Overview

Server cluster configurations manage disks on a shared storage infrastructure that are visible from multiple nodes although only one node in a server cluster can access any given disk at any point in time. In the event of the failure or corruption of a disk on the shared storage interconnect special care must be taken to restore the data and recover the applications.

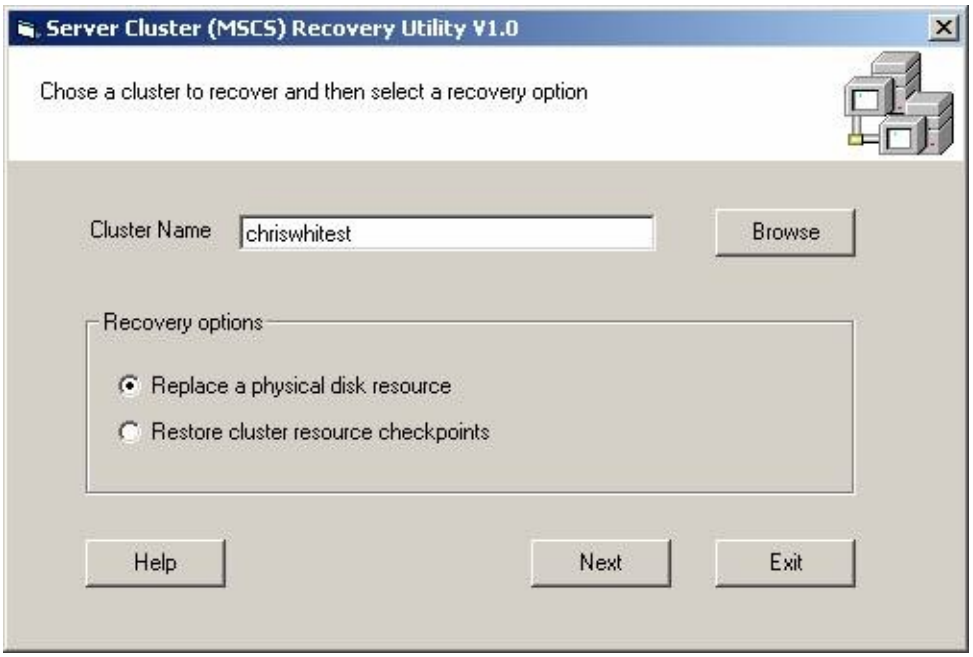
The Server Cluster Recovery Utility is a utility that collects together a number of pieces of functionality that are particularly useful in a server cluster after a disk on the shared bus has failed.

This utility is primarily aimed at the following scenarios:

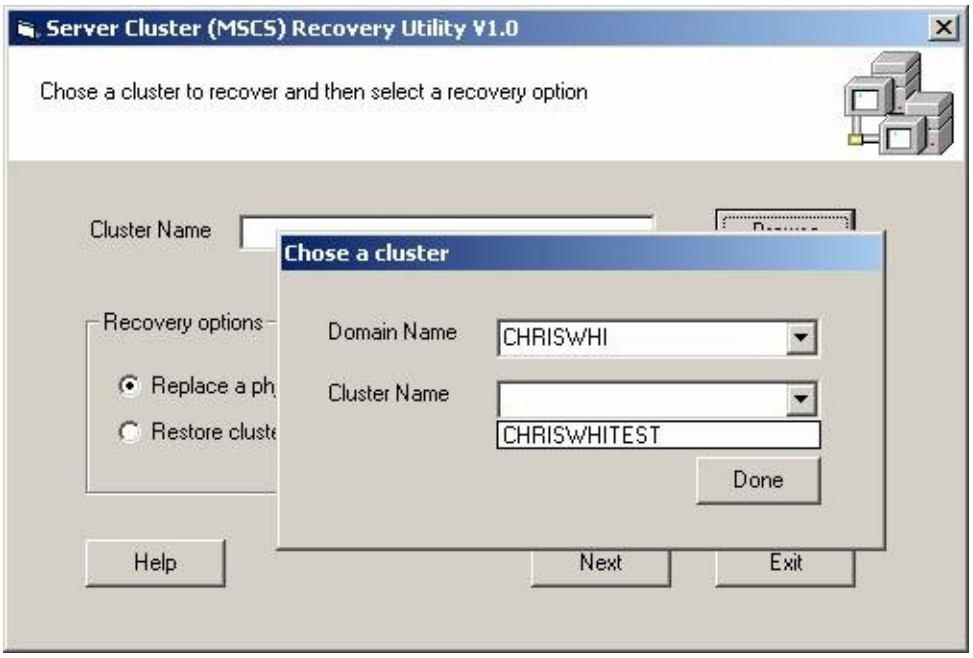
- Restoring resource checkpoint files
- Replacing a failed disk or recovering from disk signature changes
- Migrating data to a different disk on the shared bus

Connecting to a cluster

The Server Cluster Recovery Utility allows operations to be performed remotely. In other words, you can run the Server Cluster Recovery Utility on a management station and recover checkpoints or failed disks on a cluster in the data center.



