

Reversing the Requirement for Storage Growth — IBM Consolidates and Simplifies Tier-2 Storage

Analyst: David Reine

Management Summary

We live in an era where we have been told repeatedly that bigger is better. A *big* car is a symbol of success – *a limo is even better!* Our home entertainment systems have to have the biggest possible screen – with high-definition and surround sound with the largest speakers that can blast out music with 100 watts per channel. Even our fast food has to fit this image. It is not enough to have a *Big Mac*; we have to have the burger with the king-size fries and largest soft drink. Fortunately, sanity usually does prevail in restoring some semblance of order, and size, to our lives. With the cost of gas rising again, fuel economy and hybrid engines have conquered our macho ego and replaced the gas-guzzlers with a more environmentally sound means of transportation. With our waistlines expanding in response to too many requests to “supersize that”, our vanity, and doctor’s orders, lead us to the salad bar and a more sensible diet. Technology has even enabled all of the audiophiles amongst us to reduce the size of the speakers in our sound systems with something a little smaller and less damaging to our ears. In fact, technology has enabled us to replace that *portable* Boom Box with an *iPod* which can fit in a pocket.

Technology and innovation have also come to the rescue of the enterprise data center where virtually uncontrolled growth in storage requirements have forced the IT staff to supersize every storage array throughout the enterprise, centralized or distributed, in order to be able to handle the needs of the information generation. And make no mistake about it: **the data center does need to be rescued!** With increased costs for energy, and a limited amount available, the IT budget is being stretched to the limit in order to power and cool a data center containing storage arrays in every available nook and cranny. Floor space is another factor, as the IT staff tries to meet capacity requirements within the confines of an already bulging data center, staving off the impending doom, and millions of dollars in capital investment, for a new data center. The IT staff is always looking for ways to minimize space and power, while at the same time, striving to satisfy performance and capacity requirements.

IBM is one of the companies that have recognized the need to not only consolidate data center storage, but to reduce the amount of array capacity required to store multiple copies of business-critical information, both unstructured and transactional. One way to accomplish this is through the innovative technology of *data deduplication*, a subject of increasing popularity throughout the high-tech industry. Rather than attempt to reinvent the wheel and develop another set of data deduping algorithms, IBM acquired one of the leading lights in data deduplication artistry, Diligent Technologies Corporation. IBM has integrated their architecture into a framework of IBM infrastructure components to enable the deployment of the *TS7650 ProtecTIER Deduplication Appliance* as part of a dynamic infrastructure to simplify storage management and reduce the total cost of ownership (TCO) of data center storage. To learn more about the TS7650, please read on.

IN THIS ISSUE

- **Coping with Data Center Storage Pain.... 2**
- **The Advantages of Data Deduplication... 3**
- **The IBM TS7650 ProtecTIER Deduplication Appliance 3**
- **Conclusion..... 4**

Coping with Data Center Storage Pain

Server sprawl and rampant storage growth are putting a serious strain on the IT budget of every enterprise. Data centers around the globe are addressing server sprawl with a concerted effort to consolidate their underutilized application platforms on new, multi-socket, multi-core open systems servers. This enables the IT staff to reduce the TCO of the server infrastructure by reducing the number of systems in service, freeing up floor space, reducing the administrative staff, and reducing the amount of energy required to run and cool the data center. Unfortunately, server consolidation does not solve the problem of rampant growth in storage, especially in the area of the backup of unstructured data consisting of email, video, and images, such as x-rays and MRIs. In the midst of a downward economic spiral, the IT staff has to get control of storage requirements and limit the impact on the data center's budget.

Storage requirements are growing at a rate of anywhere from 50% to 100% annually, depending upon such factors as enterprise policies for backup and recovery and the requirements of your enterprise to adhere to standards and comply with government regulations regarding retention. In fact, IBM has estimated that the world is now generating **15 new petabytes** of information *daily*. Traditional data centers have been using tape as the media to accomplish backups and recoveries, as well as archive data to preserve history and protect the enterprise in the event of litigation. In recent years, however, many enterprises have moved to a disk-to-disk (D2D) architecture for their backup and recovery requirements in order to meet enterprise SLAs for performance. Unfortunately, the unprecedented growth in storage requirements has put the IT budget in jeopardy from both acquisition and operational standpoints.

What are the significant storage pains points that are threatening the data center and giving storage administrators *Excedrin* headache #1? First, and foremost is the fact that enterprise backups take too long – not only the full weekly backups that can consume the better part of a weekend, but also the incremental backups that can cause significant delays in the nightly mission-critical activity. Not far behind the backup dilemma is the unmistakable grief caused by slow recoveries. If your servers are

down because of lost data and you can not process customer transactions, you may lose those customers to competitors who are operational. In addition to these stresses, the data center staff also faces the constant battle with the human element, i.e., the additional staffing required to manage the process and media, as well as the issues surrounding the measurement of backup and recovery success.

