



TiBS Disk Library Interface Improves Backup and Recovery Performance

Teradactyl LLC.
2301 Yale Blvd. S.E., Suite C-7
Albuquerque, NM 87106
U.S. & International Sales: (505) 242-1091

info@teradactyl.com

www.teradactyl.com

Introduction

The True incremental Backup System® (TiBS) backup server capabilities have been extended to include a new Disk Library Interface (DLI) for storage of backup data on disk. This paper discusses the benefits to backup and recovery processing using TiBS with the new DLI. The DLI extends the backup and recovery processing capabilities of the patented multiple-level backup consolidation technology, TeraMerge®. For more detailed information about how this technology works, refer to the Teradactyl white paper, “TeraMerge® Enables Multiple Level Backup Consolidation for the Full Version of the True incremental Backup System”. Below are some basic definitions about the types of backups available with TiBS:

Network Full

In a full or epoch network backup, all of the primary storage on a target system is backed up. The data backed up in a Full Backup includes file data and meta data such as directories, folders, filenames, file attributes, access control lists and other security information.

True incremental

A True incremental backup will only backup the files that have been modified or created since the most recent backup from the client. Note: The last client backup could be at any level.

Network Synthetic Cumulative Incremental

This unique type of backup takes a True incremental backup from a backup client and consolidates it with previous cumulative incremental backup data on a backup server to produce a new cumulative incremental backup.

Synthetic Cumulative Incremental

A special type of Cumulative Incremental that will consolidate some or all of the Backup Volumes that have been created since the most recent backup at level n-1 or lower on the Backup System. This process is performed by a Backup System with no interaction with the Primary Storage on the target system. The resulting backup volume contains the same data as one that would have been taken from the target system for that period of time.

Synthetic Full

A type of Full Backup that consolidates Incremental Backups with a previous Full Backup to produce a current Full Backup. This process is performed by a backup system with no interaction with the target system. The resulting backup volume contains the same data as one that would have been taken from the target system at that point in time.

Expanded Disk Storage Capabilities for Backup Servers

TiBS has been traditionally modeled as a Disk-to-Disk-to-Tape (D2D2T) solution. Data is backed up from primary storage disks to disks on a backup server that implements a backup cache. Data in the backup cache is then copied to tape and, in most cases, deleted to make room for new backups. The amount of disk allocated to the backup cache is typically about 10% of the total primary storage size. Additional disk space can be added to the backup cache to allow for large file systems of several terabytes in size to be backed up as a single volume.

The DLI extends the use of the TiBS backup cache by allowing any backup volume to be stored on disk in addition to or instead of tape. Significantly larger amounts of disk storage may be added to a backup server to store backup data. The introduction of ever cheaper and larger disk drives means that backup server disk storage can be expanded to provide many benefits over backup caching alone. This new capability enhances processing for traditional D2D2T backup servers. The DLI also provides for a pure Disk-to-Disk (D2D) backup server implementation. In either case, the DLI allows for a high granularity of data on disk. This may include a site’s most critical data, which can now be stored entirely on disk and optionally on tape for offsite storage and true disaster recovery.

Improved Data Recovery from Disk

One of the greatest benefits of the DLI is the faster restore time that disk can provide over tape. The random, high-speed, and instantaneous access of hard disk trumps the load delays, seek times, and typically slower read performance of tape.

The DLI works with the TiBS backup classes to provide any number of storage policies for disk and/or tape. Critical site data can be retained for a longer period on disk for faster restore processing. Less critical data can be retained for a shorter period on disk or even just on tape. This flexibility allows site administrators to manage their recovery time requirements more cost effectively.

Data from very large partitions may be split to provide for more manageable backups using TiBS. The additional restore time required to restore a complete partition that has been split can now be mitigated by using the DLI to keep more of this backup data on

disk to improve restore time for each subsection of data. For the Andrew File System (AFS), the ability to restore multiple volumes in parallel from the disk library will significantly improve the restore time of a vice partition recovery. TiBS file checksum analysis works with data in the disk library to ensure proper restore of all file data.

Reduced Server Processing of Network Incremental Backups

The TiBS backup cache stores the most recent incremental data as a Network Synthetic Cumulative Incremental backup volume. This unique type of backup takes a True incremental backup from a backup client and consolidates it with a previous Network Synthetic Cumulative Incremental backup on the backup server disk cache to produce a new backup volume. The True incremental backup will only backup the files that have been modified or created since the most recent backup from the client. The DLI allows each new True incremental backup to be maintained as a separate backup volume on disk and optional tape. True incremental backups differ from Network Synthetic Cumulative Incremental backups in that they do not have to consolidate existing data in the backup cache during each new network incremental backup.

True incremental backup volumes typically consume about 50% less storage compared with Network Synthetic Cumulative Incremental volumes for the same data retention period. This results in faster backup cycles. In a D2D2T configuration, there is reduced wear and tear on tape devices. Reduced backup server processing allows the same backup hardware to support more backup clients and/or more data.

The elimination of Network Synthetic Cumulative Incremental processing can reduce network backup processing time significantly. A typical cumulative incremental backup contains about 50% new, network data, and about 50% data from the previous incremental backup data. Not having to read from disk and write to tape the 50% previous cache data will result in at least a 50% reduction in the overall workload for nightly network backups. The actual reduction in workload is actually even higher to do the reduction in read/write contention on the backup server disks.

Reduced Processing Time of Lower Level Backup Consolidations

In addition to storing True incremental backup volumes on disk, TiBS can also store the most recent lower level Synthetic Cumulative Incremental backup volumes in the DLI. Lower level backup consolidation processing can be performed entirely on disk. In a D2D2T configuration, freeing tape resources allows more data to be written to tape at a faster rate.

Full Integration with Tapes and Tape Library Storage

Sufficiently large disk libraries will have the ability to store lower level backup volumes, including full backups on disk. This allows a TiBS backup server to implement a Disk-to-Disk (D2D) backup solution. While this is a popular mechanism for providing backup services today, Teradactyl currently maintains that Disk-to-Disk-to-Tape (D2D2T) is still a superior solution for many reasons including:

1. **Data Retention:** Tape is still the most cost effective solution for long term storage of data. Keeping multiple backup levels for an extended period of time in a D2D solution requires backup disk sizes that are many times larger than the amount of data that is being protected.
2. **Offsite Management:** Removable media tape is easier and cheaper to move offsite for sites that do not maintain an online, offsite tape library or backup server.
3. **Security:** Tape is immune to viruses and worms and it is easier to recover data from a failed tape than from a failed disk drive. Superior portability of tape provides security against fire, theft, and natural disaster

The DLI works seamlessly with the tape storage and the TiBS Automated Tape Library Interface (ATLI). Site administrators can define separate retention policies for disk and tape. This allows more data to be retained for a longer period of time on tape more cost-effectively than with a disk only solution. True incremental backups have the ability to extend the total amount of data that can be supported by the same tape library hardware versus cumulative incremental backups.

Conclusions

Cost difference between disk and tape storage continues to shrink. Our new Disk Library Interface will allow sites to increase their use of disk based backup storage and take advantage of all the benefits that it provides. The DLI provides efficiencies in backup and recovery processing that is well worth the investment.