



A CommVault White Paper:

Continuous Data Replicator

Centralized Management & Disaster Recovery of Remote Office Data

CommVault Corporate Headquarters
2 Crescent Place
Oceanport, New Jersey 07757-0900 USA
Telephone: 888.746.3849 or 732.870.4000

CommVault Continuous Data Replicator Product Overview

Table of Contents

Centralized Management and Continuous Data Protection of Local and Remote Office Data.....	2
<i>The Challenge</i>	2
<i>The Solution</i>	2
<i>Primary Use Cases and Key Differentiators</i>	2
Remote Office Backup Consolidation.....	3
Cost-Effective Disaster Recovery (DR).....	4
<i>Flexible Configuration Options</i>	4
Continuous Data Protection for Datacenter and Remote Office Environments.....	5
How it works.....	5
<i>Efficient Data Movement</i>	5
<i>Point-in-time Recovery Points</i>	5
A minor interruption.....	6
A major interruption.....	6
<i>Maximizing use of Available Network Bandwidth</i>	6
Compression.....	6
Encryption.....	7
Bandwidth Throttling.....	7
Out-of-Band Synchronization.....	7
Incorporating Capacity Management with Replication.....	8
Unique Benefits of CommVault Singular Information Management.....	8
Summary.....	9

Centralized Management and Continuous Data Protection of Local and Remote Office Data

The Challenge

Managing data at remote sites can be costly, time consuming, resource draining and inefficient. Relying on manual operations, remote media devices, erratic WAN connections and limited resources can result in questionable data protection and recoverability. By some estimates, 60% of critical business data resides outside the datacenter. Managing that remote data can account for 40 to 60% of a typical enterprise IT budget. With data growing exponentially year over year and sprawling beyond the datacenter, distributed organizations must find cost-effective technology to reliably protect and manage their remote office data.

The Solution

CommVault® Continuous Data Replicator (CDR) is the ideal solution designed to help businesses meet critical recovery point and recovery time objectives. This host-based, multi-platform replication technology helps centralize and simplify the administration of remote office data and provide cost-effective disaster recovery. CDR continuously captures and replicates data changes at the byte level. This minimizes the amount of data that has to go over the network and reduces impact on performance. One of the many unique features that differentiate CDR from other replication products is the use of application-consistent snapshot technology. This ensures the integrity of replicated data and its recoverability. The use of snapshots also provides an additional layer of protection against viruses or corruption.

Key Benefits of CDR:

- Reduces management costs by centralizing remote office data
- Provides cost-effective Disaster Recovery for Microsoft Exchange, SQL Server and Oracle
- Delivers fast, reliable recovery of remote office data
- Improves RTO/RPO
- Maximizes data protection and availability

This white paper provides detailed information about consolidating remote office environments and protecting critical application data for Microsoft Exchange, SQL Server and Oracle. It outlines unique capabilities that set CommVault Continuous Data Replicator (CDR) apart from alternative solutions.

Primary Use Cases and Key Differentiators

Continuous Data Replicator offers many unique advantages over other replication products currently on the market. It can be used as a standalone solution to consolidate remote office data and provide continuous data protection from site failures. Best of all, it can be added to an existing CommVault environment for simplified management and continuous data protection of local and remote environments.

When used with other components of the CommVault® Singular Information Management™ suite, CDR can be easily managed along with backup, recovery and archiving through a single policy-based user interface. This enables administrators to easily track remote backup sets to the source without time consuming manual processes. It also delivers the granularity to browse backup recovery points and perform recovery operations as if the backup was locally created.

