

BY

W. CURTIS PRESTON, VP DATA PROTECTION, GLASSHOUSE TECHNOLOGIES

WITH

JOHN HAIGHT, PRINCIPAL CONSULTANT, GLASSHOUSE TECHNOLOGIES

KEVIN SUTTLE, SENIOR CONSULTANT, GLASSHOUSE TECHNOLOGIES

It's time for some serious changes in data protection -- and they're here. The ways that we have been backing up and archiving data for the last few decades no longer meet today's requirements. Thankfully, vendors such as Hitachi Data Systems and CommVault have seen these changes coming, and have taken steps to address them.

### **TODAY'S DATA PROTECTION CHALLENGES**

Organizations today are experiencing many of the following data protection challenges:

- Meeting recovery time objectives (RTOs) and recovery point objectives (RPOs)
- Meeting backup window and application impact requirements
- Satisfying industry and regulatory requirements for retention and recoverability
- Managing increasingly complex data protection environments built upon multiple disparate tools

RTO requirements determine how quickly the organization must restore data and RPO requirements determine how old the restored data can be. Although many customers focus on their backup window, it's actually the RTO and RPO that should really drive the design. How fast you can back up is nowhere near as important as how fast you can recover. A number of challenges are making it more difficult for organizations to meet stringent RTO and RPO requirements, starting with rapidly rising data volumes and shrinking or nonexistent backup windows. The next challenge is the mismatch between the speed of tape drives and the other components in the environment. Believe it or not, tape drives are actually too fast for the job. This means that you can't just solve an RTO challenge by buying more tape drives.

It's also difficult to meet RPO requirements when backups cannot be completed within specified backup windows. It's difficult to limit your loss to only four hour's worth of data when your backup takes twelve hours to complete.

Backup windows are important as well. In addition, running backups should not significantly impact the application they are backing up. If they are impacting it, business units may require that the backup window be shortened even farther. For all of the reasons stated above, backups should finish quickly and have as little impact on the application as possible. Accomplishing that with typical data protection systems, however, can be quite a challenge.

Another challenge with today's data protection systems is that industry and government regulations have been tightened or created anew in an effort to drive sound business practices. Many of these policies have direct ties to data protection and to an organization's ability to retain and recover certain types of data within given timeframes, impacting RTO, RPO, and retention requirements.

A typical reaction to the challenges mentioned above is to purchase multiple point solutions that each address a single challenge; one product for traditional backup and recovery, another for point-in-time recovery, and yet another for archiving. This results in a more complex data protection environment that requires additional resources to manage and maintain.

The complex data protection environment can also increase the risk of a human error when recovering data, as individuals attempt to understand which solution should be used to recover the data and maintain the operational support of the various point solutions.

Meeting each of the challenges above while still maintaining a financial competitive advantage is yet another challenge. An organization's ability to recover data quickly and accurately can set it apart from the competition, but only if it can do so at a lower cost. The value of the data being protected must be appropriately matched to the cost of the product(s) protecting the data. This cost is not just product cost, but also the cost of managing the data protection solution(s). As a rule of thumb, the more complex the solution, the higher the management and operational cost. Remember also that all information in a company does not have the same value, and the value of any given piece of information changes over time. Therefore, an ideal data protection solution should be able to provide different levels of data protection to different pieces of information over time.

## **DATA PROTECTION TECHNOLOGY**

Not so long ago, data protection was equated with "backup and recovery." Now data protection encompasses not only traditional backup and recovery, but also technologies such as archiving, snapshots, replication, and security.

## **BACKUP AND RECOVERY**

Traditionally, organizations of all sizes have used tape for backups as the primary method of data protection. No doubt, the unbridled growth of data has contributed to today's data protection challenges. Data growth results in more data to back up, more data to restore, and more tapes to manage. As data grows, more resources are required to facilitate processing of the data. Tape drives, processing power, and network bandwidth are all in greater demand. This leads to the contention of these resources in meeting the data protection requirements, which further contributes to RTO and RPO challenges.

One interesting development, made possible by the wide availability of economically popular ATA disk drives, is the possibility to utilize disk storage systems to augment the existing tape-based infrastructure. This combination of disk-based and tape-based data protection offers different options that may mitigate

many of the data protection challenges experienced by an organization. As a result, many organizations are turning to disk-to-disk backup (D2D) and disk-to-disk-to-tape backup (D2D2T), which can offer improved throughput and protection.

