

Silo Storage Training Reference

Last Reviewed Date: 3/18/2009

Documentation provided by:

CommVault Technical Reference Library

www.commvault.com/library

©1999-2009 CommVault Systems, Inc. All rights reserved.

CommVault, CommVault and logo, the "CV" logo, CommVault Systems, Solving Forward, SIM, Singular Information Management, Simpana, CommVault Galaxy, Unified Data Management, QiNetix, Quick Recovery, QR, CommNet, GridStor, Vault Tracker, InnerVault, QuickSnap, QSnap, Recovery Director, CommServe, CommCell and ROMS are trademarks or registered trademarks of CommVault Systems, Inc. All other third party brands, products, service names, trademarks, or registered service marks are the property of and used to identify the products or services of their respective owners. All specifications are subject to change without notice.

CommVault Technical Reference Library
www.CommVault.com/library

The Information in this Document Is Subject to Change without Notice

Contents

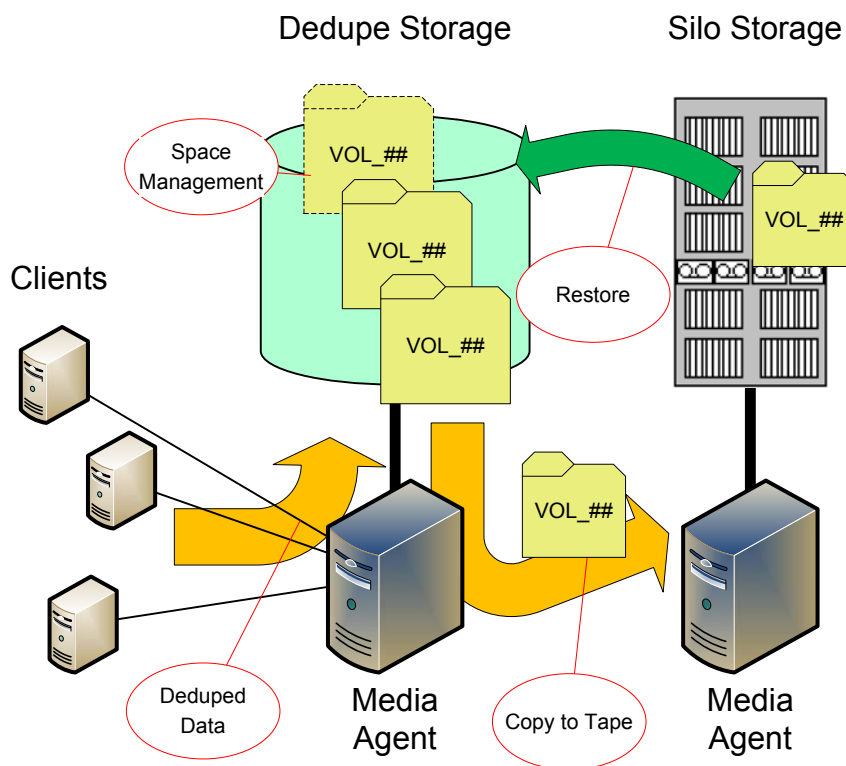
Silo Storage	i
Training Reference	i
Overview	4
Enabling Silo Storage	4
Sealing a Silo Storage	5
Backing up to Silo Storage	5
Space Management	6
<i>Data Aging</i>	7
<i>Data Compression on Silo Storage</i>	7
Auxiliary Copy with Silo Storage	7
Using Silo Storage with a Replica Magnetic Library	8
Silo Restore	8
<i>Restoring From Alternate Secondary Copy</i>	8
<i>Reserving Space for Silo Restores</i>	9
Reports	9
<i>Media Prediction Report</i>	9
<i>Silo Archive Job Summary Report</i>	9
<i>Silo Retrieve Job Summary Report</i>	9

Overview

Silo storage is a feature enabling users to copy or move deduplicated data from magnetic storage to secondary storage devices such as tape media. This enables data to be stored at offsite storage locations while preserving the space savings provided by the deduplicated format.