To select the appropriate cluster, you can either type in a cluster name directly or you can browse through the clusters that are available to you by clicking on the Browse button.


Server Cluster (MSCS) Recovery Utility V1.0
Chose a cluster to recover and then select a recovery option
Cluster Name
Recovery options
Replace a ph
Restore cluste
Help
Next
Exit
Chose a cluster
Domain Name
CHRISWHI
Cluster Name
CHRISWHITEST
Done

Recovering resource checkpoint files

Many applications and other resources store data in registry keys outside of the cluster database. Resource checkpointing is the process of associating a resource with one or more registry keys so that when the resource is moved to a new node (during failover, for example), the required keys are propagated to the local registry on the new node. This allows an application to store configuration data in the registry and have an up to date version of that data available, irrespective of where the application is hosted in the cluster.

Resource registry checkpoints are setup by defining a sub-tree in the registry (in the HKEY_LOCAL_MACHINE hive) that is to be made available on all cluster nodes. Cryptographic keys are used by applications and stored in the registry as required. The cluster keeps a resource checkpoint file associated with each checkpointed registry key or crypto key on the quorum disk. Each file is used to save and restore the contents of the checkpointed key. The checkpoint files are found on the quorum disk under the quorum path as files with names *.CPT.

The following defines how the checkpoints are maintained:

- Whenever anything changes on the checkpointed registry tree and the resource is online, the Cluster service stores a copy of the tree in a resource checkpoint file on the quorum resource.
- A change made to a checkpointed key while the resource is offline will be overwritten with the checkpointed data when the application comes online.
- If the resource moves to another node, the Cluster service restores the registry tree from the quorum resource checkpoint file to the registry on the new node before the resource is brought online.
- If the resource is deleted, the checkpoint file is deleted.

The configuration data in the registry is typically required for an application to function correctly and it is important that the checkpoint file is correctly maintained and kept consistent and up-to-date with what the application expects. There are, however, a number of scenarios where the checkpoint file may be lost or become out of date:

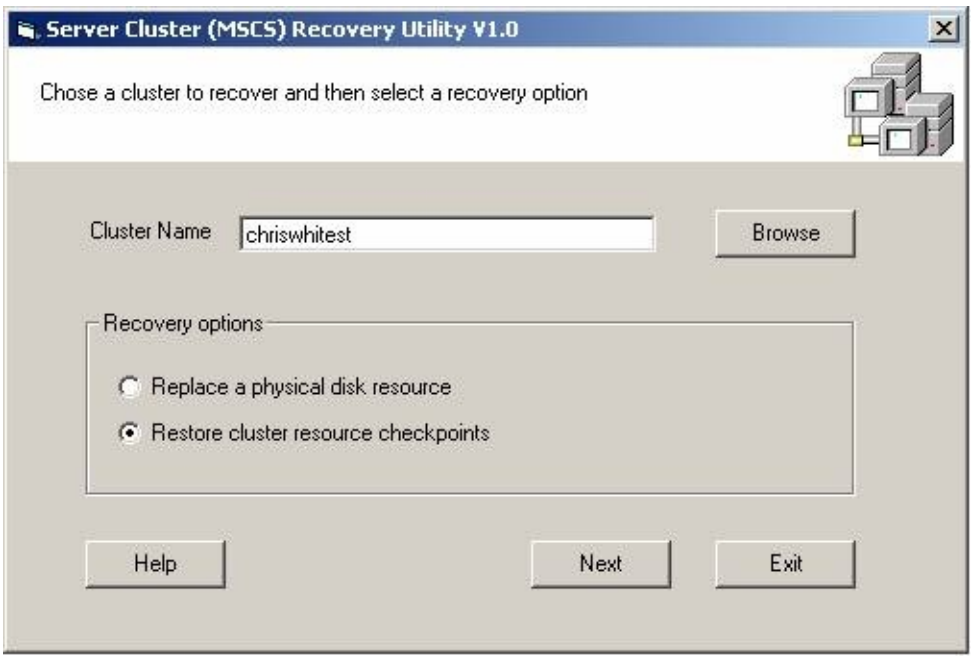
- The quorum disk fails. In this case a new quorum disk may replace the old disk. The cluster database itself can be recovered using procedures defined

in Recovery after a disk failure, however, this does not recover the checkpoint files. For applications to failover correctly with up-to-date registry data, the checkpoint files must be re-created.

