



## **Top Use-Cases for Replication Demand Integration with Backup and Restore Capabilities**

**CommVault Leads in Providing a Single Integrated  
Solution Covering Replication and Backup/Restore**

### *Abstract*

*A one-product approach that combines replication and protection is the most beneficial for organizations virtually every time they need replication. CommVault Simpana is the only product that completely integrates backup/restore and replication so that the software is managed as one.*

November 2008

By

DCIG LLC  
7511 Madison Street  
Omaha NE 68127

402.884.9594



# **Top Use-Cases for Replication Demand Integration with Backup and Restore Capabilities**

## **CommVault Leads in Providing a Single Integrated Solution Covering Replication and Backup/Restore**

### ***Table of Contents***

Executive Summary .....	Pg 1
The Piece-meal Approach to Replication is Not the Best for IT .....	Pg 2
Detailed Consideration of the Most Popular Replication Use Cases.....	Pg 2
Specialized Requirements for Data Protection when used with Replication.....	Pg 3
Requirements for an Integrated Solution.....	Pg 4
Considerations and Requirements for an Integrated ROBO Data Management Approach.....	Pg 4
Considerations and Requirements for an Integrated DR Approach.....	Pg 5
CommVault CDR and Galaxy: Replication and Backup/Restore in a Single Simpana Product.....	Pg 5
Starting with Replication .....	Pg 7
Understanding the Value of Advanced Replication Capabilities .....	Pg 7
Obtaining Coherent Application Data with Consistent Recovery Points.....	Pg 7
Recovering Data from Replicated Recovery Points .....	Pg 8
Leveraging Intelligent Storage Systems.....	Pg 8
Views from the Field: Simpana Replication and Backup/Restore in Operation .....	Pg 8
DCIG Conclusions: CommVaults Leads in Providing a Best-Fit, Single Solution Approach to Replication and Backup/Restore .....	Pg 8
DCIG Channel Analysis: CommVault's Best-Fit Replication Provides Some Compelling Channel Advantages.....	Pg 10



## **Top Use-Cases for Replication Demand Integration with Backup and Restore Capabilities**

**CommVault Leads in Providing a Single Integrated  
Solution Covering Replication and Backup/Restore**

### ***Executive Summary***

Data replication is available in a variety of forms, including storage-embedded, network-embedded, volume manager-embedded, application-embedded and system hosted software. Depending on the systems environment and rate of change of the data, each of these approaches has its own advantages and compromises to offer. Typically, however, storage- and network-embedded replication is deployed with UNIX systems while Windows and Linux environments commonly run host-based software replication.

In this paper, DCIG considers the specialized requirements of using host-based software replication in combination with data protection. Generalized statements made about replication technologies assume a host-based implementation. Most often, host-based implementations replicate data by duplicating the I/O stream being written to the host-attached disk. Although features vary, the essential technology of duplicating the I/O stream does not vary much from product to product for host-based technologies first developed on Windows platforms. Vendors developing host-based replication on UNIX systems have offered block-based implementations. However these have failed to capture much market share in these environments due to their preference for storage-embedded approaches and are therefore not considered in this analysis.

Host-based data replication is generally used in combination with data protection. Electronic vaulting of data for the purpose of Disaster Recovery, while still the most popular use of replication, moves data protection to an off-site location. Centralizing data from remote and branch office (ROBO) locations is done entirely for the purpose of moving data protection from those remote sites to a primary datacenter location. In either case, the requirements for data backup and restore must be considered when evaluating and selecting a data replication technology and vendor.

A one-product approach that combines replication and protection is the most beneficial for organizations virtually every time they need replication. Piece-meal approaches to addressing replication separate from backup and restore are insufficient because they add new complexities and costs which increase the risk of data loss.



## Top Use-Cases for Replication Demand Integration with Backup and Restore Capabilities

### The Piece-meal Approach to Replication is Not the Best for IT

When IT teams deploy replication technologies for their environments, nine times out of ten the replication is part of a larger data protection solution. The most popular use of replication is the same as it has always been: disaster recovery. Using replication to centralize and protect data from remote and branch offices is a close second. In a very distant third, some organizations also use replication for distributing data – including application files – from central offices out to remote offices and users. This high correlation between replication and the use of replicated data for backup protection means that IT organizations should consider their backup and restore requirements when selecting their replication vendor.

