



## WHITE PAPER

# Data De-Duplication Eliminates Tape for NonStop Systems

For most companies the use of magnetic tape is a necessary evil, a remnant of equipment from the early days of computing. Vast and ever-growing storage requirements for archiving to meet corporate and regulatory requirements, on-site failure recovery, and off-site disaster recovery keep tape media requirements growing each year. The majority of Fortune 1000 companies (63%) have at least 1,000 tapes offsite, according to Wave 7 of TheInfoPro Storage Study<sup>1</sup>.

Although magnetic tape has increased in capacity, speed and sophistication over the years, its use continues to come at high cost: media problems leading to recovery failures, physical loss and consequent security and privacy issues, handling and management costs, maintenance costs and headaches. Tape equipment is often one of the highest support and maintenance cost items for data centers.

Recent incorporation of disk-based backup solutions, such as **ETI-NET's EZX/BackBox™** for HP NonStop systems, has significantly reduced the need for use of tape in short-term recovery solutions, while maintaining full transparent operation with the host systems. Short-term backups are stored in disk pools, from which they can be rapidly restored when required. Upon reaching the expiration of their designated retention period, the backup data is automatically deleted from the disk pools and space freed up for subsequent use. This is a highly efficient approach, with no need for removable media handling, risks of tape media failures, operations staff activities, or their associated costs.

However, data that must be retained for longer periods still ends up on tape, even if it gets there via disk pools in the disk-based backup products. The primary reason for this is the lower storage cost. Even with continuing decreases in disk costs, tape costs are also decreasing as well. For very large data storage capacities and long retention times, tape continues to retain a perceived cost advantage. And for off-site storage, primarily for disaster recovery (DR), tape has been a mainstay. Many companies contract with DR providers to store their backup media and provide the capability to restore it onto working systems in the event of a catastrophic data center loss.

Recently this cost dynamic has been changing. **Data Domain** has introduced a range of products that implement "**Capacity Optimized Storage™**" (COS), drastically and cost-effectively increasing effective disk storage capacities for backup data and drastically reducing WAN bandwidth requirements for replication of backup data between sites. Their products are based on an approach often referred to as "data de-duplication." This relies on the fact that typically relatively little data on a host system actually changes between the time of one backup and the next. Most backup utilities, such as NonStop Backup/Restore, take this into account in "incremental" backups, where only the files that have changed are backed up. But incremental techniques only address the tip of the iceberg. Within individual files, only a few bytes may have changed and those files will still be backed up in their entirety during the incremental backup. This is equally true for databases.

NonStop TMF dumps entire database tables even though only a few records in them may have changed.

“Data de-duplication” techniques<sup>2</sup> take incremental backup to its logical conclusion by looking at the content of backup data streams and only storing the actual data elements that have changed since previous backups. For instance, for two consecutive backups of the same database, the first backup would be stored in its entirety, while the second backup would only require storage of the database elements that changed since the first backup (see figure 1.) And this is done without any direct knowledge of the data structures involved on the host system. The underlying technology to accomplish this is very sophisticated and is incorporated in Data Domain’s products.