One way to control the amount of data being backed up is to move older files and data to an archive and out of the mix of active data. **The IT staff can and should take advantage of the long-term retention characteristics of tape to take advantage of reduced acquisition and operational costs¹ for archiving purposes. Tape provides the enterprise with a cost-effective solution to deal with the capacity required for archiving and is removable and transportable to support disaster recovery requirements.** With the innovative features of *LTO-4*, such as *WORM*² and data encryption, incorporated in products such as IBM's *TS2340 LTO Ultrium 4 drive*³, the data center can achieve *long-term* compliance with financial and regulatory requirements and reduce the size of the active data. The requirements for *short-term* retention, on the other hand, are completely different. **The data center needs to commit the right balance of high-density tape for archive and high-performance disk for the backup/recovery process.**

The most typical environment for short-term data retention is enterprise backup. In this environment, performance, as opposed to capacity and cost-efficiency, is the gating factor. **Backups must be fast; recoveries even faster.** In order to avoid interference with mission-critical applications, backups must be initiated *and* completed within a specified window of opportunity. If the incremental or full backup is not completed when the mission-critical applications need to come back on line, the data center runs the risk of violating their own policies and losing a critical checkpoint.

¹ See the issue of *Clipper Notes* updated October 21, 2008, entitled *Disk and Tape Square Off Again – Tape Remains King of the Hill with LTO-4*, and available at <http://www.clipper.com/research/TCG2008057.pdf>.

² Write Once, Read Many.

³ See **The Clipper Group Navigator** dated August 2, 2007, entitled *A Tape Solution for Every Size Enterprise – IBM Adds LTO-4 to Tape Family*, available at <http://www.clipper.com/research/TCG2007077.pdf>.

One major factor that is having the biggest impact on the growth of enterprise storage is the backup of identical data segments on a nightly and weekly basis. **Redundant data is rapidly becoming one of IT's biggest headaches, and biggest opportunities.** It increases acquisition costs, wastes energy, and clogs the storage I/O. One way the IT staff can reduce the amount of data that needs to be saved, and therefore the length of time required for the backup, and the recovery, is by eliminating the backup of identical data through data deduplication. **You only back up the same data once!**

The Advantages of Data Deduplication

In order to reduce data proliferation, many enterprises are now deploying data deduplication software to eliminate redundant copies of the same data, reducing it to a single instance. In the deduplication process, duplicate copies of the same data are deleted, leaving only one instance to be stored. However, indexing of all data is still retained in the event that data needs to be recovered. Deduplication can significantly reduce the required storage capacity since only unique data is stored. For instance, if an email system containing 100 instances of the same 1MB attachment is being backed up, all 100 instances would be saved, requiring 100 MB of disk capacity. With data deduplication installed, only one instance of the attachment is actually stored. Each additional occurrence just references the one saved copy. In this case, a 100MB storage requirement would be reduced to just 1MB. Now imagine the amount of storage required if the data center is doing a full backup weekly. **The 100MB storage requirement would quickly expand to 5,200MBs!** With data deduplication deployed, only 1MB of capacity would be consumed. This savings cannot be assumed for all storage environments. The IT staff needs to examine their data structure closely to determine if data deduplication can provide them the same type of savings.

If your backups do contain duplicate data, then it is easy to see the benefits that data deduplication can provide to the enterprise in terms of reduced storage requirements, improved SLAs, and reduced TCO. With reduced acquisition costs - and reduced power, space, and cooling requirements - a D2D environment becomes more suitable for primary

backup and restore and for retention that can easily extend to months. With backup data on disk, the recovery SLA is higher, tape media handling errors are eliminated, and multiple recovery points become more attractive. Data deduplication also reduces the amount of data that must be transmitted across the network for remote backups, replication, and disaster recovery.

Data deduplication is an innovative technology that can deliver the storage consolidation and, therefore, the economic advantages of reducing the TCO of the storage infrastructure. For Enterprises backing up large amounts (more than 20TB nightly), the IT staff can choose to acquire data deduplication software from an ISV and integrate it, along with an open systems server and storage management software, themselves. However, for mid-range customers, the acquisition of a pre-integrated data deduplication appliance, such as the *System Storage TS7650 ProtecTIER Appliance* from IBM, makes more sense to enable the enterprise to hit the ground running.