### Detailed Consideration of the Most Popular Replication Use Cases

Disaster Recovery (DR), or Business Continuity, is the most often cited use of replication technology. Replication is superior in many ways to alternative methods of moving data offsite to survive site and regional disasters, such as tape rotation. Not only is data replication easier and less burdensome to manage than physical collection and tracking of tape cartridges, replication also moves data as the data is created rather than waiting for a weekly or monthly tape rotation cycle. Moving data offsite as it is created reduces the risk of data loss in the event of a site disaster, preserves data up to the moment of the disaster and makes it available almost instantaneously at the replication target.

Compare this to the norm of storing data on tape that may take a week or more before it is moved off-site and is available for use in a disaster recovery. Replicated data is more readily available for use should a disaster occur so rather than spending precious time restoring data from tape using a process that can take days or weeks to complete, replicated data is online and ready for use immediately.

DCIG analysts estimate that ~80% of the time data replicated for DR purposes is also copied onto tape at the remote location. Tape is still the lowest cost media for long-term data retention. However with replication in place, rather than performing a backup of the primary data at the production site, it can be done using the replicated copy of the data at the remote site. Backing up a replicated copy of the production data can now be done at any time without impacting application performance or end-user productivity. In a similar manner, data from remote and branch office (ROBO) sites is protected at the central datacenter location. This enables organizations to remove the cost and burden of operating backup in their remote offices while lowering the risk of data loss.

Based on their collective 30+ years of replication experience and discussions with real IT teams, DCIG analysts estimate that replication is used primarily for the following reasons:

- 70% of the time for disaster recovery
- 50% of the time for ROBO data protection (often in combination with Disaster Recovery)
- 10-20% of the time for data distribution



## Specialized Requirements for Data Protection when used with Replication

When considered separately, data replication and data protection solutions may each be adequate on their own but inadequate when used in combination. If the replication software lacks integration or ‘awareness’ of the backup process, it forces administrators to integrate the two to work in conjunction with one another. Integrating the two is prone to human error and puts data at risk.

Relying on the replication vendors to integrate their replication software with the backup software is also problematic. This approach does not take into account how troubleshooting problems between the two would be resolved and could ultimately result in the creation of unrecoverable data sets.

IT teams should consider the following specialized requirements for data replication software, when they intend to use it in combination with data protection:

***Transparency of data to the remote client level:*** Backup jobs are scheduled, prioritized, monitored and reported by the client system. When working against a replicated data copy, however, this transparency can be lost disrupting normal backup administration. In the case where transparency is lost, often all that is known is that a blob of replicated data is either protected or unprotected. When failures occur, the risk to the organization and the resulting urgency for resolving the backup issue is an unknown. Moreover, reports on the transparency of the remote client are relied upon for protection audits, governance and oversight.

***Efficient, rapid restore of individual files and email for individual remote client systems:*** When transparency to the remote client is lost, backup administration and reporting suffers and data restores are put in jeopardy. Restores are more difficult because the process requires more steps and takes longer than should be necessary. Most of the time a data restore at a granular level is required, such as for individual files or email messages.

When data is protected as a blob without awareness of which remote clients are included in the blob, the capability to find individual files is compromised. Working with blobs of data, restores must start with a thorough, tedious search through the backup data copies in order to find the required files or email. Many times, operators need to resort to a full restore of all data for an entire remote site because they cannot be sure which systems and files are required from the blob of protected data. The resulting delays place unnecessary burden on IT teams, networks and infrastructure while negatively impacting end-user and business productivity.

***Known coherent state of replicated data:*** A fact often ignored or overlooked is that a stream of replicated data taken from systems running applications, such as Microsoft Exchange, SharePoint, SQL Server or Oracle database systems, are not coherent at every point in the stream. In fact, the stream is coherent only when these applications have completed transactions and committed to disk all pending transactions held in memory.

