WHITE PAPER: DATA PROTECTION

symantec.™

Confidence in a connected world.

Veritas NetBackup™ 6.5: Designing and Implementing Backups Using Storage Lifecycle Policies

Alex Davies | November 2007

White Paper: Data Protection

Veritas NetBackup™ 6.5:

Designing and Implementing Backups Using Storage Lifecycle Policies

Contents

Contents *(cont'd)*

Abstract

Veritas NetBackup™ 6.5 introduces two new features, Data Classifications and Storage Lifecycle Policies, that are intended to simplify the way in which backed-up data is organized. This paper looks at how these features relate to recovery service levels and how they can be deployed to help ensure that backup storage is used cost-effectively while still maintaining appropriate levels of recoverability.

1.0 An overview of recovery service levels

A service level is an agreed standard (normally a minimum) for the delivery of a service between a customer and a supplier. In the world of data protection, the service level is based on recovery capability—there is no such thing as a "backup service level." Two key concepts underpin all recovery service levels:

Recovery point objective (RPO)—The most recent state to which an application or server can be recovered in the event of a failure. The RPO is directly linked to the frequency of the protection process; if the application is protected by backups alone, then it means how often a backup is run.

Recovery time objective (RTO)—The time required to recover the application or server to the RPO from the moment that a problem is detected. Many factors influence the RTO—including the provisioning of hardware and the roll-forward time for application transaction logs—but one constant factor is the time needed to restore the data from the backup or snapshot that forms the RPO.

When setting up a data protection system, it is fairly typical to define three or four different service levels, or tiers, for data protection to meet the requirements of applications that are of varying importance to the business. An example might be something like this:

- **Platinum service level**—Uses snapshot backup technologies to take frequent backups of mission-critical applications such as order processing systems and transaction processing systems; typical RPO and RTO of one or two hours
- **Gold service level**—Uses frequent backups, perhaps every six hours or so, for important but non-critical applications such as e-mail, CRM, and HR systems; typical RPO and RTO of 12 hours or less
- **Silver service level**—daily backup, used to protect non-critical (such as user file and print data) and relatively static data. Typical RPO and RTO of one or two days.

In many cases a fourth tier with longer RPO and RTO is added for data in, for example, test and development environments, where data is not critical to the business or can be easily recreated with relatively little time and effort.

The reason for using a tiered system of this kind is that the cost of storage devices that can support short periods of RTO and RPO is considerable. Applying longer RPO and RTO values to non-critical data allows the use of lower-cost backup media and longer intervals between backups (resulting in less data being held on backup storage and in further reduced costs).

Recovery service levels are generally oriented toward recovery to the last "known good state" and do not incorporate policies for retention of the data beyond the time of the next backup. However, many applications, particularly the more critical ones, are often given long backup retention periods, either for specific compliance reasons or due to some requirement of the application owners.

One problem with associating long retention periods with backups of "critical" applications is that, in order to meet the short RPO/RTO targets of the initial recovery period, a high-cost storage medium is often used to store the backup data. Long-term retention on a high-cost storage medium is capital-intensive, and it is therefore desirable to transfer the data to a lower-cost storage medium once the initial rapid recovery requirement has passed (i.e., when a more recent RPO exists). Once the initial recovery period has passed, it is reasonable to assume that the RTO requirements can be relaxed and that this relaxation can be extended as the time from when the backup was created increases. If the RTO is increased, the data can be moved to slower, lower-cost storage without contravening the service level—but often, this is not done. One main reason for this is that the cost of managing the migration of backups from high-cost to low-cost storage is often prohibitive. This is where Storage Lifecycle Policies come in, as they can be used to automate the migration process from high-cost, short-term RTO storage to low-cost, longer-term RTO storage as the backup ages.

2.0 What is a Storage Lifecycle Policy?

A Storage Lifecycle Policy is a plan or map of where backup data will be stored and for how long. The Storage Lifecycle Policy determines where the backup is initially written to and where it is subsequently duplicated to. It also automates the duplication process and determines how long the backup data will reside in each location that it is duplicated to.

A Storage Lifecycle Policy thus replaces both the duplication process and the staging process by introducing a series of storage locations or destinations that use different types of storage with different retention periods and by ensuring that data always exists at the appropriate locations at the appropriate phases of the lifecycle.

A Storage Lifecycle Policy consists of two core components: a list of storage destinations where copies of the backup images will be stored, and the retention period for each copy. The Storage Lifecycle Policy may, optionally, be associated with a Data Classification within NetBackup. This provides a simple frame of reference for managing both Backup Policies and Storage Lifecycle Policies, and sees that the correct associations between the two are enforced at all times. All storage destinations and retentions of all copies are managed and viewed via one central GUI location for the Storage Lifecycle Policy.

The Storage Lifecycle Policy is reusable by many Backup Policies. When a storage plan changes (e.g., if a new regulation is imposed on your business requiring changes to retention periods or the number of copies created), you simply need to change a small number of Storage Lifecycle Policies, and all associated backups will take the changes into account automatically.

After the original backup completes, the Storage Lifecycle Policy process creates copies of the image, retrying as necessary until all required copies are successfully created.

By implementing a Storage Lifecycle Policy you remove the need for both Disk Staging Storage Units and duplication step in Vault profiles by defining all the locations where the data resides and for how long the data is retained in each location in a single policy definition.

2.1 Improved cost-effectiveness with Storage Lifecycle Policies

As discussed in the previous section the RTO of a given backup becomes less critical as the backup ages. Storage Lifecycle Policies allow automatic creation of multiple copies of the backup on different types of storage and with different retention periods creating a set of of backup copies with different recovery times—the primary copy always offering the shortest RTO. As time passes the copies of the backup with shorter retentions (located on storage that allows faster restore) expire and the storage is released for re-use. Automating this function dramatically reduces management and storage costs associated with the longer term retention of backups (beyond the immediate recovery point requirements) while significantly reducing the risk of premature expiration of backups caused by poorly designed or managed backup policies.

Figure 1 illustrates the data flow in a typical Storage Lifecycle Policy. Backups are initially written to a Flexible Disk Storage Unit (AdvancedDisk or SharedDisk), which offers high read and write speeds but has limited storage capacity. The images are then duplicated to three separate locations: a de-duplicating disk storage pool, a tape that remains onsite, and a tape that is sent offsite.

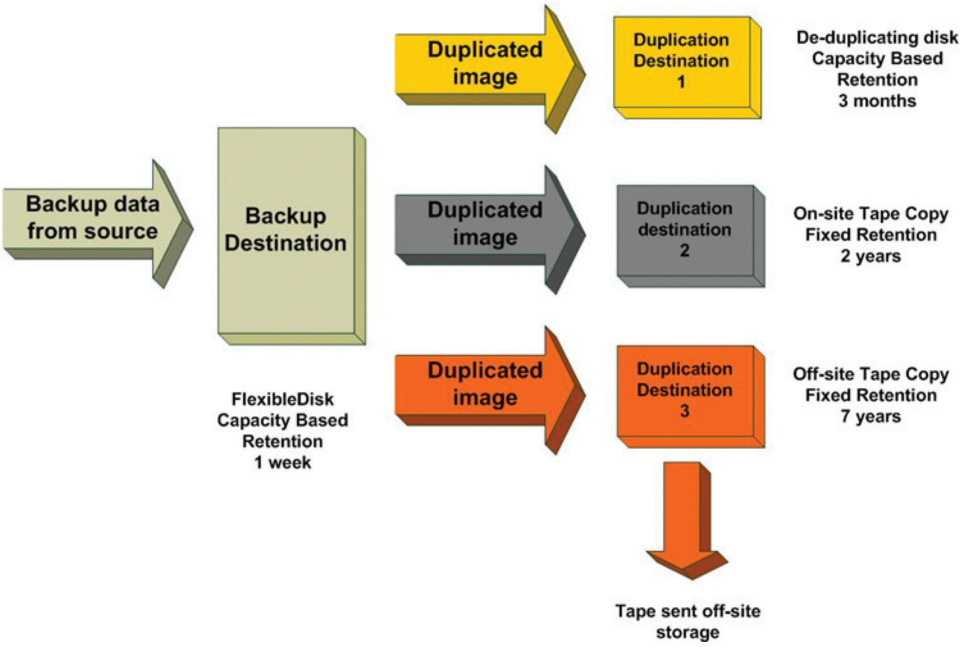

Duplicated image
Duplication Destination 1
De-duplicating disk Capacity Based Retention 3 months
Backup data from source
Backup Destination
Duplicated image
Duplication destination 2
On-site Tape Copy Fixed Retention 2 years
Duplicated image
Duplication Destination 3
Off-site Tape Copy Fixed Retention 7 years
FlexibleDisk Capacity Based Retention 1 week
Tape sent off-site storage

Figure 1. Data flow in a Storage Lifecycle Policy

There are now four copies of the backup, each with a different retention period. Note that the duplication process is not hierarchical: All duplications are done at the same time, and data resides in all four locations; it is not migrated from one location to another as the expiration in that location approaches. As time passes and the copies that allow faster restore expire, the time to restore the data (the RTO) increases, as shown in figure 2.

