An Oracle White Paper
June 2011

StorageTek In-Drive Reclaim Accelerator for the StorageTek T10000B Tape Drive and StorageTek Virtual Storage Manager

ORACLE

Introduction

This paper explains how Oracle's StorageTek In-Drive Reclaim Accelerator works in conjunction with Oracle's StorageTek T10000B tape drive and StorageTek Virtual Storage Manager to address performance and scalability challenges typically encountered in today's tape archive environment. It describes in detail the challenges of tape storage as a serial medium, the limitations created by common organization formats such as TAR, and tape performance itself.

Existing technologies define linear tape as a single serial stream of stored information and data. This serial stream winds across thousands of meters of tape in a serpentine fashion and is not designed to support efficient random access. For applications that are not sequential in nature, this fundamental characteristic presents a significant challenge, and can result in long access times and inefficient use of space.

Partitions are a tape technology that have been around for many years, and have been used to segment tape media into a few areas that act like independent tape volumes. Typically, one of these segments is dedicated to metadata or system information and the other segment is used for user data. Unfortunately, efficient partition sizing requires that applications have some knowledge of how much storage space is required before it is allocated. In today's dynamic storage environments, with virtualization, data transformations and de-duplication, this becomes an impossible task.

Existing tape media can store a TB (terabyte) of user data. In the near future, a single tape media will hold tens of terabytes, with 100 TB capacities in the foreseeable future. Managing these multi-terabyte tape cartridges requires a new approach for managing data on tape. To address this need, Oracle has developed a new tape storage format with the StorageTek T10000B tape drive, using an innovative partitioning architecture, that allows the addition or removal of storage space as needed. This new format utilizes an innovative partitioning architecture, which allows the addition or removal of storage space as needed; a capability that

will make the management of high-capacity tape much easier for future applications, accommodate new storage paradigms for the greenest storage technology available, and help solve the problems caused by the digital data explosion.

The Tape Storage Space Problem

Current technology manages tape media as a single stream of serial storage. This serial stream of data has a single beginning of data location and single end of data (EOD) location. Subsequent write operations define a new single EOD, which move down the length of the tape as more data is stored to the tape cartridge. When tape files are updated or modified, the original file is left on tape and the new modified file is added to the end of the serial stream. This process is shown in Figure 1, below. If "File B" must be modified, tape format conventions add the modified File B (designated File B') to the end of the serial stream.

Figure 1: Modified File B written to tape after F

While this approach works fine for sequentially written data that may have a common expiration date for the entire cartridge, it is not optimal for random access data or cases when multiple files are written to tape with different expiration dates. For example, this method results in wasted space since the original File B is left in place. To periodically reclaim this wasted space, the files between B and EOD must be rewritten with the updated File B' replacing the original File B. To avoid data loss, files that will be over written are read from tape (Files C-F) and stored elsewhere. Then every file from the original "File B" location to the end of the serial data stream (EOD) is rewritten with the old "File B" replaced as shown in Figure 2 below. Note: Files C-F designated with a "*" do not change but they are rewritten.

Figure 2: To reclaim space on tape, all files from B to F are rewritten

One can easily see that with tape cartridges exceeding 1 TB in capacity, this reclaim operation can consume a large amount of storage and compute resources. Typically, this process requires several hours per tape. To use tape for archive data, a better architecture is required.

The StorageTek In-Drive Reclaim Accelerator Solution

To address these problems, and to help optimize tape for archive applications, Oracle has developed an innovative new method for using tape partitions. We believe adopting this new StorageTek In-Drive Reclaim Accelerator format will revolutionize tape storage for the 21st century. This solution allows applications to break up tape's single serial data stream into smaller more manageable divisions of storage. The format is implemented using physical partitions that are linked together and managed as a doubly linked list. This simple concept supports removing or adding physical partitions as storage requirements evolve over time.

Figure 3 illustrates how several physical partitions are linked together to store Files A-I. File A is the beginning of this logical volume and is located in physical partition 1. The end of the logical volume is located in physical partition 3. The storage space used for any one logical volume is obtained by linking together any number of physical partitions.

