



STONEFLY

Implementing Storage Concentrator FailOver Clusters

Technical Brief

April 2006

All trademark names are the property of their respective companies. This publication contains opinions of StoneFly, Inc. which are subject to change from time to time.

This publication is copyright © 2006 by StoneFly, Inc. and is intended for use only by recipients authorized by StoneFly, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of StoneFly, Inc., is in violation of U.S. copyright law.

Introduction

The StoneFly Storage Concentrator™ is the interface between hosts and storage devices in an IP network. IP-based storage area networks (IP SANs) use the iSCSI protocol over an Ethernet and TCP/IP network. The Storage Concentrator FailOver Cluster is a redundant storage solution comprised of two Storage Concentrators configured so that one is actively managing storage volumes and the other is in standby mode monitoring all fault conditions. If the active Storage Concentrator fails for any reason, then a FailOver event will be initiated. FailOver to the standby unit is automatic and transparent to users.

This application note describes some typical usage scenarios, including:

- FailOver Events
- Using a FailOver Cluster with different types of back-end storage

Note: This paper is not intended to provide detailed setup instructions for a FailOver Cluster. Comprehensive setup and user instructions for FailOver Clusters can be found in the *Storage Concentrator Setup Guide* and *Storage Concentrator User's Guide*.

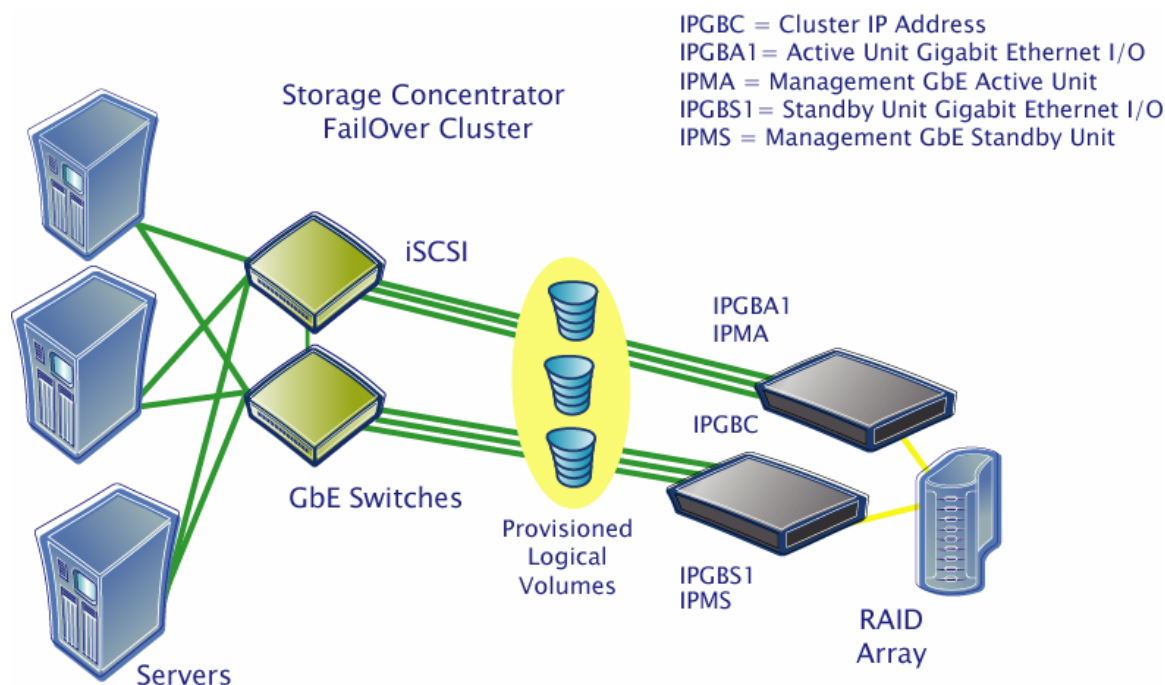
FailOver Cluster Overview

FailOver is an important fault tolerance function of mission-critical systems that require constant accessibility. A Storage Concentrator FailOver Cluster adds a high availability layer to an IP Storage Area Network. If a component of the active system fails, the FailOver Cluster automatically substitutes a functionally equivalent standby system that takes over the operations of the failed active system. FailOver automatically redirects user requests from the failed active system to the standby system.