Primary Use Cases:

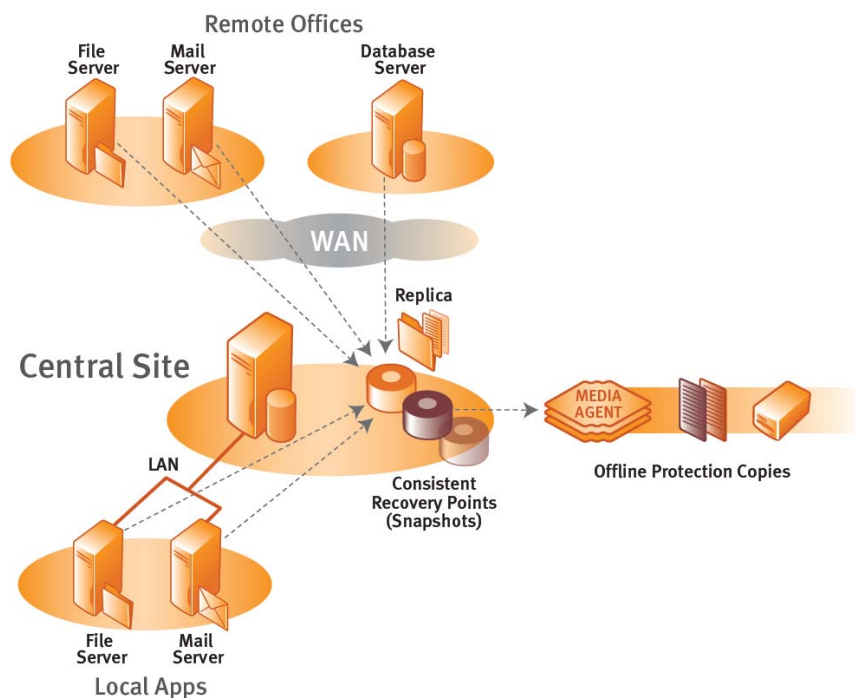
1. Remote office backup consolidation
2. Offsite application disaster recovery
3. Continuous data protection for datacenter and remote office environments

Remote Office Backup Consolidation

Continuous Data Replicator is the ideal solution for customers with various remote office locations who are looking to implement a consolidated, centralized data protection strategy for all remote office data. CDR is fully integrated with the lifecycle management capabilities of the CommVault Singular Information Management suite. This integration provides a powerful single unified interface to control the replication, data protection and archiving processes of remote office data. It also allows administrators to seamlessly extend lifecycle management policies to remote office locations from a single policy and single console.

Figure 1 shows one possible example of a scenario which may include any number of file, mail, or database servers at remote offices as well as on local servers.

As shown, application data is replicated to the central site where it can be tied in directly to the established backup policies for offline data protection. These backup sets can then be browsed and restored directly to the client as if the backups were created locally.



Key Benefits:

- Easily manage data lifecycle from a central User Interface
- Leverage existing policies to simplify management
- Use Recovery Points to Perform faster backups

These backup sets can then be browsed and restored directly to the client as if the backups were created locally. This greatly reduces the burden on the administrator to track locations of backup sets for specific remote locations.

Once replicated, the remote office data becomes part of the data protection policies as soon as the recovery point is created. The recovery point represents a copy-on-write (COW) space-efficient snapshot of the volume containing the replicated data.

These snapshots can be utilized for data recovery in a number of ways:

- A snapshot can be mounted as a read-only volume to provide an end-user browse and retrieve capability for replicated file data.
- The administrator can perform a copy-back operation of any recovery point to create a full volume of the specified point-in-time to any server within the environment.
- When the recovery point is created, it can be associated with a storage policy to automatically create a backup set. These backup sets can be browsed and restored as part of the normal restore process.

This unique integration of backup with the replication process also allows for users at remote office locations to browse and restore their backup sets as if they were created from the source. CDR automatically maintains the mapping of replicated data to backup host.

Cost-Effective Disaster Recovery (DR)

Continuous Data Replicator offers a compelling asynchronous alternative to high-cost remote hardware mirroring solutions. It is ideal for customers who are looking to implement a cost-effective disaster recovery solution for key enterprise applications including Microsoft Exchange, SQL Server and Oracle. CDR maintains updated remote copies of file system and application data at remote geographic locations. These copies can be quickly mounted and brought on-line in the event of a site outage or disaster.

Key Benefits:

- Quickly recover data and applications in the event of disaster
- Cost effective alternative to mirroring

Establishing the initial replica of these critical data sets can occur while applications are online and in use (CDR can handle locked file situations). For SQL, Exchange and Oracle, CDR automatically detects associated folders that should be replicated (logs, databases, etc.).