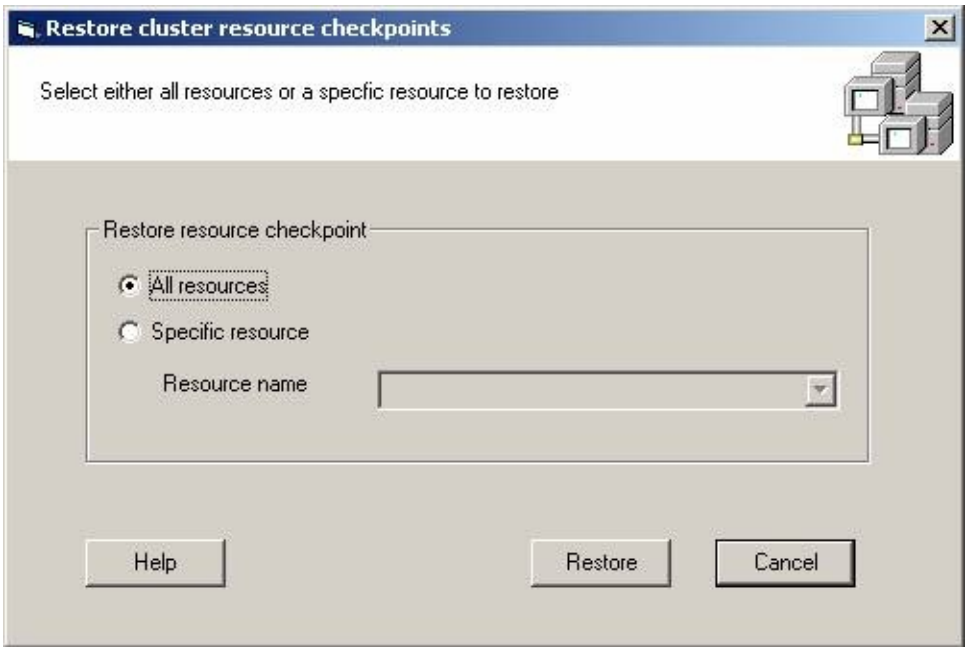
- If the quorum disk fails the cluster database and checkpoint files can be restored from a backup. (In Windows .Net Server, the cluster database and checkpoint files are part of the system state and can be saved and restored using Automated System Recovery). In this case, the checkpoint files are restored; however, the contents may not reflect the current state in the registry. When the application that has associated checkpoints is brought on-line, the data in the registry will be over-written with the data from the checkpoint files. To avoid this issue, if the quorum disk is restored from a backup, you should delete all of the checkpoint files and re-create them using the Cluster Recovery Utility.
- The checkpoint files are accidentally deleted. This can be either due to operator error or a rouge application incorrectly deleting files.
- In some extreme failure cases (e.g. of the underlying disk IO subsystem hardware or software), files on the file system may become corrupt.

The Cluster Recovery Utility allows an administrator to re-create the checkpoint files for one or all resources on the cluster. It gathers the information to re-create the checkpoint files from the node that currently owns the resource.

In the Cluster Recovery Utility, specify the appropriate cluster to recover and select the “Restore cluster resource checkpoints” option then click the Next button.



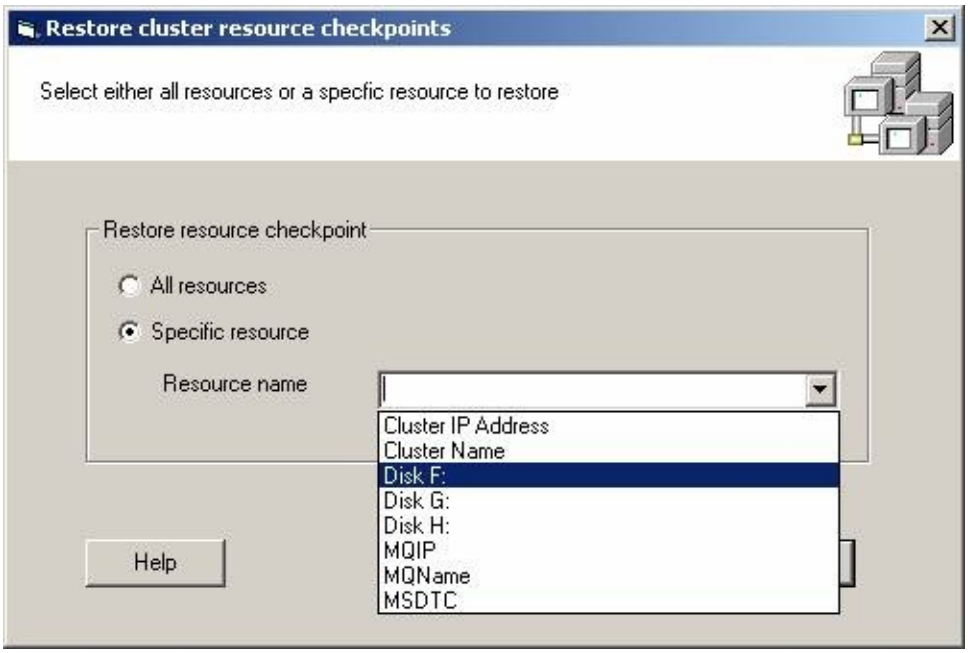
You can use the cluster recovery utility to restore the checkpoint files for either a single resource or for all resources in the cluster. To restore all checkpoint files for all resources in the cluster select the “All resources” option and click the Restore button.