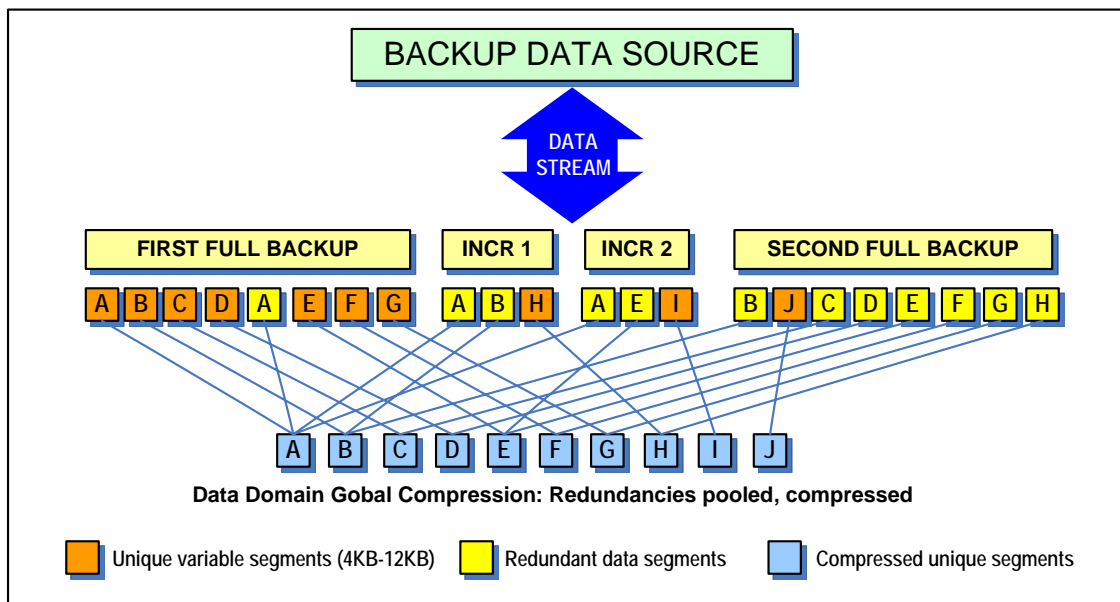


Figure 1: Data De-duplication mechanism

Speaking about data de-duplication in her Computerworld blog<sup>3</sup>, Heidi Biggar, storage systems analyst said: “Data de-duplication is exciting because, as I’ve said before, it has massive, immediate real-world implications for users. In fact, in my 10+ years in the industry, I have never seen a new technology get as much attention -- and, importantly, be incorporated into vendor product lines -- as quickly as data de-duplication.”

ETI-NET has integrated Data Domain’s Restorers™ with its EZX/BackBox™ Non-Stop-integrated virtual tape product, to reduce or totally eliminate the need for use of tape in HP NonStop data centers. By incorporating proprietary engineering designs in both products, NonStop data formats have been made transparently “compressible” by the Data Domain equipment, requiring a small fraction of their normal disk storage requirements for multiple iterations of backups.

EZX/BackBox connects to NonStop S-Series systems via SCSI and NonStop Integrity systems via Fibre Channel, transparently emulating a tape library so that applications such as NonStop Backup/Restore, BR2 and TMF can continue to be used without change. Virtual tape media are automatically cataloged in DSM/TC or TMF, and are protected against accidental deletion or overwriting. For demanding fault-tolerance and availability, the EZX/BackBox product is managed by a

NSK-based application and its catalog and metadata stored on NonStop resident files so that it can optionally be TMF-protected.

Up until now EZX/BackBox has offered the abilities to store the virtual tape media on its internal disks, SAN-based disk, file servers or forward it to enterprise storage managers (ESM) such as Tivoli Storage Manager, Veritas NetBackup or Legato NetWorker for storage in disk and tape pools. With the newly introduced product capabilities, EZX/BackBox can now take advantage of Data Domain's extreme compression capabilities to offer disk-based backup storage that is price-competitive with tape storage. NonStop data centers can now store their on-site archive data on disk, cost-effectively.

Perhaps most exciting are the new DR abilities using the EZX/BackBox-Data Domain products. Since iterations of backups may be compressed by 20-50x or more for storage on disk within the Data Domain Restorers, it is now feasible to copy this compressed data between data centers over affordable WAN telecommunications circuits. So companies with data centers in multiple locations can use them to back-up each other via WAN, rather than move tapes to a DR provider's site. This is easily accomplished by locating EZX/BackBox and Data Domain Restorer combinations at each data center, connecting them to the NonStop systems, and interconnecting them via LAN. EZX/BackBox incorporates the ability to implement one or more "active" backup domains for each data center, while also maintaining "dormant" domains for the remote sites. Backup data from each remote site can be continuously replicated between the Data Domain Restorers in the different sites. Upon a catastrophic site failure, the "dormant" EZX/BackBox domains associated with it can be activated in a designated DR site, making the failed site's backups available for restoration on its NonStop systems. Figure 2 illustrates two NonStop-based data centers serving as DR sites for each other.

