EMC RecoverPoint/Cluster Enabler

Abstract

EMC® RecoverPoint/Cluster Enabler (RecoverPoint/CE) software integrates with Microsoft Failover Cluster software to allow geographically dispersed cluster nodes to be replicated by RecoverPoint continuous remote replication. RecoverPoint/CE seamlessly manages all storage system processes necessary to facilitate cluster node failover. RecoverPoint/CE supports Windows Server 2003 and Windows Server 2008 in both Enterprise and Datacenter editions that use Node Majority and Node and File Share Majority quorum modes only.

April 2012
Copyright © 2008, 2012 EMC Corporation. All rights reserved. Published in the USA.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is.” EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

VMware is a registered trademarks of VMware, Inc. in the United States and/or other jurisdictions. All other trademarks used herein are the property of their respective owners.

Part Number h5936.5
# Table of Contents

**Executive summary** ................................................................. 4
  Audience ......................................................................................... 4

**Introduction to RecoverPoint/CE and RecoverPoint** ....................... 4
  RecoverPoint/CE .............................................................................. 4
  RecoverPoint .................................................................................. 6
  Terminology .................................................................................... 7

**General configuration guidelines** ................................................ 8
  Physical and logical requirements .................................................. 8
  Configuration rules and limits ......................................................... 10
  Installation and configuration ......................................................... 11
  Configuration Wizard ..................................................................... 13
  Create Group Wizard ..................................................................... 14

**Cluster group movement, failover, and recovery** ............................ 16
  RecoverPoint consistency groups .................................................. 17
  CE group policies ........................................................................... 19
  RecoverPoint link failure ............................................................... 20
  Node failure .................................................................................. 20
  Site failure .................................................................................... 20
  Majority node set clusters ............................................................. 21

**Architecture and communication** .................................................. 23

**Troubleshooting** .......................................................................... 25

**Conclusion** .................................................................................. 27

**References** .................................................................................. 28

**Appendix: Replication and consistency group states** ...................... 29
Executive summary

EMC® Cluster Enabler software enables geographically dispersed Microsoft Failover Clusters to replicate their data using MirrorView™, RecoverPoint, and SRDF®. When Cluster Enabler is used with RecoverPoint then continuous remote replication (CRR) can be used for replication between cluster nodes of a geographically dispersed cluster.

Geographically dispersed clusters offer increased levels of high availability, disaster recovery, and automation over non-clustered solutions. RecoverPoint/CE works seamlessly with applications designed to take advantage of Microsoft Failover Clusters, such as Exchange Server 2003 and 2007 and SQL Server, in Microsoft Windows Server 2003, Windows Server 2008, Windows Server 2008 R2, and Hyper-V environments.

Cluster Enabler does not support Microsoft Exchange Server 2010. RecoverPoint/Cluster Enabler (RecoverPoint/CE) is the version of Cluster Enabler that supports RecoverPoint and RecoverPoint/SE. RecoverPoint/CE is host-based software that integrates with Microsoft Failover Clusters that use Node Majority or Node and File Share Majority quorum modes. RecoverPoint/CE manages cluster resource failover between storage systems. RecoverPoint software is an appliance-based product that replicates data locally and remotely between storage systems. RecoverPoint/CE automates failover operations and the presentation of the latest replicated storage resource at the direction of Microsoft Failover Clusters to provide a seamless geographically dispersed solution.

RecoverPoint/CE allows existing Microsoft Failover Cluster customers to extend the protection of their current solution to include site disaster recovery. It reduces recovery time objectives (RTO) for existing RecoverPoint/SE and RecoverPoint customers by allowing Microsoft Failover Clusters to automate resource and application failover between sites.

Audience

This white paper is intended for any customers, partners, or EMC personnel wishing to gain a general understanding of RecoverPoint/CE benefits, features, and functions. It is intended to serve as a technical overview document as opposed to a step-by-step guide. For more in-depth information, please see the product guide and software release notes on EMC Powerlink®. Readers are presumed to have a basic understanding of RecoverPoint and Microsoft failover clustering technology.

Introduction to RecoverPoint/CE and RecoverPoint

RecoverPoint/CE

RecoverPoint/CE controls the RecoverPoint aspects of moving cluster resources between storage systems. When using CRR without RecoverPoint/CE, users typically
configure separate clusters at the primary and secondary sites. In this case, moving volumes between the clusters involves unmounting LUNs from the cluster at the primary location, making the replica available to the secondary cluster, and then redirecting application clients to the new cluster. By integrating with Microsoft Failover Clusters, RecoverPoint/CE enables the cluster to automatically manage resource and application failover, which greatly improves recovery time objective (RTO). Cluster Enabler has been updated to version 4.0 with the addition of support for Windows Server 2008 R2. Additionally, RecoverPoint/CE has been updated to version 4.0 and supports RecoverPoint and RecoverPoint/SE version 3.1.1 and later.