Disk-based backups make use of disk as either a permanent destination for backup data (D2D) or as a temporary destination, also called staging, prior to sending the backup data to tape (D2D2T). These solutions are typically implemented in either a disk-as-disk architecture or a disk-as-tape (Virtual Tape Library, or VTL) architecture. With a disk-as-disk architecture, backup and media servers utilize standard filesystems on SAN, NAS, or direct-attached storage where the disk behaves as disk. With a disk-as-tape (VTL) architecture, disks are placed behind servers running software that allows the disk system to pretend it is a tape library with tape drives. While disk-as-disk solutions may be cheaper than VTL solutions, the way that most backup software products use them causes all of the challenges that exist with managing and monitoring storage today. VTLs can offer better performance than disk-as-disk solutions, but they add another component to the infrastructure, which needs to be managed and monitored, thus resulting in additional complexity. What's desired is a solution at the cost level of disk-as-disk and the performance level and management cost of a VTL.

## **ARCHIVING**

Archiving involves the long-term storage of data for the retrieval of its logical components, such as all the files or emails pertaining to a particular person or project. Today this data protection method is usually associated with some form of industry or government regulatory compliance. In some cases the act of archiving data causes the source data to be deleted after an archive copy is created, thus freeing up space on the original storage device.

## **SNAPSHOTS & REPLICATION**

Snapshots use block-level data change capture technology to create local point-in-time copies of disk volumes or filesystems at predetermined intervals. Replication can be used to create remote copies of these snapshots as well. In the event of data loss or data corruption, the point-in-time copy can be accessed quickly to minimize the downtime associated with such events. Replication can be host-based or storage system-based. Host-based replication requires special software that runs as a background process on the host that owns the file system being protected. Storage system-based replication is performed directly on the storage system and requires no additional host processing.

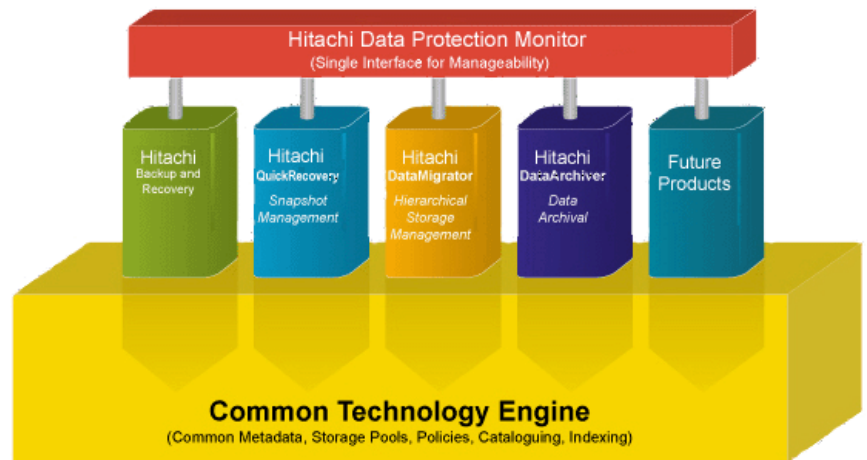
## SECURITY

In addition to making sure data is backed up and archived, today's environments must ensure that it is also protected from hackers and thieves. Several state and future federal laws require notification of customers whose data you've lost. In addition, the potential loss of intellectual property to a competitor cannot be overlooked, either. Today's data protection solutions must keep this in mind.

Currently, each of these data protection methods is typically facilitated by multiple disparate tools. The intent of this white paper is to highlight how the Hitachi Data Protection Suite, powered by CommVault®, and Hitachi storage systems can facilitate various data protection strategies in a single, integrated package, while meeting today's data protection challenges.