This commit only occurs occasionally, such as when applications are shut down, are inactive for a period of time, or when they are deliberately quiesced. A replicated copy of incoherent application data is of questionable value to protect; incoherent data is not usable for restores without considerable administrative intervention including roll backs and some data loss, if the source backup copy is usable at all.

In practice, most administrators know to wait until applications are not in use and replication streams are inactive before they begin backups of the replicated data as this practice usually prevents data from being incoherent when it is protected. However, this restricts backup to off-hour and overnight windows. As end-user activity extends into overnight hours this approach no longer ensures data coherency.



## Requirements for an Integrated Solution

Because data protection and replication are most often required in combination, a single solution approach offers the best promise for efficient, simple, reliable DR and centralized ROBO data management. Requirements for an integrated approach vary slightly for each of these popular use-cases which are considered in this section.

## Considerations and Requirements for an Integrated ROBO Data Management Approach

To ensure seamless support for all recovery scenarios, an integrated solution for ROBO data management must consider these scenarios:

- **Requirements for local, on-site recovery:** When local disks or systems fail, it is faster and easier to restore data from a local replicated copy of the data rather than relying solely on a remote copy. This requires a one-to-many replication stream, which can simultaneously duplicate data to a second disk location in the local environment as well as replicating data to a remote site.
- **Single Management Interface:** An integrated solution is easiest to manage when a single GUI interface is provided, offering full local and remote control of every step in the replication, backup and restore of ROBO data.
- **Comprehensive Remote Management:** Remote management is a requirement to enable central IT teams to manage remote sites and systems. These must include comprehensive installation and maintenance of the software. Insufficient capabilities can prompt the need for administrators to travel to remote sites, or to rely on local ROBO staff.
- **Single Policy Management:** To ensure ease-of-administration, an integrated solution must go beyond superficial GUI integration and use the same or similar policies to schedule, trigger, monitor and control replication, backup and restore processes. When technologies for replication and protection originate from separate companies, each will require more work from administration teams to learn, operate, and maintain.
- **Client Transparency through Replication Process:** Preserving knowledge or ‘awareness’ of client systems in the replication stream is critical for enabling administrative oversight through the backup process. Transparency is also critical for simplifying and speeding the restore process. This assumes, of course, that the replication and protection product includes sufficient reporting and one-step data restore capabilities.
- **Application Awareness:** To ensure data coherence and reliable recovery of data, replication technology must integrate with applications such as Microsoft Exchange, SQL Server and SharePoint along with Oracle database systems. Application awareness enables a replication product to force a quiescent state by triggering the application to commit all transactions from memory to disk. A checkpoint is then inserted into the replication stream, marking the point at which the data is coherent and suitable for backup protection.



## Considerations and Requirements for an Integrated DR Approach

In addition to the capabilities discussed for ROBO data protection, an integrated solution for DR must consider these scenarios:

- ***DR test and practice walk-through:*** Organizations must plan for DR tests and practice a walk-through at least once a year and ideally more often than that. The tests help administration teams be better prepared in the event that a real disaster occurs.
- ***Proactive avoidance of disaster:*** Some types of disasters, notably hurricanes and other types of catastrophic weather events, prompt proactive switchovers from primary datacenter locations to DR facilities. Rapid failback with re-synchronization of data is a key requirement for enabling this type of avoidance scenario.

## CommVault CDR and Galaxy: Replication and Backup/Restore in a Single Simpana Product

The CommVault Simpana product is unique in providing a single-product approach combining replication, backup and restore. Built on a single architecture of modular data management services, Simpana offers a combination of the capabilities required for efficient, simplified, and reliable data management for ROBO environments and DR.

The Simpana modules providing these capabilities include:

- Galaxy backup and restore with client modules for each file system and application that must be protected.
- Continuous Data Replicator (CDR) provides many-to-many data replication on TCP/IP local and wide area networks.
- Common Technology Engine which is the core architecture for managing data captured from clients onto secondary storage media.
- Unified Console, an intuitive GUI for managing replication, backup and restore all from a single pane with common policies, scheduling and reporting capabilities.
- Single Instance Store (SIS) data deduplication, which makes preservation and retention of backup data on disk more efficient and less costly.
- Advance Feature Pack, including GridStor capability, which keeps backup and replication running through system and network outages.