Storage Concentrator FailOver Cluster: Currently, a cluster consists of two Storage Concentrators, one active unit and one standby unit. A cluster appears as a single entity to hosts on the network. The components of a FailOver Cluster include:

- **Active Storage Concentrator:** The active Storage Concentrator is the Storage Concentrator in a cluster that has active iSCSI sessions with hosts on the network. The active Storage Concentrator configures the volumes and back-end sessions and replicates the metadata to the standby Storage Concentrator in the cluster.
- **Standby Storage Concentrator:** The standby Storage Concentrator monitors all FailOver conditions. In the event of a failure at the active Storage Concentrator, the standby Storage Concentrator takes over the active session. The standby Storage Concentrator in the cluster transparently becomes the active Storage Concentrator. The Storage Concentrator volumes are now managed by the new active unit. There can be only one standby Storage Concentrator.
- **Cluster IP Address:** The cluster IP address is the IP address of the FailOver cluster. Host server initiators view the cluster as a single entity with the Cluster IP address. Each unit in the cluster also has its own IP addresses for both the Gigabit Ethernet I/O traffic and the Management traffic. A single cluster uses either five or seven IP addresses depending on whether each SC has one or two GbE iSCSI ports, as shown below:

Implementing Storage Concentrator FailOver Clusters



FailOver Events: Each type of FailOver event is described below:

A. Management Port Link Failure

The loss of link between the Active unit's Management Port and its switch causes the Standby unit to take control of the storage devices and then become the Active unit. All future access to the storage devices is through the new Active unit. When the fault is corrected, the failed Storage Concentrator will automatically reboot itself and rejoin the Cluster as the Standby unit

B. iSCSI Gigabit Ethernet Port Link Failure

The loss of link between the Active unit's iSCSI Gigabit Ethernet Port and its switch causes the Standby unit to take control of the storage devices and then become the Active unit. When its Gigabit Ethernet connection is restored, the failed unit reboots and becomes the Standby unit in the FailOver Cluster. If you are using the dual GbE connections, you can plug in both connections or a single connections. Only by plugging in both connections to the same switch will you have automatic Adaptive Load Balancing. If a GbE port fails on an active dual-ported unit in the cluster, it will FailOver to the standby unit if it still has two functioning ports.

C. Power Failure to the Active Storage Concentrator

If the power fails to the Active unit but not the Standby unit, the Standby unit immediately takes control of the storage devices and then becomes the Active SC. As soon as power is re-applied to the failed Storage Concentrator it reboots as the Standby unit in the FailOver Cluster.

D. Power Failure to Both Storage Concentrators

Should power fail and then be re-applied to both Storage Concentrators in a FailOver Cluster, both units will reboot resulting in a negotiation between the units where one

Implementing Storage Concentrator FailOver Clusters

unit becomes the Active unit and the other becomes the Standby unit. No reconfiguration is required by the Storage Administrator.

E. Loss of Connection to Storage Devices

FailOver Clusters are established such that each Storage Concentrator has a separate connection to each storage device. These storage devices present the same Logical Units (LUNs) to each Storage Concentrator. If the Active unit loses connection to a storage device and the Standby unit continues to be connected, then the Standby unit takes control of all the storage devices and becomes the Active unit. The connection may be re-established between the failed unit and the storage device at a later time. The failed Storage Concentrator will then reboot to establish it as the new Standby unit in the FailOver Cluster.

Configuring Storage Behind a Storage Concentrator FailOver Cluster

The storage subsystems behind a Storage Concentrator FailOver Cluster should support RAID and have two SCSI ports or two Fibre Channel ports. For FailOver to operate correctly, storage must be configured so that it is symmetrical and connected to both Storage Concentrators. A system that has a spare JBOD on the active side, for example, may not FailOver as expected due to the asymmetrical configuration. The storage must be configured such that both Storage Concentrators can each view all LUNs.