## A UNIQUE APPROACH TO TODAY'S DATA PROTECTION CHALLENGES

The Hitachi Data Protection Suite, powered by CommVault®, is a unified platform comprising modules for data backup and recovery, migration, archiving, snapshots, and replication – all through a single management console and database. At the foundation of the Data Protection Suite is the Common Technology Engine (CTE). The CTE represents a common code base for the various data protection modules as well as a common set of metadata to facilitate sharing of data protection policies across the data protection modules. This results in a cohesive set of data protection solutions with unified cataloging, indexing, and data movement technology, rather than a disparate collection of standalone software. The Data Protection Suite comprises the following modules:



### HITACHI BACKUP AND RECOVERY SOFTWARE:

This is a modern backup and recovery solution for distributed, heterogeneous environments, with the features that organizations have come to expect in backup and recovery solutions, in addition to a number of advanced features not available in any other products. The Hitachi Backup and Recovery module is adept at using disk as both a permanent and temporary storage resource without the need for a Virtual Tape Library and is therefore well suited to D2D and D2D2T uses.

- **Hitachi Quick Recovery software:** Quick Recovery software creates point-in-time snapshots using its own host-based snapshot engine or by integrating with a number of snapshot providers, including Hitachi Data Systems, Dell, EMC, HP, IBM, NetApp, and XIOTech.
- **Hitachi Data Migrator software:** Using universally defined storage policies, the Data Migrator module provides automated, policy-based data migration functionality to move files from primary storage to a tiered storage environment. This facilitates HSM activities and can help in the reclamation of primary storage capacity.
- **Hitachi Data Archiver software:** This is an archiving solution for Microsoft Exchange. Data Archiver software retains all incoming and outgoing e-mail messages (using MS Exchange's journaling feature), and archives them to a secondary storage device, indexing them to allow for a number of advanced searches, including searching any headers as well as a full text search of each email.
- **Hitachi Data Protection Monitor software:** Data Protection Monitor software serves two functions. First, it operates as a single management console for geographically dispersed deployments of Hitachi Data Protection Suite across the various sites in an organization. Second, it provides global reporting, forecasting/trending, and monitoring capabilities that may be useful for help desks, storage administrators, IT managers, and IT executives.

The Hitachi Data Protection Suite offers a number of advanced areas of functionality, many of which are not available in any other products. The following is an attempt to summarize these advanced features.

**Distributed database:** The largest portion of a backup catalog is the file-level information (e.g. what file was backed up to what volume). This product distributes this information among MediaAgents, making the product much more scalable. Typically, backup environments are limited by how much metadata needs to be passed through the main backup server. Since only operations-level metadata is sent to the CommServer server, it can scale far beyond other solutions. The product does deal with the negative issues you might believe are present in a distributed database model, through the use of caching and automated reconstruction.

**Multistreaming beyond the filesystem level:** Another major limitation of many backup products is that they multistream at the client level. In other words, if there are five volumes/filesystems on a client, the software generates up to five simultaneous backups, but not more. This method worked fine when filesystems were a few GB; now they're a few TB. This product deals with this issue by allowing the administrator to specify how many simultaneous data readers it should use per filesystem or volume.

- **Full text search of all backups and archives:** This very unique feature of the Hitachi Data Protection Suite is a by-product of the Common Technology Engine. Since the software needed to provide full-text search capabilities for the archiving product, it was simple to extend this

functionality to the backup product as well, allowing administrators to perform full text searches against their backups.

- **Global storage policies:** Most backup products ask you to define a given backup's storage lifecycle every time you define a backup. An example of such a policy would be that backups should be sent to disk and copied to tape, the tapes should be sent off-site, and the backups here and there should expire at this time. The Hitachi Data Protection Suite allows you to define such policies once as a storage policy, and then apply that policy to one or more backup configurations.
- **Dynamically shared disk volumes:** Most backup products cannot dynamically share disk volumes. If you want to back up directly to a disk volume, you need to create (at minimum) one disk volume per backup server, and they cannot be shared. This, of course, leads to load-balancing, capacity, and performance challenges. The Hitachi Data Protection Suite can dynamically share disk volumes between multiple media servers.
- **Native persistent snapshots:** Without requiring volume management software, the Hitachi Data Protection Suite can create persistent snapshots of its clients and use them for quick recoveries. These snapshots can be natively created using the backup client, or it can leverage the snapshot capabilities of many storage systems.
- **Automated use of multiple copies:** Many backup products can use only one copy for restores. Additional copies are available for restores only if you tell the backup product that the first copy is no longer the primary copy. The Hitachi Data Protection Suite does not have this limitation. It will automatically use any secondary copy if it is available.
- **Synthetic backups:** The Hitachi Data Protection Suite has had synthetic backups for several years, allowing you to create a full backup without having to transfer all that data across the network.
- **Multiple-platform, customizable Java web interface:** What could be better than a single interface that's available on any platform with a web browser? In addition, with a click of a button, the product can be made to look like a Windows GUI or a KDE/Motif GUI. Pick your flavor!
- **Automated UNIX installation:** UNIX clients can be installed automatically using *ssh* or *rsh*.
- **Global & custom exclude lists:** Instead of maintaining several individual exclusion lists, this product allows you to maintain exclude lists on the server. In addition, you can create customized exclude lists for any given client, and they will automatically combine with the global exclude lists.
- **Network notification with thresholds:** Would you like to know how your backups are performing? Would you like to be notified when your tape library is full? Use pre-defined notifications or define your own. You can also customize these thresholds to do things like send an e-mail at 80%, an SNMP trap at 90%, and a page at 95% (for example).
- **Automatic configuration of default backup configuration:** When you install a client, it automatically creates a default backup configuration for it.
- **True priority settings:** Many backup products only apply priorities when backups are initiated. Therefore, backups that have already begun will not be stopped or paused for backups or

