Quantum Enhances DXi Deduplication Price-Performance with DXi 2.0 Software for SMB

Analyst: David Reine

Management Summary

I've noticed a problem recently with my local dry cleaner. They continually are shrinking my clothes, especially my pants. I've asked several friends who use the same store if they have noticed the same problem. Unfortunately, they replied, "No." In fact, some of my less sensitive friends implied that my waistline might be expanding and there was no plot against me from the dry cleaner industry. Trying to fit an expanding waist line into pants that were perfectly fine last year is becoming a serious problem. So, I am left with two options, visit the tailor to let the pants out or restrict the number of visits that I make to the refrigerator, reducing the amount of food that I eat. This problem is not totally unique to me, or any other person who enjoys a good meal or two. A similar condition arises every day in every data center around the world.

The data center is generating more and more data every day because of requirements to protect the enterprise with incremental and full backups, in order to meet corporate guidelines and government regulations. Unfortunately, nothing can be done about the amount of time available for the backup window, as there is nothing to be let out! The eight-hour, or less, block of time is fixed; when the first shift comes in at 8:00a.m., the backup must be complete. The IT staff is caught in an impossible situation; they have to backup more and more data in the same time frame that they have been allotted in the past. Therefore, they have to improve throughput, reduce the amount of data going through the pipeline to their backup devices, usually tape, or convince management to lengthen the backup window. Or, perhaps, all three! We know that the third is not going to happen any time soon, so they must figure out a way to speed up the process or reduce the amount of data going through that pipeline, or both. One way to do that is by transitioning your short-term backup requirements from tape to disk and implementing a solution that incorporates data reduction into the data center architecture. Two methods of data reduction are compression, an old standby, and data deduplication, relatively new to the IT scene.

Fortunately, for the data center, Quantum has recently announced DXi 2.0, the next generation operating software for its DXi-Series disk backup and deduplication appliances. By delivering the highest deduplication speeds for any open protocol inline appliance, with no change in price, DXi 2.0 enables users to protect more data on smaller systems, adding to the price-performance advantage of the DXi-Series. Quantum enables the data center to meet their QoS objectives. DXi 2.0 provides in-line, disk-based, variable-length block-based data reduction, integrated into your existing environment to protect your existing backup process. It provides built-in replication to provide automated remote protection for branch offices over existing WANs. DXi 2.0 also provides an easy-to-use direct link to tape systems for long-term protection and a single point of management and support. To learn more about the DXi software and hardware family, please read on.

Data Growth Challenges

Coping with a variety of data backup scenarios has become one of the most significant issues that every data center, across the board, must face on a daily, weekly, and even monthly basis. Expanding data, at a rate of 50%-100% every 12 to 18 months, is creating havoc with backup windows in SMBs and enterprise data centers alike. How does the IT staff do incremental and full backups, within the allotted backup window, reliably, efficiently and cost effectively, without interfering with mission critical applications that are the lifeblood of any enterprise? An even more significant issue: How does the data center recover lost data in the most efficient manner?

In simple terms, traditional backups have not kept pace with the requirements of the modern data center where the need for faster backups and trouble-free restores is mission-critical. Duplicate data is stealing time and money from the enterprise. If you have a 1TB file that is backed up every week, that would result in 53TBs being used for a file that may not have changed, instead of just 2TBs, the original plus one full backup, and pointers. (Imagine the proliferation of disk capacity if that file was backed up on a daily basis!) That cannot be allowed! The data center needs to implement a backup procedure that does not replicate the same file time after time. This is where data deduplication comes in. In its simplest terms, data deduplication eliminates the duplication of redundant data at a sub-file level, replacing the duplicate data with a pointer. IT departments reduce the disk space required to store the backups. Data deduplication appliances also often increase backup speeds enabling IT departments to finish within the backup window. The data center can reduce disk proliferation, and energy consumption, while at the same time, increasing the throughput and data retention periods. As a result, the data center can reduce the total cost of ownership (TCO) of the backup environment. If you have a question regarding the viability of data deduplication as a mainstream technology, the sheer number of vendors providing this technology should put those fears to rest!

