SEPATON

SEPATON

Changing the Rules of Data Protection

Scalable, Highly Automated Virtual Tape Library Technology Reduces the Cost of Storing, Managing and Recovering Data

January 2007

400 Nickerson Road Marlborough, MA 01752 • Main: 508.490.7900 • Fax: 508.490.7908 • Web: www.sepaton.com

Table of Contents

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 and increasingly stringent government regulations (SEC, HIPAA, SOX, etc.) 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 challenges, including:

- Managing continuously lengthening backup and recovery times
- Meeting business continuity objectives and regulatory requirements by ensuring data availability through failures and site-wide disasters
- 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

Until now, the options available to data managers to meet these challenges were limited. Although traditional tape-based technology is efficient for storing large volumes of data, it is highly labor-intensive and prone to failure. Restoring data from tape is also unacceptably slow. As a result, disk-based backup and restore technologies are becoming the industry standard. According to an Enterprise Strategy Group report[1], "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 Enterprise Strategy Group 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."

[1] John McKnight and Peter Gerr, ESG Research Report: Tape Replacement Realities, April 2005

Best-in-Class VTL Technology

While SATA-based technologies solve some of the challenges posed by tape, they have their own drawbacks, including performance issues, increased capacity requirements, and ongoing management complexity. Host-based virtual tape library (VTL) software relieves some of the management burden of disk-to-disk systems, but requires highly complex integration and cannot meet performance or scalability needs of today's enterprises.

SEPATON has designed a best-in-class VTL technology that combines the benefits of both physical tape and disk-to-disk technology with its own high performance I/O disk subsystem, unique Dynamic Disk File System (DFS), and innovative ContentAware™ architecture to deliver the industry's most advanced data protection solution. The result is a simple, modularly scalable, high performance appliance that reduces the cost of storing, protecting and recovering data. This system is designed to be highly efficient for backing up and restoring both large volumes of data and individual files.

SEPATON CUTS RESTORE TIME, IMPROVES TAPE RECLAMATION

A global financial services company had outgrown their physical tape library. There wasn't enough time in the day to do backups, restores and off-site tape reclamation. They ran two tape drives 95% of the day for reclamation and could not reclaim their off-site tapes efficiently (500 off-site tapes protected 175 on-site tapes). By implementing a SEPATON S2100(r)-ES2 VTL, the company achieved dramatic results:

- Cut daily backup in half (from seven hours to less than four)
- Reduced restore time for 25 GB of data from one week to just 30 minutes
- Reduced number of off-site tapes from 500 to 223

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.

The SEPATON technology delivers:

- Industry-leading performance that cuts both backup and restore times dramatically
- High availability, high reliability design that eliminates downtime and meets business continuity objectives
- Automated management functions that eliminate the complexity of implementation, configuration and ongoing allocation tasks
- 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
- Complete emulation of leading physical tape libraries and tape formats as well as integration with leading backup software applications
- Supports mainframe, open systems computing and mixed environments
- Advanced disaster protection through off-site replication functionality
- Extensible design that enables simple addition of new features and functionality

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

SEPATON S2100-ES2 Virtual Tape Library

The SEPATON S2100-ES2 VTL system is an intelligent, high performance data protection appliance that 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.

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. Unlike a physical tape library, SEPATON delivers Fibre Channel wire-speed data transfer rates, linear scalability, advanced data compression, and rock solid reliability.

Innovative Architecture

The SEPATON VTL solution was designed from the ground up to deliver maximum performance to backup and restore data in an easy-to-manage, high availability environment. Until now, few technologies could backup AND restore data efficiently. While physical tape handles large volume backups efficiently, its restore performance is poor. Physical tape has to read a full tape linearly to find specific data for restoration. 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 other 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 data protection solutions.

Optimized to Backup and Restore Both Large and Small Volumes

The SEPATON Dynamic DFS 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 Dynamic DFS 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. In addition, SEPATON appliances feature the SEPATON I/O subsystem (SiS), a comprehensive kernel-level data handling system. Other applications handle I/O in "user space," an inefficient mechanism for high-volume data applications. The SEPATON SiS handles I/O in "kernel space" allowing SEPATON appliances to achieve unmatched two- and four-Gigabit wire-speed performance.

SEPATON's unique data storage system also enhances performance. SEPATON software stores data efficiently as large 32 MB extents[2] instead of sector-sized (512 byte) or cluster-sized (646

[2] Extents are 32 MB by default. Size can be changed as needed.

byte) blocks used by other vendors. During a backup, the SEPATON Dynamic DFS distributes extents evenly to all available arrays, eliminating fragmentation and improving performance by distributing the work of handling data across all available spindles and RAID controllers.

The SEPATON Dynamic DFS, SiS, and SEPATON extent storage technology combine to deliver industry leading performance for both backup and restore. The SiS also provides an easily extensible environment into which SEPATON ContentAware application modules such as Site$^{2™}$ remote site replication software and DeltaStor™advanced data de-duplication software can be added for enhanced functionality

Another innovation is SEPATON's ContentAware architecture, which captures 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. SEPATON's next-generation DeltaStor software uses this information in conjunction with a unique pointer capability to provide the industry's most powerful data de-duplication technology.

