Continuous Data Availability and Operational Recovery: Why You Need Both

A Comprehensive Data Protection Solutions Methodology

By Hitachi Data Systems

April 2015
Contents

Executive Summary ....................................................................................................... 3

Data Protection Challenges............................................................................................ 4

  Amount of Data That Must Be Protected........................................................................ 4
  Protection of Large Objects.......................................................................................... 4
  Protection of a Large Number of Small Objects.............................................................. 4
  Long-Term Retention.................................................................................................... 4
  Service-Level Requirements......................................................................................... 4

Unified Recovery Management ....................................................................................... 5

The Service-Level Approach to Data Protection ............................................................ 7

  Prevention .................................................................................................................. 7
  Effectiveness............................................................................................................... 8
  Less Effective and Efficient.......................................................................................... 9

Business-Defined Data Protection ................................................................................. 9

Appendix: Hitachi Data Systems Data Protection Portfolio ............................................. 10
Executive Summary

This white paper discusses the challenges that application users face when trying to keep applications running continuously, and to protect data effectively from a number of data loss threats. It considers inadequacies of existing processes and industry trends that are likely to exacerbate these issues. It also profiles the comprehensive Hitachi Data Systems solution that effectively addresses these challenges. The intended audience for this document is technology acquisition decision-makers and influencers.

For the most part, data protection is presently synonymous with backup and recovery, and few organizations, if any, are happy with their existing environments. This white paper advocates taking a holistic view by leveraging other technologies, such as active-active fault tolerance, continuous data protection, snapshots, replication and archiving.

Indeed, no single technology is the right choice for every application, workload or service level requirement. The key is to build an understanding of which solution to use in each situation, and how they work together to provide a holistic approach to keeping the business running. Data protection has always been “software-defined.” Hitachi Data Systems is making it “business-defined.”
Data Protection Challenges

As noted earlier, there are many data management, availability, protection and recovery technologies on the market today. In fact, there are so many that it has become difficult to choose which solutions, combinations and locations, are required and suitable for your particular situation.

Amount of Data That Must Be Protected

Increasingly, more and newer types of data are being added to the mix that must be protected. Until recently, laptops, desktops, remote offices and testing and development environments were not generally included in enterprise data protection processes. Due to a number of government regulations, litigation and business-critical information, which is often stored on user devices, this distributed data now must be protected adequately.

Protection of Large Objects

Protecting a single large object becomes difficult, as there is no easy way to break the object into smaller pieces before copying it (backing it up). For example, an 84TB dataset over a single 10GB connection takes 24 hours, which makes daily protection of a larger dataset virtually impossible. Synchronous or asynchronous replication has been offered as a solution. However, replication alone does not offer recovery from a previous point in time, so it does not protect against data deletion or corruption.

Protection of a Large Number of Small Objects

Sequentially opening, reading, copying and closing a large number of objects (files) takes an inordinate amount of time. In file systems with myriad files (tens of millions of files), this process could take several hours and exceed the available backup window. For example, assuming 100 files can be backed up every second, over a 24-hour period only 8.6 million files can be backed up in a single stream.

Long-Term Retention

Given existing technologies, it is difficult to ensure that data can be reliably recovered after long retention periods, such as 20 or more years. The usable life of tape has improved over the years, but periodic technology refresh cycles and the need for constant environmental control make tape a challenge to manage in the long term.

Service-Level Requirements

As IT struggles with the challenges listed above, they are also being asked to keep everything running all the time. Some applications and data are more important than others, but for those elements that are critical to the operation of the business, any amount of downtime is increasingly intolerable.

There are several measures of data protection effectiveness as it relates to availability of the systems being protected. The goal is drive these metrics down to zero.

- **Backup window** is the amount of time that is necessary to stop or pause activity while the backup operation is conducted. Failing to stop write activity could result in a corrupt or inconsistent backup data set. A traditional incremental or full backup often takes many hours to complete, which makes backup window a critical metric to watch and improve.

- **Recovery point objective** (RPO) is the amount of time that elapses between backup operations. Another way to describe it is as the amount of recent data that is at risk of loss, because if it hasn’t been protected, it can’t be recovered. For example, traditional nightly backup operations result in an RPO of 24 hours.
Recovery time objective (RTO) is the amount of time it takes to restore operations following a data loss incident. It’s fair to ask, for each application or location, how long can you afford to be down before the business is negatively impacted? It is not uncommon for the downtime of critical applications to cost large enterprises millions of dollars per hour.

Unified Recovery Management

Most organizations follow a policy of full backups on the weekends and incremental backup on weekdays to protect all data. This one-size-fits-all approach is increasingly becoming inadequate, as all data is not equal in importance.