No matter which architecture the data center uses, disk to tape (D2T) or disk to disk (D2D), there are concerns: time to completion of the backup, the reliability of restores, the capability for remote protection with bandwidth savings, and, always, cost. In a D2T environment, still flourishing in many data centers, the IT staff is often worried about the recovery time objectives (RTO) and the ability to complete the backup, successfully, in the time allotted. In a D2D scenario, on the other hand, there is more concern for the TCO issues[1] in trying to save and restore more short-term data to disk, especially where facility issues are a problem in terms of floor space and energy. Ideally, the data center staff should want to protect the enterprise retaining short-term backups in a D2D environment, for faster recovery times, while reserving tape for use in a D2D2T environment for long-term protection, reducing the TCO.

Once a decision has been made to invest in a data reduction solution, the IT staff must consider a wide variety of data deduplication techniques, including how and where to do the deduplication: using file-level dedupe or block-level, in-line versus post-process, at the source or at the target. Keep in mind that each implementation carries with it its own overhead. With in-line processing, the data deduplication process is executed as data flows through the pipeline, with the deduplication being performed dynamically, so that it can minimize the I/O traffic across the network. In post-process architectures, the deduplication is executed after the data has been stored into the backup environment. On the other hand, the in-line process is a very CPU-intensive exercise, potentially causing a bottleneck between the backup server and the storage target. Today's multi-core Intel *Xeon* and AMD *Opteron* processors, along with solid-state disks (SSDs) for I/O intensive environments, should, however, provide adequate performance to deduplicate the largest files.

While compression searches for repetitive strings of data within a single file or part of a fixed block of storage (i.e., a LUN), file or block level deduplication identifies entire files or blocks that are identical and saves it only once. Unfortunately, a single change to one paragraph of each file, for example, makes them all different (i.e., unique) and saves them all. Using sub-LUN or sub-file, deduplication enables the data center to backup only the changed parts, saving both storage space, time, and cost. Some implementations of block-level dedupe compare fixed-length chunks of data, while others use a variable-length chunk to identify redundant data.

---

[1] See the issue of **Clipper Notes** dated December 20, 2010, entitled *In Search of the Long-Term Archiving Solution – Tape Delivers Significant TCO Advantages over Disk,* and available at http://www.clipper.com/research/TCG2010054.pdf.

Clearly, chunk-level deduplication is more efficient and effective than the whole file or whole LUN deduplication, and comparing variable-length chunks is preferred over the fixed-chunk method, as it is better at dealing with smaller changes. Furthermore, data deduplication can be integrated with your existing storage and backup environment to simplify operations.

Data deduplication can be performed by software within the backup application or it can be deployed as a data-deduplication appliance. With the former, less data is transmitted over the network; however, because data deduplication is a high-overhead process, it may slow down the overall performance of the operation. With the appliance method, systems can be added, or replaced, with little or no changes to the backup procedures. The same is true for the in-line method: if your backup server is already running at maximum efficiency, then turning on the deduplication process in-line would take compute resources away from the backup jobs, potentially causing further delays in your backup window. To compensate for this, the data center might have to deploy a more powerful backup engine to both do the mission-critical backups and the data deduplication. When the data reduction process is deployed on a purpose-built appliance, there is no effect on the available compute capacity of the backup client or media server. Any of these approaches will work; your choice depends upon the bottlenecks in your environment.

How much data can be placed by pointers is another matter for conjecture. Dedupe ratios vary widely with some vendors claiming 10:1 while others speculate on a 50:1 ratio. This will vary with the data and with the backup frequency. If your data center employs a daily backup of data that changes infrequently, then your savings may approach 50:1. However, if your data center does full backups less frequently, with data that is changing constantly, then you are likely to experience savings at the lower end of the scale. In fact, with some data, such as videos and geo-physical reports, you may not experience any savings at all.

Our intention here is not to explain how data deduplication works, but to describe the benefits of it and how Quantum's *DXi* Series, identifies duplicate data, and eliminates it, reducing your TCO to a size that fits your budget. Keep in mind that acquisition cost is but one factor. Others that must be considered include performance as a factor of satisfying backup window

limitations, scalability as data grows, restore time, ease-of-use for deployment and maintenance, and integration with tape for long-term data protection.

