Abstract

This white paper describes existing and new replication capabilities available in EMC® CentraStar® 4.0. It covers all replication topologies and use cases.

July 2008
# Table of Contents

**Executive summary** .................................................................................................................. 5

**Introduction** ............................................................................................................................... 5
  - Audience ..................................................................................................................................... 5
  - Terminology ................................................................................................................................. 6

**Introduction to EMC Centera Replication** ............................................................................... 7

**Theory of operation** .................................................................................................................. 8
  - Replication .................................................................................................................................. 8
  - Restore ........................................................................................................................................ 9

**Replication topologies** ............................................................................................................. 10
  - Unidirectional replication ........................................................................................................... 10
  - Bidirectional replication ............................................................................................................ 10
  - Chain replication ....................................................................................................................... 12
  - Star replication ........................................................................................................................... 13
  - Combining star and chain topologies ......................................................................................... 14
  - Supported number of clusters in chain or star setups ............................................................... 14

**Replication and virtual pools** .................................................................................................... 14
  - Replication and restore ............................................................................................................. 14
  - Pool names and pool IDs ............................................................................................................ 15
  - Only one target cluster can be defined .................................................................................... 16
  - Replicating to a non-pool aware EMC Centera cluster ............................................................. 16

**Modes of operation** .................................................................................................................. 16
  - Normal operation ....................................................................................................................... 16
  - Failover operation ..................................................................................................................... 17
  - Restore operation ..................................................................................................................... 18

**Replication of deletes** .............................................................................................................. 19
  - Delete immediate after ingest .................................................................................................... 20

**Application failover** .................................................................................................................. 21
  - The operational failover defaults .............................................................................................. 21

**Replication interoperability matrix** .......................................................................................... 22

**Network segmentation** ............................................................................................................. 23

**Conclusion** .................................................................................................................................. 24

**References** .................................................................................................................................. 24
Figures

Figure 1. Basic Replication.................................................................................................................. 7
Figure 2. How replication works .......................................................................................................... 8
Figure 3. Replication failover ............................................................................................................. 10
Figure 4. Bidirectional replication with hot standby ........................................................................... 11
Figure 5. Bidirectional replication with different applications .......................................................... 11
Figure 6. Chain replication topology .................................................................................................. 12
Figure 7. Use case for pools being replicated using chain replication ............................................... 13
Figure 8. Star replication topology ...................................................................................................... 13
Figure 9. Combining star and chain replication topology ................................................................. 14
Figure 10. Replication of specific pools ................................................................................................. 15
Figure 11. Outgoing star replication is not supported ........................................................................... 16
Figure 12. Normal operation ................................................................................................................ 17
Figure 13. Failover operation ............................................................................................................... 18
Figure 14. Restore operation ................................................................................................................. 19

Tables

Table 1. Timestamp changes ............................................................................................................... 9
Table 2. Default SDK failover behavior ............................................................................................... 21
Table 3. Replicate and restore between CentraStar versions ............................................................... 22
Executive summary

EMC® Centera® Replication provides a disaster recovery mechanism for content written to EMC Centera clusters. EMC Centera Replication can be used to create multiple protected copies of content written to a primary Centera cluster by copying the content to a replica Centera cluster. Replication runs as an asynchronous background task and can be configured in a number of topologies, namely unidirectional, bidirectional, chain, star, and star and chain. These combinations provide the replication topology that best fits customer requirements.

EMC Centera Replication provides functionality that allows the recovery of content that is either missing or simply unavailable. Applications will fail over automatically to retrieve missing or unavailable content on the primary cluster from the replica cluster.

Introduction

This white paper introduces readers to EMC Centera Replication and Restore by defining replication and restore, the various topologies that can be configured, how replication works with virtual pools, how failover and restore function, and how applications may fail over in the event of an outage.

This white paper does not discuss network bandwidth sizing, configuration, monitoring, or managing the EMC Centera Replication process.

