Understand Key Challenges, Provide Proactive Solutions

In today’s “always on” business world, fast and continuous access to applications and data is essential for success. As a result, data resilience has become an integral part of the business and IT lexicon, and it has consumed an increasingly larger portion of the IT budget.

Ensure Data Resilience

Business today operates in a highly competitive, 24/7 world of just-in-time delivery, global supply chains, around-the-clock customer demand and alternative vendors that are just a mouse click away. Delays in resuming IT operations following an unexpected disruption can have a detrimental impact on bottom line financial performance and competitive advantage. In addition to securing business data against outages,
many organizations face regulatory oversight that dictates mandatory levels of data protection. In the United States, regulations now apply to health care (HIPAA), financial services (SEC 17a-4), corporate accountability (Sarbanes-Oxley), life sciences (21 CFR Part 11) and government (DoD 5015.2-STD). Elsewhere in the world the story is similar, from the New Basel Capital Accord (Basel II) globally to RIPA and FAS in the U.K., and to COB in France. Like government regulators, investors and insurers are now insisting that businesses maintain feasible business continuity plans to protect critical information.

The fast restart of your business operations after a disruption is essential. Ideally, you will want to resume business at the point in time at which operations stopped functioning. To achieve this, the most up-to-date business data must be continually available at a location safely removed from the disaster area.

The frequently desired scenario of a remote disaster site ready to go at a moment’s notice can be costly to implement and may not be warranted for all applications. The challenge for IT is to construct a comprehensive data resilience solution that minimizes annual loss expectancy by appropriately aligning solutions from the palette of interrelated capabilities. These encompass local site replication and recovery, offsite transportation of backup data, electronic vaulting and remote data replication.

Risk Reduction and Data Resilience
Choosing a technology, or combination of technologies, to support business continuity and ensure data resilience starts with an assessment of the potential risks facing your organization. An operational risk management approach will clarify business requirements and reduce uncertainty by estimating the likelihood and loss potential of each type of incident. This systematic analysis of each business critical application will help to determine the extent of vulnerability, the business impact of downtime and any regulatory requirements. With this information in hand, you can assess technologies to achieve the optimal balance of recovery speed, data value and cost.

Two key metrics used for data resilience planning are recovery time objective (RTO) and recovery point objective (RPO). RTO measures how long it takes to resume essential operations: how long it takes to get back on your feet. RPO is a measure of data currency: how far behind you can afford to be when resuming operations after a disruption. Your RTO and RPO determine which data replication and recovery options your business needs and how much the solution will cost. Another frequently overlooked requirement that can help align cost and risk for your organization is the recovery capacity objective (RCO). RCO considers that while you need your data and applications to process it, you might not need all the throughput capacity in certain disaster scenarios, at least for some period of time following a disaster.

Evaluate the Options
Planning for business continuity and data resilience involves matching the RTO, RPO and RCO values of each critical application to the capabilities of various data protection technologies. Figure 1 illustrates the correlation between the RTO and the cost of comprehensive recovery capabilities.

Traditional Tape-based Data Protection Challenges and Solutions
Tape backup technology is well established as the historic cornerstone of data protection. While it still plays a key role in data resilience efforts, the explosion of data growth can cause two risk exposures to creep into an organization. The more obvious risk will typically be to RPO. This manifests itself increasingly in frequent challenges to successfully getting the backups done and within an appropriate time window. The more subtle risk will be to RTO. This will only be visible during tests or exercises of the recovery process, which are typically done on a far more infrequent basis. The increased adoption of disk-to-disk and disk-to-disk-to-tape backup methods, data de-duplication and virtual tape strategies is alleviating some of these pressures. Simply put, organizations need a comprehensive solution.

Cost versus Value of Data

Figure 1. Recovery time objective is based on the value of particular data, mapped to a range of technical approaches, costs and degrees of data protection.
strategy that pulls from the entire array of disk and tape solutions to optimize data protection, improve data availability and facilitate rapid access to stored information. Though the strategy must be comprehensive, it must also be cost-effective and simple to manage.