Figure 1 is a high-level depiction of the RecoverPoint/CE environment. Cluster nodes are geographically dispersed over two sites. RecoverPoint CRR is used to replicate data from the primary site, Site 1 in this case, to the secondary site, Site 2. The cluster group can be moved manually to Site 2 at the user’s discretion or automatically by Failover Cluster software in the event of a failure.

RecoverPoint/CE software is the only EMC software element required on each cluster node. RecoverPoint/CE uses the RecoverPoint software APIs to manage the RecoverPoint environment. Most users also have the RecoverPoint Windows utilities or the RecoverPoint Management Application installed on the server for managing RecoverPoint operations, but it is not a requirement for RecoverPoint/CE.

**Figure 1 RecoverPoint/CE environment before failover**

Figure 1 illustrates a RecoverPoint/CE environment before failover while Figure 2 shows this environment after failover. In Figure 2, the cluster group resources are moved to Site 2. To facilitate the move, RecoverPoint/CE instructs RecoverPoint to access, in physical mode, the latest image of the LUNs at Site 2 by initiating a RecoverPoint production failover.

Once the LUNs are accessed, they are automatically unmasked by RecoverPoint and then the cluster starts the application running at Site 2 instead of Site 1 and clients are redirected to a Site 2 node just like it was a local cluster.
**RecoverPoint**

EMC RecoverPoint provides local and remote data protection with replication of data over any distance; that is, locally within the same site, and/or remotely to another site—even halfway around the globe. RecoverPoint protects and supports replication of data that physical or virtualized applications are writing to local SAN-attached storage. RecoverPoint uses the existing SAN infrastructure to integrate seamlessly with existing host applications and data storage subsystems. For remote replication RecoverPoint uses either Fibre Channel or an IP network to send the data over a WAN. RecoverPoint/CE manages remote replicas and if local replicas are present, it does not manage them.
Asynchronous replication can occur over native Fibre Channel or Ethernet ports on each RecoverPoint appliance. Synchronous replication occurs over native Fibre Channel. When RecoverPoint replicates using the Ethernet ports, it uses standard TCP over IP and is not limited by distance. When RecoverPoint replicates using Fibre Channel distance is not technically limited by RecoverPoint, but is determined by the connectivity products and the distance extension protocols used. Synchronous replication is restricted to 4 milliseconds of round trip latency (which is around 200KM); customers that exceed this limit will only be supported with an RPQ.

RecoverPoint supports consistency groups, which ensures write-order consistency across related volumes. This is common practice for many applications such as Microsoft SQL Server, where log volumes and database volumes must be kept logically consistent with one another but may reside on different volumes or arrays.

![Diagram showing RecoverPoint Cluster group and RecoverPoint consistency group relationship](image)

**Figure 4 Cluster group and RecoverPoint consistency group relationship**

RecoverPoint/CE manages the RecoverPoint remote replicas at the RecoverPoint consistency group level. RecoverPoint/CE operates on a consistency group for each Failover Cluster group. Therefore, as cluster resources are moved between sites, the movement occurs in a consistent manner across all of the volumes in the cluster group. RecoverPoint/CE does not manage the local replica, sometimes called the CDP replica, of a RecoverPoint concurrent local and remote (CLR) consistency group.

**Terminology**

- **CE Cluster** – A Microsoft Failover Cluster managed by RecoverPoint/CE
- **CE Group** – A Microsoft Failover Cluster group managed by RecoverPoint/CE.
- **Consistency Group** – A group of one or more RecoverPoint replication sets managed together to achieve write-order consistency across the group of replicas. Consistency groups are a basic building block of RecoverPoint and are used when a single application or group of applications needs to be recovered at the same consistent point in time such as with transactional databases, where log volumes and data store volumes must be kept in sync during the replication process; they
are also used for applications like Exchange that have data and log devices with write-order dependencies.

