HITACHI
Inspire the Next

WHITE PAPER

Continuous Data Availability and Operational Recovery: Why You Need Both

A Comprehensive Data Protection Solutions Methodology

By Hitachi Data Systems

November 2016

Contents

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.

The pace of change in all industries is astounding, with many new market entrants totally disrupting the way customers acquire and consume products and services. This change is driving almost every organization to start the journey toward digital transformation in order to remain competitive or even relevant. When an organization transforms its IT environment to support changing business models, it is imperative that it also modernizes its approach to data protection. Failing to do so, and continuing to use methodologies developed 20+ years ago, will limit much of the performance, availability and agility benefits it hopes to gain in transformation efforts. Modern approaches reduce the risk of data loss and enable much faster recovery, reducing or even eliminating potential business impact.

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

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, increased litigation, and more business-critical information stored at the edge, distributed data must now be protected.

The world is becoming increasingly digital, thanks, in part, to the explosion of sensor data, video monitoring, higher resolution medical and scientific devices and the internet of things (IoT). As a result, we expect that the 40% annual data growth rates we've seen in the past to increase exponentially in the future. Where IT leaders planned around gigabytes and terabytes in the past decade, they will be designing IT solutions capable of petabytes and exabytes before the end of this decade.

Not all of this new data, such as some machine-to-machine communications, will need to be protected, but much of it will. Understanding the difference can lead to significant cost savings and competitive advantage.

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. You could invest in faster networks, but that's an expensive and temporary approach. Alternatively, 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 the availability of the systems being protected. The goal is to 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 monitor 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

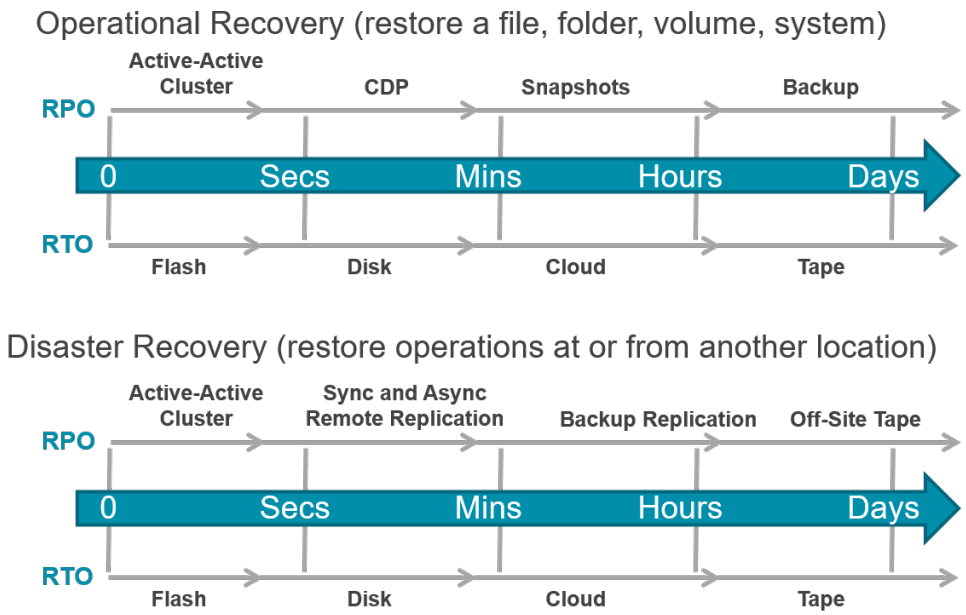
Most organizations follow a policy of full backups on the weekends and incremental backup on weekdays to protect all data. This one-size-fits-all approach is increasingly becoming inadequate, as all data is not equal in importance.

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, major storms and regional power outages. Fortunately, such recovery is fairly infrequent. It is highly difficult and expensive, and usually involves restarting operations at an alternate data center. In addition to provisioning the data and the infrastructure, planning for availability of personnel resources is also a key consideration, especially in geographic areas susceptible to widespread disruptions.
- **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, active-active storage clustering and continuous data protection (CDP) can be good choices for reducing RPO to near zero, while a flash-based or disk-based storage repository can provide the fast RTO. Additionally, private and hybrid cloud storage such as Hitachi Content Platform can provide a very cost-effective repository for long-term backup retention.

Figure 1. Organizations can choose various options for a holistic data protection and recovery solution.