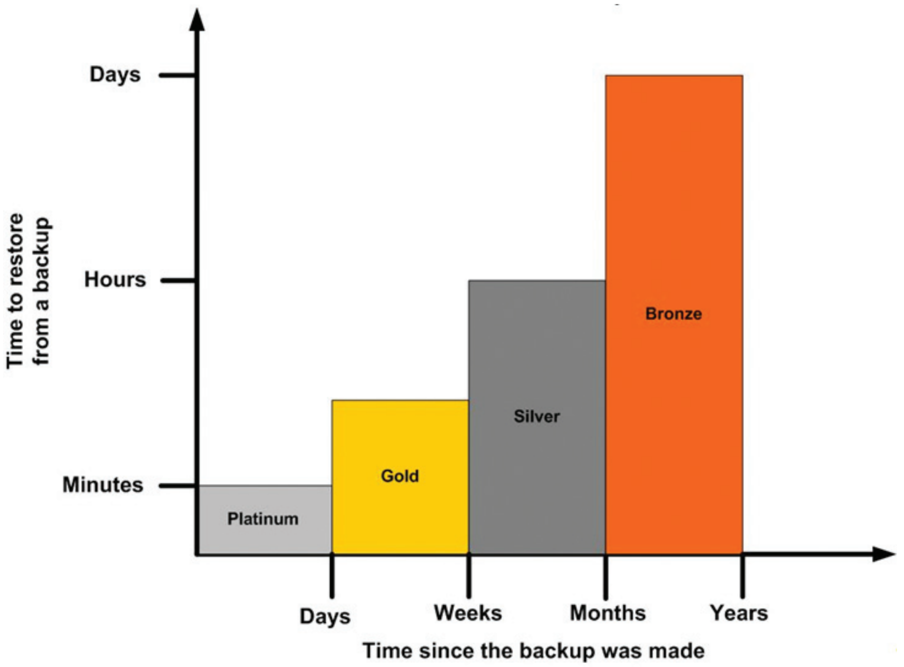


Figure 2. Recovery time vs. age of backup

Each location can be regarded as a distinct service-level tier for RTO; thus the RTO is initially a "Platinum" service level, but it degrades over time to a "Bronze" service level instead of remaining at Platinum indefinitely. As the service level degrades, the cost of storing the data decreases. This is an acceptable trade-off, as the value of backup data decreases with time. Backup data is at its most valuable immediately after the backup has been made, and it is at that time that the RTO needs to be kept to a minimum. Once a more recent backup exists, the previous backup is of less value because it does not offer the best RPO. As more time passes, the likelihood that a restore from the backup will be required decreases—and even should a restore be required, it is unlikely to be an urgent requirement. It is therefore reasonable to allow the RTO to increase.

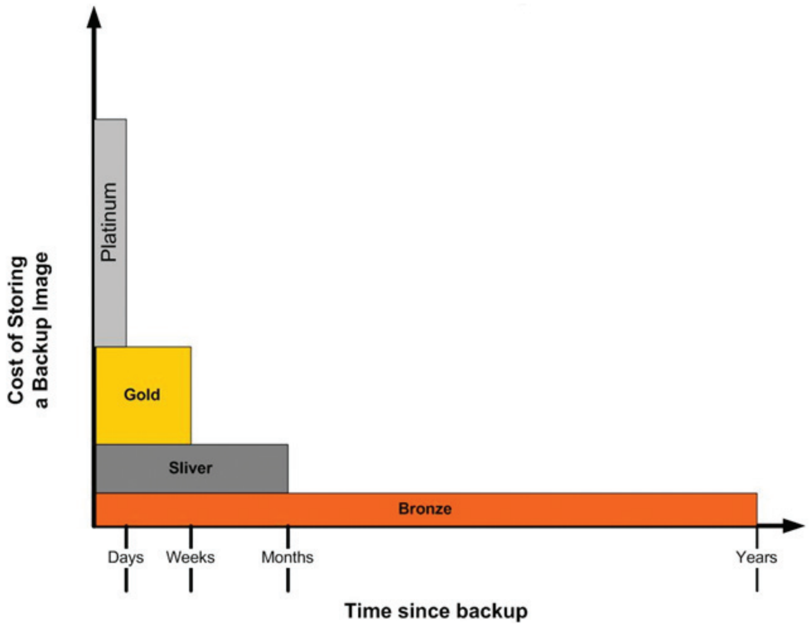


Figure 3. Cost of storing a backup image over time

Figure 3 shows how the initial cost of storing a backup image rapidly decreases as the backup ages. The total cost of storage over the life of the backup is significantly lower than it would be if the backup was held on Platinum storage for its entire life.

Less higher-cost storage is required overall, because the available storage can be reused more often with the tiered model.

2.2 Improved manageability with Storage Lifecycle Policies

A single Storage Lifecycle Policy can be applied to a large number of Backup Policies that fall within a single service-level tier or have similar retention requirements. This greatly simplifies the process of modifying the retention model compared to other existing backup products.

When a new Backup Policy is created that must conform with an existing retention model, it is a simple case of associating the new Backup Policy with the existing Storage Lifecycle Policy for that retention model rather than needing to configure individual retention periods and duplication rules for each copy of the data.

Similarly—because both the storage destinations and retention periods are associated with the Storage Lifecycle Policy and not with the Backup Policy—when a change to the retention model is required, it is a simple matter to change the retention periods or storage destinations of a single Storage Lifecycle Policy compared to the large number of discrete modifications to the attributes of a large number of Backup Policies that would normally be required to implement such a change.

This "single point of configuration" significantly reduces management overhead and the risk of errors when making changes of these types.

In practice it is likely that a Backup Policy may have two or three Storage Lifecycle Policies covering different types of backup (e.g., daily incremental, weekly full, and monthly full) but the principle of single point configuration still represents a significant saving when a large number of Backup Policies use the same retention model.

2.3 Storage Lifecycle Policies vs. Disk Staging

Most commercially available backup products support the use of disk storage as a target for backup data. The principal benefits of using disk storage as a backup target can be summarized as follows:

1) The random access nature of disks and the fact that they are always online means that restore is significantly faster than tape restore, as there are no delays in loading and positioning the media.
2) Unlike tape devices, disks have no minimum streaming speed, so the write speed of a disk device is not reduced if the incoming data fails to arrive at a high enough rate, so they are better suited than tape to backups coming from 'slow' sources.

The main disadvantage of disk is that its relatively high cost compared to tape means that disk space is always limited and backups written to disk can only have a short retention period. This is fine for meeting the high RPO/RTO requirements for a critical application service level but does not lend itself to the long retention periods often required for archival backups of such mission critical applications (i.e., periodic backups that are retained as "point in time" records of the application long after they have ceased to have value as a recovery source in the event of data loss).