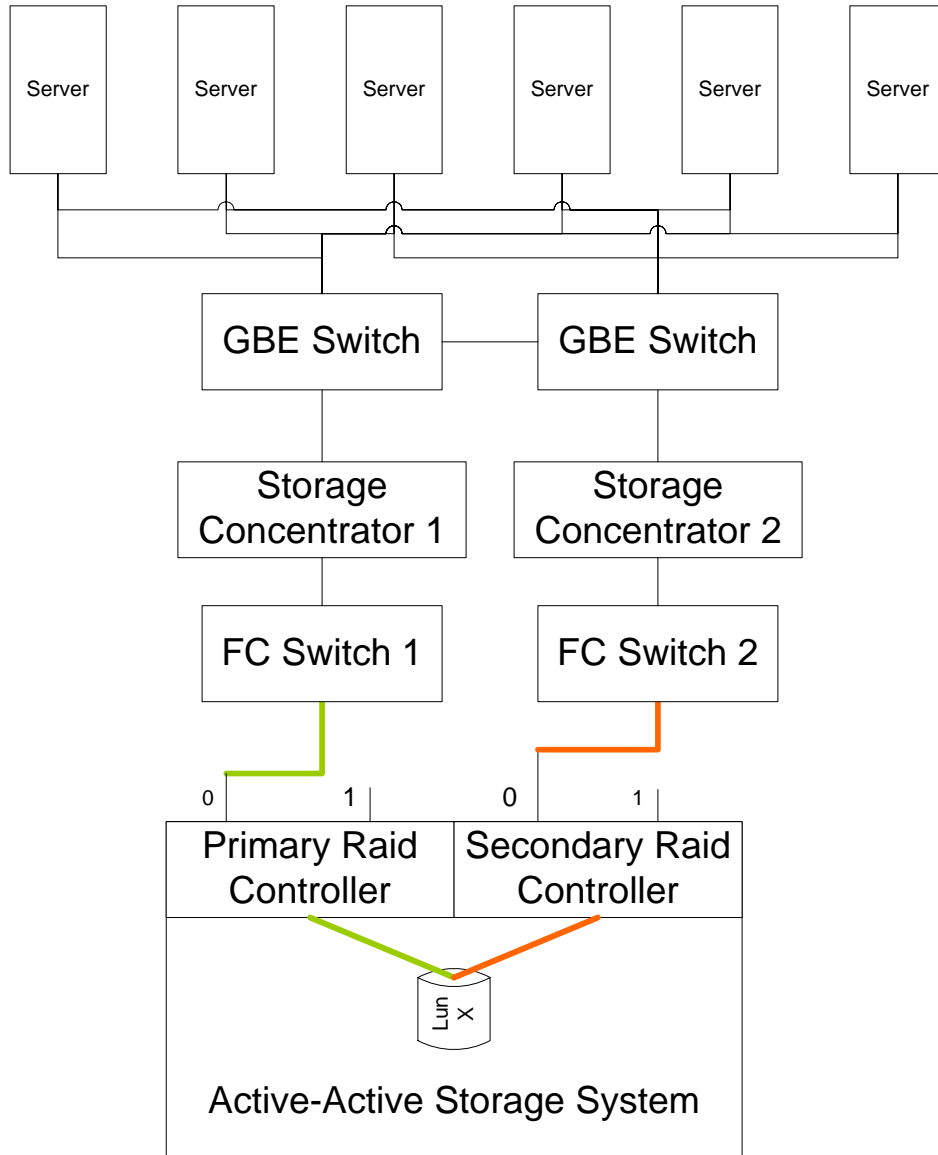
- For storage with active-active dual controllers, each Storage Concentrator should be connected to a different port on each RAID controller.
- For single controller RAID systems, the storage must have two ports on a single controller with a Storage Concentrator connected to each. In the case of active/passive configurations, there must be two ports on the active controller with the Storage Concentrators both connected to this active controller.
- Fibre Channel connections configured through a switch must be zoned correctly make the LUNS visible to both units in the FailOver configuration.
- Users can employ two Gigabit Ethernet switches and two Fibre Channel switches for full redundancy.

Note: Contact StoneFly for a list of certified storage disk arrays for FailOver

Configuring the Cluster with Active-Active Dual Controller Storage

This section is an overview of setting up Storage Concentrator FailOver configuration with a dual-controller, active/active RAID storage array that operates in a primary / secondary mode in a Fibre Channel network. On the Storage Array, a full high-availability solution can be implemented with two Fibre Channel Switches for full redundancy and no manual intervention. The same configuration could be designed with a single FC Switch and appropriate Zoning but this would create a single point of failure in the Fibre Channel network; likewise, the user could build the configuration with a single Gigabit Ethernet switch.