Audience

The intended audience for this paper is customers, including storage architects and administrators, and any others involved in evaluating, acquiring, managing, operating, or designing an EMC networked storage environment. It can also be used for guidance and development of proposals for EMC staff and partners.
Terminology

- **Centera Cluster**: One or more cabinets on which the nodes are clustered. Clustered nodes are automatically aware of nodes that attach to and detach from the cluster.
- **Primary Cluster**: The EMC Centera cluster that the application writes to.
- **Replica Cluster**: The EMC Centera cluster that is the target of the replication process from the Primary cluster.
- **Centera Pool**: Virtual pools were introduced in CentraStar 3.0 mainly to enable system administrators to segregate data into logical groups and provide more granular access control to applications accessing the same cluster.
- **Node**: Logically, a network entity that is uniquely identified through a system ID, IP address, and port. Physically, a node is a computer system that is part of the EMC Centera cluster.
- **Node with the access role**: The nodes in a cluster that communicate with the outside world. They must have public IP addresses. These nodes are known as access nodes.
- **Node with the storage role**: The nodes in the cluster where application content is written to and protected by the access nodes.
- **Content Address**: A data object’s unique identifier. A Content Address is the claim ticket that is returned to the client application when an object is stored to the archive.
- **Centera API / Centera SDK**: The EMC Centera SDK is a set of cross-platform application programming interfaces (APIs) that simplify functions for customer applications such as store, retrieve, delete, and query for data objects. The EMC Centera SDK does this in a variety of flexible and powerful ways. All applications must use this API to read and write to EMC Centera.
- **Access Profile**: Access profiles are used by applications and users of management tools to authenticate to a cluster, and also by clusters to authenticate to another cluster for replication or restore connections. System administrators can create access profiles using the CLI. Each access profile consists of a profile name, a secret (password), and a set of capabilities and roles.
- **Centera Capabilities**: Pool-bound rights granted by the system administrator to an access profile. They determine which operations an application can perform on the pool data. Possible capabilities are write (w), read (r), delete (d), exist (e), privileged delete (D), query (q), clip copy (c), purge (p), and litigation hold (h).
- **Centera CLI/ Centera CV**: The CLI is a command line interface for system administrators that enables them to manage and monitor EMC Centera.
- **CentraStar**: EMC firmware used by EMC Centera.
Introduction to EMC Centera Replication

Replication is the process whereby an EMC Centera cluster automatically copies new content to another EMC Centera cluster. As an EMC Centera cluster acquires new content from a local application, the replication mechanism ensures that this new content is automatically and transparently transferred across a WAN to a designated EMC Centera cluster presumably in another location.

Restore is a process complementary to replication, whereby an EMC Centera cluster copies some set of its content to another Centera cluster. Replication is used on an ongoing basis to keep two EMC Centera clusters synchronized with new content, whereas restore is used only as needed to populate one EMC Centera cluster with the content of another Centera.

In a typical replication setup the EMC Centera clusters are geographically separate to ensure disaster recovery or to distribute the content for access from another location. For example, a company may replicate to a second EMC Centera cluster to enable recovery from the loss of the primary Centera or to avoid multiple requests for the same content across a WAN connection.

Figure 1. Basic EMC Centera Replication

Unlike synchronous utilities such as EMC’s SRDF® or MirrorView®, which work to keep databases synchronized across different locations, EMC Centera Replication operates as an asynchronous process.

The EMC Centera Replication and Restore processes operate in a similar manner, but with an important difference: While replication is an ongoing process, the duration of the restore process is bounded by the specified selection to copy some set of the contents of one EMC Centera cluster to another. Once the restore process has successfully copied the selected content, the process will stop. In other words, replication will copy newly added content to the target system; restore will copy already existing content to the target system.

Before we proceed, we assume in this white paper that the reader has a basic understanding of virtual pools. Please refer to the appropriate EMC Centera documentation regarding virtual pools for a detailed description of virtual pools and how to configure them.
Theory of operation

Replication
How does replication work inside the box? Once replication is enabled, each storage node in the cluster is responsible for replicating newly written C-Clips. When a new C-Clip is mirrored on node A and B, both nodes independently attempt to replicate that C-Clip. Each node will thus keep track of its own progress.

Replication iterates on the store timestamp (ST) index, submits encountered C-Clips, and moves its checkpoint after successful replications. As a storage node cannot copy a C-Clip by itself, C-Clip addresses are sent to a replicating node that will do the actual copy of the C-Clip to the target cluster. The result of such a copy is then returned to the storage node.

Besides new C-Clips being written to the storage node, there are two other instances in which C-Clips may be replicated:

- C-Clips found in the replication parking that are checked once every 24 hours (or on-demand)
- Entries found in “folderlog”, a list specified by EMC Service personnel

Because C-Clips are mirrored, we need a way to avoid duplicate replications as much as possible. Therefore, replication requests are not directly sent to the access nodes. Instead each request is routed through a “concentrator” node. This node combines the replication entries from each individual storage node into a distinct list of C-Clips that need to be replicated. Making the list distinct avoids sending duplicates to the target cluster.