Designed for Performance and Reliability

SEPATON dramatically reduces backup and restore windows to seconds with its unique Scalable Replication Engine (SRE®) Technology (see figure 1). 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. SEPATON's unique I/O processing software uses two- or four-Gbit Fibre Channel technology to make each path capable of at least 300 MB/sec (2 Gbit) or 600 MB/sec (4 Gbit). The S2100-ES2 is currently tested and approved for eight nodes. However, the system's built-in clustering architecture integrates 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.

The innovative SEPATON Dynamic DFS 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/Os in the queue. The software can manage loading intelligently and deliver consistent maximum throughputs.

SEPATON VTL hardware is built for reliability using SATA RAID-based storage with hot spares and high availability failover capability. 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.

Figure 1. SEPATON software runs on the SRE Nodes, where all processing intelligence resides.. Add SREs for performance or add disk arrays for capacity.

- Built-in sparing design includes multiple floating hot-spare drives.
- 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 Delivers Maximum Control of Performance and Capacity

SEPATON offers the most scalable solution in the industry in terms of both performance and capacity. SEPATON's unique SRE and Dynamic DFS technology allow simple scaling of both performance and capacity. Each SRE hosts the SEPATON emulation application and interfaces with media servers and backup hosts. Each SRE node supports up to two 4-Gbit Fibre Channel ports, providing up to 600 MB/second aggregate bandwidth. Each S2100-ES2 can be configured with up to eight SRE nodes to reach 17.2 TB/hour. You can start with a one- or two-node system and easily increase performance as you need it by adding SRE nodes (see figure 2). 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.

Simple Capacity Scaling

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

SEPATON appliances scale capacity from 7 TB to more than 2 PB (more than 50 PB with DeltaStor data de-duplication software and compression enabled) in a single appliance. SEPATON uses virtual tape cartridges to 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 192 virtual LUNs (logical unit numbers) per node (1536 LUNs in total). Each virtual LUN can be configured as a single virtual device, such as a virtual tape drive or virtual tape library. Approximately 5.3 million cartridges can be configured per application.

Storage Pools Enable Unprecedented Management Flexibility and Control

The SEPATON Dynamic DFS is built around the concept of storage pools (See figure 2). A storage pool is a group of disk arrays and controllers that provide capacity, performance, and data availability as one collective unit.

You can assign all disks to one pool or to several pools of different sizes (containing different numbers of arrays) or even different levels of availability. For example, a user who needs to

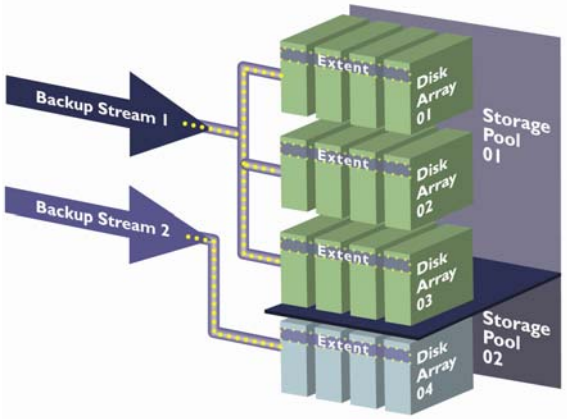

Figure 2. The Dynamic DFS distributes *extents* evenly to all arrays in a given pool to eliminate fragmentation, maximize performance, and enhance security.

guarantee data segregation can choose a set of disk arrays and group them into a storage pool. The user 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.

The SEPATON architecture supports two additional capabilities that dramatically simplify scalability: thin provisioning and capacity-on-demand. Thin provisioning allows you to create as many virtual cartridges as you want without needing 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. No end-user involvement is required.

For example, you can create a single two-TB cartridge in a VTL without mapping the virtual cartridge to physical disk storage. SEPATON's Dynamic DFS 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 for backup administration 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.

LARGE VOLUME TAPE LIBRARIES

As an added benefit, you can use SEPATON's emulation capabilities in conjunction with its clustering software to create extremely large and robust virtual tape environments. For example, you can create a single virtual library with 5.3 million cartridges and 1536 tape transports spread across eight SRE nodes. The backup application will detect a storage "silo" with sixteen Fibre Channel paths to the requisite tape transports. If the backup application supports multi-path to the robotics (to account for Fibre Channel path failures in the SAN), only SEPATON can reciprocate by providing redundant paths on the VTL appliance.

Plug-and-Play Simplicity

SEPATON products are designed to simplify and streamline all aspects of data protection from configuration and installation to ongoing administration and maintenance. SEPATON VTLs are turnkey appliances that include: one or more scalable replication engines that host the SEPATON VTL software, one or more RAID-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

SEPATON's Dynamic DFS automates all aspects of disk storage, including: performance management, space allocation, sparing, capacity growth, classes of service, and advanced features.