Veritas NetBackup 5.1 introduced the concept of the Disk Staging Storage Unit—a special type of Storage Unit that acts as a buffer on the Media Server between the incoming data from the clients and the final storage destination. Backups are written to the Disk Staging Storage Unit and later duplicated to the final destination using a separate schedule associated with the Storage Unit. The backups remain available on the Disk Staging Storage Unit, allowing a fast restore capability, until it fills up, at which time the oldest backups are removed from the disk and restores are directed to the duplicated copy (usually on tape). One of the primary benefits of Disk Staging Storage Units is that they can defer the write to tape until after the backup window has closed, increasing the amount of data backed up within the window and extending the amount of time that the tape drives are used for (and thus the ROI on the drives, which are often among the most expensive components in a backup solution).

Many competitive products also offer forms of disk staging to address the need for high initial RPO/RTO and the problem of “minimum stream speed,” which has gotten worse as tape drives have gotten faster.

In Veritas NetBackup 6.5, this “traditional” staging model is still used for BasicDisk Storage Units, but all other types of disk Storage Unit can form part of a Storage Lifecycle Policy.

One of the problems with conventional disk staging models is that they support only one storage plan per disk device; thus they require a disk array to be divided into a number of small discrete devices instead of allowing it to be configured as a single large one. For example, if “Gold” backups are to be staged to disk and then written to two tape copies, but “Silver” backups (although also staged to disk) are written only to a single tape copy, then separate disk staging areas are required for the Gold and Silver backups. This may be an inefficient use of the disk space: the amount of space required by each backup type will vary from day to day, and each disk staging area must be configured to handle a “worst case” usage.

Storage Lifecycle Policies allow staging to be implemented with far better granularity with multiple storage plans sharing the same staging area. Gold and Silver backups can be sent to Storage Lifecycle Policies that specify the same disk space as the original backup destination, with varying numbers and types of secondary storage and different retentions for each copy. Storage Lifecycle Policies also automate the duplication process, eliminating the need to specify and schedule separate duplication jobs using the Vault option or scripted solutions. The result is that storage resources are used much more efficiently, and RTO and RPO for backups are maximized in an automated fashion that reduces administrative burdens.

3.0 What is a Data Classification?

A Data Classification, introduced in Veritas NetBackup 6.5, defines the relative importance of backup images residing on the same storage.

In Veritas NetBackup 6.5 Data Classifications are used by Storage Lifecycle Policy to apply a rank to backup images written to the same Capacity Managed Storage Units so that they can be retained for different periods, overriding the traditional "first in first out" model associated with Disk Staging Storage Units; this is discussed in more detail in section 4.0.2.3.

While Storage Lifecycle Policies simplify the management of the retention model for a particular type or class of backup, associating Data Classifications with both Storage Lifecycle Policies and Backup Policies maintains the correct relationships and helps ensure that backups of a particular importance (determined by the Data Classification) will have the appropriate retention model associated with them.

Data Classifications can also be used in Veritas NetBackup 6.5 as a simple way of identifying, grouping, and reporting on Backup Policies with a common level of importance, even where Storage Lifecycle Policies are not being used.

3.1 Creating and changing Data Classifications

A Data Classification has four components:

- A unique name
- An textual description (optional)
- A unique rank
- A global unique identifier (GUID)

Any of these attributes of a classification can be changed at any time with the exception of the GUID, which is assigned automatically when the Data Classification is created.

By default, Veritas NetBackup provides four Data Classifications: Platinum, Gold, Silver, and Bronze. As with any classification, these default classifications can be renamed and reordered, and up to 21 more Data Classifications can be added through the Data Classification tab on the Master Server Host Properties window in the administration GUI.

These Data Classifications are ranked higher or lower than each other. No two classifications are of the same rank.

Note that a Data Classification can be renamed and re-ranked, but it can not be deleted. This persistence is necessary because there may be images in the catalog that are tagged with the classification. NetBackup does not want to lose track of the classification of any images.

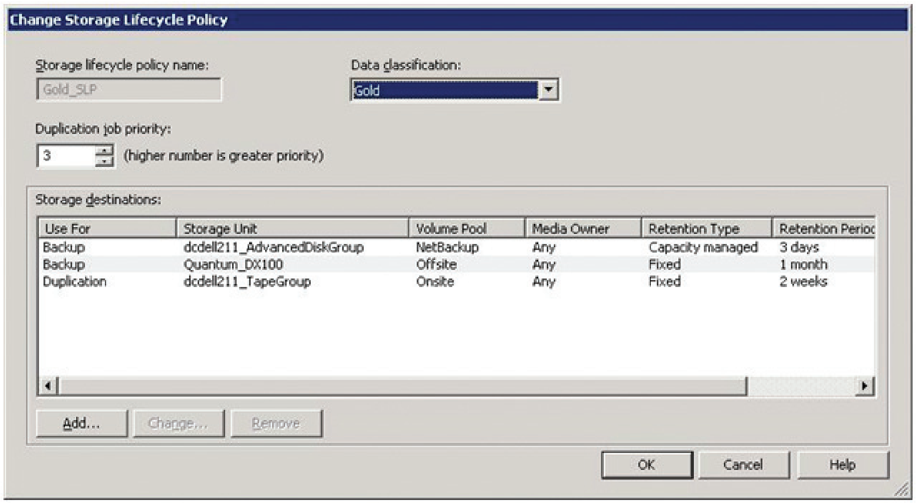
The GUID is a long alphanumeric string that uniquely identifies a classification. The purpose of the GUID is to differentiate classifications from different NetBackup domains. For example, if two NetBackup domains both have a classification named Gold and images from one domain are imported into the other domain, it may be important to be able to differentiate the source domain's Gold classification from the target domain's Gold classification.

4.0 Configuring Storage Lifecycle Policies

Storage Lifecycle Policies are configured under the "Storage" node in the Veritas NetBackup console. There are now three types of storage targets: Storage Units, Storage Unit Groups, and Storage Lifecycle Policies. A Storage Lifecycle Policy consists of the following elements:

- A Storage Lifecycle Policy name
- A Data Classification (optional)
- A duplication job priority (optional but recommended)
- A list of all Storage Destinations that the images will ultimately reach

Storage Lifecycle Policies do not support BasicDisk Storage Units because the BasicDisk model cannot handle the new retention schemes, etc., but they can be used with all other disk options and with tape.



Change Storage Lifecycle Policy

Storage lifecycle policy name: Gold_SLP

Data classification: Gold

Duplication job priority: 3 (higher number is greater priority)

Storage destinations:

Use For	Storage Unit	Volume Pool	Media Owner	Retention Type	Retention Perioc
Backup	dcdell211_AdvancedDiskGroup	NetBackup	Any	Capacity managed	3 days
Backup	Quantum_DX100	Offsite	Any	Fixed	1 month
Duplication	dcdell211_TapeGroup	Onsite	Any	Fixed	2 weeks

Figure 4. Storage Lifecycle Policy menu

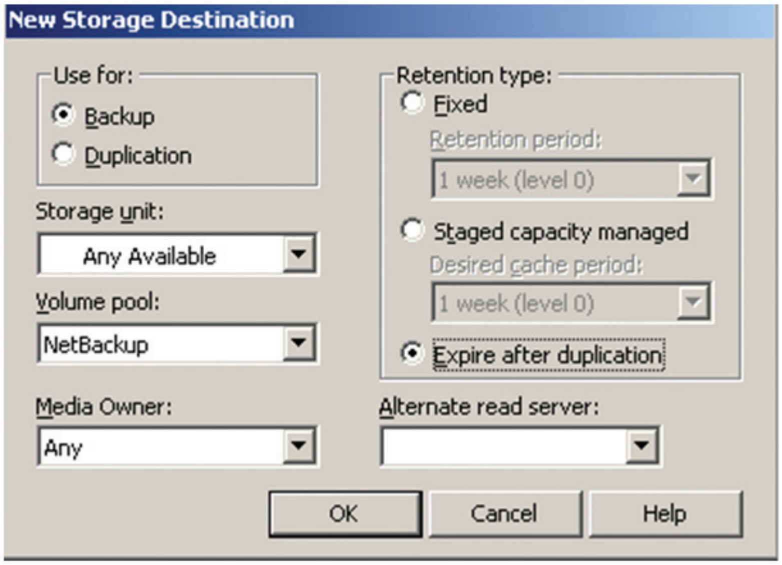
The duplication job priority determines the priority to be associated with duplication jobs created by this Storage Lifecycle Policy. This can be used to prioritize duplications from one Storage Lifecycle Policy relative to duplications from another one—or other backup jobs; this is described in more detail in section 4.0.2.3.