There is only one active concentrator service found in a single cluster. The order in which the node is selected for this service is based on the following preference: first spare, then storage only, and finally one with an access role.

Before revision 4.0 of CentraStar, only access nodes within the cluster would participate in replicating clips to the replica cluster. With revision 4.0, the storage administrator can choose which nodes in the cluster
will participate in replication. Batches of C-Clips will be distributed from the concentrator to each node with an access role. Replication requests that arrive on a replicating node are dispatched to one of the replication threads on the node. By default, there are 20 threads per replicating node. That thread will first copy over all BLOBs referenced in the C-Clip, and then the C-Clip itself. A replication is successful if all BLOBs and the C-Clip itself were copied over without errors.

EMC Global Service can change the number of threads to control the bandwidth used.

Replication processes use the store timestamp (ST) to order the C-Clips in a “best effort” FIFO (first in, first out) manner. The ST is the time when the C-Clip was actually stored on the node and is only used inside the Centera cluster.

To place ST into context, the ST can be different from the write timestamp (WT) and the creation time found in the CDF. Both ST and WT are internal to the cluster. The CDF creation time is the most visible time to the application and the user. The creation time found in the CDF is calculated by the EMC Centera SDK. The WT is calculated by the node with the access role when the C-Clips are written to the cluster. This WT will stay the same for the life of the C-Clip inside the cluster, even when it is (re)-protected. The ST is updated every time the C-Clip is written.

### Table 1. Timestamp changes

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Timestamp changes when</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDF creation time</td>
<td>The SDK creates or copies the CDF.</td>
</tr>
<tr>
<td>Write timestamp (WT)</td>
<td>C-Clips are written to the cluster by the SDK, Replication, or Restore.</td>
</tr>
<tr>
<td>Store timestamp (ST)</td>
<td>The cluster writes (copies of) the C-Clips to storage nodes. This includes writes done by the cluster’s self-healing processes.</td>
</tr>
</tbody>
</table>

**Restore**

How is EMC Centera Restore different from Replication? A restore process has much in common with replication in that it also copies C-Clips from a source to a target cluster, shares many common software components, and also has 20 threads assigned per replicating node. However, there are some key differences that have an impact on reporting.

The first major difference is the process trigger. Where replication is a continuous process, triggered by new entries in the replication queue, restore is an on-demand process, triggered by a full or partial restore request from the system administrator.

Another key difference is that replication processes C-Clips based on the store-timestamp while a restore processes all C-Clips whose write-timestamp matches the restore period. Since CentraStar 3.1.2, the restore checkpoint is centralized and maintained by the concentrator node.

To ensure that a restore starts restoring and protecting data as soon as possible, each storage node starts iterating its internal indexes and pushing the C-Clips whose write-timestamp matches the restore period to the nodes with restore roles. It does not do a full iteration to determine the queue.

In CentraStar 3.1.2, restore implementation has changed to address some of the shortcomings that were reported, such as progress made when nodes and/or disks are offline. This change resulted in a three-phase restore process, where before there were only two. These three phases are:

- Restoring available C-Clips and associated BLOBs that do not exist on the target
- Process any C-Clips that may have been missed due to offline nodes and/or disks
- Process any failed C-Clips in the parking until the retry count is exhausted
Replication topologies

Unidirectional replication
The most basic form of replication is unidirectional replication between two EMC Centera clusters. The application writes data to cluster A and that data is automatically replicated to cluster B.

Unidirectional replication provides disaster recovery (DR) capabilities where application(s) may write to a single EMC Centera cluster and automatically create an (online) copy of the data at a remote site (active/passive configuration).

In case of a disaster or when the primary EMC Centera cluster becomes unavailable, the application may fail over to the replica cluster B. Automatic read failover is a feature of the EMC Centera SDK and is enabled by default, but other failover options may be configured (Figure 3).

Data written to cluster B during a disaster needs to be restored to cluster A once it is back online. If cluster A was lost during the disaster, all lost data needs to be restored from cluster B using the EMC Centera Restore feature.

Bidirectional replication
The bidirectional replication topology available to EMC Centera allows users to write to either of the EMC Centera clusters and have data copied to the other Centera cluster. Data written to cluster A will be replicated to cluster B and visa versa.

