investments in data protection technology

| <b>Table 1: Side-by-Side Comparison of Data Protection Technologies</b> |  |  |   |                                  |
|---|--|--|---|----------------------------------|
|   | <b>SEPATON S2100 VTL</b>   | <b>VTL Software</b>  | <b>D2D</b>                                | <b>Tape Backup</b>               |
| <b>Performance</b>  | Maximum: 9,600 MB/sec  | Maximum: 90 MB/sec per stream  | Maximum: 40-50 MB/sec                     | Maximum: 80 MB/sec               |
| <b>High Availability</b>  | Grid design allows for scalability to 16 nodes (32 nodes design) with active node failover and no performance impact on failover | Limited, active cluster of two servers supported. Node failure reduces performance by half | None                                      | None                             |
| <b>Scalability of Performance and Capacity</b>                          | Fast, cost-effective, easy   | Costly, complex and typically limited  | Time-consuming, costly, complex           | Time-consuming, complex          |
| <b>Ease-of-Use</b>  | Easy, highly automated, web-enabled  | Complex, must manage disk and software separately  | Complex, manual                           | Complex, manual                  |
| <b>Security</b>   | Fully protected from end-users, hackers, or viruses. Works with industry-stand encryption  | Must manage security on multiple different devices   | Allows access to data through file system | Vulnerable to human error, theft |
| <b>Ease of Implementation</b>   | Non-disruptive appliance installs in minutes   | Complex, requires configuration of disks   | Disruptive, labor intensive               | NA                               |
| <b>Total Cost of Ownership</b>  | Low  | Medium   | High                                      | High                             |
| <b>Disaster Recovery Capability</b>                                     | High   | High   | Low                                       | Medium                           |

### Notes

<sup>1</sup>John McKnight and Peter Gerr, ESG Research Report: Tape Replacement Realities, April 2005

<sup>2</sup>Extents are 32 MB by default. Size can be changed as needed.

<sup>3</sup>Compression and deduplication results vary with data type and backup methodologies.