**Advanced Backup and Recovery Solutions**

The Hitachi Data Protection Suite, powered by CommVault®, is a unified platform that helps ensure outstanding data protection, high availability and rapid access to information through the integration of backup and recovery, migration, archiving and replication under a single point of control. Designed for disk-to-disk SAN-based data protection, the Data Protection Suite avoids the performance issues of products that have evolved from LAN-based disk-to-tape infrastructures. Hitachi Data Protection Suite has de-duplication capabilities embedded in the software; this increases flexibility and makes it ideal for virtual server environments and remote locations. The Data Protection Suite helps organizations implement the right mix of data protection components without adding complexity, reaching across the enterprise, even to the desktop.

**Virtual Tape with Integrated Data De-duplication Solutions**

Replacing slow tapes with high speed disk enables organizations to increase backup and recovery performance, optimize their storage infrastructure and achieve their data protection objectives. Data de-duplication eliminates redundant data, greatly reducing the disk capacity required by up to 25 times or more. Virtual Tape Library Solutions by Hitachi Data Systems employ IBM® ProtecTIER® technology to enable organizations to get all the benefits of backing up to disk at a drastically reduced total cost of ownership. Virtual Tape Library Solutions are industry’s first enterprise-class data protection solutions that offer a high capacity, high performance and highly scalable target for backup, archive and compliance data.

**Local, In-system Replication Solutions**

While remote replication receives much of the attention from disaster recovery planners, local, in-system copy solutions add a valuable complementary layer to a data resilience program. For example, point-in-time (PiT), in-system replication disk copies permit organizations to rapidly recover from logical data corruption introduced by human or application errors.

The Hitachi ShadowImage® Heterogeneous Replication software provides high speed, nondisruptive replication for any Hitachi storage system or pool of storage virtualized by Hitachi Universal Storage Platform® V (USP V) and VM (USP VM), and Virtual Storage Platform (VSP). ShadowImage software consistency groups allow a user defined group of ShadowImage volume pairs to be simultaneously split, at a precise moment in time, with a single command.

This method creates a consistent PiT copy of an entire system, database or any related sets of volumes. This copy can then be used for remote replication, in conjunction with Hitachi Universal Replicator software, for example, to another storage system anywhere in the world. Hitachi Copy-on-Write Snapshot software provides the same functionality, but it stores only the changed data, resulting in a significant savings over full cloning methods.

**Remote Data Replication Solutions**

Replicating data to a remote secondary site represents the most effective insurance policy against system downtime. Remote data replication offers the fastest recovery time following an outage and the lowest risk of data loss. Replication eliminates the time consuming, manual and error prone multistep recovery process required by traditional tape-based backup. It also provides a variety of productivity benefits through secondary access to repurpose data, without affecting regular production workloads.

Remote data replication increases data availability by:

- Automating procedures to reduce the duration of planned events, such as system maintenance, application testing and development, and data backups
- Allowing nondisruptive backup of current production data with no impact or acceptable impact to the production application, depending on whether asynchronous or synchronous replication is selected
- Speeding failover and data restoration in the event of an outage by replacing slow and labor intensive tape-based restores with continuously available online backups
- Allowing secondary sites to take over primary processing to eliminate scheduled downtime
- Enabling frequent, nondisruptive disaster recovery testing with an online copy of current and accurate production data

Two basic variations of remote data replication are available — synchronous and asynchronous.
Synchronous and Asynchronous Replication

Synchronous replication ensures that a remote copy of the data, identical to the primary copy, is created at the same time the primary copy is updated. In synchronous replication, an I/O update operation is not considered complete until confirmed at both the primary and remote sites, guaranteeing that the remote copy is always an exact image of the primary.