Hitachi Data Systems recommends a tiered protection approach that is based on service-level requirements of the data and is focused on recovery objectives. Organizations protect data to recover for three broad reasons. Each of these reasons requires different technologies that are optimized for that specific recovery type.

- **Operational recovery** includes recovery from operational issues, such as inadvertent deletion, malicious behavior, localized hardware failure, data corruption and so forth. It is the most common form of recovery performed in data protection operations.

- **Disaster recovery** includes recovery from catastrophic site disasters, such as earthquakes and tsunamis. Fortunately, such recovery is fairly infrequent. It is highly difficult and expensive, and usually involves restarting operations at an alternate data center.

- **Long-term recovery** provides for the discovery of and access to data that has been retained for long periods of time, such as 20 or more years. This data can include records kept for regulatory, governance, preservation or research purposes. Supporting long-term recovery requires a data life-cycle management approach. It must migrate inactive or required files to an archive repository, index the files to enable later discovery, and provide retention services, such as version control for auditing, legal hold, expiration and bit-level destruction.

Figure 1 shows some of the technology choices available to meet application-specific RPO and RTO. For example, continuous data protection (CDP) can be a good choice for reducing RPO to near zero, while a disk-based storage repository can provide the fastest RTO. A dedicated data deduplication appliance such as Hitachi Protection Platform can provide a very cost-effective backup repository while maximizing throughput performance.
Figure 1. Organizations can choose various options for a holistic data protection and recovery solution.

Operational Recovery (restore a file, folder, volume, system)

RPO
- CDP
- Snapshots
- Backup

RTO
- Disk
- Virtual Tape
- Tape

Disaster Recovery (restore operations at/from another location)

RPO
- Sync & Async Mirror
- Backup Replication
- Off-Site Tape

RTO
- Disk
- Virtual Tape
- Tape

Sync = synchronous, Async = asynchronous, RPO = recovery point objective, RTO = recovery time objective, CDP = continuous data protection

In addition to recovery objectives, organizations can take steps to prevent data loss with operational resilience to improve application availability from hardware failures, site disasters, network outages and other challenges. They can provide continuous operations across data centers.

For each of these recovery types, it is recommended that organizations protect the more valuable data more aggressively than less valuable information. This approach will help reduce the risk for the higher value data and reduce the cost of protection for data with lesser value to the organization. Figure 2 lists three possible application tiers and the technologies that could meet the RTO or RPO requirements for each. Customize this list for your organization, based on environment and needs.

Figure 2. Various tiers and technologies can meet an organization’s RTO and RPO requirements.

<table>
<thead>
<tr>
<th>Protection Objectives</th>
<th>Tier 1 Critical Data</th>
<th>Tier 2 Less Critical Data</th>
<th>Tier 3 Noncritical Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational Resilience</td>
<td>Active-active clusters</td>
<td>Active-passive clusters</td>
<td>Multiple access points</td>
</tr>
<tr>
<td>Prevent against</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hardware failure, network outage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operational Recovery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recover from</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corruption, HW crash, deletion …</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disaster Recovery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recover from</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catastrophic site-level disasters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long-Term Recovery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recover after</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very long retention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replicated content store</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content store with local protection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content store with local protection</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Service-Level Approach to Data Protection

In designing the perfect data protection, retention and recovery infrastructure for your organization, it can useful to categorize the tiers of service levels required. In Figure 3, we use “Prevention,” “Effectiveness,” “Efficiency” and “Less Effective, Less Efficient” to help set expectations and define RPO and RTO.

Figure 3. Categorize service levels requirements by recovery point and recovery time objectives.

Definitions
Recovery Point Objective: How much data can you afford to lose?
Recovery Time Objective: How long can you afford to be down?

Prevention
There is often a set of applications, workloads or processes that must always remain operational. Keeping them up and running 24/7 maximizes the effectiveness or profitability of the organization, and helps to avoid any catastrophic consequences from unavailability (such as lost customer orders, penalties or fines).

The global-active device replication and synchronization capabilities of Hitachi Virtual Storage Platform G1000 (VSP G1000) work with clustered application servers. They enable the absolute highest levels of continuous availability that these applications require. These services go above and beyond typical disaster recovery operations that require some level of manual intervention to accomplish an effective failover in another location. They truly drive RPO and RTO to zero. With an active-active cluster, there is no need to fail over following a disaster. The full set of applications and processes are already running and available in the other location.

Active-active clustering prevents downtime, and therefore provides a zero RPO and zero RTO service level.
Effectiveness

Major complaints about legacy backup operations include the amount of time that they take to perform the backup (backup window) and the time to restore (recovery time objective). Full backups require the copying of all data from a source system to a target backup system. Incremental backups require a lengthy scan of the source directory to determine what changed since the last backup.