This allows for establishing a complete DR solution whereby the application servers also are redundant. When the complete site A is lost, site B can take over immediately (hot standby, Figure 4).

Data written to cluster B by application 2 during a disaster will be held in a replication queue and replicated to site A as soon as this site comes back online. If site A was lost during the disaster, all lost data must be restored from cluster B using the EMC Centera Restore capability.
Figure 4. Bidirectional replication with hot standby

Rather than using site 2 as a hot standby, both sites can be used for production. This is useful when both sites have users who work on the local application server. All data will be kept available at both sites using the bidirectional application.

Note: Although from an EMC Centera perspective all data can be shared between sites, each local application may not be aware of data created by the other application. Application databases may need to be synchronized.

The applications used at both sites can also serve completely different purposes (Figure 5). Bidirectional replication provides users with a copy of the data on both sites for all applications at either location. In this situation bidirectional replication is a useful setup for companies that have different operations at each location, but still require replication of each application’s data.

Figure 5. Bidirectional replication with different applications
**Chain replication**
When a customer requires more than two copies of data, chain replication can provide extra protection.

EMC Centera supports chain replication as of CentraStar 3.0 (Figure 6). Chain replication allows an EMC Centera cluster to replicate not only data written by the application, but also data replicated to it by another Centera cluster. Currently three EMC Centera clusters can be configured in a chain. Chain replication is unidirectional. It is possible to do bidirectional replication between the last two EMC Centera clusters in a chain\(^1\), but then the cluster loses its chain replication behavior and will resemble star replication. C-Clips that are replicated from A to B can then no longer be replicated to C.

---

**Figure 6. Chain replication topology**

When the customer requires two clusters at the primary site, for example, for application access and failover, chain replication allows replication to a third EMC Centera cluster installed at a remote location to meet DR requirements.

This topology can also be appealing for companies that have more than two locations of production servers and require a copy of the data at all sites (Figure 7). The clusters are then configured in a ring and pools are used to avoid replicating in a ring. Circular replication is not allowed.

---

\(^1\) The reason for this is that a single cluster can only have one replication target.
Star replication

Star replication topology was introduced in CentraStar 3.0 (Figure 8). Star replication allows an EMC Centera cluster to receive replicated content from up to three Centera clusters. It is, however, not possible to replicate “out” to more than one Centera cluster. This topology is therefore also referred to as “incoming” star.

Figure 8. Star replication topology

With star replication, typically smaller “edge” EMC Centera clusters replicate to a larger “center” Centera cluster. The “edge” Centera clusters hold only local data while the “center” then holds all replicated data.
Companies that have several smaller sites can use this star topology to replicate all data to the main site. In order to allow for site-selective restore in case of a disaster, the use of virtual pools is required.

**Combining star and chain topologies**

Although the star topology provides DR attributes by having data both on the “edge” and “center” EMC Centera clusters, this may not be sufficient in all cases. One example is when the site with the “center” Centera cluster also has applications writing directly to the “center” Centera cluster. This data needs to be replicated as well to meet the customer’s DR requirements. In such cases, star and chain are combined (Figure 9).

![Combining star and chain replication topology](image)

Figure 9. Combining star and chain replication topology

It is possible to replicate only the application data written to the “center” EMC Centera cluster or to also include the replicated data from the “edge” Centera clusters. The latter ensures that all data from clusters A, B, C, and D are replicated to the target (DR) Centera cluster and will simplify the restore of a “center” Centera cluster to some extent. In order to allow for site-selective restore in case of a disaster, the use of virtual pools is required. Restore to an “edge” Centera could happen from either the “center” Centera or the target Centera (assuming that the data from the “edge” Centera has been replicated to the target Centera).

**Supported number of clusters in chain or star setups**

EMC Centera chain and star replication topologies are qualified for a limited number of clusters. The maximum qualified clusters for chain is three and with star there can be up to three “edge” clusters in addition to the “center” cluster.

**Replication and virtual pools**

**Replication and restore**

With the introduction of pools in CentraStar 3.0, users are no longer required to replicate all content that is written to an EMC Centera cluster. When replication is set up, users will be prompted to specify which
pools need to be replicated (Figure 10). Replicating all content written to the EMC Centera cluster may still be selected.

Pools that are replicated must also exist on the EMC Centera target cluster. Any pool mappings defined for the replicated pools on the EMC Centera source cluster must also be configured on the target. If these requirements are not met, replication will auto-pause as soon as an attempt is made to replicate a C-Clip from a virtual pool that does not have a matching virtual pool on the replica cluster.