The Quantum DXi Solution

With the availability of the latest generation of Quantum's software platform, *DXi 2.0*, for its DXi Series backup and data deduplication solutions, Quantum has delivered a purpose-built appliance for *in-line* data deduplication, as compared to previous solutions that employed an adaptive approach to data reduction, in order to streamline the operation. In keeping with their history as an innovator in disk-based backup and data deduplication technology, Quantum has retained its variable-length chunk dedupe, thus increasing the efficiency of the data deduplication process. Quantum now delivers even more performance and improved data reduction when compared to other inline appliances in the SMB and mid-range category. In fact, DXi 2.0 doubles the performance of the previous generation of DXi products *at the same price*, and Quantum claims up to a five times advantage over competitive systems in price-performance ratios. As a result, data centers around the world can now protect even more data on even smaller systems, establishing a lower TCO and an improved price/performance ratio, using commodity products, such as Symantec *Open Storage (OST)*[2] or a conventional NAS interface, for seamless integration into the existing architecture.

DXi 2.0 software is optimized for Quantum's newest purpose-built deduplication appliances, from the *DXi4500* through the *DXi8500*, which include the latest x86 processors and, for the *DXi6500*, which include dual quad-core *Nehalem* CPUs and SSDs[3] to manage the indexes that are used in deduplication. The DXi-Series both simplifies the data path and improves data protection performance. This enables the IT staff to meet shrinking backup window requirements and satisfy the QoS agreements with their users. It also reduces the TCO of the backup processes.

The DXi 4500, running DXi 2.0 software will deduplicate data at a rate of up to 1.4TB per hour for NAS applications and up to 1.7TB per hour for OST, faster than any other entry-level

---

[2] The OST interface works with many backup appliances to provide an additional level of operational advantages, such as increased performance and better replication management.
[3] SLC SSDs from Intel.

open-protocol appliance, raising the bar for entry-level data reduction. With a DXi6500 deployed, the data center can deduplicate at 4.3TB per hour for NAS and up to 4.6TB per hour for OST, faster than any other mid-range open protocol appliance[4], where data centers are looking to dedupe up to 25TBs in total capacity. Both of these solutions are currently available. Quantum also has announced future availability of DXi 2.0 with improved performance for their *DXi6700 VTL* and the multi-protocol, enterprise-class DXi-8500 appliance for high-performance requirements. Please keep in mind that although deduplication ratios of up to 50:1 are possible, *your mileage may vary*.

Quantum clients with the latest versions of DXi4500 and DXi6500 can take advantage of the improved performance with an on-site upgrade to DXi 2.0, *at no charge*. DXi 2.0 offers full replication compatibility with older DXi appliances, in order to protect user's prior investments. DXi 2.0 also has enhanced management features that improve the deployment process by reducing the number of screens and steps by up to 65% and by adding new installation wizards.

The DXi Series provides built-in replication for automated remote protection for branch offices over existing WANs. It has an easy-to-use link to tape systems with a single point of management and support, in order to reduce the TCO of long-term storage. The following models are available in 1Q11.

DXi4500

DXi4500 appliances with DXi 2.0 are easy-to-use and affordable. They accelerate backup and improve the restore process, reducing the TCO for data protection. The DXi4500 supports the small- to medium-sized business or branch office. It is a turnkey data deduplication solution that can be deployed rapidly and seamlessly with all leading backup applications, including *NetBackup* and *Backup Exec 2010* under the OST API. Each DXi4500 includes support for remote replication, virtual environments, and an OST interface as standard features. They automate and centralize the disaster recovery process and are capable of virtualization.

The *DXi4510*, with an entry price of only $12,500, is configured with 2.2TB of usable capacity, while the *DXi4520*, with an entry price of $22,500, has twice that at 4.4TB, with both hav-

ing a throughput of up to 1.7TB/hour. Communication between the host and appliance is supported by four 10/100/1000 BaseT Ethernet ports. The DXi4500 supports presentation layers for Symantec OST, as well as NAS backup targets using *CIFS* or *NFS*. The throughput for OST is 1.7TB per hour and 1.4TB per hour for NAS. Asynchronous replication is provided for one-to-one or multiple-to-one configurations.