restores that are more important. With this product, you can tell it to interrupt less important backups for more important backups or restores. Once the more important job is completed, the less important job will resume where it left off.

- **Role-based administration:** You can define what capabilities any particular person should be allowed to perform.
- **Single view of all restores:** Instead of being required to specify what kind of backups you wish to restore from, you simply say which client you want to restore. This product then presents to you a drill down list of the various backups available for that client (e.g. file, Exchange, Oracle, etc.)

## LEVERAGING HITACHI STORAGE SYSTEMS WITH THE HITACHI DATA PROTECTION SUITE

Hitachi Data Systems offers a wide range of storage systems in its Hitachi TagmaStore™ line of storage for both small to mid-size environments as well as storage solutions for large-scale enterprise environments. The Hitachi TagmaStore™ Workgroup Modular Storage and Adaptable Modular Storage, models were designed with small to mid-size environments in mind, while the Network Storage Controller meets the needs of entry-level enterprises and fast-growing midsize organizations, and the Hitachi TagmaStore Universal Storage Platform was designed for the needs of large enterprise customers.

The Workgroup Modular Storage model WMS100 with SATA-only drives provides an inexpensive platform for businesses to implement disk-based data protection solutions such as the Hitachi Data Protection Suite. The Adaptable Modular Storage models AMS500 and AMS200 can use both Fibre Channel and SATA drives in a single frame using the Hitachi SATA Intermix Option, allowing you to also provide a tiered approach to your data protection strategies. Data with performance requirements that can't be met with SATA can be stored on Fibre Channel disks, and the rest of the data can be stored on SATA disks – all in a single frame. Both the Workgroup Modular Storage and Adaptable Modular Storage models provide the capability to partition their internal cache. This feature makes it possible for a single storage system to be effectively shared by production applications and data protection solutions without taking performance away from high-I/O applications.

Environments interested in virtualization should look at the Network Storage Controller model NSC55 and the Universal Storage Platform models USP1100, USP600, and USP100. These models provide capabilities beyond the traditional large, enterprise-class storage system, including controller-based virtualization, logical partitioning, and universal replication. In addition to presenting its own internal storage, the Universal Storage Platform models can virtualize heterogeneous external storage systems into a single pool and present them to hosts as if they were internal, allowing organizations to leverage their existing storage assets. The Network Storage Controller provides the same virtualization capability in a small footprint package designed for entry-level enterprises and fast-growing midsize companies. This controller-based virtualization makes it possible for organizations to build a multi-tier storage

infrastructure as part of a disk-based data protection strategy. These tiers can include older, not yet fully depreciated storage hardware or economically popular SATA-based storage systems that may not satisfy production data processing requirements, but adequately meet data protection requirements.

When evaluating storage solutions for disk-based data protection, organizations should take into consideration attributes including: cache size, maximum capacity, port count, footprint, price, shared versus dedicated resources, and ease of management. As you progress through the TagmaStore line of storage solutions from Workgroup Modular Storage to the Universal Storage Platform, these attributes increase. And since the TagmaStore line of storage utilizes the same set of management tools, this particular attribute remains relatively constant across the product line. It is important for an organization to identify its requirements up-front as these will help decide which storage solution is appropriate.