Flexible Configuration Options

CDR offers flexible configuration options to meet your needs. In a one-to-many configuration, CDR allows the replication of critical datasets from a single central site to multiple disaster recovery locations. Recovery Points can be created on an ad-hoc or scheduled basis providing the ability to bring up a critical application at the disaster recovery site quickly and reliably.

Figure 2 shows one possible configuration for Disaster Recovery. The mail server and database server at the central site are both replicated over the WAN to two separate locations. As application data is replicated to the remote datacenters, Recovery Points are created at specified intervals ensuring an available snapshot that can be immediately brought online in the event of a site disaster.

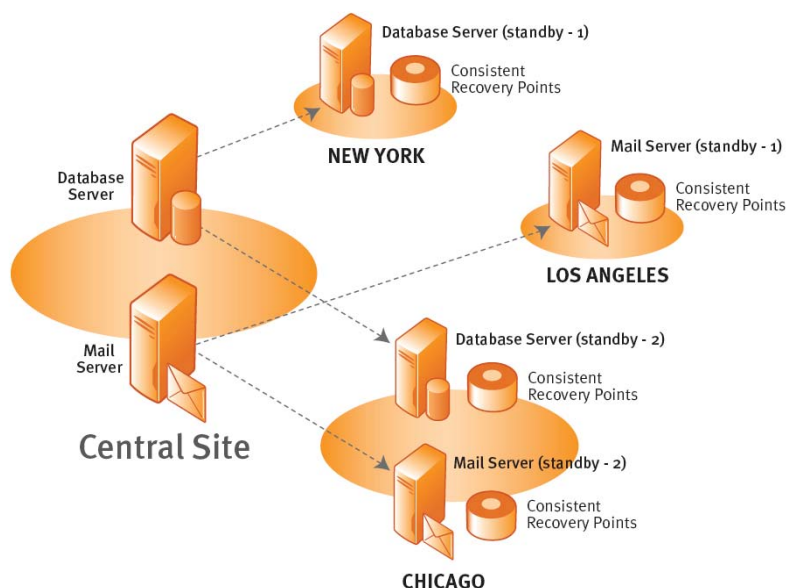


Figure 2. Disaster Recovery Configuration

Continuous Data Protection for Datacenter and Remote Office Environments

When periodic backups just aren't enough, many customers complement their existing data protection strategy with continuous data protection solutions. CDR provides continuous data protection by continuously capturing and replicating changes as they are written. These changes are then sent over the network as the amount of changes reach a size to time threshold selected by the administrator. This not only minimizes bandwidth consumption, it also enables snapshots to be taken at administrator defined intervals. These snapshots represent consistent recovery points that in the event of data or application corruption, can be used to roll back to a previous recovery point, enabling recovery to a specific point in time.

Key Benefits:

- Reduces RTO/RPO helping meet SLA's
- Expedites recovery by providing more points in time to recover from than traditional backup methods
- Reduces potential data loss
- Minimizes network bandwidth consumption

How it works

CDR protects application data and file systems, by replicating data from a source machine to a destination machine in nearly real-time mode. Key data at the source or primary site can be designated as a member of replica set pointed to a target file location across the WAN to the central site. CDR may also be configured for local replication to service data availability scenarios. In either case, replicated data is captured as it is written at the source and all byte-level changes are transferred to the target location where the writes are replicated.

Efficient Data Movement

A replication deployment moves data in two stages:

1. When creating the initial baseline replica or "mirror".
2. When performing continuous, incremental update changes to that file set

A logging mechanism is used to ensure the reliability when transferring incremental changes to the target location. Network throttling can be leveraged to limit the bandwidth that is available to the source machine during working hours. Depending on the data type, data set sizes, and rates of change, a replication scenario may provide a more efficient and cost effective alternative to protecting remote data sources.

Point-in-time Recovery Points

CDR offers a richer degree of total data protection for replicated sources as compared to point replication products in the market. Unlike competitive products that create snapshot images at the source and replicate these potentially very large images over the network, CDR is able to provide a snapshot by leveraging the data that has already been replicated. Once the data has been replicated to the target location, CDR integrates the use of snapshots to create point-in-time recovery points for the data. These recovery points, residing on the destination or target machine, can also serve as content sources for DR backups. This seamlessly integrates the replication configuration directly with your centralized backup strategy. Figure 3 shows the replication process and recovery point creation.