Synchronous replication offers very fast recovery. After a disruption, business operations can resume immediately at the remote site, starting from the exact point the primary site stopped. Only I/Os that were in flight at the instant of disruption will be lost. And, because neither the primary nor remote site will have a record of these transactions, rolling back a database to the last confirmed state offers full data integrity.

The drawback to synchronous replication is distance. Fibre Channel, the primary enterprise storage transport protocol, can theoretically extend a remote copy as far as 300 kilometers (km) or 180 miles from the primary site. However, as distance increases, so does latency. Delays propagating updates to the secondary site can affect performance. Depending on the application’s sensitivity and the communications technology between the two sites, the effectiveness of synchronous copy can begin to degrade at 32km to 160km or 20 miles to 100 miles. This may not be far enough to clear a wide area disaster zone. However, Hitachi TrueCopy® Synchronous software can also be leveraged for “no data loss” out-of-region solutions in three data center configurations from Hitachi Data Systems.

Asynchronous replication provides a mechanism for mirroring data across any distance, usually involving IP networks. With asynchronous replication, the primary write operation is disconnected from the remote write operation. After a successful write of data to the primary storage system, the application continues processing without waiting for an acknowledgement from the secondary storage system.

<table>
<thead>
<tr>
<th>Asynchronous</th>
<th>Synchronous</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overview</strong></td>
<td>Replication across any distance without significant host performance impact</td>
</tr>
<tr>
<td></td>
<td>No acknowledgement dependencies from secondary site</td>
</tr>
<tr>
<td></td>
<td>Potential for some data lag between remote and primary sites, particularly at longer distances</td>
</tr>
<tr>
<td><strong>Needs Addressed</strong></td>
<td>Data is secured at secondary or tertiary sites across long distances.</td>
</tr>
<tr>
<td></td>
<td>Each I/O update operation waits until completion is confirmed at both the primary and mirrored sites. Any incomplete operation is rolled back at both locations; thus, the remote copy is always an exact mirror image of the primary.</td>
</tr>
<tr>
<td><strong>Data Loss</strong></td>
<td>Asynchronous mode may or may not lose some committed transactions in the event of an unplanned failover to the secondary site. However, a rapidly restartable data image is ensured because the remote database will be in an I/O-consistent state, resulting in the need to reapply only the most recent transaction logs to recover to the point of the outage.</td>
</tr>
<tr>
<td></td>
<td>Maximum distance for synchronous mode theoretically extends to 200 kilometers, depending on channel extender specifications. But latency quickly becomes a problem as propagation delays (the time spent waiting for the update to travel to the remote site and confirmation to come back) lengthen with increased distance. The practical distance for synchronous replication of a busy transaction system is usually about 30 to 50 kilometers (20 to 30 miles) — possibly further, depending on an application’s tolerance for delayed response and other factors.</td>
</tr>
<tr>
<td><strong>Distance</strong></td>
<td>Asynchronous mode can span virtually any distance, because there is no propagation delay involved in confirming transactions at the remote site. Remote sites can be up to thousands of miles from the primary site, ensuring that the replicated copy of data is safely outside any likely disaster zone.</td>
</tr>
<tr>
<td></td>
<td>Synchronous mode has a greater impact on performance than asynchronous mode, because a write from the host must wait for acknowledgment from the secondary storage system.</td>
</tr>
<tr>
<td><strong>Performance Impact</strong></td>
<td>The performance impact on the host is minimal.</td>
</tr>
<tr>
<td></td>
<td>Synchronous mode has a greater impact on performance than asynchronous mode, because a write from the host must wait for acknowledgment from the secondary storage system.</td>
</tr>
<tr>
<td><strong>Data Integrity</strong></td>
<td>Several mechanisms ensure the remote copy is made in precisely the same write sequence as the primary copy, including the use of sequence numbers and time stamps in the data packets.</td>
</tr>
<tr>
<td></td>
<td>With synchronous replication, data integrity issues caused by out-of-sequence writes do not arise because “dependent writes” are not initiated until prior writes on which they depend complete.</td>
</tr>
<tr>
<td><strong>Network Infrastructure</strong></td>
<td>When coupled with bridging technology from Cisco or Brocade, Hitachi Data Systems replication technologies support a wide range of network infrastructures, including IBM® ESCON® and FICON®, IP, SONET, ATM and others.</td>
</tr>
</tbody>
</table>

Table 1. Consider the pros and cons of asynchronous and synchronous replication.
acknowledgement of a successful write at the secondary storage.