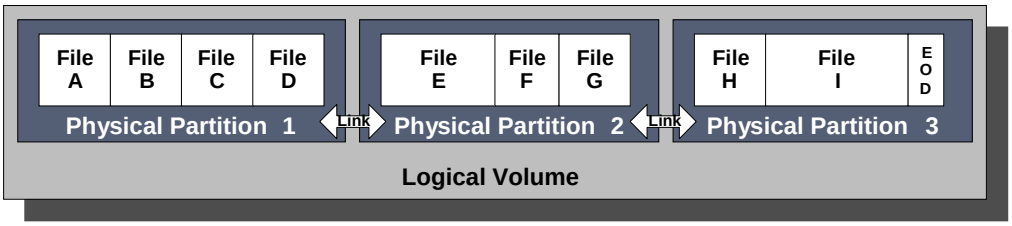

Figure 3: Three partitions linked together

Since this is tape, file access within each logical volume must be handled as a single serial stream. However, each physical partition is accessed randomly and dynamically linked to the logical volume during write. If Files E-G are obsolete and we wish to free that space; that can be accomplished by the application changing the partition map it maintains, as shown below.

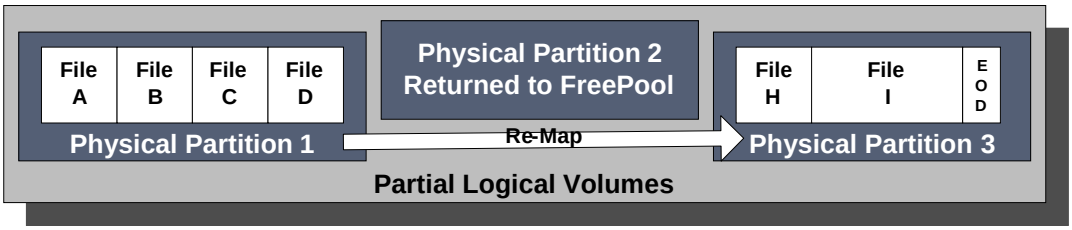

Figure 4: Physical Partition 2 is removed from the logical volume

If we are increasing the size of the logical volume that is accomplished by adding a link from physical partition 3 to physical partition 4, as shown in Figure 5.

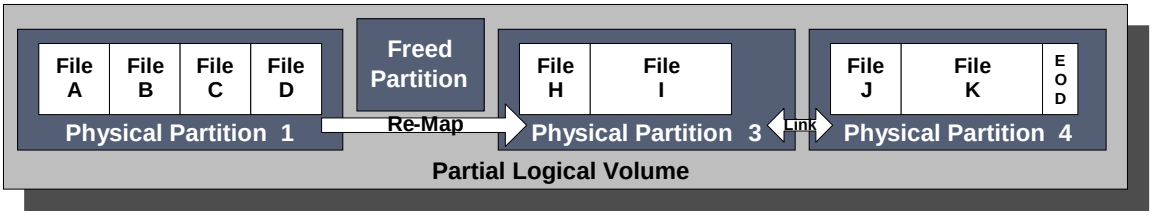

Figure 5: Physical Partition 4 is added to the logical volume

Linking together relatively small physical partitions and creating logical volumes changes that way tape is managed. To grow any logical volume, the application continues writing past the end of physical partition 3. This is shown in the transition between Figure 4 and Figure 5. The next available physical partition is automatically linked by the tape drive into the logical volume space.

Note, the actual link on tape does not change once a physical partition has been linked to a logical volume. This information is embedded in the data format of each physical partition and does not change until that partition is rewritten. For this reason partition maps need to be retained by the application so that it can skip over partitions that have been rewritten.

Tape Reclamation With The StorageTek In-Drive Reclaim Accelerator

With Oracle's new StorageTek In-Drive Reclaim Accelerator feature, there is no need to leave old files on tape. Reclaiming space is as simple as re-linking a single physical partition, as shown in Figure 4. The example is interesting, but usually obsolete files don't fit exactly within a physical partition boundary. The more likely scenario is addressed in Figures 6 through 9, which illustrate modifying only File F and then reclaiming a partition with a new File J.