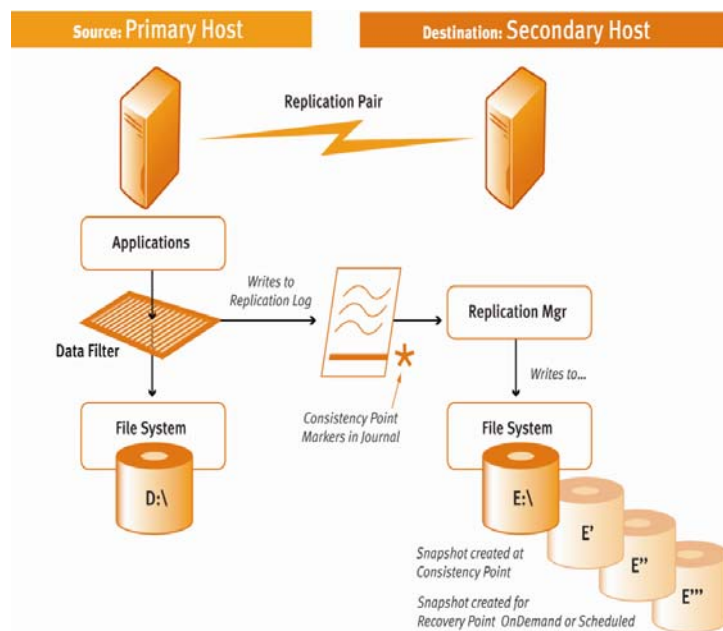


Figure 3. Replication Process and Recovery Point Creation

CDR supports two types of Recovery Point snapshots. In the case of SQL, Exchange or Oracle, application

integration enables consistent recovery points. This guarantees application consistency of the snapshot and extends point-in-time recovery capabilities.

When taking snapshots of applications, it is important that the application data is in a consistent state prior to taking the snapshot. Using consistent recovery points, CDR interacts with the application and signals it to put its data into a consistent state. When the application data is in a consistent state, a marker is placed into the replication journal which signals the point where a consistent snapshot can be taken. When replayed by the replication manager at the destination, a recovery point snapshot of the replica volume is created ensuring application data consistency.

Handling Network Disruptions

The reliability of any remote replication solution is at the mercy of the underlying network. Network interruptions do and will happen. The ability to recover from network outages is critical to the viability of the overall disaster recovery solution. CDR is unique in its ability to quickly and reliably recover from network outages ensuring that a replicated data set is brought back into sync in an efficient manner.

There are two basic network failure scenarios that impact replication:

A minor interruption

A minor interruption can be considered a network disruption that is restored within a few minutes to a few hours. Depending on the duration of the outage and the amount of data being replicated, the replication system can simply continue to capture file changes within the replication log area of the source machine and send the log to the destination when the network connectivity is restored. Like most replication solutions, CDR will continue to log changes on the source machine during a network outage. Similarly, on a target machine, any replication logs will continue to be replayed (written to the replica). Normal replication and application activities occur uninterrupted.

A major interruption

A major interruption is one where the duration of the outage is so long that the captured replicated changes exceed the available log space on the source machine. When this occurs, the replication changes that occurred during the outage are lost along with the integrity of the replica on the target host. One method to re-establish the replica on the target host is to re-initialize the replica by copying all the source data (again) to the target host. For large datasets this can be prohibitive. Because of this, some replication solutions instead initiate a file by file comparison between the source and target file structures. While this method eliminates the need to re-copy all of the data to the replica, it is a time and resource consuming process that increases in time relative to the size of the replica set.

Key Competitive Advantage:

- Automatically re-establish the integrity of a replica

CDR has a smart synchronization mechanism for link recovery that provides the ability to automatically re-establish the replica integrity without the need to perform a full re-sync of the destination replica or endure a time-consuming file-by-file comparison process. Through an innovative mechanism that involves the change journal of the file system, CDR quickly determines which files were modified during a network outage. This is regardless of the size of the replica set and without a time and resource-consuming comparison process. This is an important consideration when working with the size of today's enterprise-class datasets.