- **Group Policy** – CE Group parameters that specify failover behavior for various conditions.
  - Restrict Group Movement is a Group Policy for RecoverPoint/CE.
- **Lateral Nodes** – Cluster nodes at the same site. Lateral nodes mount the same physical LUNs.
- **Peer Nodes** – Nodes at different sites that are part of the same cluster. One cluster node mounts a RecoverPoint primary image and its peer node mounts the corresponding replica image.
- **Primary Image** – The production volume of a RecoverPoint replication set. This volume is read/write accessible to the cluster node. All writes to the primary image are replicated to the secondary image.
- **Promote** – A RecoverPoint production failover operation that converts the latest copy of a secondary image to a primary image. Once converted to a primary image, the volume is read/write accessible. Since RecoverPoint consistency groups are used, all the secondary images for a given consistency group are converted to primary images at the same logical point in time. RecoverPoint production failover also causes the original primary images to be converted to secondary images, so that replication can continue in the reverse direction.
- **Restrict Group Movement** – Group Policy that prohibits movement between peer nodes (between RecoverPoint sites) when RecoverPoint replication is disabled for that RecoverPoint consistency group. If RecoverPoint replication is down and failover of cluster groups is required, then additional recovery steps are required beyond just moving peer nodes while the RecoverPoint link is down. This is discussed in the “Cluster group movement, failover, and recovery” section on page 14.
- **Secondary Image** – The RecoverPoint remote replica volume, which resides at a different site from the production volume. The secondary image is not read/write accessible during normal operation.
- **Site** – A location that consists of one or more storage arrays, two or more RecoverPoint appliances, and any cluster nodes that are connected to the arrays. RecoverPoint remotely replicates data between arrays at different sites and locally replicates data inside one array or between one or more arrays at the same site.

**General configuration guidelines**

This section provides an overview for supported environments. Please consult the *EMC Support Matrix* for detailed software compatibility requirements and the *Microsoft Knowledgebase* for Microsoft cluster requirements.

**Physical and logical requirements**

Figure 5 illustrates the physical connections in a RecoverPoint/CE environment. Each site has one or more cluster nodes. Each cluster node has a network connection to a
private interconnect and the enterprise LAN/WAN. For Windows 2003 Failover Clusters, all private interconnect IP addresses must be on the same subnet. Therefore, a VLAN is typically set up between the two sites. Windows 2008 Failover Clusters do not have this requirement. IP addresses of the private interconnect can be on different subnets. Nodes at each site are also connected to the local storage system. Storage systems are connected for replication purposes.

Figure 5 RecoverPoint/CE physical connections

Figure 6 represents a logical view of the environment. At each site the local cluster nodes are zoned to the local storage system. The nodes at each site are assigned to the same volumes. At Site 1, both Node 1 and Node 2 are assigned to the primary image LUNs. Node 1 and Node 2 are lateral nodes with respect to one another. Microsoft Failover Clusters control ownership of the LUNs between the lateral nodes at Site 1. Nodes 3 and 4 are peer nodes to Nodes 1 and 2. Node 3 and Node 4 are masked to the secondary image LUNs.

The cluster group is active on Node 1. To move the cluster group to either of the peer nodes, RecoverPoint/CE promotes the corresponding consistency group and brings the secondary images online at Site 2. Nodes 3 and 4 are also lateral nodes to one another. With the images active on Site 2, Failover Clusters manage movement between the lateral nodes. In this example, the RecoverPoint appliances on Site 1 are zoned to the CLARiiON array in Site 1 and P1 and P2 are masked to the appliances in Site 1. This is repeated for the RecoverPoint appliances in Site 2.
Figure 6 RecoverPoint/CE logical view

Configuration rules and limits


- Supported cluster quorum types include majority node set (MNS) and MNS with File Share Witness (FSW). RecoverPoint/CE does not support disk-based quorum or disk witness.

Table 1 RecoverPoint limits for storage systems supporting RecoverPoint/CE

<table>
<thead>
<tr>
<th>RecoverPoint parameters</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum replicated LUNs</td>
<td>2048</td>
</tr>
<tr>
<td>Maximum consistency groups</td>
<td>128</td>
</tr>
<tr>
<td>Maximum LUNs per consistency group</td>
<td>2048</td>
</tr>
<tr>
<td>Maximum cluster nodes per RecoverPoint cluster</td>
<td>16</td>
</tr>
</tbody>
</table>

- All RecoverPoint/CE cluster disk resources must be replicated. Replication must be performed by RecoverPoint and can be synchronous or asynchronous. Only the number of replication sets and/or consistency groups supported by the RecoverPoint cluster limits the number of resources.

- RecoverPoint/CE is managed by RecoverPoint cluster. One RecoverPoint cluster system can host nodes from multiple RecoverPoint/CE clusters.

- The same RecoverPoint cluster must manage all replicated physical disk resources for a cluster node. A cluster node cannot have resources replicated by multiple RecoverPoint cluster systems.
- Cluster nodes can only be zoned to the storage system at the local site. Nodes cannot be zoned to storage systems at multiple sites.
- Bi-directional replication is supported if you have multiple active cluster groups. A RecoverPoint/CE cluster can have active cluster groups at Site 1 replicating to Site 2 and have other cluster groups active on Site 2 replicating back to Site 1.
- Host connections must be Fibre Channel connected except in the case of a VNX Series or CLARiiON® CX3 or CX4 array using the array-based splitter, where the host connection to the VNX series or CLARiiON can be iSCSI.
- RecoverPoint can replicate asynchronously over Fibre Channel or IP. When replicating over IP, it is suggested that the bandwidth available for replication should be 3 megabits per second (Mb/s) or more.
- RecoverPoint can replicate synchronously over Fibre Channel. When replicating synchronously the round trip latency must be below 4 ms.