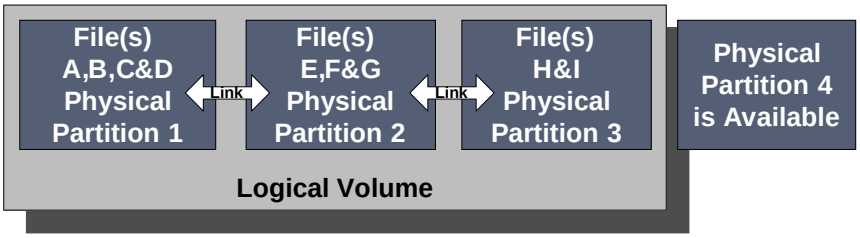

Figure 6: Files A-I are in one Logical Volume, the same as Figure 3

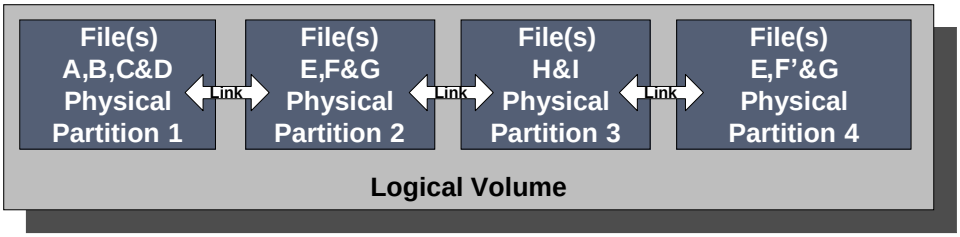

Figure 7: Files E&G are copied with a modified File F' to the end of the Logical Volume

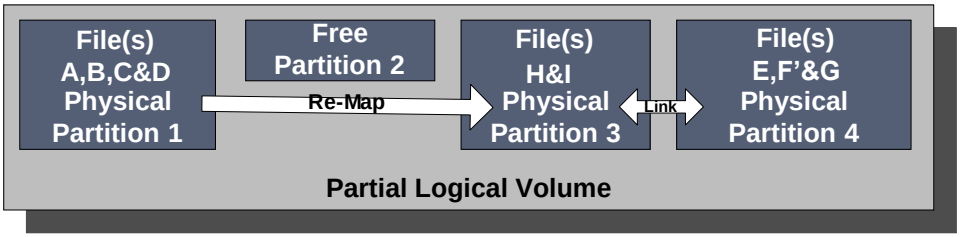

Figure 8: Physical partition 2 is removed from the Logical Volume

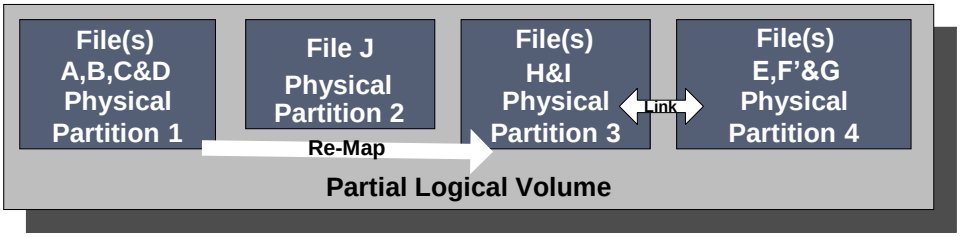

Figure 9: Physical partition 2 is reclaimed with File J

While not as simple as disk, automatically linking during write, and remapping these links manually, allows applications to manage tape using randomly accessible physical partitions. File modification and storage space recovery are only constrained by the time and resources needed to copy and re-map a single physical partition. The reclaim operation may be performed incrementally, and limits the impact on valuable business resources.

StorageTek In-Drive Reclaim Accelerator Implementation

Oracle developed this partitioning format for ease of use and flexibility. A set of vendor unique commands support the FICON interface on the StorageTek Virtual Storage Manager (VSM) system.

The magic of Oracle's partitioning format is that logical volumes do not need to be completely allocated up front. A partitioned tape drive links a physical partition to the logical volume only when it is used. Any unused physical partition can be re-linked at any time to any logical volume.

Implementation of Oracle's partitioning format is based on maps defining which physical partitions are used to build a logical volume. To build a logical volume, physical partitions are structured as a doubly linked list. These maps tell the drive which physical partitions to automatically link as a logical volume grows during write operations. For example, to construct the yellow logical volume in Figure 10, the application provides a partition map reserving physical partitions 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11. Note, this original partition map includes all the partitions on the first two wraps of tape. As the drive writes data to the yellow logical volume, it automatically links together physical partitions in ascending order starting with the lowest numbered physical partition in the map. The logical volume starts in physical partition 0. When physical partition 0 fills the drive creates a link from physical partition 0 to partition 1. After physical partition 1 fills, another link it created from partition 1 to partition 2, and so on. In this example, writing ends in physical partition 6 and an EOD is written. Once the logical volume is complete, any unused physical partitions can be removed from the map and allocated to other logical volumes. The map defining the completely written yellow logical volume shown in Figure 10 contains partitions 0, 1, 2, 3, 4, 5 and 6. Partitions 7, 8, 9, 10 and 11 can be returned to a free pool and used in another logical volume.