CommVault is well-known as a source of world-class backup and restore technology, and the Galaxy product is the flagship of CommVault's Simpana suite. Adding CDR replication capabilities to Galaxy deployments is relatively easy as adding replication does not require new servers or networks but can be enabled – turned “on” – with a license key change to the Simpana software.



## Views from the Field:

### Simpana Replication and Backup/Restore in Operation

At a high level, the benefits of using both Simpana Replication and Backup/Restore are intuitively obvious. Administrators can select the best data protection option for each application and then use a common console to create and put in effect the appropriate management policies for either product. Further, all of the protected data is then stored and indexed by Simpana. But those in the field who are already testing and using Simpana Replication and Backup/Restore are finding this integrated solution is changing their perspective on data protection in other ways.

InfoReliance's John Chirhart is already testing the new Simpana Replication and Backup/Restore in his current role as a consultant to the US Government. As he performs these tests, he is finding that it is forcing him to re-examine what data protection strategy that he should recommend to his clients and how they should manage data protection going forward.

One of the biggest changes that he sees coming out of the merger of these two technologies is the creation of a common team that manages both backup and disaster recoveries. Companies now often use different teams to manage the replication software and the backup software since the replication software is used primarily for off-site disaster recoveries while backup software is used in the day-to-day recoveries of emails, databases and files. But by bringing both of these technologies together, it eliminates the labor associated with managing each one independently and makes it possible for companies to combine the two teams normally required to manage these products into one.

A distinct advantage Chirhart also sees is the elimination of the need to train people on the other product. He has worked with other replication products that required either him or his staff to attend week-long boot camps to learn how to configure and manage the product. Even once the training was complete, he still needed to purchase that vendor's professional services to configure and install the product plus it took a substantial amount of time to manage the product once it was installed.

He has not found that true with Simpana Replication and Backup/Restore. Since he is already using Simpana Backup/Restore, he understood Simpana's architecture and could deploy the Replication software without the need to get CommVault professional services or support directly involved. Then because the two products are integrated, his existing backup team could manage Replication jobs just like Backup/Restore jobs using the existing Simpana management console.

Chirhart does not view the combined Simpana Replication and Backup/Restore as a package that can solve every replication need that a company may have. Rather he sees Simpana's host-based replication solution as providing a broad range of benefits to companies that can help them "stop the bleeding" associated with their day-to-day backups while giving them a more effective way to recover their data. As this occurs, he sees the day rapidly approaching where the industry comes to view integrated replication and backup/restore not as an option but as the standard way that companies protect their data.



When both replication and backup are configured with Simpana software, disk targets can be easily shared and managed collectively between backup and replication tasks. Organizations who want to run replication on a separate server also have this option due to the flexibility of the Simpana product architecture.

#### Starting with Replication

For organizations that already have other backup software in place, CommVault enables easy addition of just the replication piece of the Simpana software. CommVault does this by packaging and selling CDR as a full-featured stand-alone product in addition to selling it as a module for its full backup suite. Organizations obtaining CDR receive all of the capabilities required to replicate data to and from a server, as well as manage that replication through the Unified Console interface. Adding on to CDR with backup capabilities sometime in the future is still an option.

### **Understanding the Value of Advanced Replication Capabilities**

Although designed as a part of the Simpana architecture, the CDR replication capabilities offer many advanced features considered essential for efficient, cost-effective network performance.