DXi6500

The DXi6500 Series with DXi 2.0 installed is a family of mid-range disk backup appliances that scale from 8TB up to 56TB in usable capacity for the medium-sized enterprise, with an entry-price of only $64,000. Like their smaller brethren, they provide an easy to deploy and easy to use solution for the data center backup environment, and they are affordable, with all software licenses included in the base price[5]. With DXi 2.0 installed, they provide 2X or higher throughput than their predecessors, with an ingest rate of 2.5TB per hour and in-line deduplication performance of up to 4.6TB per hour with an OST interface and 4.3TB per hour for NAS. Quantum's deduplication software allows data centers to cost effectively retain backups on disk for an extended period of time, providing faster recoveries and increased recovery points. The DXi6500 provides value for a wide variety of enterprise data, including databases, email, and user data. This provides the medium-sized enterprise with a unique level of price-performance.

The DXi6500 also can be deployed rapidly and seamlessly with all leading backup applications, including NetBackup and Backup Exec 2010 under the OST API. The DXi6500 replicates encrypted backup data between sites using global deduplication to reduce network bandwidth requirements for disaster recovery. They have advanced connectivity options that include 10GbE (optical and copper) and 8Gb Fibre Channel (FC). Furthermore, the DXi6500 provides direct tape creation in conjunction with the Symantec Open Storage API, providing long-term data protection. All DXi6500 models support both virtual environments and conventional servers. Configurability of the five members of the DXi-6500 family is as follows.

- **DXi6510** supports a usable capacity of 8TB, with two 1GbE ports in a single system node. This is Quantum's

---

[4] According to Quantum.

[5] This includes replication and OST.

most affordable mid-range deduplication appliance.

- *DXi6520* supports a usable capacity from 8TB up to 32TB, with six 1GbE ports in a system node and up to three expansion modules.
- *DXi6530* supports a usable capacity from 24TB up to 56TB, with six 1GbE ports in a system node and up to six expansion modules.
- *DXi6540* supports the same configuration as the DXi6530 along with a direct path to tape using two 8Gb FC ports (and this appliance may be upgraded in the field to 10GbE connectivity).
- *DXi6550* supports the same configuration as the DXi6540 along with 10GbE connectivity.

Conclusion

Data deduplication has become a mainstream technology in the data center. It provides the IT staff with a viable method of data reduction beyond simple compression. However, data deduplication can be far from simple, as we have seen in the number of methods and technologies that have been applied to it. Trying a do-it-yourself solution is not a recommended course of action. Many vendors have deployed integrated solutions for the largest enterprise data centers. Quantum, however, has also paid strict attention to the SMB and mid-range data center in this release and has announced plans to port the new operating software to its Enterprise appliance early in the summer of 2011.

With a lower TCO (including acquisition costs of the data deduplication appliance) and cost of DXi 2.0 software (including deduplication, replication and interfaces), Quantum is making a concerted effort to provide a better price-performance benefit for the mid-sized data centers. By reducing the number of disk devices required, saving floor space and energy, Quantum has enabled the data center to improve their bottom line while maintaining the highest QoS to their customers. Quantum is also delivering to the data center an efficient, affordable, easy-to-use deduplication solution. Overall, the DXi Series can be counted on to save the enterprise both time and money.

The DXi platforms are reliable, available, and scalable. This is especially true for the larger DXi6500 platforms that include a tape interface for long-term data retention. No matter how often some vendors tell us that tape is dead, it remains an important component of any backup and recovery architecture. Data deduplication appliances provide the ideal environment for short-term recovery; however, while many data centers talk about going tapeless, tape remains their best alternative for long-term storage.

All of these factors play an important role in positioning Quantum and their DXi Series as a major player in the backup and recovery appliance marketplace. If you are in the market to improve the price/-performance of your backup environment, the Quantum DXi Series may be the solution that your enterprise seeks.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

➢ *The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.*

About the Author

David Reine is a Senior Contributing Analyst for The Clipper Group. Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. In 2002, he joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

➢ *Reach David Reine via e-mail at dave.reine@clipper.com or at 781-235-0085 Ext. 123. (Please dial "123" when you hear the automated attendant.)*

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, **The Clipper Group Voyager**, Clipper Notes, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosures

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.