Implementing Storage Concentrator FailOver Clusters



Configuration and setup:

1. Allocate LUN X on the Storage Array and assign the LUN to be associated with both the primary and secondary controller.
2. Configure the FC switches so that the Storage Concentrator SC1 can see LUN X on the primary controller. Bring up SC1 and verify that LUN X is visible as a resource. This is exactly the same for an existing single SC configuration.
3. Zone FC Switch 2 to make LUN X visible to SC2 on the secondary controller.
4. Bring up SC2 and verify that LUN X is visible as a resource. Do not have this Concentrator configured to allow access by any of the connected servers. No I/O should be allowed at this time on SC2.

Implementing Storage Concentrator FailOver Clusters

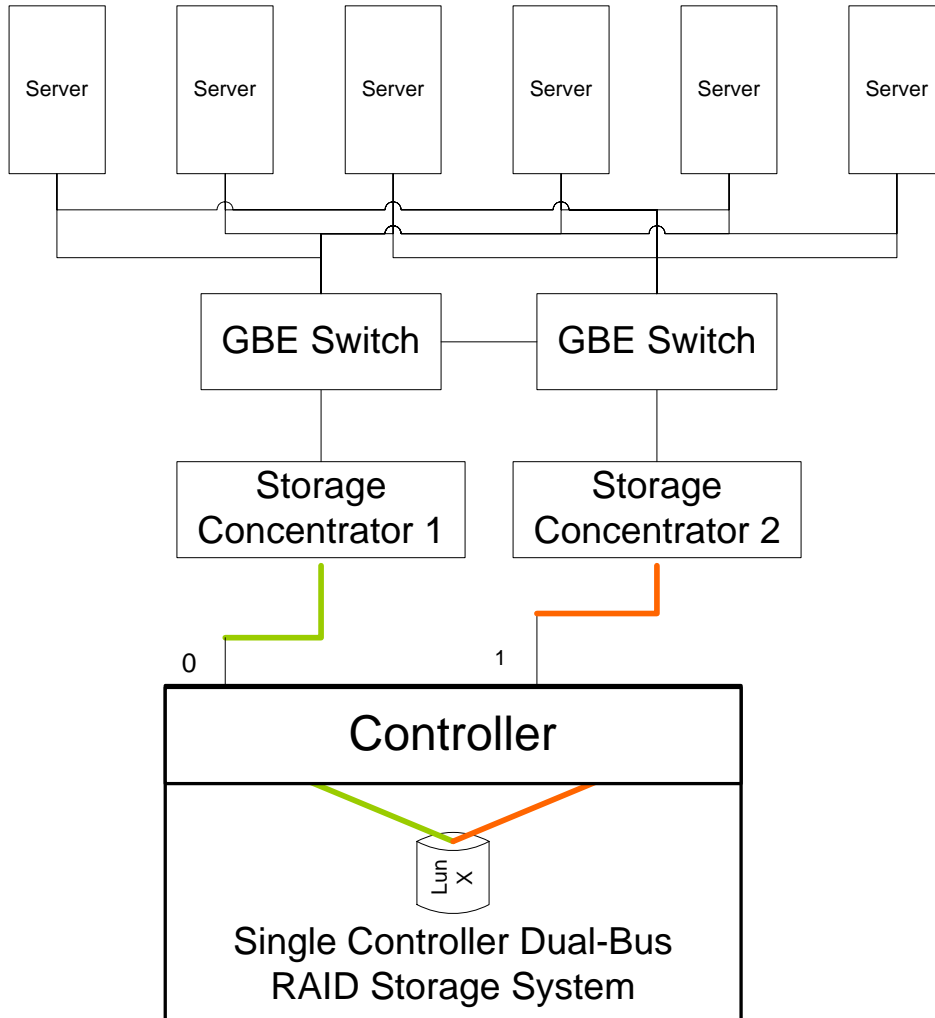
5. Open the user interface of SC1 and configure it to accept SC2 as a member of the FailOver Cluster.
6. Open the user interface of SC2 and configure it to join SC1 as a member of the FailOver Cluster.
7. The two systems will now form the FailOver Cluster. SC2 will receive a copy of the metadata specifying all attached storage, allocated volumes and access control lists. Effectively SC2 will inherit the full configuration of SC1 (except for the management port IP address – each system will maintain their own management port address).
8. SC2 will attempt to open LUN X and read block 0. This will allow the system to verify that the label on the storage device matches on both sides and check that the same LUN is accessible on both controllers. The Storage Array will migrate the LUN whenever a read request is received. There is a short delay while all pending I/Os are completed on the primary LUN controller before the read operation can be fulfilled on the secondary controller. While the secondary controller read is underway the LUN is not accessible on the primary controller. During this time I/O operations on the primary are returned with a busy status and will be retried a few seconds later, which will cause the LUN to migrate back to the primary. This read is only used during system startup and not during normal operations. The impact at the system level will be barely noticed. No I/O operations are lost, only delayed for a few seconds.
9. The operator should now verify that the Active / Standby Storage Concentrator Cluster has formed and that all resources are now redundant, dual pathed and not critical. This status is indicated on the user interface of SC1.
10. During a FailOver operation the iSCSI IP address will be migrated from SC1 to SC2. SC2 will cause the Storage Array to migrate the LUN to its secondary controller and become available for I/O operations. The Ethernet network is redirected to the MAC address of SC2 with an ARP response packet and finally the servers will reconnect to SC2 and retry any uncompleted I/O operations. This whole transition should be complete in less than 15 seconds and will be accomplished with no lost I/O operations.
11. SC1 is now available to attempt to restart and reform the Cluster with SC2. Operator intervention may be required to repair any physical failures.