![Figure 10. Replication of specific pools](image)

Restoring the content places it back in the same pool in which the C-Clip was found. The pool must exist on the cluster that the C-Clip is restored to. If a mapping is required to place it in the same pool, that must then exist too. If this requirement is not met, restore will auto-pause.

**Pool names and pool IDs**

Application pools have a global unique identifier (pool ID). The pool name is just a human readable label and by no means identifies the same pool on another EMC Centera cluster. For replication, it isn’t sufficient to create the pool with the same name on the target Centera cluster. The pool configuration must be copied to the target EMC Centera cluster to create a pool with the same pool ID. This is done by exporting the relevant pool configuration and then importing this into the target Centera cluster.

During import, if a pool ID already exists, it is considered the same pool. There is an option to update the name of the pool with what the import file specifies if they differ. For the EMC Centera cluster, the pool ID is always used as ruling reference. The pool name is only used for user reference. The following prompt is presented when the pool ID already exists:

```
A pool with this ID already exists on the cluster.
Replace existing pool name (X) with imported name (Y)?
```

When a pool name already exists, but the pool ID in the import file is different, it is considered a different pool. During import the user will be asked for a different name for the pool definition that is imported:

```
A pool with this name already exists on the cluster.
Import the pool with another name?
```
**Only one target cluster can be defined**

Outgoing star topology is not supported. EMC Centera Replication can only address one Centera cluster. All pools selected will be replicated to that single EMC Centera cluster. It is not possible to have one set of pools replicate to cluster A and another set of pools to cluster B at the same time (Figure 11).

![Figure 11. Outgoing star replication is not supported](image)

**Replicating to a non-pool aware EMC Centera cluster**

EMC Centera cluster(s) that do not support pools will store all received replicated content without any data segregation. However, since the C-Clip is immutable, the original pool and profile information is persisted. This information will be used again when users restore or replicate to a pool-aware EMC Centera cluster.

Be aware that a non-pool-aware EMC Centera cluster has no ability to do a pool selective restore or replication. This can limit the options for a replication or DR setup with chain and star topologies. For example, a non-pool-aware EMC Centera cluster as a “center” in a star replication setup will not allow users to restore an “edge” with only the data that was originally on this EMC Centera cluster.

**Modes of operation**

In this section, we explain the different modes of operation in an EMC Centera Replication setup. In the following examples, we use unidirectional replication for simplicity.

**Normal operation**

Application 1 connects to Centera cluster A to store and retrieve content. Cluster A stores the content and queues all stored content for replication to cluster B (Figure 12).
In the event the connection to the target cluster is lost or there is a loss of the target cluster itself, the source cluster will queue the replication entries. The replication queue will continue to grow as the application continues to store content on the primary cluster. In this scenario, there is no danger of the replication queue reaching a maximum size where it could no longer add content for later replication. The replication queue is only limited by the available capacity.

A second application 2 may be co-located with Centera cluster B that can be activated in case of the catastrophic loss of application 1 and cluster A. However, the application databases need to be synchronized to be able to address all content.

![Diagram of normal operation](image)

**Figure 12. Normal operation**

**Failover operation**

If a power or network failure occurs that affects cluster A and application 1 cannot write to or read from cluster A, the application can use cluster B (Figure 13).

Connection failover must be configured by adding the target cluster to the connection string. Once connected, the default failover behavior is that application 1 will automatically fail over read operations (including query and exist) to cluster B. Write and delete operations will not automatically fail over. For more information on the different failover types see page 21.

Note that content on cluster A that was not yet replicated to cluster B will not be available to the application when reading from cluster B.

When cluster A comes back online, application 1 will try to reconnect to cluster A and resume normal operation again.

Content that was written to cluster B while cluster A was offline needs to be restored to A using restore if replication was unidirectional.

In the event cluster A is lost due to a disaster or other event, application 1 may be lost as well. If application 2 is available at the target site, then it can be activated and configured to connect to cluster B while application 1 and cluster A are rebuilt. After the rebuild is completed, cluster A can resume its responsibilities again while lost data is restored from cluster B.

It is advisable to hold off taking cluster A back into production as the primary cluster if the application relies heavily on read operations. Read failover for content not yet restored may delay read operation. We
recommend restoring most urgently needed content first if possible.