Silo is a term used to describe the self-contained set of copied or migrated deduplicated data. Users can define the size of data used for a Silo by setting time or size (Default 5 TB) limits. Larger silos provide better space savings through more efficient deduplication.

Silo Storage also provides space management options to handle the space reclamation on the source disk for data that has been moved to Silo storage.



Enabling Silo Storage

Silo storage is enabled in the storage policy copy property's *Deduplication->Silo Options* tab of a storage policy copy. The option is only available when deduplication has been enabled and no data has been written to the storage policy copy.

To complete enabling Silo Storage, a library data path (called a Silo data path) must be added to the storage policy copy. This path will be used to backup Silo data containers. To add a data path, click on the *Add Data Path* control found in the storage policy copy property's *Deduplication->Silo Options* tab. The added data path will appear in the storage policy copy's *Data Paths* tab. Multiple data paths for Silo storage can be added allowing for both the load balancing and failover features of *Gridstor* to be used.

Prerequisites

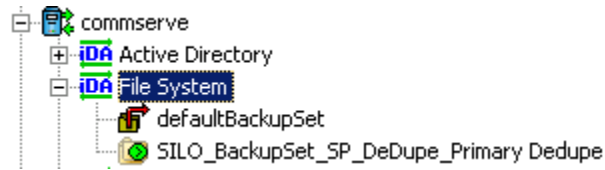
- A Windows File System *iDataAgent* must be installed on the CommServe host.
- Silo Storage can be enabled only on Storage Policy Copies with deduplication enabled and no existing backup data.

License Requirements

While there is no license specifically for Silo Storage, the feature requires the following licenses:

- **Block Level De-Duplication** for deduplication of data on the magnetic media. This license is required on Media Agents hosting the Deduplication Store.
- **Tape De-Duplication** license for providing Silo Storage feature. This license is required on Media Agents hosting the Deduplication Store.
- **GridStor** license for the additional data path required for Silo Storage.
- **FileSystem** license on the CommServe computer for creating Silo Storage.
- **Advanced File System iDataAgent Options** license to store and retrieve silo storage data.

When Silo storage is enabled on a storage policy copy, a new *Silo Storage Set* is automatically created under the *File System iDataAgent* of the CommServe® client.



The Silo Storage Set is essentially an On Demand backup set that acts as proxy management control for Silo Storage. No data is actually moved through the CommServe client. All data movement is between the source copy's Media Agent and the Media Agent managing Silo Storage. These can be the same or different Media Agents on any platform host except Netware.

The Silo Storage Set and default subclient must not be modified other than what is described here. Silo Data is retained until the Silo storage is disabled and deleted.

Once enabled, Silo storage cannot be disabled from a storage policy copy. You can however create a new storage policy without enabling Silo storage and re-point the necessary subclients to that storage policy.

Sealing a Silo Storage

At some point for space management and performance reasons a new Silo needs to be started. Closing the current Silo and starting a new one is called *Sealing*. Automatic sealing can be configured using time and size parameters that can be set in the storage policy copy. Sealing a Silo is also a user task available on the short cut Task menu at the storage policy copy level. Manually creating a silo container rather than waiting for size/time limits may be required for compliance reasons. Note that new Silos will always start on new media. Multiple Silo containers can exist with the most recent Silo container identified as the "Active" container.

Consider the following when configuring the deduplication store:

- Larger deduplication stores provide better deduplication benefits and higher space savings. However, very large deduplication stores increases data dependency within silo storage, and might affect Silo Storage performance.
- Smaller deduplication stores translates to less number of media accessed for data retrieval providing faster restore operations and better manageability of Silo Storage data.
- Deduplication of data is self-contained within the deduplication store so data between two stores cannot be deduplicated.