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.

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.

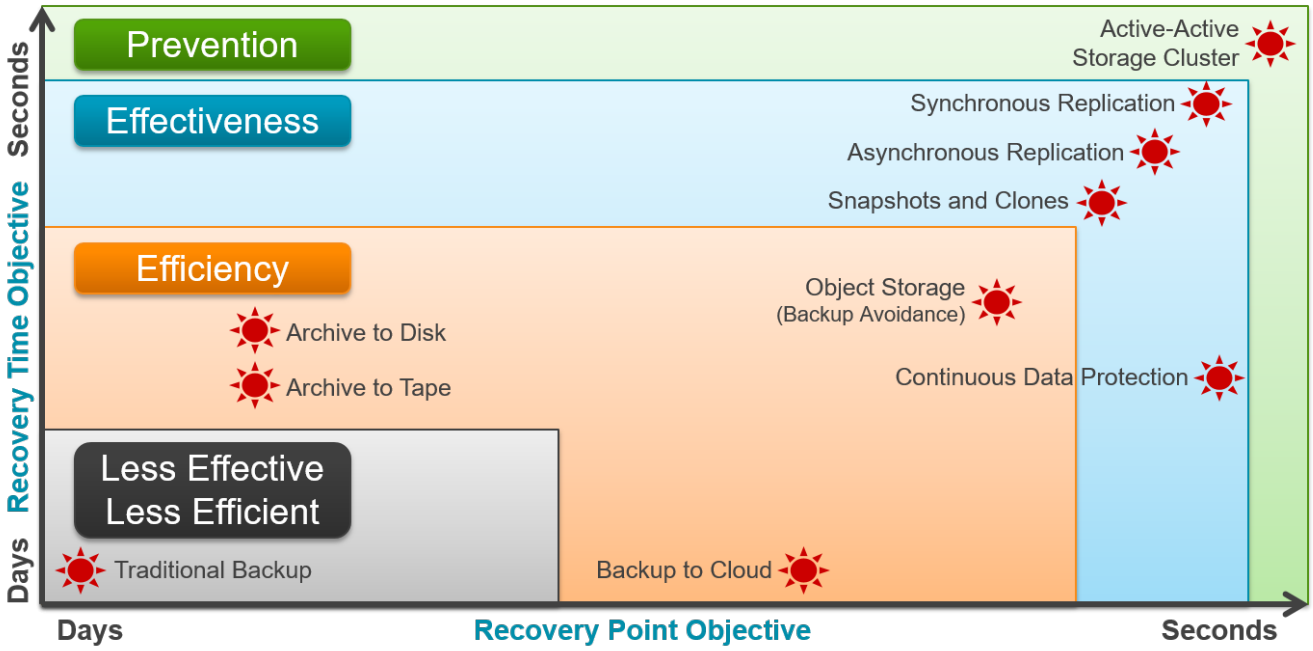
Figure 2. Various tiers and technologies can meet an organization's RTO and RPO requirements.

Protection Objectives	Tier 1 Critical Data	Tier 2 Less-Critical Data	Tier 3 Noncritical Data
Operational Resilience Prevent against: hardware, network, building outage	Active-active clusters	Active-passive clusters	Multiple access points
Operational Recovery Recover from: corruption, hardware crash, deletion …	Frequent copies on arrays	Copies on-site, on-disk	Copies on tape, on-site or off-site
Disaster Recovery Recover from: catastrophic site-level disasters	Real-time off-site copies	Near-real-time off-site copies	Periodic off-site copies
Long-Term Retention Retrieve for: e-discovery, analytics, reference …	Replicated object store	Object store with built-in protection	Object store with built-in protection

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 level requirements by RPO and RTO.



Definitions
Recovery Point Objective: How much new 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 reputation loss).

The global-active device replication and synchronization capabilities of Hitachi Virtual Storage Platform G series and F series work with clustered application servers. They enable the absolute highest levels of continuous availability that these applications require. 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.

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, the VSP family includes state-of-the-art hardware-based active-active storage clustering, as described earlier, plus local in-system and remote 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. These snapshots are very space-efficient, and up to 1 million snapshots can be maintained on large enterprise VSP storage systems.

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 enables application-consistent copies for repurposing, including decision support, information processing and software testing and development. These copies can be used concurrently while business or production applications are online.