**Installation and configuration**

This section discusses the basic steps to install and configure a RecoverPoint/CE environment, and provides best practices to help you in the early stages of planning your configuration. This section assumes you have an existing cluster-aware application setup under Microsoft Clusters. For step-by-step instructions and for details on how to configure RecoverPoint/CE with existing applications that are not yet configured for a cluster please see the *EMC RecoverPoint/Cluster Enabler Version 4 Product Guide* on EMC Powerlink.

You can implement RecoverPoint/CE on new or existing failover cluster environments. RecoverPoint/CE software needs to be installed on any existing or potential cluster nodes. You must have local administrator rights to install and configure RecoverPoint/CE. The installation of RecoverPoint/CE requires a reboot. When they are installed into an existing environment, cluster resources can be moved to other cluster nodes during the reboot.

LUNs that are or will be cluster resources must be replicated before RecoverPoint/CE manages them. Prior to running the EMC Cluster Enabler Manager GUI, you should perform the following actions using the Microsoft Cluster Administrator GUI and the RecoverPoint Management Application:

1. Before installing EMC RecoverPoint/Cluster Enabler on all host nodes, Microsoft host clusters must be operating at both sites.

2. RecoverPoint version 3.4.1 must be installed and replicating all volumes to be included in the cluster group between the two sides.

3. Ensure that the consistency groups that will be configured for RecoverPoint/CE have finished synchronizing and enable image access before continuing. If image access of the consistency groups is not enabled, Microsoft Cluster Services will not allow the volumes to be added to the cluster resource groups.

4. Before specifying the cluster groups in Microsoft Cluster Services, be sure you have enabled image access at the remote site for each of the consistency groups.
that contain volumes to be included in cluster groups using the RecoverPoint Management Application. To be replicated by RecoverPoint/Cluster Enabler, each consistency group must have exactly the same name as the corresponding cluster resource group and must contain exactly the same volumes. Names of consistency groups and the corresponding cluster groups may contain only alphanumeric characters and underscores.

5. Define all cluster groups and resources using Microsoft Cluster Services.

6. Disable image access of the RecoverPoint consistency groups that were enabled in step 3.

7. In the RecoverPoint Management Application, select the consistency group. In the Components pane, select the Policy tab. In the Stretch Cluster Support area, select the checkbox Use RecoverPoint/CE. Ensure that **Group is managed by CE, RecoverPoint can only monitor** is selected. Repeat for each consistency group to be replicated by RecoverPoint/CE.

8. After installing RecoverPoint/CE on all host nodes in the cluster the RecoverPoint Access Settings automatically appear. In the RecoverPoint Access Settings, enter the requested security account information for each host node in the cluster. You must enter the RecoverPoint Management IP address information of the site where the node is located. The default user/password is plugin/plugin and it is recommended that you change the default password. To change the default password, log in to the RecoverPoint CLI with a User of **plugin** and a Password of **plugin**. Run the `set_password` command and exit the CLI. You will need to run the RecoverPoint Access Settings again to change to the new password. To run the RecoverPoint Access Settings again, from the Microsoft Windows desktop, select Start > Programs > EMC > Cluster Enabler > RecoverPoint Access Settings.

9. In the EMC Cluster Enabler Manager GUI, use the CE Configuration Wizard to configure a CE cluster. The configuration wizard will automatically discover the storage volumes and perform the required configuration.

10. Once volumes are mirrored and available to the cluster nodes, use the EMC Cluster Enabler Manager to configure the cluster and storage system environment. The EMC Cluster Enabler Manager GUI is a Microsoft Management Console (MMC) Snap-In. The GUI offers easy-to-use wizards to guide you through the processes of discovery, conversion, and configuration. Figure 7 shows where the Configure CE Cluster, Manage CE Cluster, and Recover CE Cluster wizard launch links are located within the EMC Cluster Enabler Manager. The Recover CE Cluster wizard is not used for the RecoverPoint/CE product, but is used by its sister product, SRDF/Cluster Enabler.
Figure 7 RecoverPoint/CE user interface

Configuration Wizard

The Configuration Wizard is the first wizard you will use and clicking the Configure CE cluster link launches it. The Configuration Wizard validates that the proper software components exist on the cluster nodes, and discovers existing cluster nodes, their corresponding storage systems, and the device assignments. It then converts the cluster to a CE cluster and converts any existing cluster groups to CE groups. The Configuration Wizard can also be used to add additional nodes to the cluster as part of the conversion process. This is most useful for adding the geographically dispersed nodes to the cluster if they are not already present.