Backing up to Silo Storage

New deduplicated data written to the source storage policy copy is backed up to Silo storage automatically or through scheduled jobs. Automatic backup occurs for each volume folder when it is closed. The automatic backup of closed volumes occurs within 8 hours by default and this time can be configured in the Control Panel's Media Management parameters. Volume size can also be controlled by

Copy Data Path Candidates

Only tape libraries will appear in the Copy Data Path Candidates list. To allow magnetic library data paths to be used for Silo storage you need to add to the CommServe host registry a DWORD registry value name of HKLM\SOFTWARE\CommVault Systems\Galaxy\InstanceXXX>\MediaManager\AllowMagneticStorageForSilo with data value of 1.

Silo Storage Set

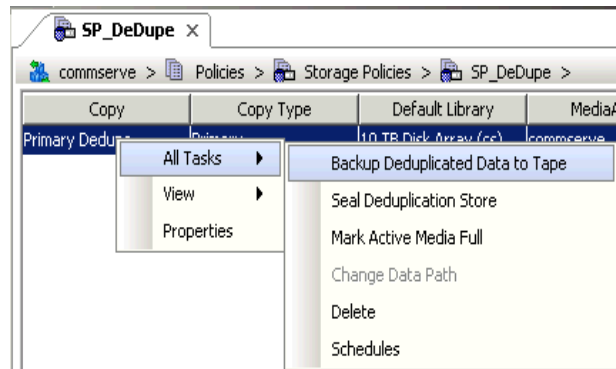
A Silo Storage Set is a special type of On Demand Backup Set managing data for the Silo storage copy. This Silo Storage Set follows the naming convention - SILO_BackupSet_<Storage Policy name>_<Storage Policy Copy name>.

Create new Store every

The size or frequency in which Silo data containers are created can be configured on the *Deduplication->Store Information* tab of the storage policy copy properties page. If both options for time and size are checked, the Silo container will be created whenever the first of either criterion is encountered.

Media Management settings or Data protection job options.

The user can also schedule or execute an immediate backup to silo storage using the *Backup Deduplicated Data to Tape* task that is available at the storage policy copy level. Note that only the Full Backup option is available when configuring a back up of data to Silo Storage. This is because a Silo Storage Set is a special type of On Demand Backup Set and the content of the backup is provided dynamically to each job. As a result of this dynamic content, only new chunks or volumes are backed up with each job. If no new data is found, the backup will still succeed, but no data will be backed up.



Silo backup jobs can be committed when the job is in progress. A **Commit** option completes the backup job by committing the data in the Silo storage as available for restore at that point-in-time. The next Silo job will resume the backup where the previous backup ended. For example, if the job is to transfer 10 volumes to Silo storage, and you chose to Commit when transferring volume 6, the first 5 volumes transferred to Silo storage will be committed, and the job will complete. The next backup job to Silo storage will include any data not transferred by the previous job.

Space Management

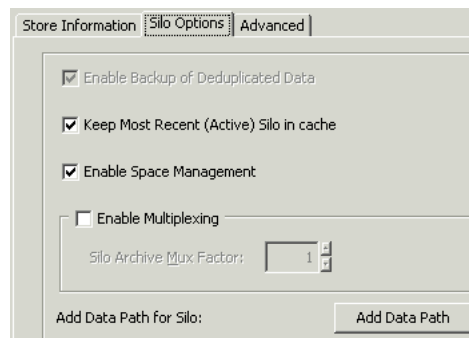
Silo Storage also provides space management options to handle the space reclamation on the source disk for data that has been copied to or retrieved from Silo Storage. The *Enable Space Management* option is checked by default and should only be unchecked to preserve data on the source disk during troubleshooting.

When space management is enabled, data that has been backed up to Silo storage will be removed from magnetic storage using the same space management threshold settings used for the magnetic library's *Managed Disk Space* option. Frequency of space management can be set using the ComMCell Console Control Panel's Media Management Configuration Parameter – *Interval (in minutes) to free the Silo Cache*.

