Best Practices Guide for Tape with NearStore™ Appliances

Nicholas Wilhelm-Olsen | NetApp | TR-3149

TECHNICAL REPORT

NetApp, a pioneer and industry leader in data storage technology, helps organizations understand and meet complex technical challenges with advanced storage solutions and global data management strategies.

NetApp Inc.

Table of Contents

1. Introduction

The NearStore product family is a part of NetApp's revolution in data protection. SnapVault™ provides a simple, efficient, and reliable method of performing backup-and-recovery operations of NetApp™ storage systems, as well as storage attached to open systems, utilizing NearStore appliances. In addition, products from NetApp's data protection partners can also be used in conjunction with NearStore to protect data on open and closed systems clients and servers. SnapVault and these partner products dramatically reduce the total cost of data protection and improve overall backup-and-recovery performance. For an overview of applications utilizing NearStore, please reference Nearline Applications for the Enterprise Utilizing NearStore.

While the use of NearStore reduces the total number of backup-and-recovery operations using traditional tape technology, it is essential to understand it is not a complete replacement for tape-based data protection. Tape technology remains the most cost-effective means of providing long-term data protection and data archiving. It should also be an essential part of all disaster recovery plans.

Since the NearStore product family utilizes the Data ONTAP™ operating system from NetApp, it shares all of the same tape-based data protection capabilities as NetApp storage systems. For more information about these capabilities, please reference Data Protection Strategies for NetApp Storage Systems and Data Protection Solutions Overview. The purpose of this paper is to highlight specific concerns and recommendations with respect to tape-based data protection of NearStore appliances.

This paper opens with recommendations for initial configuration of the NearStore appliance. It then proceeds with factors relevant to the performance and scalability of tape-based backup and recovery. The paper closes with a recap of these issues and presents overall recommendations on tape-based data protection for NearStore appliances.

2. Initial Configuration Recommendations

The NearStore family from NetApp began with the R100, and nowincludes the R150. In order to establish optimal performance and scalability of these systems with respect to tape-based data protection, it is essential to first discuss a few key architectural elements.

The R100 and R150 achieve their low-cost advantage in part by utilizing ATA-based disk drives in UltraWide SCSI drive enclosures. Each drive enclosure holds twelve ATA disk drives on three SCSI-ATA bridges for a total of four drives per bridge. The R100 has either four or seven of these enclosures, resulting in approximately 7 or 12TB of raw storage capacity. The R150 has either five or nine of these enclosures with larger ATA-based drives, resulting in approximately 12 or 24TB of raw storage capacity.

In order to maximize the performance of these systems, RAID groups within a volume should be constructed so that drives are balanced as evenly as possible across separate SCSI controllers. In addition, if more than one drive in a RAID group must reside on the same SCSI controller, it should be on a separate SCSI-ATA bridge. This not only increases performance, but eliminates the chance a bridge failure will take an entire RAID group offline. Balancing drives across controllers and bridges is performed automatically during volume creation when using Data ONTAP 6.4 and newer. *Figure 1* illustrates an eight-disk RAID group constructed to these recommendations.