	Section 0	Section 1	Section 2	Section 3	Section 4	Section 5	
	Partition 49	Partition 48	Partition 45	Partition 44	Partition 43	Partition 42	
	Partition 36	Partition 37	Partition 38	Partition 39	Partition 40	Partition 41	
	Partition 35	Partition 34	Partition 33	Partition 32	Partition 31	Partition 30	
BOT	Partition 24	Partition 25	Partition 26	Partition 27	Partition 28	Partition 29	EOT
	Partition 23	Partition 22	Partition 21	Partition 20	Partition 19	Partition 18	
	Partition 12	Partition 13	Partition 14	Partition 15	Partition 16	Partition 17	
	Partition 11	Partition 10	Partition 9	Partition 8	Partition 7	Partition 6	
	Partition 0	Partition 1	Partition 2	Partition 3	Partition 4	Partition 5	

Figure 10: Layout of the StorageTek T10000B partitioning format.

To later free space, when files become out of date in this red logical volume, physical partition 1 can be removed from the map. The new partition map would start with a partial logical volume at physical partition 0 and a second partial logical volume containing partitions 2, 3, 4, 5 and 6. At any time the application can read a copy of the associated partition map and links between physical partitions, and dynamically determine which partitions will be added or deleted.

Any StorageTek T10000B tape cartridge can be formatted as either a non-partitioned cartridge or as a partitioned cartridge. The capacity of a non-partitioned cartridge is 1 TB. The capacity of a partitioned cartridge is about 929 GB. The capacity of a partitioned cartridge is reduced because of the servo guard bands that are added between each section. On the partitioned tape there are 6 sections, each containing 32 physical partitions. The partitioned tape has 192 physical partitions available. The

average capacity of a partition is about 4.84 GB. Figure 11 shows the layout of a T10000B partitioned tape.

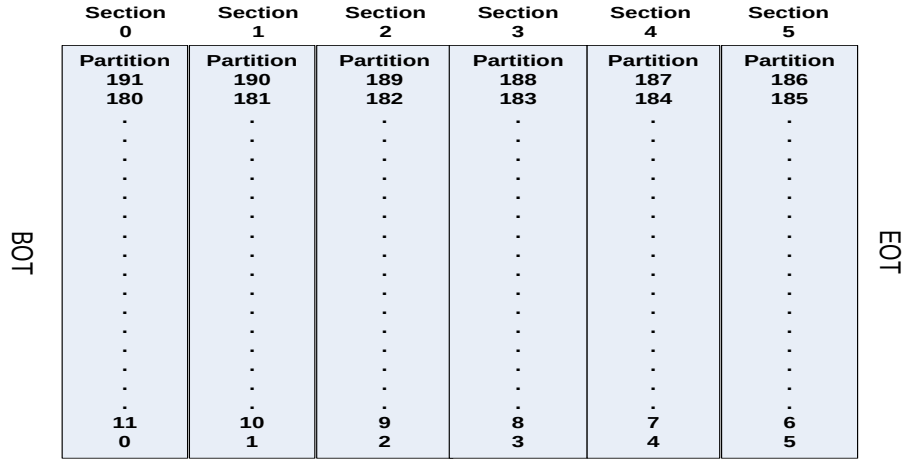

Figure 11: Layout of the T10000B partitioned format.

Like the representative example in Figure 11, the physical partitions serpentine through the tape starting on the lower left hand corner of tape and ending on the upper left hand corner of tape.

Conclusion

By breaking up tape's single serial data stream into smaller, more manageable divisions of storage, the StorageTek In-drive Reclaim Accelerator simplifies the reclamation process and eliminates the need to leave obsolete files on tape. In conjunction with the StorageTek T10000B and StorageTek Virtual Storage Manager, the StorageTek In-drive Reclaim Accelerator enables users to overcome the serial nature of data management on tape, making it possible to fully realize its inherent total cost of ownership advantage.

StorageTek In-Drive Reclaim Accelerator for the
StorageTek T10000B Tape Drive and
StorageTek Virtual Storage Manager
June 2011

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com

Oracle is committed to developing practices and products that help protect the environment

Hardware and Software, Engineered to Work Together