The “All resources” option loops through each resource in turn (regardless of whether the resources are online, offline or failed) in the cluster and re-creates each checkpoint file. Each checkpoint file is populated with the current value of

the appropriate registry key from the node that currently owns the resource (in the case of an online resource, this is the node that is currently hosting the resource, in the case of an offline resource, this is the node that is defined as the current owner – visible through Cluster Administrator or the cluster.exe utility).

In some cases, re-creating the checkpoint files for a single resource is more appropriate. The Server Cluster Recovery Utility allows a single resource to be restored using the “Specific resource” option. You can either type in a resource name or you can select from all the resources in the cluster using the drop-down menu. Once you have selected a resource, hit the Restore button to re-create the checkpoint files for that resource.



Recovering from a disk failure

Recovering a cluster disk (i.e. disks managed by the cluster on a shared storage interconnect) can be a complex process. The Server Cluster Recovery Utility aims to simplify the process and make it less error prone so that in the event of a disk failure, a shared disk can be brought back quickly and application availability can be restored.

Note: Microsoft recommends that shared disks on the cluster provide some level of redundancy to protect against disk failures (such as RAID-1 or RAID-5), however, in catastrophic cases disk recovery may be necessary.

A server cluster has a special disk known as the quorum disk. This disk is used to store cluster configuration and such things as resource checkpoint files. The quorum disk needs additional work to recover in the event of a failure and we will cover that separately.

Note: This recovery procedure is ONLY valid for disks that are managed in the cluster using the physical disk resource type. The recovery procedures for other types of cluster disk will vary depending on the configuration and the features. You should contact your vendor if you are using disks that are not exposed to the cluster as physical disk resource types.

Related topics:
Recovering a shared disk; Recovering a quorum disk;

Recovering a shared disk

The server cluster physical disk resource uses the disk signature to identify a disk and to map the real device to a physical disk resource instance. When a physical disk fails and is replaced, or when a physical disk is re-formatted with a low-level format (may be required if the IO subsystem information on the disk becomes corrupt for any reason), the signature of the newly formatted disk no longer matches the signature stored by the physical disk resource. There are other reasons that the disk signature may change, for example, a boot sector virus or a malfunctioning multi-path device driver can cause the signature to be re-written (see kb article Q293778 Multiple-path software may cause disk signature to change). In all of these cases, the physical disk resource cannot be brought online and action is required to get the applications using that disk up and running again.

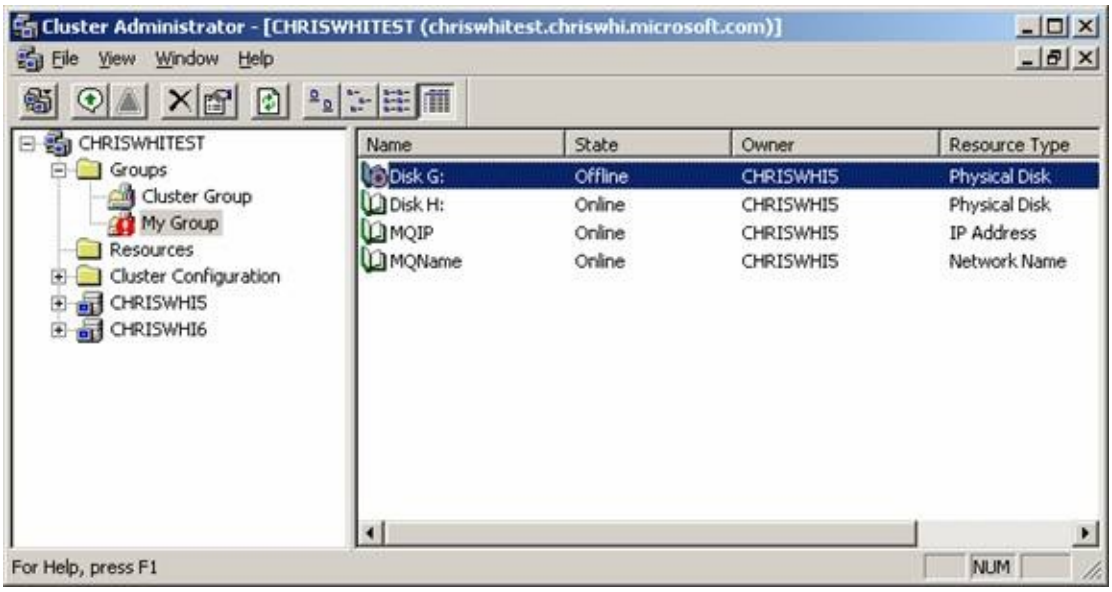
The cluster recovery utility allows a new disk, managed by a new physical disk resource to be substituted in the resource dependency tree and for the old disk resource (which now no longer has a disk associated with it) to be removed.

