EMC Data Domain Boost for Symantec NetBackup and NetBackup Storage Unit Group Failover

Applied Technology

Abstract

Service interruptions that result in failed or suspended backups can be detrimental in a backup environment. However, many commercially available backup applications are designed to accommodate service interruptions and can automate failover using policy-based algorithms. This white paper explores how to leverage Storage Unit Group failover to achieve N+1 redundancy when using EMC® Data Domain® Boost software with Symantec NetBackup OpenStorage.

June 2010
Table of Contents

Executive summary .............................................................................................................. 4
Introduction .......................................................................................................................... 5
    Audience .......................................................................................................................... 5
Storage Unit Groups ............................................................................................................. 5
    Storage Unit Group selection criteria .............................................................................. 6
    Cost-effective system failover .......................................................................................... 7
Failure scenarios ................................................................................................................. 8
    Testing with Storage Unit Groups only ............................................................................. 8
    Testing with Storage Unit Groups and storage lifecycle policies ...................................... 9
Standby Data Domain system recovery .............................................................................. 10
    Examining failures in the primary Data Domain system .................................................. 11
    Returning a standby system to an idle state .................................................................... 12
    Automating the recovery process .................................................................................... 13
Conclusion .......................................................................................................................... 13
Executive summary

Symantec NetBackup has been architected such that it is able to accommodate backup device failures without the need for backup devices clustered for high availability. In fact, NetBackup goes a step further and accommodates media server failures with its ability to bypass an offline media server so that backups can continue through available media servers. This functionality is made possible by a combination of product features that include tighter integration between media servers and storage systems using OpenStorage, backup device status tracking, Storage Units, Storage Unit Groups, and automated media server load balancing.

EMC® Data Domain® Boost software and NetBackup Storage Unit Groups can be configured to take advantage of the N+1 failover configurations architected into NetBackup. An N+1 configuration allows administrators to configure a standby EMC Data Domain system to protect against failures on any number of primary Data Domain systems. In the unlikely event of an unplanned outage on a Data Domain system, a standby Data Domain system is able to accommodate backups that are automatically restarted or resumed by NetBackup. Overall, the more active systems that can share the same standby system, the more cost-effective the solution will be compared to backup devices clustered for high availability.

Figure 1 depicts an environment that employs multiple NetBackup media servers, as well as Data Domain systems configured for N+1 failover. The loss of a NetBackup media server would have a negligible impact on backup or restore jobs as an offline media server is bypassed when performing backup or restore jobs. Likewise, the loss of a primary Data Domain system would have minimal impact on backup jobs as a standby Data Domain system would automatically be selected for use. Restore requests could be fulfilled from a duplicate backup image copy residing on an optimized duplication destination Storage Unit.

The true impact of a failure or failover event on backup processes should be understood before considering the use of storage systems clustered for high availability as backup devices. Often, a failure for any one system in a clustered storage system results in a failed or suspended backup job for the entire cluster. In these cases, it may be advantageous to leverage the architectural features included with NetBackup instead of dealing with the cost and complexity of something that does not really provide transparent high availability or seamless cluster failover from the perspective of the backup application.
Introduction
This white paper outlines how to leverage Symantec NetBackup Storage Unit Group failover to achieve N+1 redundancy when using EMC Data Domain Boost software with Symantec NetBackup OpenStorage.

The paper describes NetBackup Storage Unit Groups and how to use selection criteria and storage lifecycle policies with Data Domain deduplication storage systems to provide cost-effective system failover. Numerous failure scenarios are presented to demonstrate this failover capability. Finally, an examination of these scenarios and operational recovery to a standby Data Domain system in such failure scenarios is discussed.

Audience
This white paper is intended for EMC customers, technical consultants, partners, and members of the EMC and partner professional services community who are interested in learning more about Data Domain Boost for Symantec NetBackup and NetBackup Storage Unit Group failover.

Storage Unit Groups
Within the context of NetBackup, a Storage Unit is an abstraction of a physical storage device or devices. A simple example of a Storage Unit might be a tape library containing two tape drives. By referencing a Storage Unit rather than the devices the Storage Unit may contain, backup policies are able to select an available resource to fulfill a backup request instead of being hard-coded to use a specific device. In the case of our example tape library Storage Unit, this might equate to using the second tape drive when the first tape drive inevitably fails for one reason or another. When using DD Boost for NetBackup, Storage Units map to disk pools in NetBackup and each disk pool maps to a Logical Storage Unit (LSU) on Data Domain systems.