4.0.1 Storage Destinations

Storage Destinations are the key elements of a Storage Lifecycle Policy. Each Storage Destination specifies whether the copy will be made by the backup operation or by a subsequent duplication operation; it also specifies the retention period of the data in that Storage Destination.

4.0.2 Storage Destination fields

Figure 5 shows the fields associated with a Storage Destination.



New Storage Destination

Use for:
● Backup
○ Duplication

Storage unit: Any Available

Volume pool: NetBackup

Media Owner: Any

Retention type:
○ Fixed
Retention period: 1 week (level 0)
○ Staged capacity managed
Desired cache period: 1 week (level 0)
● Expire after duplication

Alternate read server:

Figure 5. Setting up a new Storage Destination

The meaning/usage of these fields is as follows:

Use for—This field defines what the Storage Destination is intended to do. There are two types: Backup Storage Destinations and Duplication Storage Destinations. Backup Storage Destinations are the primary targets to which the initial copies of the backup are written; it is possible to have more than one Backup Storage Destination in a Storage Lifecycle Policy, but all such destinations should be on the same Media Server (see section 4.0.3 later in this document). Duplication Storage Destinations are secondary targets where backup images are duplicated to from the first (primary) Backup Storage Destination. Duplication occurs as soon as possible after the backup completes (see section 4.0.4).

Storage Unit—Each Storage Destination must have a Storage Unit or Storage Unit Group associated with it. As previously mentioned, BasicDisk Storage Units cannot be used with Storage Lifecycle Policies and cannot be selected when setting up a Storage Destination. Storage Destinations control the retention period of the images they hold, overriding any retention set on the Backup Policy schedule. The same Storage Units can be used by multiple Storage Destinations.

Volume Pool and Media Owner—These setting apply when a tape Storage Unit is selected and provide the necessary media attributes that would normally be applied at the Backup Policy and Backup Schedule levels.

Alternate Read Server—This setting determines the server that is used to do the duplication (by default this is the server on which the primary copy of the backup was written). Note that this setting applies on the source Storage Destination, not the target Storage Destination, as duplications are always made using the primary backup copy this setting is only effective when specified on the first Backup Storage Destination.

Retention type—There are three different retention types available for a Storage Destination. These are discussed in the following sections.

4.0.2.1 Fixed Retention

Fixed Retention is the traditional Veritas NetBackup retention. When a copy of an image is to be retained for X months, it will be retained in the Veritas NetBackup catalog for X months—no more and no less than X months. The one exception to this rule is that the primary copy of a backup will be kept longer than the retention if all copies specified in the Lifecycle have not yet been created.

At least one Storage Destination with a fixed retention period must exist in each Storage Lifecycle Policy, and no copy will be kept longer than the longest Fixed Retention specified in the Lifecycle Policy.

4.0.2.2 Expire After Duplication

This is a new retention type introduced to support staging to secondary and tertiary storage and is similar in nature to the 'Expire original disk backup' option in the Vault option—the primary difference is that in this case the image expires immediately when the duplication completes, not at a later time. This option also supports the expiration of tape images as well as the disk images supported by the Vault option.

4.0.2.3 Capacity Managed Retention

Also a new retention type, Capacity Managed Retention has been introduced to support smart cleanup of backup images staged to disk devices. This type of retention works only on disk devices that Veritas NetBackup recognizes as disk, and that tell Veritas NetBackup to keep the image copy until the disk capacity it is using must be made available to an incoming backup. If a storage device cannot be configured in Veritas NetBackup to have high and low “water marks” (e.g., virtual tape libraries or other devices that present disk storage as an emulation of tape storage), it cannot do capacity-based retention.

With Capacity Managed Retention, the retention period is the time the image should ideally be kept on the storage device. This time period is known as the “try-to-keep time.” Veritas NetBackup will keep the image copy until such time as the space is required for other backups, which may be a longer or shorter period than the try-to-keep time, depending on the demand for space.

Because capacity management takes place at the level of the underlying DiskPool, it may be applied to multiple Storage Destinations and Storage Units at the same time. When the space limit (high water mark) on the DiskPool is hit, the images will be removed in a specific order based on whether or not they have passed the try-to-keep time and, if implemented, what their Data Classification level is.

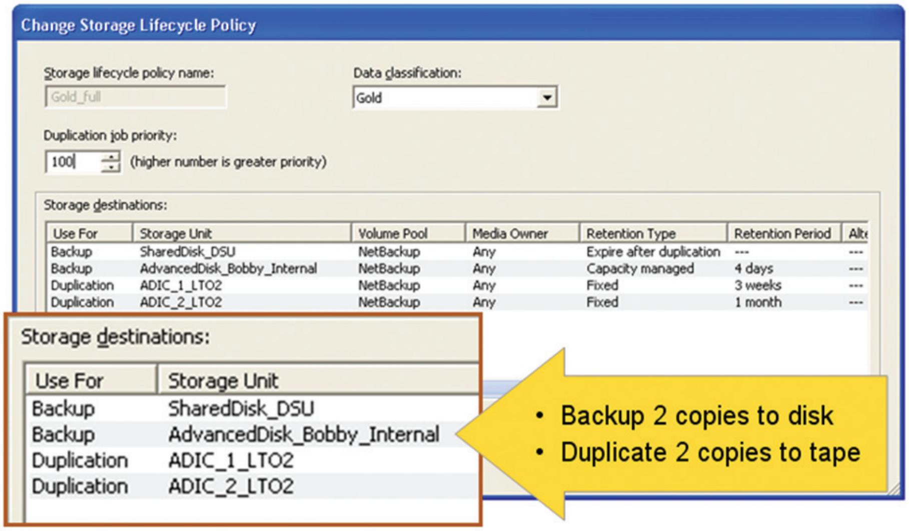
Thus when the high water mark is met, the images will be removed—starting with the images with the lowest Data Classification that have passed the try-to-keep time and then working up the data classifications until either the low water mark is met or there are no more images that are past their try-to-keep time. If the low water mark has not been met, images that have not yet reached their try-to-keep time will be removed, again starting with the oldest images and lowest Data Classification, until the low water mark is reached.

No image will be removed if the Storage Lifecycle Policy has not completed all of the specified duplications for that image (see section 4.0.4).

Note that the overall expiration date for a backup is determined by the longest fixed retention, and Capacity Managed Retention periods should not exceed this figure (see section 6.4.2).

4.0.3 Inline Copy

If more than one Backup Storage Destination is listed in the Storage Lifecycle Policy, an Inline Copy operation is implied. In the following example, two copies will be made by the Backup operation (using Inline Copy), and two copies will be made by subsequent duplication operations.



Change Storage Lifecycle Policy

Storage lifecycle policy name: Gold_full

Data classification: Gold

Duplication job priority: 100 (higher number is greater priority)

Storage destinations:

Use For	Storage Unit	Volume Pool	Media Owner	Retention Type	Retention Period	Alt
Backup	SharedDisk_DSU	NetBackup	Any	Expire after duplication	---	---
Backup	AdvancedDisk_Bobby_Internal	NetBackup	Any	Capacity managed	4 days	---
Duplication	ADIC_1_LTO2	NetBackup	Any	Fixed	3 weeks	---
Duplication	ADIC_2_LTO2	NetBackup	Any	Fixed	1 month	---

- Backup 2 copies to disk
- Duplicate 2 copies to tape

Figure 6. Multiple Backup Storage Destinations

Inline Copy will only work when all the copies are directed to devices that are accessible via the same Media Server. If more than one Media Server is associated with the Backup Storage Destinations the backup will only run to the first Storage Destination and will then be duplicated to the other Storage Destinations, so they effectively become Duplication Storage Destinations.

The Storage Lifecycle Policy will always apply the “If this copy fails, continue the job” option to an Inline Copy operation, whether it is during the original backup job or during a subsequent duplication job.