Configuring the Cluster with Single Controller Dual-Bus Storage

In a single controller Storage Array, each of the Storage Concentrators must be connected to the same controller, but to different buses. This will allow both Concentrators to view the same LUN.

1. Bring up SC1 and verify that LUN X is visible as a resource. This is exactly the same for an existing single SC configuration.
2. Bring up SC2 and verify that LUN X is visible as a resource. Do not have this Concentrator configured to allow access by any of the connected servers. No I/O should be allowed at this time on SC2.

Implementing Storage Concentrator FailOver Clusters



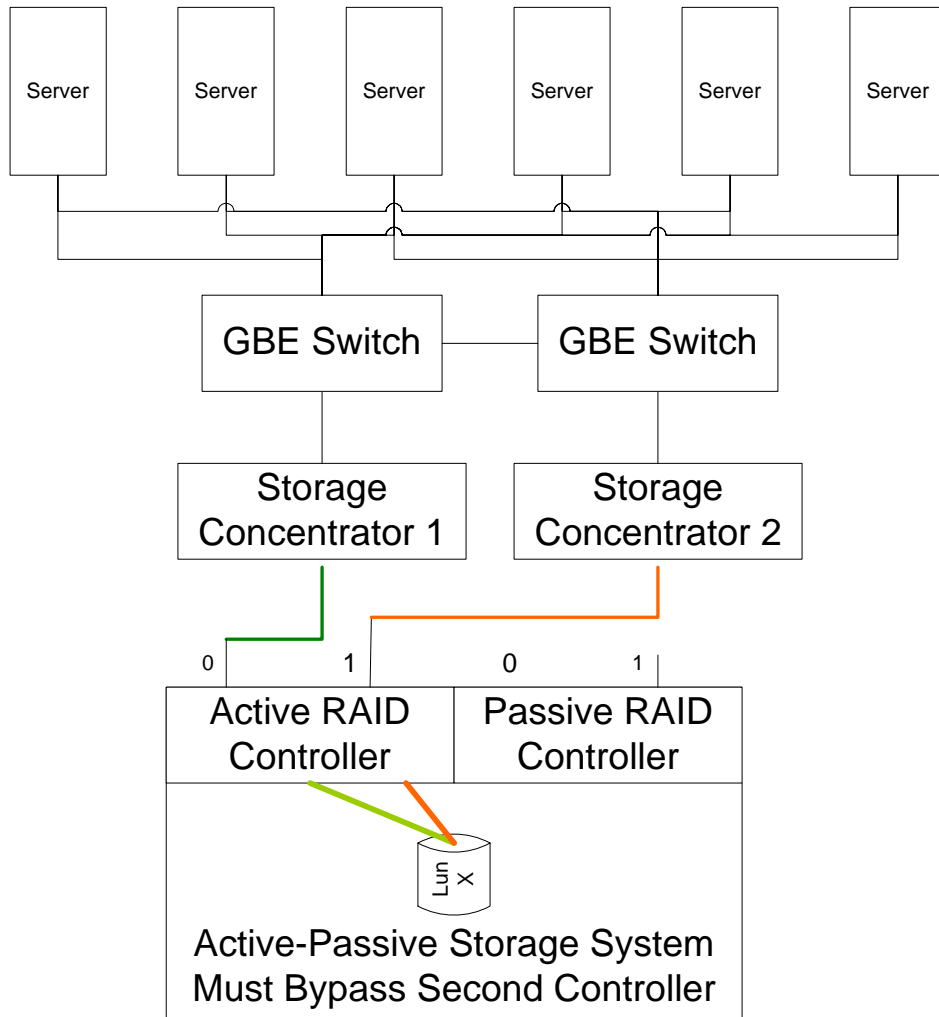
3. Open the user interface of SC1 and configure it to accept SC2 as a member of the FailOver Cluster.
4. Open the user interface of SC2 and configure it to join SC1 as a member of the FailOver Cluster.
5. The two systems will now form the FailOver Cluster. SC2 will receive a copy of the metadata specifying all attached storage, allocated volumes and access control lists. Effectively SC2 will inherit the full configuration of SC1 (except for the management port IP address – each system will maintain their own management port address).
6. SC2 will attempt to open LUN X and read block 0. This will allow the system to verify that the label on the storage device matches on both sides and check that the same LUN is accessible. This read is only used during system startup and not during normal operations.