To replace a failed disk use the following procedure:

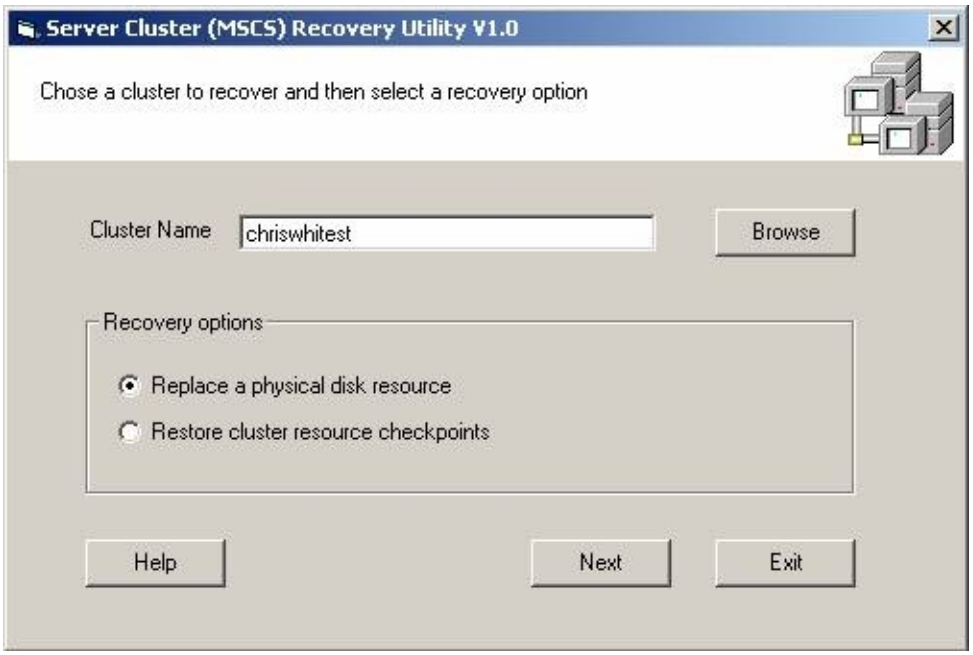
- Add a new disk drive to the cluster. In a storage area network environment, adding a new disk drive may involve creating a new logical unit and exposing it to the server cluster nodes with appropriate LUN masking, security and zoning properties.
- Make sure that the new disk is only visible to one node in the cluster. Until the Cluster service takes control of the new disk and a physical disk resource is created, there is nothing to stop all nodes that can see the disk from accessing it. To avoid file system issues, you should try to avoid exposing a disk to more than one node until it has been added to the cluster. In some cases (such as with low-end fiber channel RAID devices or devices in a shared SCSI storage cabinet) there is no way to avoid multiple machines from accessing the same disk. In these cases, a CHKDSK may run when the disk resource is brought online in step 5 of this procedure. Although this situation is recoverable through CHKDSK, you can avoid it by shutting down the other cluster nodes, although this may not be appropriate if the cluster is hosting other, currently functioning applications and services.

- Partition and format the new disk drive as required. Note: For a disk drive to be considered as a cluster-capable disk drive, it must be an MBR format disk and must contain at least one NTFS partition. Assign it a drive letter other than the letter it is replacing for now.
- Create a new physical disk resource for the new disk drive using Cluster Administrator (or the cluster.exe command line utility).
- Make the disk drive visible to the same set of nodes as the disk drive that it is replacing (in a typical configuration, a disk driver is visible to all nodes in the server cluster). In the event that the device does not appear to the cluster nodes, you may perform a manual rescan for new hardware using the device manager. At this stage you should try to bring the disk resource online and then fail it over all nodes of the cluster in turn to ensure that the new physical disk is correctly configured and can be viewed from all nodes.
- Use the Server Cluster Recovery Utility to substitute the newly created physical disk resou rce for the failed resource. Note: The Server Cluster Recovery Utility ensures that the old and new disk resources are in the same resource group. It will take the resource group offline and transfer the properties of the old resource (such as failover policies and chkdsk settings) to the new resource. It will also rename the old resource to have "(lost)" appended to the name and rename the new resource to be the same as the old resource. Any dependencies on the old resource will be changed to point to the new resource.
- Change the drive letter of the new physical disk to match that of the failed disk. Note: The new physical disk resource must be brought online first and then the drive letter can be changed (on the node hosting the physical disk resource) using the Disk Management snap-in available via Computer Management.
- Once you have validated that the new resource is correctly installed, you should delete the old physical disk resource as it no longer represents a real resource on the cluster.
- Once the cluster is configured, you should restore the application data to the new disk drive.