Storage Unit Groups add another layer of abstraction to this concept. Multiple Storage Units can be added to a Storage Unit Group and organized in a prioritized list to increase the likelihood of successful backups.

Figure 2 shows the hierarchy of objects in NetBackup. LSUs on a Data Domain system are mapped to a disk pool in NetBackup, which is then mapped to a Storage Unit. Multiple Storage Units can then be aggregated into a Storage Unit Group and backup policies can then be directed to the aggregated Storage Unit Group.
Applied Technology  6

Figure 2. Hierarchy of objects in NetBackup

**Storage Unit Group selection criteria**

Storage Unit selection criteria within a Storage Unit Group determines the order and conditions under which individual Storage Units are selected for use. Storage Unit Groups allow multiple selection criteria for choosing Storage Units within a group. As of NetBackup version 6.5.4, any of the following selection criteria can be used:

- **Prioritized**: Choose the first Storage Unit in the list that is not busy, down, or out of media.
- **Failover**: Choose the first Storage Unit in the list that is not down or out of media.
- **Round Robin**: Choose the least recently selected Storage Unit in the list.
- **Media Server Load Balance**: Choose the best available media server that can access the Storage Unit in the list.

The scope of this paper is purposely limited to reviewing the use of the Failover selection criteria. Failover is defined as:

- NetBackup chooses the first Storage Unit in the Storage Unit Group list that is not down, out of media, or full.
- If the first Storage Unit is busy, the backup policy waits for the resource to become available.
- In the event of a failover, the other Storage Units in the Storage Unit Group list are used.
- If the maximum concurrent job limit has been reached, NetBackup queues subsequent jobs to wait for the first Storage Unit to become available again.

Figure 3 shows an example Storage Unit Group. Two Data Domain systems have been added to the Storage Unit Group. The first Storage Unit in the group is the primary backup target and the second Storage Unit in the group has been designated as a standby Data Domain system. Storage Unit selection criteria has been set to Failover, indicating that the standby Data Domain system will be used only in the unlikely event of a failure on the primary Data Domain system.
As a best practice, EMC recommends using only the Failover Storage Unit selection criteria with Data Domain systems. Leveraging the selections Prioritized or Round Robin will not consistently send the same backups to the same system repeatedly, which will yield inferior deduplication ratios. With Failover, backups would be sent only to a standby Data Domain system in the event of a failure, resulting in optimal deduplication ratios. The efficiency of optimal deduplication ratios combined with N+1 failover for critically important backups is easily configured using Storage Unit Groups and the Failover selection criteria.

**Cost-effective system failover**

EMC recommends creating a unique OpenStorage disk Storage Unit to act as a failover Storage Unit. This Storage Unit is active in a standby role such that it will be used only in the event of failure of a primary Data Domain system. A single standby Data Domain system can be added to any number of Storage Unit Groups to provide N+1 failover for a number of primary Data Domain systems.

Figure 4 shows how to set up the same single Data Domain system (dd120b_Standby) as a standby system for multiple primary Data Domain systems. Each of the four Storage Unit Groups shown has been configured to use the Failover Storage Unit selection criteria.
EMC Data Domain Boost for Symantec NetBackup and NetBackup Storage Unit Group Failover
Applied Technology

Failure scenarios
This section examines a variety of failures injected into a NetBackup and Data Domain environment configured for N+1 failover. The investigation results show the impact of the injected failure on NetBackup, as well as the action taken by NetBackup as a result of the failure event.

Testing with Storage Unit Groups only
Table 1 summarizes testing with Storage Unit Groups only where various types of failures were manually injected to simulate different failure scenarios and highlights the corresponding NetBackup response and result. In all cases, NetBackup was able to overcome the injected failure and used the standby Data Domain system, whether the backup jobs restarted automatically or manually. In Cases 1, 2, and 5, NetBackup automatically selected the standby system prior to backup job execution, resulting in zero production impact. In Cases 3 and 4, backup jobs failed but were retried or restarted. In all cases, when the injected failure was removed subsequent backup jobs used the primary Data Domain system without operator intervention.
Table 1. Testing results with the Storage Unit Group only