Implementing Storage Concentrator FailOver Clusters

7. The operator should now verify that the Active/Standby Storage Concentrator Cluster has formed and that all resources are now redundant, dual pathed and not critical. This status is indicated on the user interface of SC1.
8. During a FailOver operation the iSCSI IP address will be migrated from SC1 to SC2. SC2 will become available for I/O operations. The Ethernet network is redirected to the MAC address of SC2 with an ARP packet and finally the servers will reconnect to SC2 and retry any uncompleted I/O operations. This whole transition should be complete in less than 15 seconds and will be accomplished with no lost I/O operations.
9. SC1 is now available to attempt to restart and reform the Cluster with SC2. Operator intervention may be required to repair any physical failures.

Configuring the Cluster with an Active-Passive Controller Storage Array

Some storage arrays with active-passive dual controllers do not allow viewing the same LUN on both the active and passive controller at the same time. If this is the case, then the Storage Array is compatible with the Storage Concentrator FailOver Cluster **only** if it allows simultaneous connection to two buses on the active controller. If there are not two buses on the active controller, then it will not be possible to use the Storage Array with the Storage Concentrator FailOver Cluster.



Implementing Storage Concentrator FailOver Clusters

In an active-passive Storage Array, each of the Storage Concentrators must be connected to the same controller, but to different buses. This will allow both Concentrators to view the same LUN.

1. Bring up SC1 and verify that LUN X is visible as a resource.
2. Bring up SC2 and verify that LUN X is visible as a resource. If LUN X is visible to SC2, then continue to set up this storage device with the FailOver Cluster. The remaining steps are identical to those for setting up a single controller dual-bus storage array.

Conclusion

The Storage Concentrator FailOver Cluster forms the basis of a redundant and reliable IP-based Storage Area Network. The Storage Concentrator FailOver pair features redundant CPUs, power supplies, hard drives, port connections and operating system. All customer configuration data from the active Storage Concentrator is replicated to the standby unit for unparalleled data protection. Redundant Storage Concentrators ensure storage volumes are continuously available.

Storage Concentrators fit seamlessly into existing storage and TCP/IP data networks and provide storage provisioning for heterogeneous servers and storage systems in a SAN. Setting up and managing the provisioned storage is an administrative task that is easily accomplished with the Storage Concentrator.

With StoneFly Storage Concentrators, large, medium, and small organizations alike can finally leverage the benefits of storage networking using IP, including security, manageability, and quality of service, and use of their existing infrastructure.