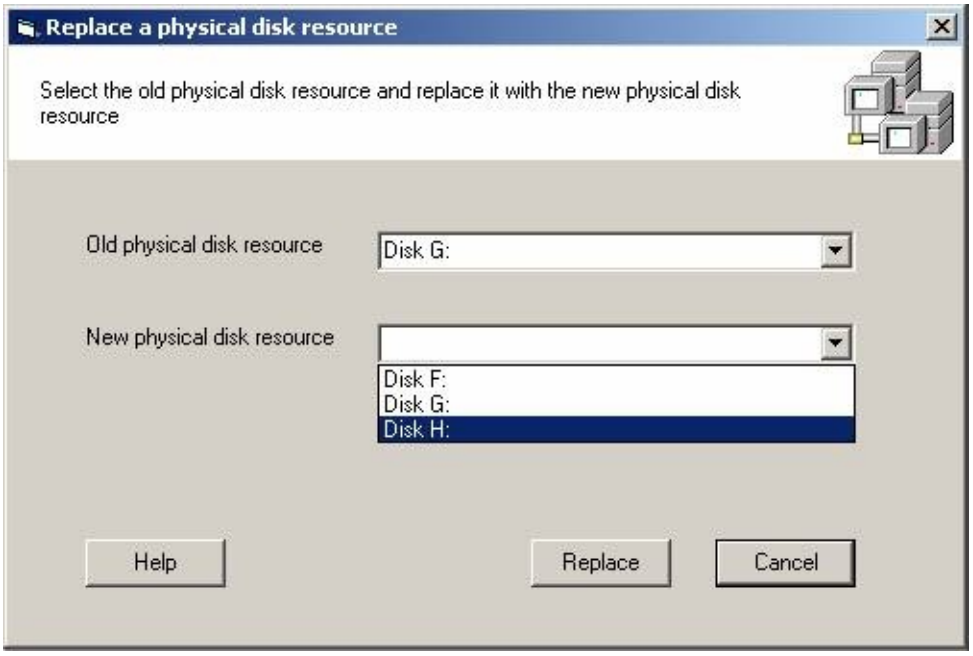
The following shows the affect of running the Server Cluster Recovery Utility. Start with the following cluster configuration. In this case, Disk G has failed and is to be replaced by Disk H. (Note, if the failed disk was in the cluster group, you may have to start the Cluster Administration tool on the cluster node itself using “.” as the cluster to connect to).



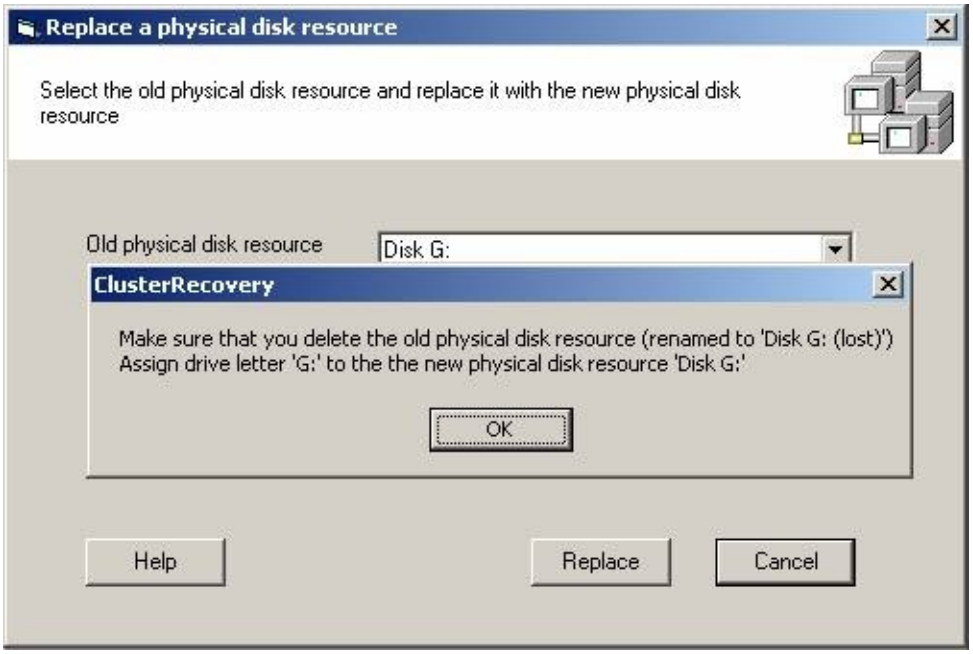
In the cluster recovery utility specify a cluster and select the “Replace a physical disk resource” option and click the Next button.