Once the cluster is converted to a CE cluster, you can see the cluster environment in the GUI. Use the EMC Cluster Enabler Manager GUI to add additional cluster nodes, create/delete cluster groups, and create/delete cluster disk resources.

Figure 8 shows the EMC Cluster Enabler Manager GUI after a cluster has been converted to a CE cluster. The GUI has the same look and feel as the Failover Cluster Management GUI within Windows 2008. The left pane of the GUI offers a tree view of the environment. In this example, a cluster named CE is being managed. The cluster has a CE group named RP-CE. As the user highlights various elements, more detailed information about that element becomes available in the middle pane. Also, right-click object menus are available for all the objects in the tree view. The Actions pane, which is the right-most pane, has quick links to the object menus of the item that is currently highlighted. The object-based right-click menus and the Action pane offer the same management options.
After the initial configuration, select the Manage CE Cluster link when opening the GUI. The EMC Cluster Enabler Manager GUI refreshes the cluster information and presents the user with the view shown in Figure 8.

Some applications, like Exchange 2007, may create their own cluster group as part of the installation. In these cases, use Cluster Administrator (or Failover Cluster Management for Windows 2008) to add the mirrored physical disk resources to the cluster group created by the application. Then use the Configuration Wizard to convert the cluster group to a CE group.

For applications that do not create a cluster group automatically, like Exchange 2003, you can create the cluster group within Cluster Administrator then use the Configuration Wizard to convert the group, or you can use the Create Group Wizard within the GUI to create the new group.

**Create Group Wizard**

The Create Group Wizard creates additional CE groups. For volumes to be eligible, RecoverPoint must replicate them, with the production LUNs and the replicated LUNs assigned to nodes at their respective sites. The Create Group Wizard can be launched from the right-click menu of the Groups icon in the tree view shown in Figure 9.

---

**Figure 8. EMC Cluster Enabler Manager GUI object menu**

**Figure 9. Create Group object menu**
In the Create Group Wizard, do the following:

1. Enter a name for the new group.
2. Select devices to be included in the group.
3. Select a Group Policy (Restrict Group Movement).

Figure 10 shows the GUI after the Create Group Wizard created the “FileShare” group. The new group appears in the left-pane tree view. The middle pane shows the details of the new group. It contains cluster disk resource Disk F. The Failover Behavior is “Restrict Group Movement”, which is the only setting.

![Figure 10. New group in RecoverPoint/CE Management GUI](image)

Within Cluster Administrator, the group appears with an EMC Cluster Enabler resource as seen in Figure 11. The resource is a dependency of each physical disk resource in the cluster group.

![Figure 11. RecoverPoint/CE group shown in Cluster Administrator](image)
RecoverPoint/CE requires that RecoverPoint have a consistency group for each cluster group with the same name as the cluster group name. Figure 12 shows the consistency group configuration in the RecoverPoint Management Application. In this example, the RecoverPoint storage administrator created two RecoverPoint consistency groups, one for each of the two cluster groups: FileShare and RP-CE. The RecoverPoint consistency group names are identical to the cluster group name. The only valid characters in either name will be A-Z, a-z, 0-9, dash (-), and underscore (_).

Figure 12. RecoverPoint Management Application consistency groups

Cluster group movement, failover, and recovery

Once groups are configured, they can be moved between nodes using Cluster Administrator. Groups can also be moved between nodes by Failover Cluster software in response to a failure. RecoverPoint/CE will automatically manage the underlying RecoverPoint replicas to facilitate the move if it is across the link to a peer node. In Figure 13, Nodes 1 and 2 are lateral nodes. Nodes 3 and 4 are peer nodes to Nodes 1 and 2. Group movement between lateral nodes has no impact on the underlying mirrors. Movement between peer nodes requires manipulation of the underlying mirrors. The following sections discuss RecoverPoint and Failover Cluster behavior during group movement under normal conditions and under failure conditions.
RecoverPoint consistency groups

RecoverPoint/CE manages LUNs at the consistency group level. RecoverPoint consistency groups are a collection of LUNs that are managed as one entity. Consistency groups ensure write ordering is maintained across the set of LUNs in the group. Error! Reference source not found. shows an environment that has two consistency groups, FileShare and RP-CE. The FileShare consistency group has one replication set (RSET1), and the RP-CE consistency group has two replication sets (RSET2, RSET3). Each replication set has the primary or production LUN on the Site1 storage system and secondary or replica LUN on the Site2 storage system.