Asynchronous replication can span any distance without affecting application performance. Remote sites can be hundreds, or even thousands, of kilometers or miles from the primary site, ensuring critical data is stored safely outside a disaster zone.

To protect against rolling disasters, a data replication technology must be able to freeze remote replicas at a point in time prior to the onset of the outage. The ability to create frozen point-in-time images of data is what differentiates remote copy technology from simple mirroring.

Asynchronous replication introduces potential write sequence fidelity problems because packets of data can arrive at the remote site out of order. This is especially true of technologies that send disk track updates, which cannot preserve the original write sequence. To ensure the integrity of asynchronously replicated data in a rolling disaster, replication technology must employ techniques to automatically preserve write sequence fidelity at the remote site.

Choosing between Synchronous and Asynchronous Replication

Clearly, remote storage replication for recovery and business continuity requires more than just shipping data over a network. The selection process starts with an assessment of the potential risks and their probability. If your organization cannot tolerate any data loss and operations must be resumed quickly following an outage, synchronous replication is likely to be the best choice. Of course, the decision must also factor in how far the data has to be replicated to clear the slight time difference between data being at the primary and remote sites. Asynchronous replication can span any distance without affecting application performance. Remote sites can be hundreds, or even thousands, of kilometers or miles from the primary site, ensuring critical data is stored safely outside a disaster zone.

Because the remote and local I/O of a synchronous replication succeed or fail together, this replication approach does not introduce data inconsistencies following a disaster. Rolling disasters are primarily a challenge for remote asynchronous replication, and one of the principle problem areas is write sequence fidelity.

Write Sequence Fidelity

Database and file managers maintain very complex internal data structures, including indices, structured data tables, directories, logs and so forth. To preserve the integrity of these internal structures, each write is carefully sequenced so that, at any point in time, a correct file system or database state can be recreated.

During a rolling disaster, not all systems, storage and network connections fail at precisely the same moment. In this situation, a system may still be able to process transactions and issue updates to primary storage devices, but due to earlier failures, not all updates may replicate successfully to the secondary site. Rolling disasters pose a challenge because they often result in corrupted and unusable data at the remote site, requiring difficult and very lengthy recovery processes.

To protect against rolling disasters, a data replication technology must be able to freeze remote replicas at a point in time prior to the onset of the outage. The ability to create frozen point-in-time images of data is what differentiates remote copy technology from simple mirroring.

SOLUTIOn PROFILE

The Hitachi Data Protection Suite, powered by CommVault®, is a unified platform that helps ensure outstanding data protection, high availability and rapid access to information through the integration of backup and recovery, migration, archiving and replication under a single point of control.
benefits of both synchronous and asynchronous replication: the three data center option or the four data center option.

Three and Four Data Center Options

Two data center replication strategies are viable for most in-region recovery sites, such as, for example, serving as a hot site for a campus-level or metro-level server cluster. They are also viable for out-of-region recovery sites where propagation delays are not an issue. Synchronous replication provides very fast recovery time (low RTO) and good data currency (low RPO). However, asynchronous replication provides better protection against regional disasters, albeit with a less favorable RPO. A three data center strategy offers the best of both worlds: fast recovery and excellent data currency for local site failures, combined with advanced protection from regional disasters.

For increased flexibility and scalability among large, distributed information systems, Hitachi Data Systems also supports the four data center configuration for both open system and mainframe storage environments. This allows you to implement synchronous replication at two sites within a short distance using Hitachi TrueCopy Synchronous as well as asynchronous replication across long distances using Hitachi Universal Replicator.