Maximizing use of Available Network Bandwidth

CDR provides a number of enterprise-class capabilities that allow for maximum use of available network resources. Whether configured to run within a dedicated corporate network or over the public internet between remote office locations, CDR allows the administrator to control and protect the data in transit.

Compression

When network bandwidth is at a premium, the administrator can choose to enable compression of the replication stream. When enabled at the source machine, CDR compresses all data within the replication log before transferring it to the destination. At the destination, the data is decompressed as it is written to the target volume. The available space savings is highly dependent upon the type of data within the replication stream.

Encryption

When replicating critical business data between remote offices, a dedicated network is not always available or cost effective. In order to protect the privacy of the data when replicating over open networks, the administrator can select the encryption option which uses the blowfish encryption algorithm to encrypt the replication stream at the source. Upon reaching its destination, the replication stream is decrypted as it is written to the replica volume.

Bandwidth Throttling

Network resources are expensive. Rarely is a dedicated network available solely for the replication process. In most operating environments, the network used for replication is also used for conducting normal business operations. For these environments, it is important for the replication processes to co-exist with other resources competing for network bandwidth. CDR allows the administrator to explicitly define how much of the available bandwidth should be allocated to replication activity. Additionally, the bandwidth allocation can be automatically adjusted on an administrator-defined schedule.

A good example might be a typical office environment where the administrator chooses to utilize 90% of available bandwidth for replication during evening hours while using only 40% during business hours.

In this case, Bandwidth Throttling delivers benefits that can be leveraged for increased network productivity. However, it is important to note that the rate of change must be such that the changes that occur in a day can be transferred within that same day otherwise, the target will not be able to catch up with the changes on the source.

Also, if bandwidth is reduced to a percentage where the transfer of replication logs is delayed, additional storage space on the source will be consumed. This may exceed storage space and thus cause replication to stop.

CommVault Technical Services can assist with the configuration and implementation of your CDR environment.

Out-of-Band Synchronization

The overall size of a replication data set is typically larger than the amount of data that changes on a daily basis. Therefore, available network bandwidth sufficient for ongoing replication, might not be sufficient for the initial replica creation – taking multiple days to complete.

CDR simplifies the creation of large replica sets by leveraging a built-in utility that works in conjunction with the replication process. This utility helps capture the state of the initial replica while the administrator uses an alternate, out of band, option to quickly transfer the large initial dataset to the target host. Alternate options might include using a backup set or a disk clone sent via overnight shipping to the remote location. While the data is in transit to the remote location, CDR captures and logs ongoing changes to the source machine. Once the initial replica copy is complete, CDR then transfers the changes to the remote replica and the normal replication process continues from that point.

This capability is also an effective way to perform a fail-back operation in a remote DR configuration. If a primary site fails and the secondary site is used for production, the Out-of-band synchronization can be used to quickly identify and bring back to the primary data that was modified on the secondary site during the outage. This allows the administrator to quickly resume a normal operating environment when recovering from a failure at the primary site.

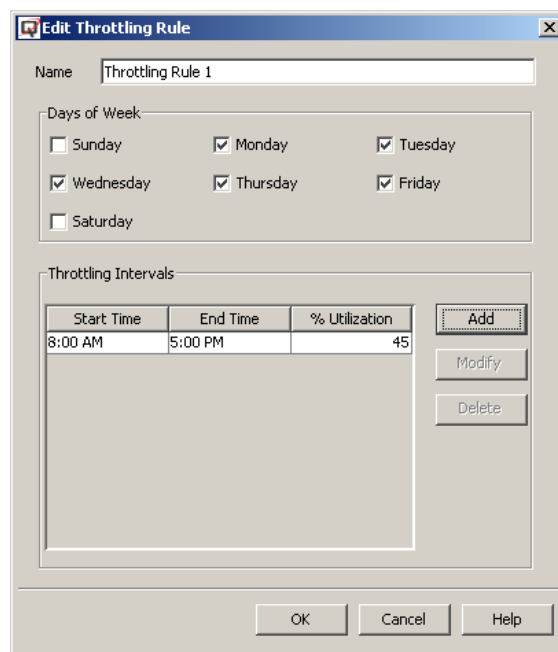


Figure 4. Bandwidth Throttling Interface -shows options administrators can use to control network bandwidth.