In Veritas NetBackup 6.0, an Inline Copy job will not begin unless all storage destinations for all copies are available. This behavior applies even if the “If this copy fails, continue the job” option is specified. This restriction has been addressed in Veritas NetBackup 6.5: If one of the Storage Destinations is healthy but busy, the job will wait for the device to become available—but if one of the Storage Destinations is unavailable, then the job will begin anyway and will not make the copy for that storage destination. Instead, this copy will be added to the list of duplication tasks and the backup will be duplicated to this Storage Destination when it becomes available.

4.0.4 Duplication and Persistent Duplication Retry

By default a Storage Lifecycle Policy checks every five minutes for backup images that have recently completed and require duplication jobs. The Storage Lifecycle Policy groups batches of similar images together for each duplication job, to optimize the performance of duplication. When there is enough data (8 GB by default) to warrant a duplication job, duplication is started.

These default settings of 5 minutes and 8 GB can be varied by setting values in the LIFECYCLE_PARAMETERS file on the Master Server, which is described in Appendix A.

If a duplication job fails to make a copy of an image, that image will be added to a subsequent batch of images to be duplicated with the next five-minute sweep of images that need to be copied. This is done three times for any single image. After three failures, the Storage Lifecycle Policy will wait two hours (by default) before trying to create that copy of that image again. This retry will continue once every two hours (by default) until either the user intervenes or the time of the longest retention specified for the image comes to pass.

A Storage Lifecycle Policy is incredibly persistent when it comes to making all specified copies of all applicable images. Until all copies of an image are made, the Storage Lifecycle Policy will not stop trying to make the unmade copies, unless the user intervenes or the longest retention specified for the image comes to pass.

If a Storage Lifecycle Policy specifies that two copies are to be made during the original backup job and no other copies are to be made, then if one of those copies fails such that the original backup job creates only one of the two copies, the Storage Lifecycle Policy will recognize that the copy does not exist and will create that copy with a subsequent duplication job.

When all copies that were specified in the Storage Lifecycle Policy at the time of the original backup of an image exist, the image is marked as complete, and the Storage Lifecycle Policy will stop processing it.

Not until all copies of an image exist, will any of those copies be deleted for purposes such as draining a disk device down to low water mark to make space for incoming backups.

4.1 Using Data Classifications with Storage Lifecycle Policies

In Veritas NetBackup 6.5, Data Classifications are used by Storage Lifecycle Policies to apply a structure to backup images written to the same Capacity Managed Storage Units so that images can be retained for different periods, overriding the traditional “first in first out” model associated with Disk Staging Storage Units. This is discussed in more detail in section 4.0.2.3.

It is important to understand the difference between the Data Classification ranking and priority.

In Veritas NetBackup, priority is applied to Backup Policies and controls the assignment of storage resources to a new job when it hits the job queue. A queued job with a higher priority will be given a resource before another queued job with a lower priority, and it will thus execute first. When jobs of different priorities are started on a server with spare resources, the highest-priority jobs will be assigned the resources first, and lower-priority jobs will be queued if insufficient resources are available.

The Data Classification has absolutely no effect at the time of the execution of the backup job. Data Classifications only control the way in which images are expired to create space in Capacity Managed Storage Destinations. Images associated with a lower classification are always expired first.

If a Storage Lifecycle Policy is associated with a particular Data Classification, it will only be available for use by Backup Policies that are associated with the same Data Classification.

Note that while Data Classifications can be used with Lifecycle Policies and Backup Policies, the use of a Data Classification is optional. Backup Policies and Storage Lifecycle Policies do not need to be associated with a Data Classification.

4.2 Example configuration of a Storage Lifecycle Policy

Figure 7 shows an example of a typical requirement for a Storage Lifecycle Policy. The initial backup is made to high-speed disk in the Storage Destination SDB1, and is then duplicated to three different targets—SDD1, SDD2, and SDD3—each with a different retention period.

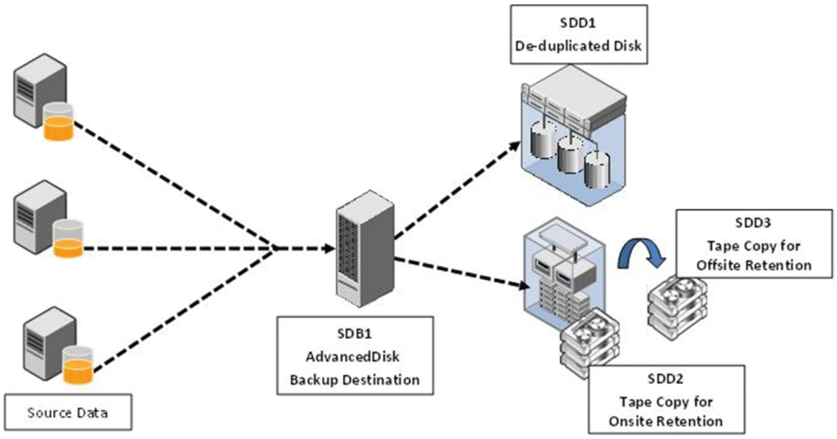


Figure 7. Example Strorage Lifecycle Policy

The following attributes would be associated with each Storage Destination within the Storage Lifecycle Policy:

SDB1

Use for—Backup (This is the primary target from which all duplicates will be made.)

Storage Unit—ADV1 (This is an AdvancedDisk Storage Unit.)

Volume Pool and Media Owner—Not applicable

Retention Type—Capacity Managed, 1-week retention (This retention aims to ensure that restores can be made quickly and efficiently during the first few days after the backup is created, when it represents the most recent RPO.)

Alternate Read Host—Not set (This could be set to another Media Server if a SharedDisk Storage Unit was used.)

SDD1

Use for—Duplication

Storage Unit—PDSU1 (This is a PureDisk Storage Unit, so the contents of this Storage Unit may be replicated to a DR site.)

Volume Pool and Media Owner—Not applicable

Retention Type—Capacity Managed, 3-month retention (This retention aims to ensure that restores can be made from this Storage Unit when the RTO requirement is still high but not critical.)

Alternate Read Host—Not applicable

SDD2

Use for—Duplication

Storage Unit—TAPE1 (This is a conventional tape Storage Unit using LTO3 media.)

Volume Pool—ONSITE

Media Owner—This could be used if media sharing is configured.

Retention Type—Fixed, 2-year retention (This retention allows restores directly from tape for two years after the backup is created.)

Alternate Read Host—Not applicable

SDD3

Use for—Duplication

Storage Unit—TAPE1 (This is a conventional tape Storage Unit using LTO3 media—note that as this a tape Storage Unit; both SDD2 and SDD3 can use the same Storage Unit.)

Volume Pool—OFFSITE (This tape will later be ejected and sent to offsite storage.)

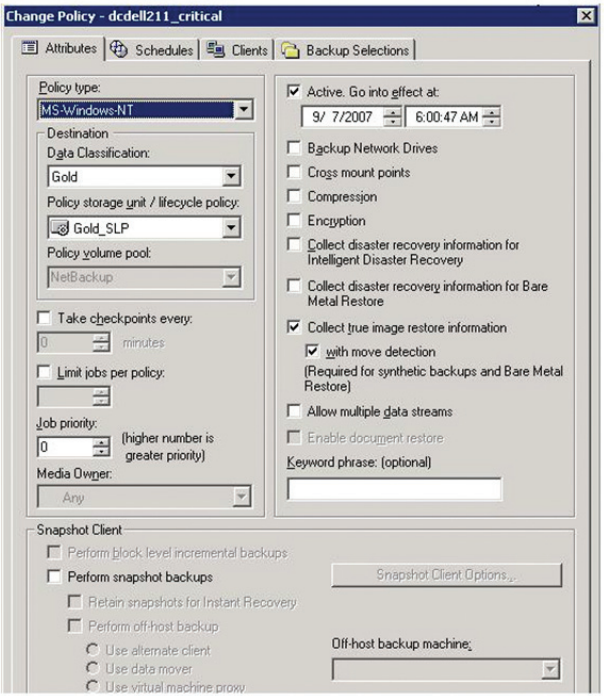
Media Owner—Can be used if media sharing is configured

Retention Type—Fixed, 7-year retention (This retention allows restores directly from tape for seven years after the backup is created.)