The removal of data on disk is based on the data's restore access frequency using a LRU (Least Recently Used) algorithm.

Removing Silo stored data from disk allows more efficient use of the disk space. However, the response time for a restore request requiring data from a Silo container no longer on disk will be slower. For faster restore response, the option to *Keep Most Recent (Active) Silo in cache* is checked by default.

Additionally, if the storage policy copy data is required for other operations such as Auxiliary Copy, Data



Silo Data Encryption

Silo Storage data can be encrypted for transmission over networks and for storage on media. Enable Encryption on the properties for the CommServe Client and then select the encryption location on the Silo Storage Set default subclient.

Silo Data Multiplexing

If the source copy has multiple data streams, you can use the data multiplexing feature to combine the streams during the Silo Storage backup operation. The number of data streams that are to be backed up concurrently to the same media is called the mux factor.

Tape Usage

Migration of Silos to tape by default uses a new tape for the start of each stream in the silo. This configuration might result in some unused tape storage, but provides clean boundaries between Silos, reducing tape recall dependencies.

Spool Copy

Spool copies are not allowed from Storage policy Copies that have been enabled for Silo Storage.

Verification, or Offline Content Indexing, then space management does not remove the data until the dependency has been satisfied.

During a restore operation, if the backup data is available in the source copy, it is restored. If the backup data has been moved to Silo Storage, then the necessary volumes are automatically restored to any mount path on the source copy disk that has sufficient space and the data made available to complete the initial restore operation. Space Management will subsequently remove this staged data.

Data Aging

When an active Silo store has been sealed and moved to storage, all the backup jobs that went to that store must meet the retention rules (defined by the source storage policy copy) for the store to become aged (prunable). Once all of the jobs have met their retention criteria, the entire store is considered prunable, and the Silo (tape) backup jobs are then aged. The tape designated for the Silo storage is then refreshed and available for re-use.

Data Compression on Silo Storage

Deduplicated data on the source copy is by default stored in compressed format but can also be stored in original form. If compressed, then data copied to Silo Storage will also be compressed effectively increasing the capacity of Silo Storage. If the source copy data is in original form, you can enable either software or hardware compression of data from the source copy. Hardware compression is enabled on the data path and overrides any software compression setting. Software compression for Silo storage is enabled in the properties of the Silo Storage Set subclient.

Auxiliary Copy with Silo Storage

With protected data moved into Silo storage we strongly recommend you maintain another copy for disaster recovery purposes. There are two methods for configuring another secondary copy. The first method retains the deduplicated storage environment. This is enabled by creating a secondary copy and using the storage policy copy's *Association* tab to associate that copy only with the Silo Storage Set created under the File System iDataAgent on the CommServe Client. As the copy is associated only with the deduplicated Silo Storage set, it copies the silo storage data directly, thus preserving data deduplication.

The second method does not retain the deduplicated storage environment. This method is the same as the first except it requires that the secondary copy NOT be associated directly with the Silo Storage Set. Any or all other associations can be used.

If the secondary copy was created after the data was already moved to silo tape, the auxiliary copy operation will bring back the data to the magnetic media, unravel, and complete the copy operation.

Unlike the initial movement to tape, there is no automatic auxiliary copy associated with Silo Storage. An Auxiliary copy job must be scheduled or executed. Data on the source copy cannot be deleted if there is data waiting for an auxiliary copy job.

Data Retention

Since the data in the deduplicated source copy, deduplicated Silo Storage Set, and a deduplicated secondary copy all follow the same deduplication store they must all have same data retention criterion. An expanded (not duplicated) secondary copy can have different retention.

Using List Media

Before performing Silo restore operations, you can use the List Media (Media Prediction) options to determine the Silo Storage media that is required. The list media prediction must be performed on the Silo Storage Set under the File System iDataAgent of the CommServe host. Using List Media is particularly useful when you have Silo storage in offsite locations.