Either of these methods can take many hours to perform both backups and restores, during which the applications being protected are usually unavailable to users. This timeframe was acceptable when businesses shut down at 5 p.m. and reopened the next morning, but that isn’t the case any longer.

To provide more effective data protection for important data sets, VSP G1000 includes state-of-the-art hardware-based active-active storage clustering, as described earlier, plus snapshot, cloning and replication capabilities.

**Hitachi Thin Image** (HTI) snapshot technology provides logical, change-based, point-in-time data replication within Hitachi storage systems for immediate business use. Business usage can include data backup and rapid recovery operations, as well as decision support, information processing, and software testing and development.

**Hitachi ShadowImage Replication** software is a nondisruptive, host-independent data replication solution for creating copies of any IT administrator-accessible data within a single Hitachi storage system. ShadowImage also maintains the availability of host applications by enabling backup operations to run concurrently while business or production applications are online.

The snapshot and replication software noted earlier can be provisioned, scheduled, managed and monitored with **Hitachi Replication Manager** (HRpM), which is part of the **Hitachi Command Suite** (HSC).

Hardware-based snapshots and clones, however, are not application-aware. When an application, such as a customer resource management (CRM) system processes a transaction, it writes to a number of different files and tables. All of those updates must be captured together to create an application-consistent copy and enable a reliable recovery.

For Microsoft® application environments, including Exchange and SQL Server®, plus Oracle databases on Linux, HDS provides **Hitachi Data Instance Director (HDID)** software. HDID sets these applications into a snapshot-ready state before calling the HTI or ShadowImage services on the VSP family of storage systems. HDID also orchestrates the directory clone feature of Hitachi NAS Platform systems. For other environments, including the Microsoft applications, Hitachi Data Protection Suite with its CommVault IntelliSnap technology, performs the same function.

Hitachi also offers a zero-worry data protection solution for VMware vSphere environments in Hitachi Virtual Infrastructure Integrator, which simplifies data management for file and block storage with an easy-to-use, business-defined policy engine for backup and recovery. It helps you meet backup and recovery service level agreements (SLAs) at virtual machine (VM) level granularity while improving resource utilization. It allows VM administrators to manage application-consistent data protection from the VMware vCenter console, leading to simplified IT operations. Virtual Infrastructure Integrator helps organizations to reduce business risk with quick application recovery improving RPO and RTO.

With hardware-based snapshots and replication, you can create fast and effective copies of your data. Create them more often to reduce the amount of data traditionally at risk between backup operations.

Another option to minimize data loss and downtime for critical data sets is host-based CDP. These software solutions capture each change as it is written to disk, eliminating the need for a backup window. The data changes can be sent to the backup repository continuously or on a scheduled basis. Since more data will be captured than with other methods, CDP is often deployed for short-term operational recovery, coupled with periodic snapshots or backups for longer-term retention. HDS offers Hitachi Data Instance Director for CDP of Microsoft Windows® environments.
For environments where the existing backup and recovery solution just needs a performance boost to complete operations on time, HDS offers Hitachi Protection Platform (HPP). This purpose-built data deduplication appliance provides industry-leading scalability and performance.

Data growth is at the heart of most data protection problems. It takes too long to back it all up, which results in unacceptable downtime during the process. And it takes too long to recover when something goes wrong, which also results in downtime. Also, consider that you often need 3TB to 5TB of backup storage for every 1TB of primary storage. With such a requirement, you can see how equipment, maintenance, software, management and environmental costs can be a major pain point.

One of the key things that can be done to rein in the undeniable growth in data is to reduce the amount of data that must be protected. Control can be accomplished with effective data lifecycle management policies that automatically move inactive data out of the primary storage system. If the inactive data needs to be retained, it gets migrated to a self-managed and self-protected tier of archive storage. When the data is no longer needed, it is automatically deleted, either from the primary or archive storage.

Hitachi Data Systems offers a leader in self-managed, self-protected archive storage with Hitachi Content Platform (HCP). Many archive, data life-cycle management and content management software solutions can send data to HCP for long-term management. These solutions include HDID, Hitachi Data Protection Suite (powered by CommVault), Rocket Arkivio AutoStor and Symantec Enterprise Vault.

Less Effective and Efficient

There is still a place for the traditional full and incremental backups that have been with us since the dawn of computing. For data sets that don’t need to be available continuously, such as overnight, and can tolerate lengthy recovery times following any failure, backup can be cost-effective. Backup has maintained its cost edge with the advent of data deduplication to eliminate the massive amounts of duplicate data each new full backup creates.

IDC predicts that sales of traditional enterprise backup software will continue to grow at about 5% per year. This estimate indicates that backup, like tape, is not going away anytime soon. Although it is not meeting the requirements for important or critical data sets and applications, it still has a place in most IT shops.