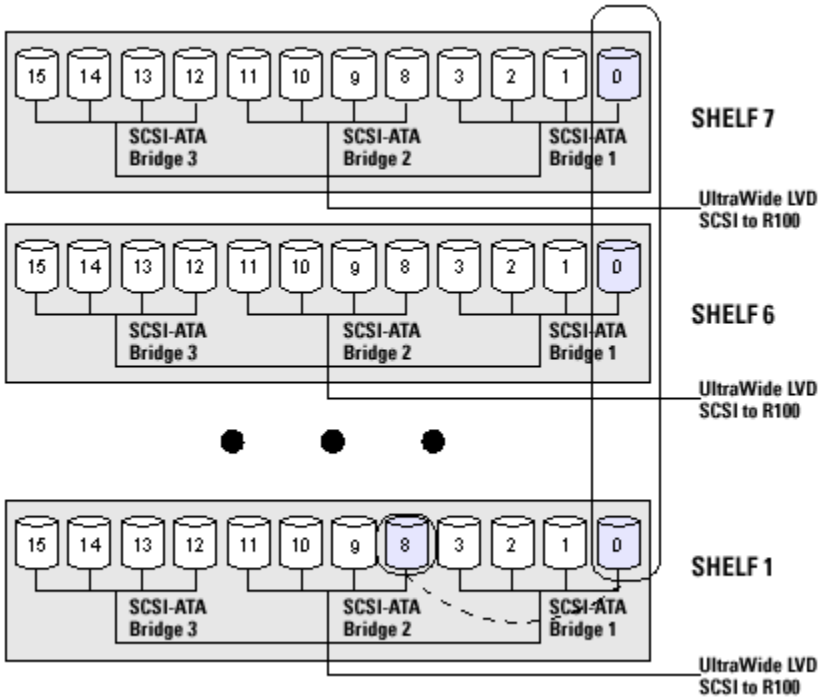

Figure 1. Recommended layout for volumes containing eight-disk RAID groups.

When constructing volumes on a NearStore system, it is essential to maximize read and write performance while minimizing the time to reconstruct a failed disk. For this reason, volumes that will be backed up to tape-based devices should contain one or more eight-disk RAID groups. Volumes containing eight-disk RAID groups offer the best balance between read and write performance and keeping RAID reconstruct times to an acceptable length.

Volumes consisting of a single eight-disk RAID group are slightly less than 1TB on an R100 and approximately 1.3TB on an R150. An R100 with ten or R150 with thirteen of these volumes maximizes the total usable storage space as well as overall read and write performance. The largest possible volume that can be constructed on a NearStore running Data ONTAP 6.4 or newer consists of twelve eight-disk RAID groups for a volume size of approximately 15TB. While adding multiple eight-disk RAID groups to a volume will increase overall read and write performance of the volume, this will not be sufficient to compensate for the extra time required to back up the additional capacity of the volume to tape. As will be discussed in the next section, writing to more than one tape drive per volume concurrently with NearStore might provide better performance, depending on the drive transfer rate. Therefore, only volumes of one or two eight-disk RAID groups are recommended.

3. Backup Recommendations

All methods applicable to NetApp storage systems for backing up and restoring data to and from tape are also applicable to NearStore. However, these methods can exhibit very different performance characteristics on NearStore than on traditional NetApp storage systems. In addition, proper data protection strategies for NearStore are dependent on the characteristics of the data stored upon it. This section will discuss both of these points.

When performing a tape backup of a volume on a NetApp storage system or NearStore, the volume can be backed up at one of three levels, shown in Figure 2. The first is backing up the entire volume in one operation, also referred to as a full-volume backup. The second is backing up by individual qtree within the volume, known as a full-qtree backup. The third is backing up individual subdirectories, known as directory-level backup. While a full-volume backup is usually the quickest way to back up an entire volume, it is sometimes possible to use multiple concurrent full-qtree operations to meet or exceed full-volume backup performance. Directory-level backups generally exhibit substantially slower performance and as such should be avoided.