Heterogeneous Replication: A Major Breakthrough

With heterogeneous storage, a feature of most midrange and enterprise data centers, the ability to replicate between systems from different vendors, locally and long distance, is critical to cost-effective business continuity planning. The downside to multivendor storage business continuity is the cost and complexity of the solution. Software, in-house expertise and licenses must be duplicated for each vendor’s replication solution.

Tiered Storage and Hitachi Replication Software

Hitachi USP V, USP VM and VSP provide enterprise and midsized storage environments with a consistent interface to pooled multivendor storage. Using industry leading controller-based virtualization, Hitachi storage systems enable a single replication tool set to operate against all heterogeneous storage resources in a tiered infrastructure. This significantly reduces the complexity and cost of replicating data, both locally and across long distances.

Improve Your Asynchronous Replication Integrity and Performance While Lowering Costs

Hitachi Universal Replicator software provides advanced, controller-based replication among all of the storage systems certified for external attachment to Hitachi USP V, USP VM and VSP systems. Universal Replicator software’s disk-based journaling increases replication hardness, reduces the impact of replication on production applications and lowers replication costs by more efficiently utilizing bandwidth, particularly when there is flexibility in the management of RPO from the business perspective.

When collecting data to be replicated, the primary Hitachi storage system writes the designated records to a special set of performance enhanced journal volumes. The remote Hitachi storage system then reads the records from the journal volumes, offloading the primary system by pulling them across the communication link instead of making the primary system push them, as in most other approaches. By writing records to journal disks instead of keeping them in storage system cache, Universal Replicator software no longer consumes a large percentage of available cache, freeing resources for production transactions. With only one set of procedures, tools and skills to be mastered, regular and comprehensive disaster recovery testing of complex heterogeneous environments becomes more feasible and affordable.

As shown in Figure 2, combining Universal Replicator software with USP V, USP VM and/or VSP systems enables three data center business continuity configurations to be achieved with less complexity, scripting and overhead than in the past and therefore makes three data center business continuity solutions more affordable for a broader range of enterprises and applications.

Example: Three Data Center Configuration

Online Commerce Traffic with Multisite Disaster Recovery Solution

When a burgeoning Internet retail company began experiencing errors and interruptions in its database systems, it sought out a premier business continuity plan.

The e-commerce marketplace has created a paradigm shift in how goods are purchased and sold. In one Asian country, this company had become the first online shopping mall to achieve US$100 million of sales within a half a year period. With more than 9.3 million registered users and nearly two million products currently available, this company’s site had become the most visited e-commerce website.

Universal Replicator software efficiently unifies the three data center solution with advanced capabilities, such as disk journaling and asynchronous replication driven by the remote site. Rather than using cache memory, the software employs time-stamped disk journals of replication data to minimize effort by the primary production site.
The key to its success: the ability to facilitate quick, secure and convenient transactions between buyers and sellers of a constantly growing inventory of products. The company pioneered the open market business model, in which supply and demand come together in a single place and participants engage in instant fixed price or negotiable deals, rather than having to wait for auction bids to close.

Despite continued efforts to expand infrastructure and maintain stability, the company’s exponential growth rate always exceeded hardware reinforcements. Performance was an issue in the company’s existing database infrastructure, with chronic errors and interruptions occurring in the transfer of data between servers and complicating back-end processes. Also, slow response times or power outages could have devastating results for the online marketplace, such as transaction interruptions and loss of data or revenue. The database architecture needed a checkup, but this was risky and time consuming. Replacing the entire architecture to resolve chronic system problems also carried a good deal of risk. Still, preventing disruption of operations was vital to the company’s continued success.