Figure 13. Failover operation

**Restore operation**

The content that was replicated from Centera cluster A to cluster B, plus the content that was written to cluster B by application 1 during repair, may need to be restored to cluster A after the outage of cluster A is resolved (Figure 14). Content on cluster A that was not yet replicated to cluster B may not be recoverable.

If cluster A is replaced by a new cluster, for example, in the case of a catastrophic loss of the data center, all of the content from cluster B will need to be restored to the new cluster A. Before you can start to restore, the new Centera must be correctly configured, including setting up pools, mappings, and profiles required for the content that will be restored.

Restore is used to rebuild the cluster A content from cluster B. The restore process is initiated via the CLI. During the restore of content to cluster A, cluster B can also replicate the new content written by the application. Once the rebuild of cluster A is complete, the normal configuration can be restored.

Alternatively, normal application 1 operation can be resumed before restore completes in case the application does not rely heavily on read operations. To make such a decision, users will need to consider all WAN traffic constraints and latencies associated with failover reads, restoring content and replication of new content. Knowing the application operations profile (for example, reads vs. writes, file size, files/second) will be a significant help to optimize and expedite a user’s restore. If possible, restoring most urgently needed content first is advisable.
Figure 14. Restore operation

Replication of deletes

EMC Centera keeps a trace of all deletes performed on the primary (source) cluster using a subset of information from the CDF called a reflection. When a C-Clip is deleted, the CDF copies of the C-Clip are physically deleted but a pair of reflections is created containing audit information about the deleted C-Clip.

Reflections have the same Content Address as the deleted C-Clips. Reflections can be queried for reporting and backup applications.

With replication, a delete can be replicated too. When this option is selected, a delete request will be replicated and will delete the same C-Clip on the target cluster. Note that this is not the same as replication of the reflection.

As we assume this to be the reflection on the source:

```
<reflection version="1.0">
  <meta name="principal" value="anonymous"/>
  <meta name="incomingip" value="10.68.20.12"/>
  <meta name="creation.date" value="2005.06.16 23:27 GMT"/>
  <meta name="deletedsize" value="1024830"/>
  <meta name="reason" value="DeleteClip"/>
  <meta name="clienttype" value="user"/>
  <meta name="options" value="0"/>
  <eclipdescription>
    <meta name="type" value="Standard"/>
    <meta name="name" value="C:\temp\myfile.txt"/>
    <meta name="creation.date" value="2005.03.14 5:31 GMT"/>
    <meta name="modification.date" value="2005.03.14 GMT"/>
    <meta name="numfiles" value="1"/>
    <meta name="totalsize" value="1024000"/>
    <meta name="refid" value="8O4H40M85MVCJ46IGBBC8"/>
    <meta name="prev.clip" value=""/>
    <meta name="clip.naming.scheme" value="MD5"/>
    <meta name="numtags" value="1"/>
    <meta name="retention.period" value="0"/>
  </eclipdescription>
</reflection>
```

Sample 1. Reflection on the source cluster
Then this is the reflection as a result of a replicated delete:

```xml
<reflection version="1.0">
  <meta name="principal" value="anonymous"/>
  <meta name="incomingip" value="10.68.10.12"/>
  <meta name="creation.date" value="2005.06.20 23:44 GMT"/>
  <meta name="deletedsize" value="1024830"/>
  <meta name="clienttype" value="replication"/>
  <meta name="reason" value="DeleteClip"/>
  <meta name="options" value="0"/>
  <eclipdescription>
  *** DESCRIPTION COPIED FROM ORIGINAL REFLECTION ***
  </eclipdescription>
</reflection>
```

Sample 2. Reflection on the target cluster

**Delete immediate after ingest**

When a C-Clip is deleted immediately after the write on the source cluster before the cluster had a chance
to replicate the original C-Clip to the target cluster, then there will be no reflection on the target cluster.
Application failover

EMC Centera integrated applications use the EMC Centera SDK to connect to a Centera cluster and perform actions such as write, read, delete, and so on. The EMC Centera SDK function that enables this connection is FPPool_Open().

Failover is the process that the SDK uses to handle errors in a multiple cluster environment. There are two types of failover:

- **Connection failover**
  Occurs when establishing a connection. Update the connection string to include the target cluster if you want to automatically fail over when the source cluster becomes unavailable.