<table>
<thead>
<tr>
<th>Injected failure</th>
<th>NetBackup response and result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Power disconnected from the primary Data Domain system</td>
<td>NetBackup automatically used the standby Data Domain system</td>
</tr>
<tr>
<td></td>
<td>No impact to production</td>
</tr>
<tr>
<td>2. File system disabled on the primary Data Domain system</td>
<td>NetBackup automatically used the standby Data Domain system</td>
</tr>
<tr>
<td></td>
<td>No impact to production</td>
</tr>
<tr>
<td>3. DD Boost disabled on the primary Data Domain system</td>
<td>Backup failed</td>
</tr>
<tr>
<td></td>
<td>Restarted backup used the standby Data Domain system</td>
</tr>
<tr>
<td></td>
<td>No impact to production</td>
</tr>
<tr>
<td>4. Network disconnected on the primary Data Domain system</td>
<td>Backup failed</td>
</tr>
<tr>
<td></td>
<td>Restarted backup used the standby Data Domain system</td>
</tr>
<tr>
<td></td>
<td>No impact to production</td>
</tr>
<tr>
<td>5. Disk pool high water mark parameter set to 1%, simulating an out-of-media condition (disk full condition) on the primary Data Domain system</td>
<td>When available capacity was less than the high water mark, backup automatically used the standby Data Domain system</td>
</tr>
<tr>
<td></td>
<td>No impact to production</td>
</tr>
</tbody>
</table>

**Testing with Storage Unit Groups and storage lifecycle policies**

Using NetBackup Storage Unit Groups and storage lifecycle policies in conjunction with Data Domain Boost is a powerful combination. With Data Domain systems as backup and duplication destinations, storage lifecycle policies enable easy management of scheduled, optimized duplication tasks by leveraging DD Boost managed file replication.

Figure 5 shows a test in which the initial backups are sent to Data Domain systems configured in a Storage Unit Group with the Failover selection criteria. The duplication destination in this configuration is set to a Data Domain system configured as a Storage Unit since Storage Unit Groups cannot be used as duplication destinations. When a failure was simulated on the primary Data Domain system in the Storage Unit Group, subsequent backups and optimized duplication operations continued executing successfully after the failure by leveraging another system in the Storage Unit Group.
Table 2 summarizes testing with Storage Unit Groups and storage lifecycle policies where failures were manually injected to simulate different failure scenarios and highlights the corresponding NetBackup response and result.

**Table 2. Testing results with Storage Unit Groups and storage lifecycle policies**

<table>
<thead>
<tr>
<th>Injected failure</th>
<th>NetBackup response and result</th>
</tr>
</thead>
<tbody>
<tr>
<td>None, no failure was injected</td>
<td>Backup used the primary Data Domain system</td>
</tr>
<tr>
<td><strong>Note: In this test, a Storage Unit Group was being used as a backup destination within a storage lifecycle policy</strong></td>
<td>Optimized duplication executed normally</td>
</tr>
<tr>
<td></td>
<td>No impact to production</td>
</tr>
<tr>
<td>The file system was disabled on the primary Data Domain system within the backup destination Storage Unit Group</td>
<td>Backup used the standby Data Domain system</td>
</tr>
<tr>
<td></td>
<td>Optimized duplication executed normally</td>
</tr>
<tr>
<td></td>
<td>No impact to optimized duplication</td>
</tr>
</tbody>
</table>

The test cases presented here show that a Storage Unit Group can be used as a backup destination in a storage lifecycle policy. However, it is important to note that NetBackup does not allow Storage Unit Groups to be used as a duplication destination.

**Standby Data Domain system recovery**

Standby Data Domain systems should be configured within Storage Unit Groups for failover use only. In the event that a primary Data Domain system failure caused a standby Data Domain system to be used, backup images from the standby Data Domain system would be fully recoverable and could be used for optimized duplication.

However, the challenge a failover situation creates is that backup images typically retained on a primary Data Domain system are now utilizing storage resources on the standby Data Domain system. Leaving
backup images on the standby Data Domain system may be less than optimal, as it will result in a reduced
deduplication ratio on the standby Data Domain system. It may also impact the ongoing ability of the
standby system to serve its ultimate purpose, which is to be utilized only in the event of a failure.

EMC recommends the use of manually invoked optimized duplication to address this situation. Once a
primary Data Domain system has been returned to an operational state, duplicating any backup images
residing on a standby Data Domain system to the original primary system can be accomplished via the
NetBackup user interface. Once any backup images residing on a standby Data Domain system have been
duplicated to the correct primary Data Domain system, they should be expired from the standby system.
Subsequent to expiration, the Data Domain file system on the standby system should be cleaned. These
actions return the standby system to its original state, ready for use again in the unlikely event of a primary
Data Domain system failure.

Figure 6 shows the result of a primary Data Domain system failure where backup images written to a
standby Data Domain system are duplicated back to the original primary Data Domain system after it has
returned to an operational state. This step assists in returning the standby Data Domain system to its
intended state.