Alternate Read Host—Not applicable

5.0 Configuring Backup Policies to use Storage Lifecycle Policies

Once the Storage Lifecycle Policies and Data Classification have been set up, it is a simple matter to define Backup Policies to use them. The “Policy Storage Unit/Storage Lifecycle Policy” field on the “Attributes of a Backup Policy” definition now provides a pick-list that includes all three types of storage locations (Storage Units, Storage Unit Groups, and Storage Lifecycle Policies). There is also an optional “Data Classification” field on the same screen.



Change Policy - dcdell211_critical

Attributes | Schedules | Clients | Backup Selections

Policy type: MS-Windows-NT

Destination

Data Classification: Gold

Policy storage unit / lifecycle policy: Gold_SLP

Policy volume pool: NetBackup

□ Take checkpoints every: 0 minutes

□ Limit jobs per policy:

Job priority: 0 (higher number is greater priority)

Media Owner: Any

☑ Active. Go into effect at: 9/ 7/2007 6:00:47 AM

□ Backup Network Drives

□ Cross mount points

□ Compression

□ Encryption

□ Collect disaster recovery information for Intelligent Disaster Recovery

□ Collect disaster recovery information for Bare Metal Restore

☑ Collect true image restore information

☑ with move detection

(Required for synthetic backups and Bare Metal Restore)

□ Allow multiple data streams

□ Enable document restore

Keyword phrase: (optional)

Snapshot Client

□ Perform block level incremental backups

□ Perform snapshot backups

□ Retain snapshots for Instant Recovery

□ Perform off-host backup

○ Use alternate client

○ Use data mover

○ Use virtual machine proxy

Snapshot Client Options...

Off-host backup machine:

Figure 8. Backup Policy attributes

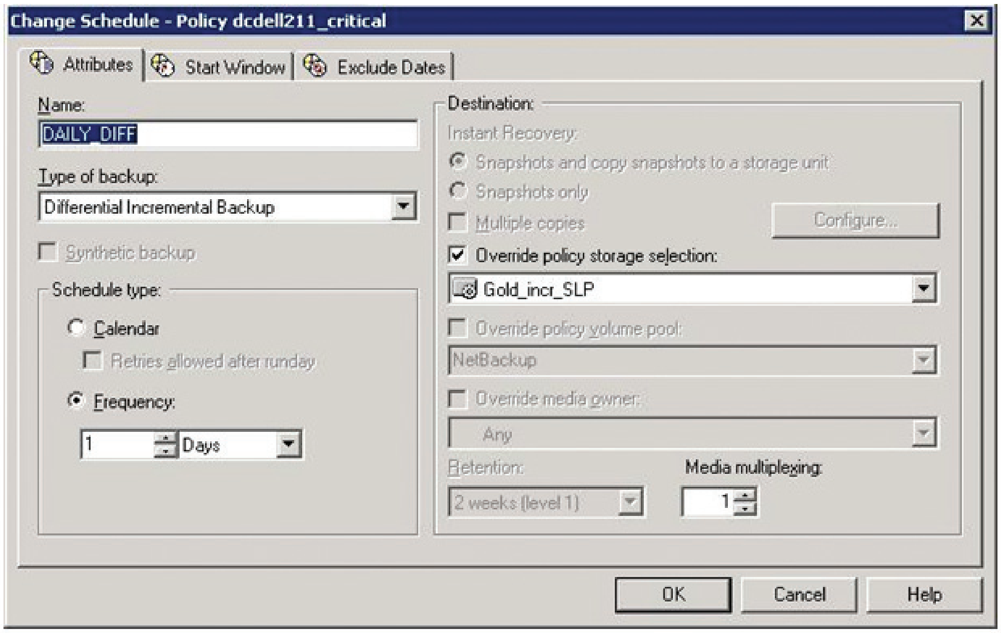
By default the “Data Classification” field is set to <No Data Classification>, and the “Policy Storage Unit/Storage Lifecycle Policy” field allows the selection of any Storage Unit, Storage Unit Group, or Storage Lifecycle Policy.

Once a Data Classification is selected, the “Policy Storage Unit/Storage Lifecycle Policy” list filters out Storage Lifecycle Policies that do not match the Data Classification. All the Storage Units and Storage Unit Groups will still be displayed as these are not currently linked to Data Classifications but if a BasicDisk Storage Unit or Storage Unit Group is selected, the policy validation process will fail and you will not be able to save the changes.

5.1 Using multiple Storage Lifecycle Policies within the same Backup Policy

Most Backup Policies contain multiple schedules which create different types of backup with different retention requirements (for example daily incremental backups that are retained for 2 weeks and weekly full backups that are retained for one month).

The primary Storage Lifecycle Policy specified in the Backup Policy Attributes sets on only the storage destinations but also the retention periods to be used for all the backups created by that Backup Policy. However, as with conventional Storage Units and Storage Unit Groups, it is possible to override the Storage Lifecycle Policy used by the Backup Policy at the schedule level. Thus the same Backup Policy may use different Storage Lifecycle Policies for full and incremental backups.



Change Schedule - Policy dcdell211_critical

Name: DAILY_DIFF

Type of backup: Differential Incremental Backup

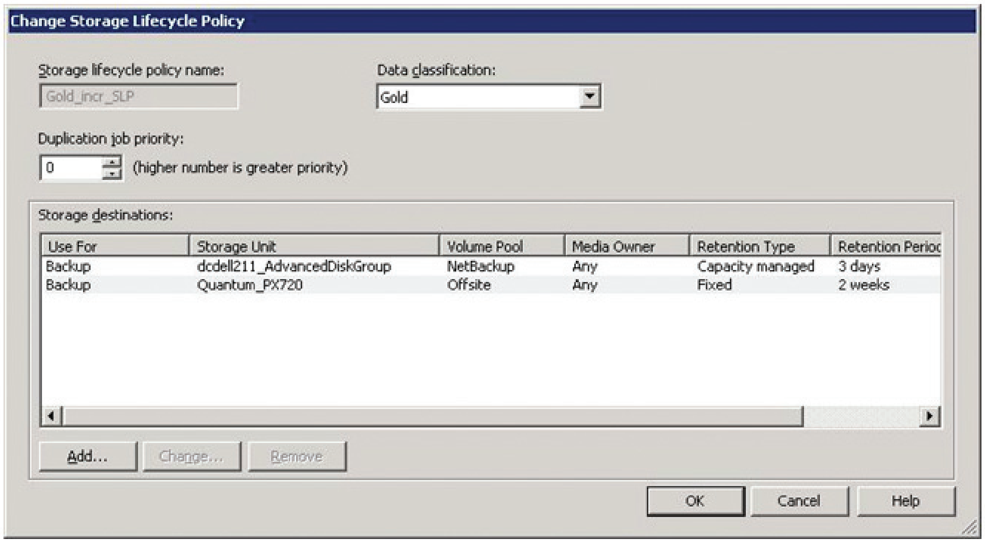
□ Synthetic backup

Schedule type:
○ Calendar
□ Retries allowed after runday
● Frequency: 1 Days

Destination:
Instant Recovery:
● Snapshots and copy snapshots to a storage unit
○ Snapshots only
□ Multiple copies
☑ Override policy storage selection: Gold_incr_SLP
□ Override policy volume pool: NetBackup
□ Override media owner: Any
Retention: 2 weeks (level 1)
Media multiplexing: 1

Figure 9. Override Policy Storage Selection in schedule

In figure 9 above the default Storage Lifecycle Policy used for the Backup Policy is overridden in the DAILY_DIFF schedule by the Gold_incr_SLP Storage Lifecycle Policy shown in figure 10 below:



Change Storage Lifecycle Policy

Storage lifecycle policy name: Gold_incr_SLP

Data classification: Gold

Duplication job priority: 0 (higher number is greater priority)

Storage destinations:

Use For	Storage Unit	Volume Pool	Media Owner	Retention Type	Retention Perioc
Backup	dcdell211_AdvancedDiskGroup	NetBackup	Any	Capacity managed	3 days
Backup	Quantum_PX720	Offsite	Any	Fixed	2 weeks

Figure 10. Incremental Storage Lifecycle Policy

6.0 Considerations when setting up Storage Lifecycle Policies

This section looks at some of the things that should be considered when setting up Storage Lifecycle Policies and the way in which they will affect the backup and duplication operations.