- **Operational failover**
  Occurs during an operation on an open connection. In order to change the default failover behavior the application must call Centera SDK FPPool_SetGlobalOption() with the appropriate parameters or set the appropriate options using environment variables (requires SDK 3.0 or later). Any global pool option may be set as an environment variable. Options that are set as environment variables override those that an application sets. This allows changing the application behavior without additional code recompilation.

During an operation, the following failovers are possible:

- **Network failover**
  Occurs when the connection to the primary cluster fails.

- **Content failover**
  Occurs for operation-specific reasons. Content failover is not necessarily caused by a failure, so the term failover can be misleading in some cases. Instead, the state of the content or configuration settings may cause the operation to propagate to other clusters.

**The operational failover defaults**
The default behavior of the SDK operational failover is as follows in Table 2.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Network failover</th>
<th>Content failover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read (clip/BLOB)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Write (clip/BLOB)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Delete (clip)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Exists (clip/BLOB)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Query (clip/reflection)</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Replication interoperability matrix

EMC Centera can replicate and restore to different versions of CentraStar, however, there are limitations. The set of replication pairs is sufficiently large that it becomes ineffectual to describe each relationship in narrative. Therefore, Table 3 presents the data in a 2D matrix. The keys below the table clarify the notations used in the table.

Table 3. Replicate and restore between CentraStar versions

<table>
<thead>
<tr>
<th>Target</th>
<th>4.0</th>
<th>3.1.3</th>
<th>3.1</th>
<th>3.0(3)</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td></td>
<td>3.1.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>R/r</td>
<td>R/r</td>
<td>R/c</td>
<td>-c(2)</td>
<td>-c(2)</td>
</tr>
<tr>
<td>3.1.3</td>
<td>R/r</td>
<td>R/r</td>
<td>R/c</td>
<td>R/c(2)</td>
<td>R/c(2)</td>
</tr>
<tr>
<td>3.1</td>
<td>R/c</td>
<td>R/c</td>
<td>R/c</td>
<td>R/c(2)</td>
<td>R/c(2)</td>
</tr>
<tr>
<td>3.0</td>
<td>R/c</td>
<td>R/c</td>
<td>R/c</td>
<td>R/r</td>
<td>R/c</td>
</tr>
<tr>
<td>2.4</td>
<td>R/c(1)</td>
<td>R/c(1)</td>
<td>R/c(1)</td>
<td>R/c(1)</td>
<td>R/c</td>
</tr>
</tbody>
</table>

Table keys:
- R  Denotes Replication is legal
- r  Denotes Restore is legal
- c  Denotes CSGReplication should be used
- (1) Data will end up in a pool, according to current home pool assignments and active mappings.
- (2) Replication/Restore of mutable metadata (EBR/RH) is not supported. CentraStar will auto-pause when this is attempted.

CentraStar 3.0 is End of Service Life (EOSL) and is no longer supported.
Network segmentation

Revisions of CentraStar previous to 4.0 required that all application, replication, and management traffic use the same physical network. CentraStar 4.0 gives the storage administrator the capability to segregate the network traffic onto multiple physical networks based on traffic type.

For EMC Centera Replication this means that nodes other than or including the access nodes can be configured to participate in the replication of C-Clips to the replica cluster. There must be at least two nodes with the replication role in the cluster. For clusters with more than eight nodes, the number should not exceed half the number of nodes to a maximum of eight.

Network segmentation is discussed more fully in the white paper *EMC CentraStar 4.0 Network Segmentation.*
Conclusion

EMC Centera Replication allows customers to add another level of protection for their EMC Centera stored content by replicating all content written to their primary Centera cluster to a replication cluster (in, say, another data center). This provides customers the capability to restore access to their content should a disaster occur at the primary site (or the primary Centera becomes unavailable).

In the event of an outage of the primary cluster, any application reading from the primary will automatically fail over to the replica cluster and read from it. This ensures continued access to application content. Application writes do not automatically fail over, but if the application writes to the replica and bidirectional replication has been configured, then content written to the replica cluster during the outage will automatically be replicated back to the primary cluster when it comes back online. For other replication topologies, the restore process can be used to copy content from a replica to the primary.

Storage administrators can choose between five replication topologies (unidirectional, bidirectional, star, chain, and star and chain combined) to suit their preferred replication strategies.

References

- EMC Centera online help
- EMC Centera SDK API Reference Guide
- EMC Centera SDK Programmer’s Guide
- EMC Centrastar 4.0 Network Segmentation white paper