The IBM TS7650 ProtecTIER Deduplication Appliance

Pre-configured for rapid deployment into an existing backup environment, the TS7650 ProtecTIER Deduplication Appliance integrates a server, storage, data deduplication software, and all of the cables and switches needed to improve backup and recovery operations for turnkey operation. Taking advantage of best-of-class technology from IBM, the TS7650 was designed to dramatically improve backup and recovery operations, with up to 500 MB/s or more of inline⁴ data deduplication performance. It consists of the following IBM infrastructure components:

- IBM's ProtecTIER software with patented *HyperFactor* deduplication technology;
- A System x server configured with (18) 2.66 GHz Xeon cores, 24GB of RAM memory and 4GB Fibre Channel (FC) HBAs;
- An IBM Storage Controller with 450GB FC drives running at 15K RPM; and
- An IBM TS3000 System Console for call home and remote support.

⁴ Inline processing is more efficient than the post-process model used by many data deduplication ISVs.

The TS7650 ProtecTIER Deduplication Appliance is a scalable solution, with up to 1PB of physical storage⁵, and is available in four configurations (see Exhibit 1, to the right) to meet a variety of needs from a mid-sized IT environment to an enterprise data center. The TS7650 can emulate up to 12 virtual tape libraries (VTLs), 256 virtual drives, and 125,000 virtual cartridges. It can scale from 7TB to 36TB, without changing the frame. Two ProtecTIER nodes can be clustered together to increase availability and to meet the most demanding data center performance requirements, scaling up to 1000MB/s.

ProtecTIER is based on IBM's *HyperFactor* deduplication technology, which is a patented pattern recognition algorithm designed to reduce storage needs by up to a factor of 25, significantly reducing TCO for the data center. This unique technology provides enterprise-class data integrity, while not relying on risky hashing algorithms. ProtecTIER requires no client or host agent, enabling more processor time for critical applications.

The various configurations can be sized and priced to serve a host of enterprise requirements. An SME can install a low-cost entry solution with 32 Fibre Channel spindles and up to 7B of capacity to deliver the precise level of performance and capacity required. That configuration is available for approximately \$100k. A larger enterprise can scale the performance and capacity to meet their specific needs.

As an alternative, IBM can also deliver data deduplication in a pure software form, integrated within the Tivoli storage product family. They also have a network of strategic partners who can be engaged to assist in the deployment of any complex IT integration.

Conclusion

With the acquisition of Diligent Corporation, IBM has put itself in the rather unique position as the only vendor that can offer to the enterprise a complete backup/recovery solution with servers, disk, tape, software, and data deduplication technology. They can support any deduplication environment, with a previously announced *TS7650G ProtecTIER Deduplication Gateway*, the new TS7650 appliance or with deduplication software from the

⁵ 1PB of physical capacity can support up to 25PB of user data.

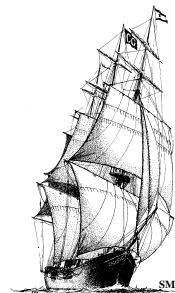
Exhibit 1 – TS7650 ProtecTIER

- **Entry-Level** – An appliance with a single System x server and up to 7TB of capacity and up to 100MB/s or more inline data deduplication;
- **Mid-Level** – An appliance with a single System x server and 18TB of capacity, scalable to 36TB, and up to 250MB/s or more inline data deduplication;
- **High-End** – An appliance with a single System x server and up to 36TB of capacity and up to 500MB/s or more inline data deduplication;
- **Cluster Configuration** – This is a high-availability model with two System x servers in an active-active configuration with up to 36TB of capacity and up to 500MB/s or more inline data deduplication

Source: IBM

Tivoli Storage Manager portfolio. IBM does not have to squeeze your data center pain into any single solution, as some of their competitors have to do. They have multiple solutions to solve the specific problems of your data center, and with a record of innovation in both server and storage technology, they can keep your enterprise current with the latest advancements in IT.

ProtecTIER is a true enterprise-class data deduplication solution designed to meet the data protection requirements of any dynamic infrastructure. It provides industry-leading deduplication performance, capacity, data integrity, reliability, and scalability with the use of FC drives as opposed to the SATA devices found in many competitive solutions. It can lower the TCO of any enterprise infrastructure significantly, compared to the costs of a standard VTL solution without data deduplication. Acquiring a preintegrated appliance from IBM enables the data center to control costs and take advantage of IBM's experience in solutions and services. IBM's TS7650 ProtecTIER Appliance might be the answer to your data center's pain as well.



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 Director, Enterprise Systems for The Clipper Group. Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. 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.

Disclosure

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.

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.