6.1 Storage considerations

One of the principal underlying themes of the Storage Lifecycle Policy concept is “the right data in the right place at the right time.” As we have seen, Storage Lifecycle Policies automate the duplication process, allowing data to be written to one or more primary locations and then copied to secondary and tertiary locations. It is important to remember that this is not a hierarchical model: Data is not migrated from one location to another as it approaches the end of its retention period; it is duplicated at the first possible opportunity and occupies all the storage locations simultaneously.

6.1.1 Backup Storage Destinations

In most cases the primary (first) Backup Storage Destination will be a high-speed storage device that allows fast restores, typically an AdvancedDisk/SharedDisk Storage Unit running on a fast disk array. This primary copy is likely to have a short retention period (either fixed or capacity managed), as the relative cost of the storage is high compared with other types of storage. Other Backup Storage Destinations may be physical or virtual tapes (in libraries or virtual tape libraries), which have fast write speeds but relatively slower read speeds. These destinations will have longer retention periods associated with them.

While space-optimized or de-duplicating Storage Units (such as PureDisk or OpenStorage devices) provide significant economies in terms of disk usage, they are generally not suitable for use as Backup Storage Destinations, because the write speed to these devices is often lower than that of conventional devices due to the overhead of the de-duplication process.

For optimal performance and space saving the Backup Storage Destination should be configured to use an AdvancedDisk/SharedDisk Storage Unit with a retention setting of “expire after duplication” and the use of space optimized Storage Units should be limited to Duplication Storage Destinations.

6.1.2 Duplication Storage Destinations

Because duplication takes place “off-line” and does not impact the initial backup time, it is possible to make use of more space-efficient technologies such as PureDisk and OpenStorage devices—which generally have lower write speeds than AdvancedDisk/SharedDisk and conventional tape—as Duplication Storage Destinations. As previously discussed, the Storage Lifecycle Policy will carry out all duplications as soon as possible after the backup completes, and therefore consideration must be given to the location of the Duplication Storage Destination relative to the primary Backup Storage Destination (“Will the duplication process create lots of network traffic?”) and the availability of resources on the Duplication Storage Destination (“Will there be tape drives available?”).

6.1.3 Prioritizing backup and duplication

Each Storage Lifecycle Policy can have a "duplication priority" associated with it that determines the priority of the duplication process. These priorities are the same as the priorities associated with Backup Policies. In most cases the priority field on a Backup Policy is left set at the default value of 0 (the lowest priority); this is the same default value used for duplication priority in Storage Lifecycle Policies, and all backups and duplications will, by default, have the same priority.

The impact of this is that, as Storage Lifecycle Policies attempt to start duplication as soon as possible after a backup completes, both backups and duplications will compete directly with each other for resources such as tape drives and disk space; suddenly, backup priority becomes more important.

When introducing Storage Lifecycle Policies into an existing environment, the priorities associated with Backup Policies should be reviewed and increased where appropriate to make sure that the correct priorities are associated with both backup and duplication operations. When planning a new deployment, consideration should be given to setting the right priorities for backups and duplications.

6.2 Using Storage Lifecycle Policies with the NetBackup Vault Option

Storage Lifecycle Policies automate the duplication processes that are normally scheduled through the Vault Option, but the Vault Option can still be used with Storage Lifecycle Policies to handle the other functions related to Vault—catalog backup, reporting, and media ejection. Simply define the Vault profiles to skip the duplication step and schedule the Vault jobs to run at around the time that the eject phase of a normal Vault job would occur.

6.3 Storage Lifecycle Policies and the Media Server Encryption Option

It is not possible to specify the use of the Media Server Encryption Option on specific Storage Destinations within a Storage Lifecycle Policy, but encryption can still be used with Storage Lifecycle Policies in the following ways:

- By specifying encryption as a global default
- By specifying an encryption policy that works on a specific copy number, in the same way that can be done for the Vault Option (Note that this approach is not guaranteed, as the persistent

duplication retry mechanism in Storage Lifecycle Policies means that the copy number associated with a given Storage Destination cannot be 100% guaranteed.)

Refer to the "Media Server Encryption Option" documentation for more details on encryption policies.

6.4 Some considerations when defining Storage Destinations

A storage destination within a Storage Lifecycle Policy may use either a specific Storage Unit or a Storage Unit Group. Validation checks within the Storage Lifecycle Policy setup are not exhaustive, and there are a number of potential things to watch out for.

6.4.1 Duplication considerations

Duplication is always done from the primary backup copy, even when there is more than one Backup Storage Destination. Duplication always takes place as soon as possible after the backup completes and, as discussed in the last section, is repeated until it is successful. It is important to remember this when defining Duplication Storage Destinations, as poor design may lead to excessive network traffic and other resource contention.

6.4.2 Global limitations to image retention

Where a mixture of fixed and capacity managed retention periods are used a Storage Lifecycle Policy, all copies of the backup are expired when the longest fixed retention period is reached—even if a capacity managed storage destination has a try-to-keep time that is longer than the fixed retention period. The longest retention period used in any Storage Destination within a particular Storage Lifecycle Policy must be a fixed retention period, not a capacity managed retention period.

In the initial releases of Veritas NetBackup 6.5, there are no checks in place to prevent the selection of a capacity managed retention period that is longer than the longest fixed retention period, and it is important to take care to avoid this situation as images may be expired earlier than expected if this rule is not followed.

6.4.3 Inline Copy does not work with different Media Servers

As mentioned in section 4.0.3, Inline Copy will occur if multiple backup type storage destinations are specified that have Storage Units located on the same Media Server but will fail if the Storage Units are located on different Media Servers. There are no checks in the Storage Lifecycle Policy configuration to ensure that all the Storage Units associated with backup storage destinations are on the same Media Server, and care must be taken to ensure that they are. Otherwise all but the first copy of the backup will fail initially, and the Storage Lifecycle Policy will subsequently duplicate the image over the LAN to correct the initial failure.

Note that if Storage Unit Groups that include Storage Units from several different Media Servers are used, this particular problem is avoided as the selection process within the group will always ensure that Storage Units on the same Media Server are selected.

6.4.4 BasicDisk Storage Units in Storage Unit Groups

A Storage Unit Group that contains just one BasicDisk Storage Unit will behave as a BasicDisk Storage Unit Group. Although Storage Lifecycle Policies do not work with BasicDisk Storage Units there are no checks in the Storage Destination set up to prevent you from selecting an Storage Unit Group that contains a BasicDisk. Care should be taken when using Storage Unit Groups in Storage Destinations to ensure that there are no BasicDisk Storage Units within the Storage Unit Group.

6.4.5 Considerations for the "Alternate Read Server" setting

The "Alternate Read Server" setting for a storage destination applies on the source destination, not the target destination. This means that the only Storage Destination on which the "Alternate Read Server" setting has any effect is the first Backup Destination (as this is the source used for all duplication).

Be careful when using this setting to ensure that the source Storage Unit or tape media can be presented to the alternate restore host. There are no checks on the type of Storage Unit when setting up the Storage Lifecycle Policy, and it is possible to configure an AdvancedDisk Storage Unit in a configuration with "Alternate Read Server" set; this error will not become apparent until the Backup Policy is run and the duplication phase fails.

Again, care should be taken when using Storage Unit Groups to ensure that a) the source data can be presented to the Alternate Read Server; and b) the target Storage Unit is on the Media Server that has been specified as the Alternate Read Server.

6.4.6 Duplication to the same location

It is possible to configure multiple AdvancedDisk/SharedDisk Storage Units within the same DiskPool. When configuring Storage Lifecycle Policies, it is important to ensure that each backup and duplication target is actually using a different DiskPool and that you are not duplicating images between Storage Units that share the same underlying storage.