RecoverPoint image failover is the process of making a secondary LUN read/writeable so that it is available to a peer node at the last point-in-time image available and then reversing the replication direction so that the peer node becomes the new production
copy. RecoverPoint/CE performs the copy failover for the RecoverPoint consistency group; therefore it is an atomic operation across all the replica volumes in the consistency group. This provides write-order consistency across all of the group members.

The ability of RecoverPoint/CE to initiate a Replica copy failover for a specific consistency group depends on the data transfer state of the consistency group. Consistency group states include Active, Init, Error, Paused, and Ready.

RecoverPoint consistency groups can be failed over as long as they are in the Active state. When in this state, the primary and secondary replicas have been synchronized and changes are being replicated from the primary to the secondary replicas. This is the ideal condition for performing a failover operation. A failover operation effectively reverses the roles of primary and secondary copies and replication continues in the reverse direction. No resynchronization is needed.

As long as there is communication between storage systems over the RecoverPoint link, RecoverPoint/CE executes a normal promote when devices are moved between peer nodes. As the cluster brings the resource offline on the active node, RecoverPoint/CE waits briefly for the consistency group state to become Synchronized, then it performs the promote operation. If there is no communication between storage systems over the RecoverPoint link, RecoverPoint cannot conduct a normal promote. This and other failure cases are discussed in the “RecoverPoint link failure” section on page 20.

Figure 15 shows the same set of consistency groups before and after RecoverPoint/CE conducts the failover. Prior to the promote operation, the FileShare and RP-CE groups were in the Active state, however, for the FileShare group the FS-CRR image was active and accessible and for RP-CE the Production image was active and accessible. When RecoverPoint/CE performed the failover the copy roles were reversed so the FS-Prod and CRR images became active and accessible. Their roles were simply reversed and replication will continue in the reverse direction (from Sant_Clara to London).
Figure 15. RecoverPoint consistency groups before and after failover

CE group policies

This section describes the CE Group Policy setting and its effect on failover behavior. RecoverPoint/CE moves groups between peer nodes in a manner that requires no resynchronization. However, there are a few failure conditions where it is not possible for RecoverPoint/CE to move the groups. The factors that determine promote and resynchronization behaviors are:

- The current state of the RecoverPoint consistency group(s)
- The status of the RecoverPoint link

A CE group’s recovery policy is set Restrict Group Movement, while other products that use Cluster Enabler (such as SRDF/CE or MirrorView/CE) support Automatic Failover, which is not supported with RecoverPoint. This means that if the RecoverPoint links are down or errors were discovered in the consistency group and you have a cluster node fail, the failover to a peer node will not be automatic.

The Restrict Group Movement only allows a group to move or fail over to a peer node if the RecoverPoint link is up and a normal promote will be successful. If a normal promote will not be successful, RecoverPoint/CE tries to bring the resource back online at the original site. If it cannot do this, it brings the resource offline. The Restrict Group Movement Policy prevents movement to a peer node when:
• The RecoverPoint link is down (including the case of site failure).
• The RecoverPoint link is up, but the consistency group is in an error state.

**RecoverPoint link failure**

A link failure is an interruption in the RecoverPoint link between RecoverPoint appliances that own the specific RecoverPoint consistency group. The result is that the consistency group state is unknown. Writes can continue on the primary image, however, they are not replicated while the link is down. Instead, RecoverPoint enters marking mode such that the blocks of the LUN that have changed are marked in the local journal for the impacted consistency group. This may occur during high-load conditions or it may occur if the IP bandwidth for replication drops below 3 Mb/s.

If the link is restored, and there was no peer-to-peer group movement, the remote replicas will be automatically resynchronized using marked blocks and the consistency group will transition from the Sync state to the Active state and will resume normal operation.

To move a group to a peer node while the link is down, the administrator must manually issue a “move group” from within Cluster Administrator and a failover at the direction of Failover Clusters in response to a failure. Additionally, the administrator must utilize the RecoverPoint Management Application to change the group to “maintenance mode” with RecoverPoint/CE monitoring the status. Then the administrator would enable image access for the appropriate RecoverPoint consistency groups, perform the RecoverPoint failover operation, and then return control over the consistency group back to RecoverPoint/CE. It is important to realize that RecoverPoint/CE does not support a peer-to-peer failover while the RecoverPoint replication or link is down.

**Node failure**

RecoverPoint/CE automatically edits the preferred owners list of each group, so that lateral nodes precede peer nodes. That way, in the event of a localized failure (such as a node failure), the lateral nodes assume ownership of the group. If no lateral nodes are available, then the group may be moved to a peer node. Movement of the group to a peer node depends on the same RecoverPoint link status as described in the previous sections. Therefore, if the link is up under a node failure condition, RecoverPoint/CE can move the group to the peer node. If the link is down, then groups must be manually moved to peer nodes.