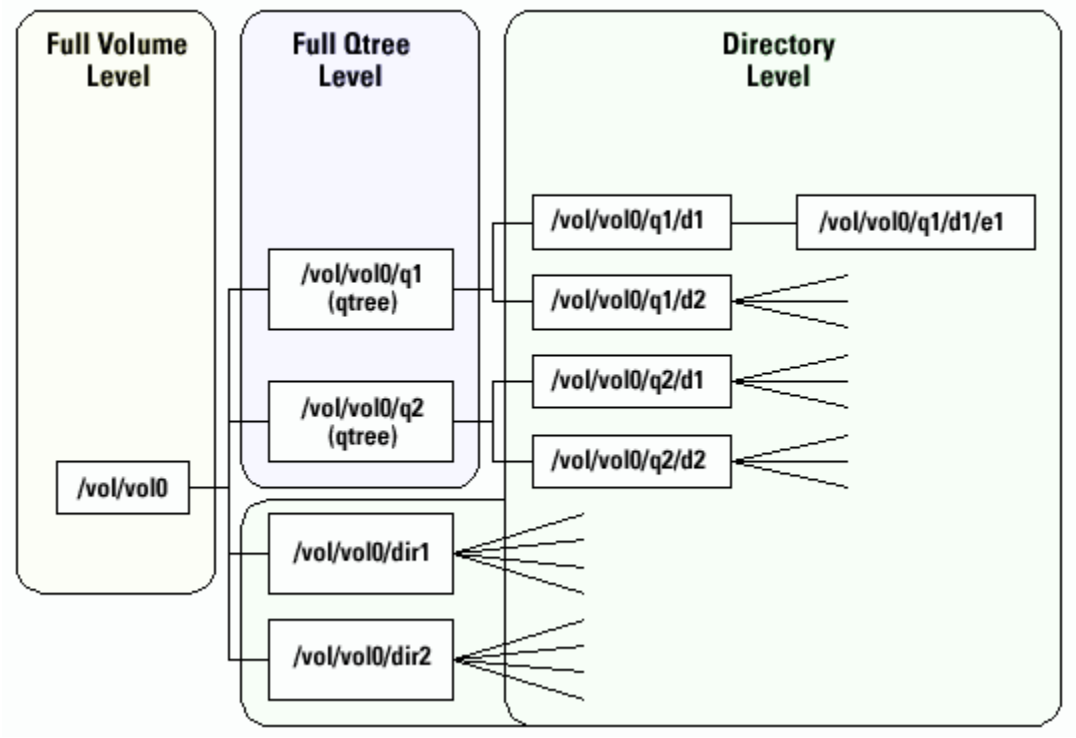

Figure 2. Backup levels.

Due to the inherent performance characteristics of NearStore volumes as outlined in the previous section, recommendations for backup levels are more complicated than for other NetApp storage systems. If using tape drives with a native transfer rate less than 10MB per second such as DLT 8000 and AIT-2, the best method of backing up a volume is using between two and four concurrent full-qtree backup operations. If using tape drives with native speeds over 10MB per second such as LTO and AIT-3, multiple concurrent qtree backup operations from one volume will not transmit data quickly enough to the tape drives to keep them operating at their minimum required transfer rate, causing the drives to pause and drastically reducing overall performance. Therefore, with tape drives over 10MB per second, only full-volume backups are recommended.

Data on a NearStore system will fall into one of two classifications. The first classification is data that is either copied or replicated to the device via SnapMirror, SnapVault, or some other data replication software. This data is logically organized on a NearStore system in a manner that

makes it meaningful and easy to use. Incremental tape backups of data in this category will be a valid option as long as the rate of change is low. However, daily incremental tape backups should be avoided in favor of online Snapshot™ backups.

The second classification is virtual tape images from third-party backup software. These images do not have a structure or organization comprehensible without significant knowledge of the internals of the software application that recorded them. They generally consist of several very large files. Incremental tape backups of this data are essentially full backups and therefore not a valid option.

The best method for recording virtual tape images to tape is through the backup application itself via cloning or duplication. Cloning or duplication will record the data onto tape in a format readable natively by the backup application, essentially transferring the virtual tape image into a real tape image. This may require some scripting or manual intervention, but will completely eliminate the need to restore these images back onto the NearStore device before restoring the data back to the original source, thereby dramatically improving the performance and simplicity of recovering data from these images after they are recorded to tape.

With the notable exception above, restoring data from tape to NearStore is no different from restoring data from tape to NetApp storage systems. It is generally appropriate when preparing for a large restore operation to allocate twice the amount of time required for the backup of the data. For this reason, whenever possible restore operations should be made from online Snapshots. Restoring from a Snapshot is nearly instantaneous. Entire volumes can be restored in seconds, regardless of how large they are.

4. Summary

If the NearStore is properly configured as outlined in this document, backup-and-restore performance will approach that of other NetApp storage systems. Special consideration should be taken when designing volumes initially. Volumes should consist of multiples of 8 disks evenly distributed across the available SCSI busses and ATA-SCSI bridges. Backup operations to tape should generally be full backups at either the full-volume level or the qtree level if slower tape devices are in use. When recording virtual tape images to tape, use the cloning or duplication features within the backup software application whenever possible.

The recommendations made in this paper are targeted at providing best practices for the majority of environments. It does not and cannot provide the best practices for all environments. In general, following the guidelines presented herein will result in one of the most flexible and scalable backup-and-recovery environments in existence today.

5. Resources

Data Protection Solutions Portal
http://www.netapp.com/solutions/data_protection.html

NearStore Product Information
http://www.netapp.com/products/nearstore/

Data Protection Solutions for NetApp controllers
http://www.netapp.com/tech_library/3131.html

Data Protection Strategies for NetApp controllers
http://www.netapp.com/tech_library/3066.html

Nearline Applications for the Enterprise Utilizing NearStore
http://www.netapp.com/tech_library/3187.html

Network Appliance, Inc.
495 East Java Drive
Sunnyvale, CA 94089
www.netapp.com