These capabilities include:

- Simple Data Transfer, which improves replication speeds with fan-in ratios of up to 100:1 for replication data from multiple production sites into a single off-site DR system, or from multiple remote sites into a single datacenter system
- A built-in Replication Predictor tool for Windows can be used to estimate the amount of data throughput required per hour, day and so forth, and thus enable accurate planning for network bandwidth requirements and allocation
- For Windows file system data, Out-of-Band Synchronization enables an initial transfer of data and resynchronization back to the original system without requiring additional network
- Threshold monitoring and automated disk space handling help to manage snapshots and retain them, without exceeding disk space
- Built-in throttling and robust scheduling options enable data to be moved on shared networks with other types of business traffic
- Comprehensive reporting enables easy validation and analysis of data replication volumes and network requirements over time
- Periodic log flushing ensures that replication streams are completed at specified time intervals, even without sufficient activity to fill the replication log

### **Obtaining Coherent Application Data with Consistent Recovery Points**

Applications hold their data partially in memory, and must be quiesced to force all data to be written to disk. Simpana CDR is integrated with the use of SmartSynch Scan to force a momentary quiescence of supported applications, insert a checkpoint marker in the replication stream, and thus enable the creation of a Consistent Recovery Point at the destination system. The Consistent Recovery Point, in this case, is nothing more than a snapshot of the replicated data made at the point that the data is known to be consistent, or coherent.



Simpana CDR also provides auto-discovery for supported applications, to ensure that all log files and directories associated with the application are captured in the replication stream. This avoids human error during deployment and configuration, and helps ensure that application data is complete and coherent at the destination system.

### **Recovering Data from Replicated Recovery Points**

Simpana replication includes a Copyback feature, for restoring data from the replicated data copy on the destination system. Individual files, folders and volumes can be browsed, selected, and copied back to their original system or to an alternate location on the network. Recovery Points can also be recovered using the Copyback capability.

To make this easy for remote systems, administrators can mount a Recovery Point for easy browse from the destination computer. On Windows, a Recovery Point can be made available on the network as a shared volume.

Recovery of data can also be done from backup copies made from Recovery Points on the destination system. File system data which has been backed up from a Recovery Point can be restored using the Simpana File System iDataAgent, which is a simple file system client agent. SQL Server and Exchange data backed up from a Consistent Recovery Point can also be browsed and restored, using the same Windows File System iDataAgent on the client system.

### **Leveraging Intelligent Storage Systems**

NetApp storage systems running OnTap offer better advantages as replication targets when used in combination with Simpana replication software, than other types of disk. Simpana software has been built with integration to recognize OnTap snapshots as replication recovery points. These recovery points can be accessed through a UNC mount path to a NetApp file server running OnTap.

Once created, OnTap snapshots are ideal backup targets. Recovery Points can be configured to be backed up automatically when they are created; for CDR on Windows there is no need to mount the snapshots that comprise the Recovery Point in order to back them up; for CDR on UNIX, the snapshots do need to be mounted, but CDR does this automatically. In addition, you can select any existing Recovery Point and perform an immediate full backup on-demand. This backup capability is only available if the source and destination computers have the appropriate File System iDataAgent installed.

All types of backups are supported using the Recovery Point, including full, differential, incremental and synthetic full backup. The restore process offers granular browse of the individual files and folders captured in the backup, selection of those which are required for restore, and full volume restore.

### **DCIG Conclusions: CommVault Leads in Providing a Best-Fit, Single Solution Approach to Replication and Backup/Restore**

In the data protection space, there are large numbers of replication products, backup/restore products and even a number of vendors that offer both replication and backup/restore products as part of their suite of products. However what sets Simpana apart from other products and product suites is that Simpana is the only one that completely integrates backup/restore and replication so that both its policies and data stores are managed under the same umbrella. So as companies select a replication solution for disaster recovery, ROBO data protection, data distribution or some combination of all three, here are five key points to consider:



- **Integration with backup software.** Replication software cannot and will not replace backup software as backup software still performs some tasks that replication software does not, such as copying data to tape. Using a product such as Simpana that natively integrates replication and backup/restore eliminates the time and expense of trying to do the integration yourself or relying on someone else to do it for you. Plus it keeps current and future versions of both products in sync.
- **Familiar user logins and security permissions.** Backup software and replication products each tend to require their own user names and passwords with the security permissions that do not align with what your company needs. Simpana integrates with Microsoft Active Directory so the user IDs and associated security permissions found in Microsoft AD roll right into Simpana and administrators do not have to compromise on user and data security due to product limitations.
- **Common management console.** When using different products from different vendors or even different products from the same vendor that are not integrated, companies must learn each product and management interface. Then to manage the replication and backup/restore jobs, they need to jump back and forth between each product to make sure they work in conjunction with one another. Simpana eliminates those concerns as replication and backup/restore jobs are managed through the same interface without the need to relearn a new product or manage multiple management consoles.
- **Extend enterprise-level recovery and availability to the application servers that can least afford it but need it the most.** Linux and Windows servers represent probably the majority of servers in most enterprises, have lower administrator to server ratios and the lowest budget for data protection. Yet the users of these applications expect and may need the same speed of recovery (30 minutes or less) that they see on their high-end application servers. Using Simpana Replication in conjunction with Backup/Restore, companies can now deliver these types of restore benefits with minimal increases in software and hardware costs or administrative intervention. If anything, Simpana Replication will decrease administrative time associated with the protection of data since it is now better protected and easier to recover.
- **Companies no longer have days or weeks to devote to recovering data – they have minutes or hours.** Trying to manage Replication and Backup/Restore software as two different products lines within the business is inefficient and cumbersome plus it precludes companies from rapidly recovering their data. Forward-thinking companies already recognize that business applications cannot withstand recoveries that take days or weeks to occur nor does their IT staff have the time to spend on these tasks. Rather they need to restore data within hours or even minutes without breaking either the bank or the backs of those responsible for managing the solution. Simpana Replication and Backup/Restore gives companies the flexibility to deliver and manage near-real time data recoveries as part of their day-to-day data protection package.



## **DCIG Channel Analysis:**

### **CommVault's Best-Fit Replication Provides Some Compelling Channel Advantages**

CommVault channel partners should find several low-hanging-fruit opportunities to offer Simpana Continuous Data Replicator (CDR) software, including:

- ***Easy, low-cost Disaster Recovery for Windows and UNIX systems.*** DR is still in high demand, and for many organizations the high costs and complexity implied puts DR out of reach for many small and medium businesses. To say that VARs sell exclusively to organizations in the SMB is wrong. However, VARs often sell into the SMB and can find many underserved DR requirements in that segment of the market into which Simpana CDR is an excellent fit. Moreover, because Simpana CDR supports any type of disk, VARs have the added incentive and advantage of being able to place any variety of storage systems at the remote DR site.
- ***Capture more data and lower end-user costs with Remote Site Data Centralization.*** As organizations continue to locate offices and systems closer to their customers, the continued proliferation of remote and branch sites is inevitable. With Simpana CDR, VARs can offer easy, low-cost data centralization at a lower cost to end-user customers than proliferating backup infrastructure into those environments. VARs who need a solution to offer into competitive backup markets should find this high-value offer to be a viable approach to capturing new customers.

Once placed into an environment, CommVault Simpana software offers VARs easy follow-on opportunities to build onto the Simpana deployment with new modules and capabilities for protecting growing data, systems and sites. Simpana software is more difficult to displace than other types of software, because of its multi-purpose capability set: it's much harder to replace the software that handles both your backup and your replication, than to replace one or the other. VARs working with CommVault should see this as an advantage to helping them to manage and control account activity within their customer base.

Moreover, as data managed under Simpana software continues to expand, the need to sell and deploy new storage systems at the customer site likewise continues to expand. Simpana software supports a wide variety of market-leading storage systems, which makes it easy for VARs to offer the storage system of their choice – according to how they are compensated, and with the flexibility to respond to ever-changing incentives and opportunities over time. CommVault Simpana software offers a better approach with data replication, for both VARs and end-user customers alike.

NOTICE: The information, product recommendations and opinions made by DCIG LLC are based upon public information and from sources that DCIG LLC believes to be accurate and reliable. However since market conditions change, the information and recommendations are made without warranty of any kind. All product names used and mentioned herein are the trademarks of their respective owners. DCIG LLC assumes no responsibility or liability for any damages whatsoever (including incidental, consequential or otherwise) caused by one's use or reliance of this information or the recommendations presented or for any inadvertent errors which this document may contain. Any questions please call DCIG LLC at (402)884-9594.