Data Verification

Data verification will automatically restore any necessary volumes or data from Silo Storage to complete the Data Verification job. Space Management is responsible for removing the restored data.

Using Silo Storage with a Replica Magnetic Library

A Replica Magnetic library uses replication to synchronize a primary read/write mount path on one Media Agent with a read-only mount path on another Media Agent. This configuration is primarily used in a remote office situation where a local (source) copy of the data is necessary for fast backup and restore and another (replica) copy is maintained as a centralized site for disaster recovery or other additional data management action.

A tape library data path enabled for Silo Storage that is accessible (LAN-free) from the Media Agent managing the magnetic library's replica copy will use that copy as its source. Automated restores from Silo Storage will always go to the source copy – not the replica. If replication is enabled, a volume restored to the source copy from Silo storage will be replicated to the replica copy.

Silo Restore

Storage policy copy data enabled with Silo storage can be restored using a *Restore By Job* task or by using the normal *Browse Backup Data* method. Browse the contents of the storage policy copy and seamlessly restore the desired data from the source copy or the Silo storage.

When a copy with a Silo backup is restored, it follows one of the following sequences:

- If the backup data is available in the source magnetic media, it is a normal restore operation. The restore operation fetches the data from the magnetic media and restores it to the restore location.
- If the backup data is not available in the source copy storage, then it performs a Silo restore which happens in two phases. This Silo two phase restore is an entirely seamless operation, and does not require user intervention. First, the original restore operation goes to Pending state, and kicks off another job called the Silo Retrieval job to retrieve data from Silo storage to the source copy storage. Once the data is available in the source copy, the original restore operation automatically resumes restore of the data to the restore location. The Job Controller displays the Silo restore as well as the original restore operation.

Restoring From Alternate Secondary Copy

When data is restored from the Silo enabled storage policy copy data, and the data to be restored is moved to silo tape, the copy operation brings back the data from the Silo tape storage. If for any reason the Silo tape is not available, the storage policy copy can be configured to restore data from another secondary copy. To enable restores from a copy other than primary and initial Silo storage copy, you must modify the copy precedence for the storage policy. Identify the secondary copy to want to restore from and move it to the first precedence. This setting enables the recovery operations to restore data from the secondary copy, instead of the original silo tape copy.

Best Practices

The following best practices apply when using Silo Storage.

- If possible, use a local path for the source magnetic mount path.
- Do not uncheck Space Management option unless there is a need to preserve data on the source copy during troubleshooting or if directed to do so by support.
- Maintain an updated copy of Silo Storage offsite for disaster recovery.
- Before performing large restores using data in Silo Storage ensure sufficient space exists in the source copy magnetic library.
- Keep the Active Silo on source copy storage for best restore performance and if Data Verification or Offline Content Indexing jobs will require access to the data.
- Use the default option settings and parameters for best performance and support. Consult with your Customer Engineer or Support before changing any option.

Reserving Space for Silo Restores

When data from Silo Storage is restored, the Silo is retrieved to the primary media. By default, the silo is retrieved to the original mount path from which the data was migrated to silo storage. However, during the restore if the original mount path does not have enough space, the Silo is automatically copied to any other mount paths with sufficient space, available in the library.

To guarantee successful Silo restore operations, ensure that a mount path with sufficient space is available in the library at all times. It is recommended that the storage on this mount path is equivalent to the size of the average Silo in the Silo Storage.

Reports

Media Prediction Report

Media prediction report can be used to identify the backup tapes that are required for the Silo restore jobs.

Silo Archive Job Summary Report

Silo Archive Job Summary provides job summary details of Silo archive jobs moving backup data from magnetic media to Silo Storage.

Silo Retrieve Job Summary Report

Silo Retrieve Job Summary provides job summary details of retrieve job operations recalling data from Silo Storage.