**Site failure**

In the event of a total site failure, no lateral nodes are available, and the RecoverPoint link is assumed to be down. Therefore, assuming the cluster remains online, groups would be brought offline.

Both node failure and site failure bring into question the validity of the cluster itself. Majority node set clusters have specific requirements for the number of nodes that must be available and able to communicate for the cluster to remain running. The
following section provides a general technical overview of majority node set clusters, their requirements, and best practices.

**Majority node set clusters**

Microsoft recommends the majority node set (MNS) quorum model for use in geographically dispersed clusters. RecoverPoint/CE only supports MNS. MNS uses a distributed quorum resource that is stored locally on each node. As changes are made to the cluster configuration, the quorum information is updated on all of the nodes. Changes are accepted to the cluster configuration only after a majority of the nodes have the changes. The same is true for cluster resource management. A majority of the nodes must be available and agree on which node owns each resource. To start cluster resources and for them to remain running, MNS clusters require that:

- N/2+1 cluster nodes be running
- The nodes can communicate with each other

A two-node geographically dispersed cluster requires that 2/2+1 = 2 nodes are available and communicating to keep cluster resources online. Figure 16 depicts a site failure that leaves only one node running. The cluster loses quorum in this instance and the cluster resources are brought offline.

**Figure 16. Site failure causing loss of quorum**

Loss of communication between the sites causes the cluster to become partitioned. This is commonly referred to as a “split brain” scenario. Cluster resources are brought offline on partitions that do not have a majority of nodes. In the case shown in Figure

---

1 Microsoft TechNet, “Server Clusters: Majority Node Set Quorum,” January 1, 2004,

In both cases, there are methods for bringing the cluster resources back online, such as force quorum\textsuperscript{2}, while the outages remain. However, it requires manual intervention and there may be limitations\textsuperscript{3} to cluster operation and configuration options while running in this temporary mode.

File Share Witness (FSW) is a feature that allows a file share to act as a voting member of an MNS cluster\textsuperscript{3}. This is helpful for maintaining majority quorum in a number of failure scenarios such as site failure or partitioning where N/2 nodes remain running and communicating with one another. The File Share Witness will act as the +1 Node in this example, ensuring that a quorum can be formed from the remaining (N/2) nodes and the File Share Witness, which meets Microsoft’s requirements of requiring (N/2+1) operating nodes to form a quorum.

Using the same example, when a File Share Witness resource is added to the environment, the cluster effectively has three voting members of the cluster even though there are still only two cluster nodes available for resource ownership. Majority in this case equates to two voting members, so resources will remain online as long as there are two or more of the voting members available and able to communicate. In the case of the loss of one node, cluster resources will remain online. Figure 18 depicts this environment.


\textsuperscript{3} Microsoft Help and Support, Article ID 921181, April 24, 2007, http://support.microsoft.com/kb/921181
Figure 18. MNS cluster with a File Share Witness

The File Share Witness should reside on a server in the same Windows domain as the cluster nodes. For high-availability reasons, it should not reside on a cluster node and should be at a location that would not be affected by site loss. Windows 2003 supports eight-node clusters when File Share Witness⁴ is used. Windows 2008 can support up to eight nodes on 32-bit systems and 16 nodes on 64-bit systems for all quorum models, including MNS with File Share Witness. RecoverPoint/CE supports all MNS cluster configurations supported by Microsoft.

Architecture and communication

RecoverPoint/CE integrates with Microsoft Failover Clusters to provide seamless function with cluster operations. RecoverPoint/CE uses the Windows Management Instrumentation (WMI) framework to interact with Microsoft Failover Clusters. An EMC WMI Provider and the CE Service manage cluster and RecoverPoint/CE-specific objects in the WMI Repository. RecoverPoint/CE monitors the resource library to pick up and react to any changes made within Cluster Administrator or the RecoverPoint/CE Management GUI. All RecoverPoint/CE configuration information is stored in the cluster database. If users back up the cluster DB, then the RecoverPoint/CE configuration will be backed up as well. Figure 19 is a high-level depiction of RecoverPoint/CE and Failover Clusters on a cluster node.

Figure 19. RecoverPoint/CE architecture

The EMC Cluster Enabler Manager GUI is a Microsoft Management Console snap-in. All communication between cluster nodes is facilitated by Microsoft Failover Clusters, so no additional network ports need to be opened between nodes. The EMC WMI Provider uses the RecoverPoint API to manage the RecoverPoint systems. Communications occur between cluster nodes and RecoverPoint over a secure connection. By using the Microsoft Failover Framework, RecoverPoint/CE has minimum impact to existing environments.
Troubleshooting

There are several places users can investigate issues in their environment. Within the RecoverPoint/CE Management GUI, users can view the cluster node’s application log as seen in Figure 20. The application log contains events pertaining to management of group and disk resources on the cluster node.