![Figure 6. Standby system recovery](image)

**Examining failures in the primary Data Domain system**

Messages, alerts, and reports all serve to inform the user of a failover condition in the backup environment.
Depending on the type and timing of a failure, NetBackup may automatically retry a failed job, which will
subsequently run to completion using a standby Data Domain system. In other cases, there may not have
been a failed or retried job because NetBackup detected the failure and automatically selected the standby
Data Domain system in the Storage Unit Group for use before a failure occurred.
As long as there are no failover events, the NetBackup Images on Disk report executed against the standby Data Domain system in the Storage Unit Group should indicate that no backup images are present. If backup images are present, this indicates that the standby system was utilized as the result of failover activity. The report provides details sufficient to identify the backup policy or policies that wrote the image(s). This information can be used to determine which backup policy or policies failed and which backup image(s) need to be moved back to the primary system(s) using optimized duplication to return the primary system(s) to their pre-failover state.

**Returning a standby system to an idle state**

The process of returning a standby system to an idle state consists of three steps:

1. Manually invoke optimized duplication to copy the backup image from the standby Data Domain system to the primary Data Domain system.

   When the source and destination images reside on Data Domain systems configured with DD Boost, optimized duplication will be used to create the duplicate backup image. The NetBackup catalog utility can be used to locate backup images residing on a standby system for the purpose of duplicating them to the appropriate primary system.

   ![Figure 7. Manually invoking optimized duplication](image)

   Figure 7 shows the NetBackup dialog box used to invoke duplication of a given backup image.

2. Expire the images written to the standby Data Domain system. This step should be performed only after all backup images on the standby system have been duplicated to the correct primary Data Domain system(s).

   - Backup images on the standby system can be expired individually with the NetBackup catalog utility.
   - The entire disk pool on the standby system can be expired with the NetBackup “bpexpdate” command.

3. Clean the file system on the standby Data Domain system.

   - The Data Domain “filesys clean start” command initiates the cleaning process.
   - The Data Domain “filesys clean watch” command can be used to monitor the cleaning process.
   - Alternatively, the Data Domain Enterprise Manager GUI can be used to monitor an in-progress cleaning operation.

When completed, the standby Data Domain system has been returned to an idle state and is ready to resume its role as a standby system.
Consider the following additional notes:

- In cases where optimized duplication is used as part of the production workflow, all optimized duplication jobs should be allowed to complete prior to expiring the associated backup image.
- The secondary copy of the backup image will automatically be set as the primary copy of the backup image after the initial copy written to a standby system is expired. In cases where optimized duplication is used as part of the production workflow, this may mean that any manually duplicated copies on the primary system will need to be manually set as the primary copy.

**Automating the recovery process**

Although it is conceptually simple, the process of returning a standby system to an idle state can create significant administrative overhead if:

- Tens or hundreds of backup images have been written to a standby system as the result of a primary system failure.
- Multiple primary systems failed (during a power interruption, for example) and each backup image on a standby system needs to be duplicated to different primary systems.
- Images written to a standby system are being duplicated as the result of a storage lifecycle policy and cannot be expired.
- The primary copy property needs to be set on a specific backup image copy number.

Some users may seek to create customized scripts that automatically detect failures and take corrective actions. Users seeking an automated process to return a standby Data Domain system to an idle state should contact their EMC sales team for information regarding pre-packaged script-based solutions.

**Conclusion**

Although many backup storage devices claim high availability, the operational continuity these systems provide should be carefully examined before making an investment. For example, some of these systems introduce interruptions to service that result in failed, restarted, or retried backup jobs as there is no coordination between the backup application and storage system. Therefore, these systems may not offer any strategic advantage compared to the use of failover Storage Unit Groups.

Clustered solutions that do not have tight integration with backup applications may also introduce interruptions to service that result in failed, restarted, or retried backup jobs. Likewise, active/active cluster products may experience performance degradation if one of the active nodes goes down. Although active/passive cluster products may minimize this performance degradation, they increase the overall cost of the solution by requiring an additional passive node for each active node.

In comparison, Data Domain Boost software allows tighter integration between NetBackup and Data Domain systems. This enables Data Domain – both single-controller systems and the EMC Data Domain Global Deduplication Array – to take advantage of the N+1 failover features available in NetBackup to eliminate the need to deploy specialized clustered solutions. Data Domain Boost and NetBackup Storage Unit Groups configured for Failover selection criteria represent a viable, cost-effective alternative for many environments.