The TagmaStore line of storage solutions and the Hitachi Data Protection Suite complement each other in disk-based data protection environments:

- The Backup and Recovery and Quick Recovery modules can be integrated with the Hitachi ShadowImage™ In-System Replication bundle. This integration results in a cohesive, efficient, nondisruptive point-in-time backup solution. Recent snapshots can be stored on Fibre Channel disk, and older snapshots can be stored on less expensive SATA disk.
- Since the Hitachi Data Protection Suite can dynamically share SAN-connected disk volumes, you can create optimally-sized RAID volumes and present them to backup servers. Other data protection products that cannot share disk volumes (or that don't use multiple volumes well) require you to create RAID groups much larger than the optimum size.
- You can use Hitachi products in conjunction with a VTL, but you are not required to do so.
- The partitioning and virtualization capabilities of the TagmaStore line of storage make it possible to have a shared environment that uses the most appropriate storage for the given requirements. Data protection and production applications can share a storage system while being provided with the performance characteristics that meet their individual needs.

## THE ECONOMICS OF THE HITACHI DATA PROTECTION SUITE

The Hitachi Data Protection Suite is licensed per module. For example, if the Hitachi Backup and Recovery module and the Hitachi Quick Recovery module are to be deployed for a particular client then two licenses would be required, one for each module. The licenses are broken down in the following table.

Component	Description	Price info
CommServe	Central management point for controlling interactions between various software components and provides the Hitachi Data Protection Monitor	Priced per server class
MediaAgent	Responsible for managing the movement of data and storing file-level index information	Priced per server class
Media Management (Library and Drive Management)	Manages the media in a library	Priced per library and # drives
Media Management (Direct to Disk)	For disk-disk backup/migration/archival	Disk license is tiered priced based on number of TBs
Client	Specific to the operation and type of data it manages	Priced per module per client

The architecture of the Hitachi Data Protection Suite allows customers to reduce their total cost of ownership (TCO) in a number of ways.

- The file-level database being distributed across multiple MediaAgents allows for a much larger environment behind a single CommServe server. Not only does this reduce the purchase price, a single server is less expensive to maintain and configure than a multiple server.
- Since only operations-level data is sent to the CommServe server, even remote environments can be backed up using a single CommServe.
- While customers are free to purchase VTL software, there is much less of a business justification to do so with this product, since it can use disk-as-disk much more effectively than many other backup products.

- Advanced client functionality does not require the purchase or installation of products other than the Hitachi Data Protection Suite.
- All backup, archive, snapshot, and replication functions are managed through a single, integrated system with a common database and scheduling mechanism. This simplicity significantly reduces management cost.

## CONCLUSION

The Hitachi Data Protection Suite, powered by CommVault, offers a strong answer to today's data protection challenges. It starts with UNIX and Windows servers, MediaAgents for UNIX, Linux, NetWare, and Windows, and clients for UNIX, Linux, NetWare, MacOS, and Windows. They, of course, offer agents for today's database and mail systems, including Oracle, DB2, SQL Server, Informix, Sybase, Exchange, Notes, and GroupWise. All of these agents are managed with a single pane of glass that's available on any platform, and can be made to look like what the customer prefers (e.g. Motif style or Windows style).

Hitachi Data Protection Suite combines effective use of tiered disk and tape with global storage policies, which makes for a very strong offering. In addition to making backup and recovery faster by using full-text indexing and disk, Hitachi Data Protection Suite also makes use of local, persistent snapshots for even faster recoveries.

Its scalability and reporting features also allow this suite to scale to meet the needs of environments ranging from the workgroup to the enterprise. Supported by the first-rate reputation of Hitachi Data Systems customer support, Hitachi Data Protection Suite presents a very strong story. If you're in the market for a new data protection system, it's definitely worth a look.



### **ABOUT GLASSHOUSE TECHNOLOGIES, INC.**

GlassHouse Technologies, Inc. is the leading global provider of independent storage services and consulting. GlassHouse offers strategy, design and implementation, operations and support services in the following areas: Data Protection and Recovery, Storage Management, and Storage Business Practices. For more information, visit [www.glasshouse.com](http://www.glasshouse.com).