![Figure 20. RecoverPoint/CE Management GUI - Display Events](image)

A more detailed log is maintained specifically by RecoverPoint/CE in its database. The CE_EventTraceDump.exe utility is used to extract the events. CE_EventTraceDump.exe is a command line utility located in the Cluster-enabler install directory that is normally located in %SystemDrive%\Program Files\EMC\Cluster-Enabler\. The default output location for CE_EventTraceDump.exe is the /logs directory within the %SystemDrive%\Program Files\EMC\Cluster-Enabler\ directory. The default naming convention of the log file is ce_event_trace<timestamp>.txt. Optional flags are available for use with CE_EventTraceDump.exe to allow the user additional dump options. The following command dumps the logfile to a specific location:

```
CE_EventTraceDump.exe -outfile <logfile.txt>
```

The following command creates a logfile in a specific location using local time:

```
CE_EventTraceDump.exe -localtime -outfile <logfile.txt>
```

The following command creates a logfile in a specific location using local time, but only includes events since the last time CE_EventTraceDump.exe was run:

```
CE_EventTraceDump.exe -sincelast -localtime -outfile <logfile.txt>
```
If requesting support, users should collect the CE_EventTraceDump logs from all cluster nodes, along with the output from the RecoverPoint Diagnostics log collection process. The RecoverPoint log collection process is run under an ssh session to the RecoverPoint appliance using the boxmgmt login. Once you log in select **Collect system info** from the **Diagnostics** menu to collect system information for later processing and analysis. You are prompted for the following information:

- Time frame (date and time) for system logs collection, in GMT. GMT is not adjusted for daylight saving (summer) time.
- Whether to collect information from the other site also.
- Whether to send results to an FTP server.
- Whether you want to perform full collection. Full collection provides additional technical system information.
- Whether you want to limit collection time.
- Where to collect logs from:
  - From RPAs only
  - From splitters only
  - From RPAs and splitters

If you are collecting splitter logs, you will be asked to identify the splitter types to collect logs, including SANTap, KDrivers, Brocade, and CLARiiON.

Lastly, there are the RecoverPoint event logs. RecoverPoint logs events that occur in the RecoverPoint system. The event log can be viewed and is particularly helpful in understanding if RecoverPoint is operating correctly. A sample of events from the RecoverPoint Management Application is shown in Figure 21.
Figure 21. RecoverPoint event logs

**Conclusion**

RecoverPoint/CE is a software package that integrates with Microsoft Failover Clusters to provide users geographically dispersed clusters over RecoverPoint links. Geographically dispersed clusters offer low RTO for failing applications between sites by allowing Failover Clusters to manage application failover. Users accustomed to managing Failover Cluster environments can easily incorporate RecoverPoint/CE into their environment to add disaster recover protection. Current users of RecoverPoint/SE and RecoverPoint can add geographically dispersed clusters to their environment and reduce their RTO.
References

For more information, please see the following on Microsoft’s TechNet and Help and Support websites:


Please also refer to the following EMC documentation available on EMC Powerlink:

- EMC RecoverPoint/Cluster Enabler Version 4 Product Guide
- EMC RecoverPoint/Cluster Enabler Version 4 Release Notes
- Disaster Recovery for Windows using EMC RecoverPoint/Cluster Enabler
- EMC RecoverPoint Family Overview
- Improving Microsoft Exchange Server Recovery with EMC RecoverPoint
Appendix: Replication and consistency group states

Two parameters, *states* and *conditions*, report the current status of each remote mirror and secondary image. Each consistency group also has a state and condition parameter. The values for states and conditions are listed in the following tables.

**Table 2. Replication states**

<table>
<thead>
<tr>
<th>State</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>The replication is running normally.</td>
</tr>
<tr>
<td>Paused by system</td>
<td>The replication process has been paused by RecoverPoint or RecoverPoint/CE. This will occur when a peer failover is being performed.</td>
</tr>
<tr>
<td>Init</td>
<td>A peer failover has occurred and the new production site needs to perform an initialization as part of reversing the replication direction.</td>
</tr>
</tbody>
</table>

**Table 3 Consistency Group States**

<table>
<thead>
<tr>
<th>State</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synchronized</td>
<td>All the replica images are in the Synchronized state.</td>
</tr>
<tr>
<td>Anything but synchronized</td>
<td>The group may be paused, it may be in maintenance mode, or other errors may have occurred. Administrative action may be required to return the consistency group to having the ability to perform a peer failover. The user should open the RecoverPoint Management Application to see further details on the status of the consistency group(s).</td>
</tr>
</tbody>
</table>