Hitachi TrueCopy synchronous remote replication keeps a zero-RPO copy of data in a metro-area location for fast failover and failback in the case of a system or site-level outage.

Hitachi Universal Replicator (HUR) asynchronous remote replication provides the ability to failover to another location, anywhere in the world, providing protection from widespread events. HUR can be combined with either global-active device or TrueCopy to enable robust 3-data-center business continuity and disaster recovery capabilities, in either cascade or multitarget topologies.

The snapshot and replication software noted earlier can be provisioned, scheduled, managed and monitored with Hitachi Replication Manager (HRpM) or Hitachi Storage Advisor (HSA).

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, SAP HANA and other databases, 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. Snapshots and clones can also be created on the replicated remote copies, enabling nondisruptive repurposing of the data for secondary operations.

Hitachi also offers a zero-worry data protection solution for VMware vSphere environments with 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 also helps organizations to reduce business risk with quick application recovery, improving RPO and RTO. Virtual Infrastructure Integrator works with HDID. VMware snapshots can be created, viewed and restored from either interface.

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. CDP of Microsoft Windows® environments is included in Hitachi Data Instance Director.

Efficiency

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. 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 life-cycle management policies that automatically move inactive data out of the primary storage system. If the inactive data needs to be retained, it can be migrated to a self-managed and self-protected tier of object storage. When the data is no longer needed, it is automatically deleted, either from the primary or long-term storage.

Hitachi Data Systems offers a leader in self-managed, self-protected object storage with the Hitachi Content Platform (HCP) portfolio, forming the foundation for secure private and hybrid cloud infrastructure. HCP users can easily support diverse use cases, such as enterprise file sync and share, remote office file services, endpoint backup, and spanning archive to cloud and beyond, all with a flexible storage ecosystem. HCP achieves data governance and data mobility in a single platform.

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 organizations.

Hitachi Data Systems can meet your enterprise backup and recovery needs with Hitachi Data Protection Suite or Veritas 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

	Product	Functionality	Platform Support	Web
Snapshot and Replication	Global-Active Device	Active-active storage clustering	Hitachi Virtual Storage Platform (VSP G1000)	
	Hitachi NAS File Clone	Hardware-based data cloning	Hitachi NAS Platform (HNAS)	
	Hitachi NAS Replication	Hardware-based replication	HNAS	
	Hitachi Thin Image	Hardware-based snapshot	VSP family	
	Hitachi ShadowImage (in-system replication)	Hardware-based data cloning	VSP family	
	Hitachi TrueCopy (synchronous remote replication)	Metro-area mirroring and failover	VSP family	
	Hitachi Universal Replicator	Hardware-based replication	VSP family	
Snapshot Management	Hitachi Data Instance Director	Application-consistent snapshot and replication management	Hitachi Thin Image and Hitachi ShadowImage replication; Hitachi TrueCopy, Hitachi Universal Replicator	
	Hitachi Data Protection Suite with Commvault IntelliSnap feature	Application-consistent snapshot management	Hitachi Thin Image and Hitachi ShadowImage replication; broad range of application support	
	Hitachi Virtual Infrastructure Integrator	Granular snapshot backup and recovery of virtual machines	HNAS, VSP family	

	Product	Functionality	Platform Support	Web
Continuous Data Protection	Hitachi Data Instance Director	Continuous data protection, archiving, replication	Microsoft® Windows® Servers	
Backup	Hitachi Data Protection Suite with Hitachi Content Platform	Enterprise-scale backup, snapshot, archive, deduplication	Broad support for operating systems, applications and storage	
	Veritas NetBackup with Hitachi Content Platform	Enterprise-scale backup, snapshot, archive, deduplication	Broad support for operating systems, applications and storage	
Archive	Veritas Enterprise Vault with Hitachi Content Platform	Enterprise-scale archive and discovery	Broad range of platform and application support	
	Rocket Arkivio Autostor	File life-cycle management	Windows, Linux	

Hitachi Data Systems

Corporate Headquarters
2845 Lafayette Street
Santa Clara, CA 95050-2639 USA
www.HDS.com | community.HDS.com

Regional Contact Information
Americas: +1 866 374 5822 or info@hds.com
Europe, Middle East and Africa: +44 (0) 1753 618000 or info.emea@hds.com
Asia Pacific: +852 3189 7900 or hds.marketing.apac@hds.com

WP-478-D R. Vining November 2016.