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 managing disk-specific items such as capacity provisioning or performance balancing.

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 also 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

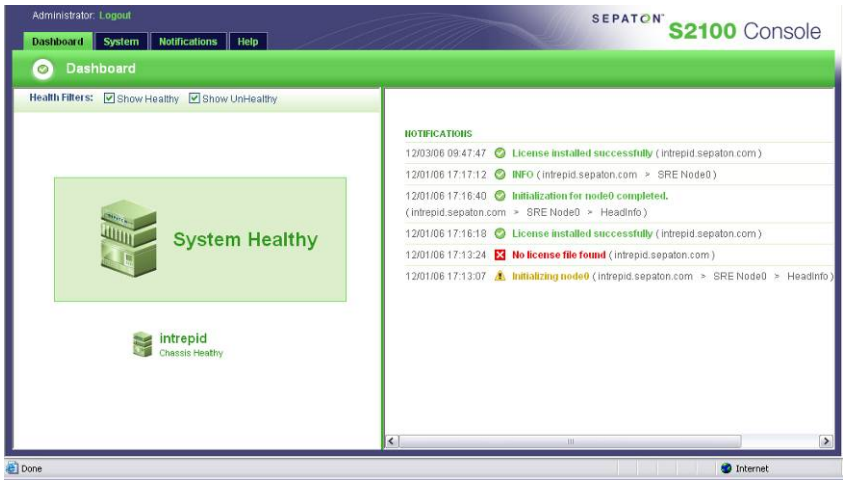

Figure 3. S2100-ES2 management console makes managing the VTL simple.

down to any aspect of the VTL's hardware status including disk drive manufacturers, firmware levels and cooling status. If status data on a component 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 administrators. The management system also supports SNMP trapping to integrate with leading 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®
- NetBackup™ by Symantec
- EMC NetWorker ™
- HP Data Protector
- IBM® Tivoli® Storage Manager
- Oracle® Secure Backup
- SyncSort®

Emulation Protects Investment

SEPATON software emulates the functionality of 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 physical 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 in such a way that data sets cannot be reconstructed from individual disk drives. SEPATON also works with leading third-party encryption systems, such as Decru® DataFort™ or NeoScale™ CryptoStor®, to encrypt data that is moved to physical tape for off-site vaulting or that is replicated across a WAN.

Mainframe Compatibility

The S2100-ES2 is fully compatible with both mainframes and open systems computing environments to enable you to protect your entire enterprise.

ContentAware Applications for Advanced Functionality

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

Site^2 Remote Replication Software

Powerful Site^2 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^2 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^2 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^2 to perform a single, consolidated remote copying of the data in these systems. The Site^2 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^2 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^2 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.

DeltaStor De-Duplication Software

DeltaStor software is next-generation data de-duplication 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 de-duplication reduces data volume, capacity is automatically made available and managed through its built-in self-management functions.

DeltaStor boasts an industry-leading de-duplication ratio of up to 50:1 (with compression) for a typical combination of business application data including email, files and database[3]. DeltaStor

[3]Compression and de-duplication results will vary with data type and backup methodologies.

software also increases the accessibility of backup data by letting you store more on disk for longer periods of time. DeltaStor software performs data de-duplication outside of the primary data path, enabling the S2100-ES2 appliance to conduct backups without loss of 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.

Database De-duplication
DeltaStor software is designed specifically to separate and eliminate storage of duplicate data in leading database and email technologies, including Oracle®, Microsoft® SQL and Microsoft® Exchange. DeltaStor software can help you control exponential email growth with de-duplication ratios of up to 225:1 for Microsoft Exchange environments.

Significant Cost Savings

The S2100-ES2 hardware compression feature is designed to enable maximum performance. When data compresses at the industry standard of 2:1, power, cooling and space saving costs can be reduced by half, and when data compresses at 3:1, costs can be reduced by up to two thirds. When DeltaStor is added, the system saves 62% in space requirements and 85% in power and cooling.

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

Table 1: Side-by-Side Comparison of Data Protection Technologies

	SEPATON VTL Appliance	VTL Software	D2D	Tape Backup
Performance	Maximum 9,600 MB/sec	Maximum 90 MB/sec per stream	Maximum 40-50 MB/sec	Maximum 80 MB/sec
High Availability	Clustered architecture allows for scalability to 8 nodes (tested) 32 nodes (design) with active node failover and no performance impact on failover	Limited, active cluster of two servers supported. Failure of a node reduces performance by half	None	None
Scalability of Performance and Capacity	Fast, cost-effective, easy	Costly, complex and typically limited	Time-consuming, costly, complex	Time-consuming, complex
Ease-of-Use	Easy, highly automated, web-enabled	Complex, must manage disk and software separately	Complex, manual	Complex, manual
Security	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
Ease of Implementation	Non-disruptive appliance installs in minutes	Complex, requires configuration of disks	Disruptive, labor intensive	NA
Total Cost of Ownership	Low	Medium	High	High
Disaster Recovery Capability	High	High	Low	Medium