Hitachi Data Systems can meet your enterprise backup and recovery needs with Hitachi Data Protection Suite or Symantec NetBackup (see Appendix). Our expert customer engineers will help you decide on the right choice for your environment.

Business-Defined Data Protection

As information technology has evolved and become ubiquitous in almost everything we do, the importance and complexity to keep it available have grown exponentially. To meet these challenges, which are often rated as the No. 1 pain point of enterprise IT worldwide, HDS has assembled a portfolio of robust, market leading hardware, software and services offerings. Our experts help you design and deploy a comprehensive solution that is tailored to your specific needs, matching application availability and recoverability requirements with the most cost-effective solution available.
Appendix: Hitachi Data Systems Data Protection Portfolio

<table>
<thead>
<tr>
<th>Product</th>
<th>Functionality</th>
<th>Platform Support</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global-Active Device</td>
<td>Active-active storage clustering</td>
<td>Hitachi Virtual Storage Platform (VSP G1000)</td>
<td></td>
</tr>
<tr>
<td>Hitachi NAS File Clone</td>
<td>Hardware-based data cloning</td>
<td>Hitachi NAS Platform (HNAS)</td>
<td></td>
</tr>
<tr>
<td>Hitachi NAS Replication</td>
<td>Hardware-based replication</td>
<td>HNAS</td>
<td></td>
</tr>
<tr>
<td>Hitachi Thin Image</td>
<td>Hardware-based snapshot</td>
<td>VSP family, Hitachi Unified Storage VM (HUS VM)</td>
<td></td>
</tr>
<tr>
<td>Hitachi ShadowImage Replication (in-system)</td>
<td>Hardware-based data cloning</td>
<td>VSP family, HUS VM</td>
<td></td>
</tr>
<tr>
<td>Hitachi TrueCopy Extended Distance</td>
<td>Wide-area mirroring and failover</td>
<td>VSP family, HUS VM</td>
<td></td>
</tr>
<tr>
<td>Hitachi TrueCopy (synchronous remote replication)</td>
<td>Metro-area mirroring and failover</td>
<td>VSP family, HUS VM</td>
<td></td>
</tr>
<tr>
<td>Hitachi Universal Replicator</td>
<td>Hardware-based replication</td>
<td>VSP family, HUS VM</td>
<td></td>
</tr>
<tr>
<td>Hitachi Replication Manager</td>
<td>Snapshot and replication management</td>
<td>Hitachi Command Suite</td>
<td></td>
</tr>
<tr>
<td>Hitachi Protection Platform</td>
<td>Data deduplication appliance, virtual tape library</td>
<td>Symantec NetBackup, IBM® Tivoli® Storage Manager</td>
<td></td>
</tr>
</tbody>
</table>

**Snapshot Management**

<table>
<thead>
<tr>
<th>Product</th>
<th>Functionality</th>
<th>Platform Support</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hitachi Data Instance Director</td>
<td>Application-consistent snapshot and replication management</td>
<td>Hitachi Thin Image and Hitachi ShadowImage Replication; Hitachi TrueCopy, Hitachi Universal Replicator</td>
<td></td>
</tr>
<tr>
<td>Hitachi Data Protection Suite, powered by CommVault, with CommVault IntelliSnap feature</td>
<td>Application-consistent snapshot management</td>
<td>Hitachi Thin Image and Hitachi ShadowImage Replication; broad range of application support</td>
<td></td>
</tr>
<tr>
<td>Hitachi Virtual Infrastructure Integrator</td>
<td>Granular snapshot backup and recovery of virtual machines</td>
<td>HNAS, VSP family, HUS VM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hitachi Data Instance Director</td>
<td>Continuous data protection, archiving, replication</td>
<td>Microsoft Windows® Servers</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------------</td>
<td>---------------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td><strong>Backup</strong></td>
<td>Hitachi Data Protection Suite</td>
<td>Enterprise-scale backup, snapshot, archive, deduplication</td>
<td>Broad support for operating systems, applications and storage</td>
</tr>
<tr>
<td></td>
<td>Symantec NetBackup</td>
<td>Enterprise-scale backup, snapshot, archive, deduplication</td>
<td>Broad support for operating systems, applications and storage</td>
</tr>
<tr>
<td><strong>Archive</strong></td>
<td>Hitachi Data Instance Director</td>
<td>Granular file and email archiving to Hitachi Content Platform</td>
<td>Windows files and Microsoft Exchange email</td>
</tr>
<tr>
<td></td>
<td>Symantec Enterprise Vault</td>
<td>Enterprise-scale archive and discovery</td>
<td>Broad range of platform and application support</td>
</tr>
<tr>
<td></td>
<td>Rocket Arkivio Autostor</td>
<td>File life-cycle management</td>
<td>Windows, Linux</td>
</tr>
</tbody>
</table>