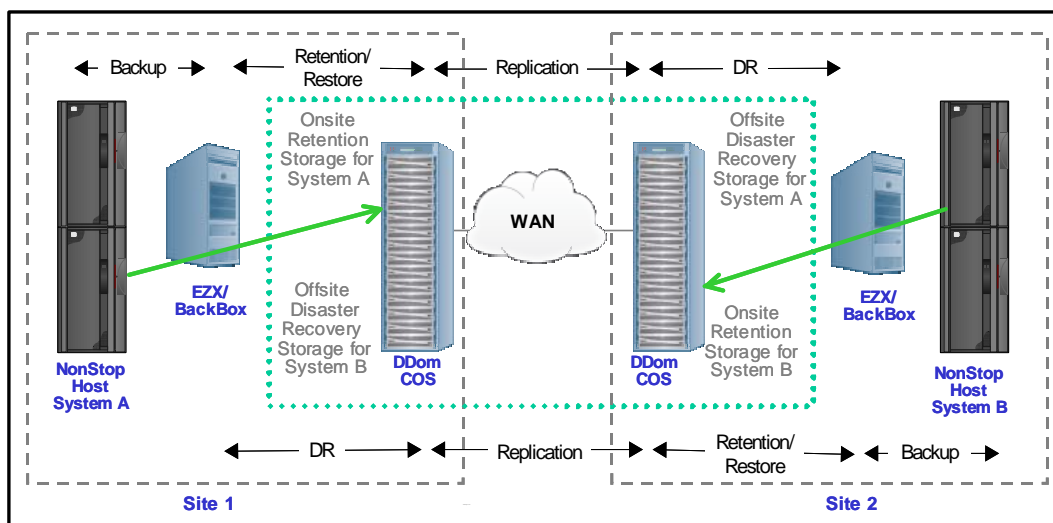


Figure 2: Bi-Directional and Selective Remote Vaulting

Another interesting possibility enabled by this new technology is use of only "full" backups instead of weekly full and daily incrementals, for example. This traditional methodology was adopted to reduce the storage media requirements as well as minimizing the backup window during peak activity periods. But it created a problem in restore time. Where all files, or a significant number of them, must be restored, this often required consecutively restoring the prior full backup and all subsequent incremental backups. The elapsed time for this, including media handling, could be quite long. Since data de-duplication can effectively eliminate the unchanged file data, as well as the

unchanged data within each modified file, the disk storage penalty for use of consecutive full backups is minimized. Of course, full backups on a daily basis will still take longer to complete than incrementals. But for systems where restore time is the primary consideration, consecutive full backups are now a cost-effective option.

### **Summary**

The combination of EZX/BackBox's virtual tape technology for NonStop systems with Data Domain's data de-duplication technology brings together two of the key technologies in "next generation data protection" as cited by InfoStor<sup>4</sup>. While for some uses such as legacy data interchange between companies, tape may remain, the new technology in the EZX/BackBox-Data Domain products now available to HP NonStop users makes elimination of tape use possible and cost-effective for most purposes, as well as enabling new approaches to disaster recovery protection.

### **Company Information**

ETI-NET develops products that help customers manage the complexities of multi-vendor computer systems. ETI-NET products are designed to easily integrate storage resources from dissimilar computers and provide cost-effective consolidation and management of backup and archiving operations. With a product development center in Montreal, Canada and field operations centers in Boca Raton, Florida and San Mateo, California, ETI-NET has been shipping products for HP NonStop systems since 1987. ETI-NET supports customers worldwide and can be contacted by phone at 1-800-546-9101 or by email at [information@etinet.com](mailto:information@etinet.com). To learn more about ETI-NET, visit [www.etinet.com](http://www.etinet.com).

---

<sup>1</sup> *TheInfoPro Storage Study Wave 7*, 2006, [www.theinfopro.net](http://www.theinfopro.net)

<sup>2</sup> *Cut data down to size*, Storage Magazine, July 2006

<sup>3</sup> Heidi Biggar, Computerworld.com blog, April 24, 2006

<sup>4</sup> *Data reduction, VTLs, CDP drive NGDP*, InfoStor, online article, June 6, 2006