You now need to select the old (failed) physical disk resource and the new physical disk resource. You can either type the name of the resource or select it from the set of physical disk resources on the cluster (Note: The old and new physical disk resources MUST be in the same resource group for the replacement to succeed).

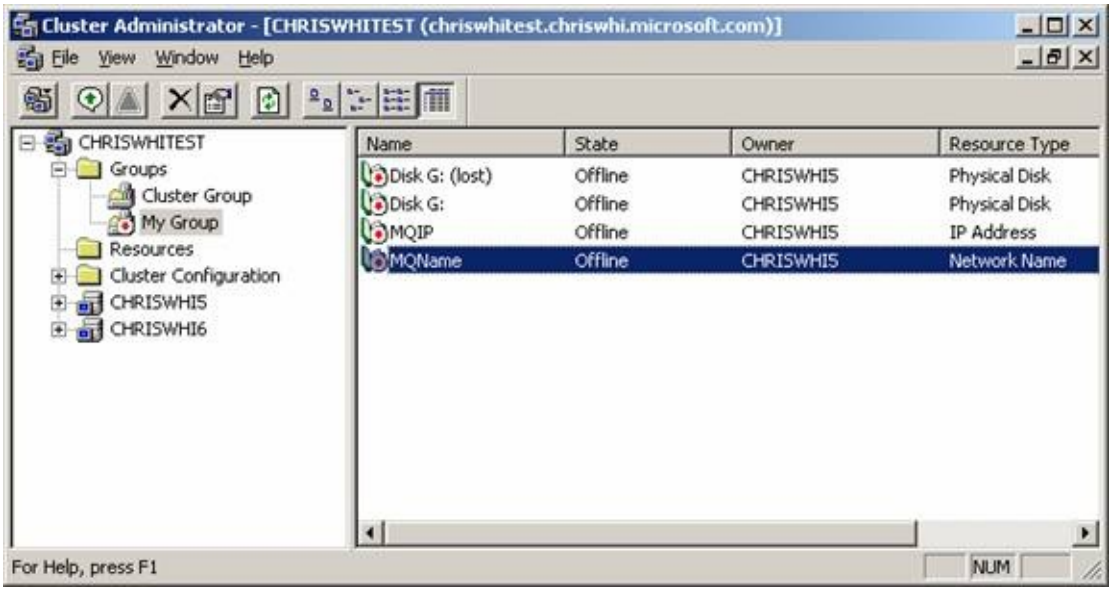


Once you have selected the resources, click the Next button. If the substitution was successful, the following message will appear to remind you of the procedures to ensure correct application operation.



All appropriate public and private properties from the old resource such as failover policies, timeouts, chkdsk attributes etc. are carried from the old resource and applied to the new resource. Any dependencies and/or dependents on the old physical disk will be transferred to the new physical disk and the new

resource is renamed to match the old one and the old resource is renamed with the suffix “(lost)”. After running the cluster recovery utility, the configuration above looks like:



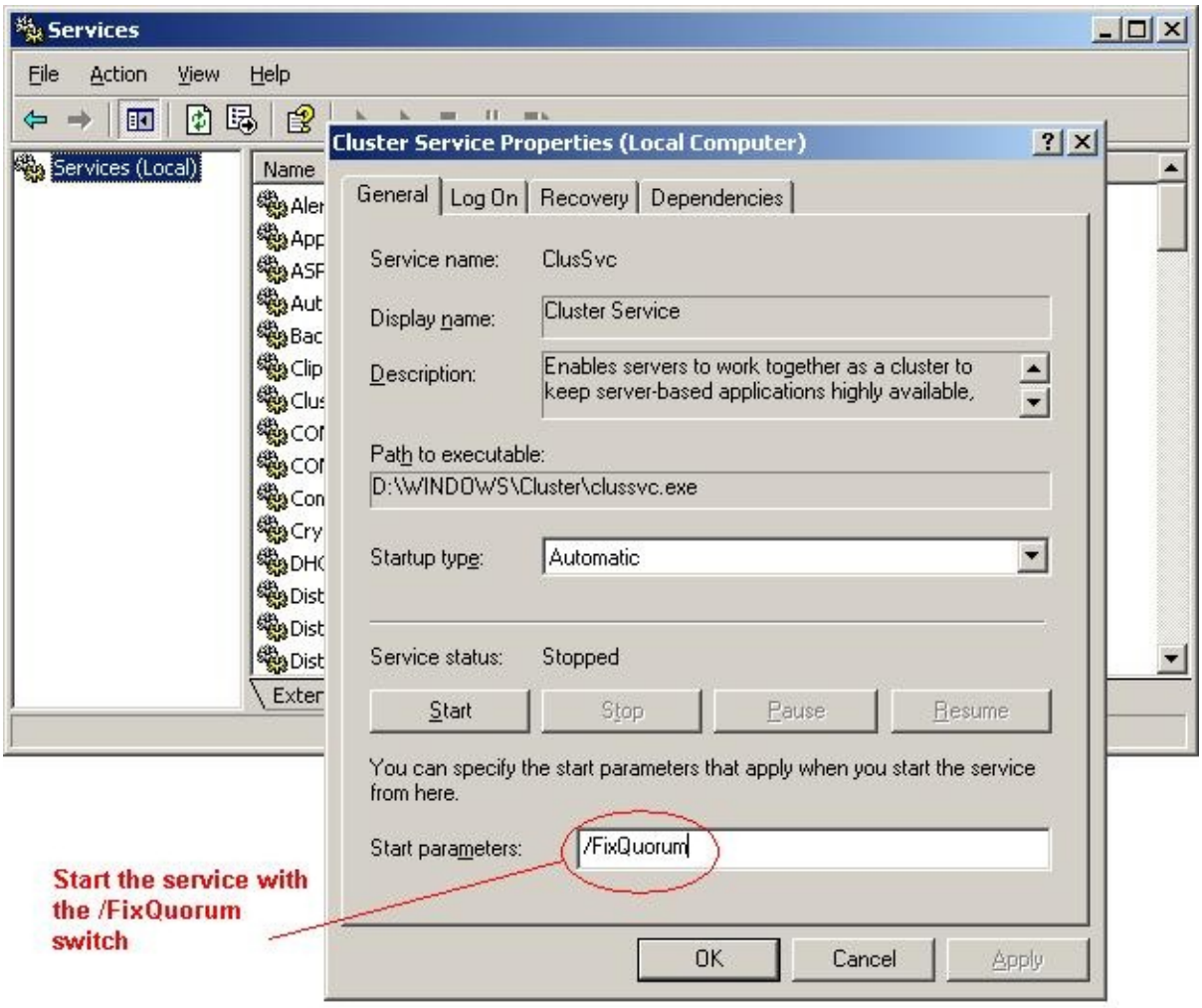
To complete the replacement, you should bring the new disk resource online and use the disk management snap-in to change the drive letter to match the old disk resource (this is necessary because applications that are using the disk will typically reference files on the disk via a drive letter). Once you are happy that the new resource has the cluster properties that you want from the old resource, you can delete the old physical disk resource as it is no longer required.

Recovering the quorum disk

The procedure for actually replacing the quorum disk is identical to replacing a data disk on the shared bus, however, since the cluster is not ordinarily available if the quorum disk is offline, a special procedure is required to startup the cluster in the event of a quorum disk failure.

If the quorum disk fails, the cluster service on ALL nodes in the cluster will stop since none of the nodes will be able to successfully arbitrate for the quorum resource. If this situation occurs, you should use the following procedure to restart the cluster so that it can be repaired using the cluster recovery utility:

- Shutdown all but one of the cluster nodes
- Using the Services Management snap-in (part of Computer Management), start the Cluster service with the "/FixQuorum" switch on the remaining cluster node



This procedure will start up the Cluster service, however, the cluster group (including the quorum resource, the cluster IP address and the cluster network name) will be offline. Other groups may be brought online. You MUST make sure that all resources in the cluster group are offline BEFORE running the cluster recovery utility.

Once the cluster service is up and running, you can use the disk replacement procedure outlined in Recovering a shared disk to replace the old quorum physical disk resource with the new quorum physical disk resource.

Notes:

- Since the cluster IP address and network name are offline, you must use the name of the cluster node when connecting to the cluster

- The replacement may take time, this is due to the Cluster service trying to access the old quorum resource to recover any information.

Once the quorum disk replacement has succeeded, you should restore any resource checkpoint files using the procedures defined in Restoring resource checkpoint files

Related topics:
Recovering a shared disk; Recovering a quorum disk; Restoring resource checkpoint files

Migrating data to a new disk

Although the Server Cluster Recovery Utility was designed to make recovering from disk failures easier, it can be used in other ways. In particular, the cluster recovery utility can be used to migrate data from one cluster disk to another.

Consider, for example, a cluster that has a set of disks attached to it. Initially, a small disk array may have been sufficient for the application, however, over time the application requirements may grow and a new, high performance disk array may be required. A new disk drive or logical disk can be exposed to the cluster and a physical disk resource created. This physical disk resource can then be substituted into the application dependency tree using the same procedures described in section Recovering a cluster disk.