The company determined that it would need to restructure its disaster recovery configuration in order to reduce costly errors, maintain system stability and support expanding operations. The plan was to design a three data center configuration and deploy a core technology of one source, two targets. Unlike a two data center disaster recovery system, which provides backup from a primary site to the secondary site, the unique model would support highly stable, continuous backup of the same data from both sites. This ensures availability and integrity of critical information.

Universal Replicator software efficiently unifies the three data center solution with advanced capabilities, such as disk journaling and asynchronous replication driven by the remote site. Rather than using cache memory, the software employs time-stamped disk journals of replication data to minimize effort by the primary production site. And data can be copied between any supported systems, regardless of operating system or protocol differences, allowing maximum flexibility for data distribution and failover options.

Achieve “No Data Loss” Replication across Any Distance
To ensure availability of current data copies in each company’s disaster recovery site, Universal Replicator software partners with the synchronous functionality of TrueCopy Synchronous software for continuously available online copies. This software combination helps Hitachi to deliver “no data loss” replication across any distance, while guaranteeing data integrity. The company is now positioned to continue large scale growth with confidence that its IT infrastructure will deliver durable, uninterrupted service. Because the company now employs advanced replication capabilities and performance, complications no longer follow back-end work processes such as data transfer and backup.

With the help of Hitachi Data Systems, the company has built the ultimate disaster recovery model in its country, based on the resiliency of the Universal Storage Platform models. Their shopping cart is piled high with benefits to support its growing e-commerce success: storage virtualization, universal replication and even a 30 percent savings in cabling costs.

Business Case for Solving the Data Resilience Problem
Adopting a sound business continuity strategy produces clear business benefits.

- **Minimize downtime, gain better ROI:** Provide continuous access to business information and eliminate or minimize both planned and unplanned downtime; provide immediate remote access to time

---

![Figure 2. A three data center, multitarget replication configuration reduces the impact of replication on production applications and lowers replication costs.](image)
Critical information for other data mining and decision support systems with no impact on production systems; and improve business confidence by enabling more frequent and nondisruptive testing and disaster recovery scenarios with current production data.

- **Protect valuable information assets:** Prevent loss of business information from site disasters, viruses, human error, and application and system failures by providing alternate remote and local copies of data and enabling rapid restart and recovery.

- **Use resources and personnel efficiently:** Offload processing and data to alternate systems and locations; use advanced software tools to minimize the labor intensive nature of backup and recovery procedures.

- **Standardize IT practices on fewer storage platforms:** Gain improved performance and greater scalability and availability; reduce burdens on IT staff.

- **Simplify disaster recovery:** Achieve rapid restart and recovery over virtually any distance with minimal impact on application response time; eliminate inefficient and resource intensive tape-based replication or retrieval systems.

Implement Your Data Resilience Solution

This Hitachi Data Systems Solution Profile offers you a basic understanding of what is involved in implementing a data resilience solution for your organization and what the return on investment can be. There are many choices among the software and services available to simplify disaster recovery, minimize downtime, speed recovery and protect your information assets while maximizing the use of resources and personnel.

Consider engaging Hitachi Data Systems Global Solution Services professionals before you undertake data resilience planning. Global Solution Services aligns infrastructure strategies to business and application requirements, using an end-to-end methodology, as shown in Figure 3.

Global Solution Services can help you:

- Understand your organization’s current risks and requirements with a Risk Analysis Workshop.

- Identify and analyze your business goals in terms of the exposure risks your company faces and the regulatory requirements your company needs to meet, as well as the length of time your business can afford to be down and how much data it can afford to lose.

- Identify your technical goals, based on which replication strategy makes sense, the nature of your existing infrastructure, the topologies of your storage layout and placement, and traffic patterns by application and host.

- Create a migration plan to a new solution that includes test and verification components.

- Document the design and create a detailed implementation and test plan.

- Implement the technologies successfully and execute the plan within scheduled time and budget.

- Manage your infrastructure with onsite residency services.

- Provide proactive reporting on the health of your data resilience solution with a Replication Scorecard.