Incorporating Capacity Management with Replication

The power of CDR's integration with the CommVault Singular Information Management suite extends beyond backup. When used in conjunction with Data Archiver (DA), CDR can leverage rule-based, capacity management capabilities to easily move large or infrequently accessed files to secondary storage. Data Archiver leaves behind a small stub file of the migrated data. CDR can replicate that stub file without needing to recall the underlying data. After the stub is replicated, it can be accessed normally.

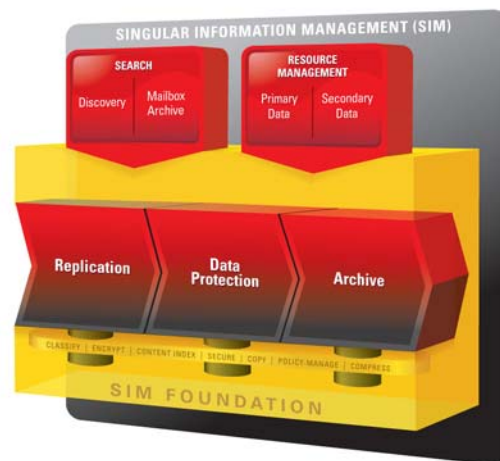
Data Archiver can be deployed at either the replication source or the target. At the source, DA can be deployed to reduce the size of the replication data set. This significantly reduces the time it takes to create the initial replica. At the target, DA can be deployed as a capacity management tool on the replica. This is particularly useful in remote office fan-in environments to manage the overall capacity of multiple remote office data sets from the central location.

Unique Benefits of CommVault Singular Information Management

Continuous Data Replicator is a component of the CommVault Singular Information Management suite, which includes archive management, data protection, recovery management, replication and data resource management. When used in combination, these capabilities provide unique and compelling benefits for managing data – all from a single unified console.

This provides additional granularity that can be used to enhance protection and simplify management. It also reduces the time associated with learning, deploying and managing replication.

Another unique advantage of CommVault integration is the ability to track replicas from a single unified console along with backup sets and snapshots. Integration with CommVault Galaxy Backup and Recovery provides data backup and one step object-level restore of individual files, e-mail messages, folders and mailboxes – all through the same easy to manage graphical user interface console provided with CDR.



CommVault Unified Console	Unification Point	Benefits
	Continuous Data Replicator + Galaxy Backup & Recovery	<ul style="list-style-type: none"> Automatically schedule backup of file system and application recovery points Browse backup set from the perspective of the original client to simplify restore and configuration in complex environments Leverage near real-time data protection to reduce interval of vulnerability helping meet RPO
	Continuous Data Replicator + Data Archiver	<ul style="list-style-type: none"> Extend your datacenter retention and lifecycle management to remote office data Reduce the size of the Exchange Information Store or file system data Extend the life of primary storage devices Improve replication performance Improve restore of Exchange Information Stores in Disaster Recovery Improve speed of file system recovery Reduce backup time and storage space when combined with Galaxy Backup

Summary

CommVault Continuous Data Replicator is a multi-platform replication solution ideal to help centralize and simplify the administration of remote office data, deliver cost-effective disaster recovery and provide continuous data protection for datacenter and remote office environments. It delivers enterprise-class features such as compression, encryption, advanced link recovery, and the ability to synchronize initial replica sets out-of-band. Unlike competitive offerings of point products, CDR can be used as a stand-alone replication solution or as an integrated component of the CommVault Singular Information Management suite. This integration provides the industry's only policy-based GUI from which to easily manage replication along with backup, recovery and archive. It also provides the granularity to track remote backup sets to the source without time consuming manual processes and allows administrators to browse backup recovery points and perform recovery operations as if the backup was created locally.