Appendix A: The LIFECYCLE_PARAMETERS file

The size and the frequency of duplication jobs requested by the Storage Lifecycle Policy can be specified in the LIFECYCLE_PARAMETERS file. Five parameters can be specified in this file.

The file is located at: /usr/openv/netbackup/db/config/LIFECYCLE_PARAMETERS

The five parameters are as follows:

- **MIN_KB_SIZE_PER_DUPLICATION:** This is the size of the minimum duplication batch (default 8 GB).
- **MAX_KB_SIZE_PER_DUPLICATION_JOB:** This is the size of the maximum duplication batch (default 25 GB).
- **MAX_MINUTES_TIL_FORCE_SMALL_DUPLICATION_JOB:** This represents the time interval between forcing duplication sessions for small batches (default 30 minutes).
- **IMAGE_EXTENDED_RETRY_PERIOD_IN_HOURS:** After duplication of an image fails three times, this is the time interval between subsequent retries (default 2 hours).
- **DUPLICATION_SESSION_INTERVAL_MINUTES:** This is how often the Storage Lifecycle Policy service (nbstserv) looks to see if it is time to start a new duplication job(s) (default 5 minutes).

If this file does not exist, the default values will be used.

Not all parameters are required in the file, and there is no order dependency in the file. Any parameters omitted from the file will use default values.

The syntax of the LIFECYCLE_PARAMETERS file, using default values, is as follows:

```
MIN_KB_SIZE_PER_DUPLICATION_JOB 8192
MAX_KB_SIZE_PER_DUPLICATION_JOB 25600
MAX_MINUTES_TIL_FORCE_SMALL_DUPLICATION_JOB 30
IMAGE_EXTENDED_RETRY_PERIOD_IN_HOURS 2
DUPLICATION_SESSION_INTERVAL_MINUTES 5
```

Appendix B: The nbstlutil command

The Veritas NetBackup "Storage Lifecycle Policy Utility" (nbstlutil) command allows the user to control some of the detailed behavior of a lifecycle or of a copy within a lifecycle.

Storage Lifecycle Policies operate on an image work list that is stored in the EMM database. The nbstlutil command operates on images that are stored in that work list.

The most interesting of the options offered by the nbstlutil command might be the abilities to:

- **Deactivate and reactivate operations on selected image copies** (For instance, should a tape library break down and need a day for repair, you can deactivate the Storage Lifecycle Policy operations for those copies until the device has been repaired.)
- **Cancel pending operations** on selected image copies

The operations available are summarized in the following table:

Operation	Applicable nbstlutil command options
Deactivate pending and future lifecycle operations on selected image copies	deactivate [-destination <name>] [-lifecycle <name>] [-backupid <id_value>]
Activate lifecycle operations on selected image copies	activate [-destination <name>] [-lifecycle <name>] [-backupid <id_value>]
Activate lifecycle operations on selected image copies	cancel [-destination <name>] [-lifecycle <name>] [-backupid <id_value>]
Initiate image expiration on a disk volume	initiate_exp [-dp <disk_pool>] [-dt <disk_type>] [-mediaid <value>] [-mediaserver <name>] [-storageserver <name>] [-volumeid <value>]
Show contents of image list	list [-l] [-U] [-backupid <value>] [-client <name>] [-mediaid <value>] [-storageserver <name>] [-mediaserver <name>] [-state <value>]
Start a new lifecycle management session	new_session
Remove all image list entries	remove_all [-force]
Remove expired fragment entries from image list	remove_exp [-force]

Refer to the Veritas NetBackup command-line interface documentation for details on the usage of this command.

Appendix C: Throttling different classifications of backups

The information covered in this appendix is not directly related to Data Classifications and Storage Lifecycle Policies as they are implemented in Veritas NetBackup 6.5; however, applying Data Classifications to Backup Policies using this approach simplifies the management of those policies and makes them easier to identify.

The DiskPool concept that forms a key component of the Flexible Disk Option in Veritas NetBackup 6.5 allows multiple Storage Units to share the same pool of disks. This feature can be used in conjunction with the “Maximum Concurrent Jobs” setting on the Storage Units to limit the number of jobs using each Storage Unit and thus prioritize one set of jobs (using a Storage Unit that allows a large number of concurrent jobs) over another (using a Storage Unit that allows a smaller number of concurrent jobs).

For example, a large AdvancedDisk or SharedDisk DiskPool can be configured as two Storage Units called, respectively, Gold_DSU and Silver_DSU. Gold_DSU has maximum concurrent jobs set to 10, and Silver_DSU has maximum concurrent jobs set to 5. Backups are configured so that the more critical backups run to Gold_DSU, and the less critical run to Silver_DSU. All the backups have a start window between 9:00pm and 5:00am, and all have the same priority. As the Gold_DSU allows twice as many concurrent jobs as Silver_DSU, more of critical backups will run at once—thus they will complete first.

Taking this a step further, suppose you have a SharedDisk DiskPool with two fast fiber–connected Media Servers, and three ordinary Media Servers. You could configure a Platinum_DSU to use the two fast servers, with a higher "Maximum Concurrent Jobs" setting, and a Gold_DSU and a Bronze_DSU that use the ordinary Media Servers. Figure 9 shows how this configuration could be laid out.

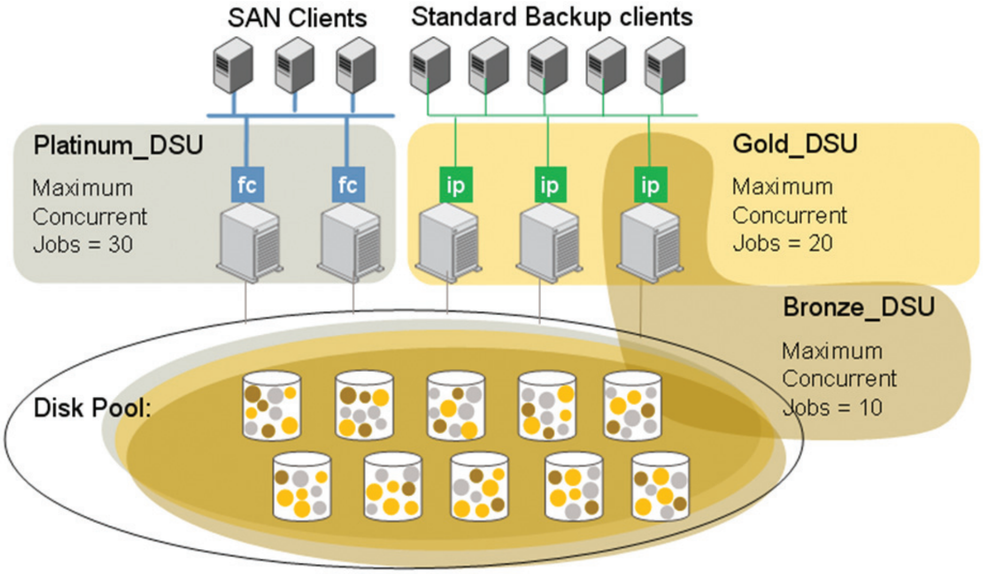

SAN Clients
Standard Backup clients
Platinum_DSU
Maximum Concurrent Jobs = 30
fc
fc
ip
ip
ip
Gold_DSU
Maximum Concurrent Jobs = 20
Bronze_DSU
Maximum Concurrent Jobs = 10
Disk Pool:

Figure 11. Jobs prioritized by concurrency

About Symantec

Symantec is a global leader in infrastructure software, enabling businesses and consumers to have confidence in a connected world. The company helps customers protect their infrastructure, information, and interactions by delivering software and services that address risks to security, availability, compliance, and performance. Headquartered in Cupertino, Calif., Symantec has operations in 40 countries. More information is available at www.symantec.com.

For specific country offices and contact numbers, please visit our Web site. For product information in the U.S., call toll-free 1 (800) 745 6054.

Symantec Corporation
World Headquarters
20330 Stevens Creek Boulevard
Cupertino, CA 95014 USA
+1 (408) 517 8000
1 (800) 721 3934
www.symantec.com

11/07 13036392