Quantum Tames Big Data
for the Enterprise Data Center

Analyst: David Reine

Management Summary

When it comes to money, many if not most of us have a hard time saving enough. Whether it is saving for a house, college, retirement, or even something more modest, putting that money away is a challenge. When it comes to preserving old memories, however, we do not have that same problem. Whether it is old ticket stubs from that concert in 1985 or photos from college (or high school, or middle school, or elementary school), we save them. We put them in a box and shove them onto a shelf in the bedroom closet, or in the basement, or in the attic. Sometimes, we even put a label on that box so that we can identify the contents (or not). When you have cause to look for something that you have stored in this fashion, perhaps years ago, you open the box, *if you can identify and find it*. Now, let's multiply that box by 10, or by 100, or even by 1000. How can you find the picture from your prom? The idea of opening all of those boxes is daunting. *However, to some degree, we all are hoarders.* It can become addictive, and in some cases, expensive, as additional storage space needs to be acquired. We save memories in order to ensure that we don't lose them. Unfortunately, in many cases where disorder and large volume are the rule, they may be as good as lost.

Now, let's transition this scenario to the enterprise data center where hoarding can be an even more serious issue. [1] Saving *everything* is nothing new here. With the collection and preservation of data out of control, the volume of data being stored somewhere in the enterprise is growing, and growing rapidly, sometimes on the wrong type of media. In fact, it is doubling in size every 18 months, or so, with no end in sight. Traditional approaches cannot keep up with this growth; they do not maximize the business value of this rich content. Policies established by governmental agencies, C-level executives, or by the legal staff, require the IT staff to deploy, manage, protect, and archive a never-ending volume of data, some of it active, most of it not so active. Some of these policies are very reasonable; others not so. Regardless, the data is being stored and, hopefully, catalogued for future use or to recover from file corruption or disastrous event, which have become all too familiar. In addition, we now find ourselves in the age of Big Data, where "big" can have many connotations, which will be discussed. It is most important that all of this data be stored on the "right" type of media at the proper tier (i.e., at the lowest cost where the qualities of service are acceptable minimally).

One company that has been innovating in data storage solutions for the past three decades is Quantum Corporation, a company dedicated to preserving the world's most important data, yours. With solutions encompassing high-performance and high-capacity disks, and high-capacity tape (yes, even tape), Quantum is dedicated to helping the data center preserve and protect enterprise data in the most cost-conscious manner, based upon the "value" of the data. To learn more about Quantum's archiving solutions, please read on.

---

[1] See the issue of *Clipper Notes* dated February 29, 2012, entitled *Regaining Control Over Hoards of Enterprise Data – Why You Need an Archive,* available at http://www.clipper.com/research/TCG2012005.pdf.

Exploring the Phenomena of Big Data

The enterprise data center has been experiencing a prolonged period of accelerated data growth unlike any period in the history of digital data. In fact, the vast majority of the world's data world has been created within the past two years. A wide variety of enterprises and government agencies have been doubling their data storage requirements every 12 to 18 months, mostly for use in everyday application requirements, from the creation of animated movies in the entertainment industry to the video surveillance applications being deployed by the TSA and a variety of military agencies. In addition to online use within the day-to-day operational workflow, vast petabytes of digital information are being saved to fast disk media, then archived within nearline libraries, and then sent offline to secure vaults in remote locations, both for retrieval for normal operational uses and for disaster recovery. The use of the phrase "Big Data", however, is not only defined by a large quantity of data that is stored. Here are some additional characteristics. More than one may be applicable to a specific use or application.

- **First, it might be large by the *size* of the data objects.** For example, each digital movie tends to be large (when compared to office files, like text documents, spreadsheets, and PDFs).
- **Second, it might be large by the *number* of objects** (like millions of tax returns), even if the size of the record is pretty small.
- **Third, it might be large by the *number and/or complexity of the relationships* between objects** (e.g., many things connected or "mashed" together, say to make recommendations for an online customer's additional purchase).
- Nonetheless, while each of these contributes to "bigness" and the problems that come with storing and retrieving vast quantities, often **it is the use of Big Data that defines the bigness.** Some business processes need to analyze a lot of data in a relatively short period of time. This often is called "analytics" or "business intelligence". The name of this analytics game is deriving insight from all of the Big Data.

What is Big Data and where does it come from? Big Data is, in fact, the lifeblood of many businesses. It is all of the data that we create, input, validate, manage, preserve, and analyze to understand better the needs of the enterprise and of the enterprise's customers. It is not limited to any one category, however. Basically, it can consist of both structured and unstructured digital assets, usually with a penchant for tremendous growth.

Big Data comes from a variety of origins, such as updates to social media websites, digital pictures and videos posted online (some for 15 minutes of fame), sensors from a wide range of sources (such as security cameras, transit transactions, and streamed research input), and some for enterprise functions such as the transaction records of online purchases and the recording of oil and gas exploration. Even if the high value of this data is fleeting, in many cases it must be preserved and protected for a very long time. This is especially true with the results of oil and gas exploration, where the value of the data may be even greater in the future, or in astrophysics, where stellar recordings are unique to the moment and can't be reproduced.

Big Data is more than merely a physical storage problem for the data center, although that can present serious technical and TCO challenges. Just like finding the right box of photos from a collection of hundreds or thousands of boxes can be an overwhelming chore, finding what you need for operations or analysis from all of that Big Data can be a recurring nightmare. That is why you need an *archiving solution* to manage your Big Data,[2]

Quantum's Archiving Solutions

Quantum has been innovating with storage infrastructure in the data center since 1980. Now, it has developed a set of products that have been purpose-built for Big Data. Quantum's Big Data solutions have been crafted around four key requirements:

- Low cost,
- High performance,
- Ease of use, and
- Flexibility.

As a result, its systems have been designed to provide the highest performance possible, along with optimized file sharing, and have been designed for high scalability to meet existing and future Big Data growth and to protect the investment that the enterprises makes in Quantum

---

[2] See, again, the paper referenced in Footnote #1, on the first page.

solutions. Quantum has done this while still focusing on the low-cost requirements that enterprises need to fit Big Data into their always-constrained IT budgets.

Quantum storage platforms are easy to configure, provide the flexibility required to meet changing business requirements, and are able to scale capacity and performance independently. They are easy to deploy, and easy to manage, enabling the enterprise to limit administrative costs. They enable the IT staff to concentrate on maximizing data collection and use, increasing workflow productivity, and quickening time-to-market, instead of toiling inside the dark recesses of storage systems.

Quantum has developed an automated tiered storage management system to enable the transparent migration of data between three separate and distinct tiers of storage, including:

- **Standard SAS drives** for high-performance, tier-1 mission-critical requirements,
- **Nearline SAS disk drives** for tier-2, high capacity storage for business-critical applications, and
- **Highly-expandable tape libraries, with LTO-5 tape drives**, as the third tier, in support of long-term archiving. These libraries come with built-in data protection and self-healing capabilities to ensure that the data center staff can protect their valuable data and ensure accessibility over time while maintaining the lowest cost per terabyte, and preserving natural resources by lowering power consumption.

Quantum's high-performance shared file system enables sharing across server platforms, using a single namespace for multi-PB systems. Its solutions provide data access across both SANs and LANs and also feature replication and data deduplication for improved utilization of all storage and data protection.

These solutions are available in the *StorNext* family and include the following.

- *StorNext M330 Metadata Appliances* **and** *StorNext G300 Gateways* that are purpose-built for Big Data applications with high-performance file sharing requirements and easy access to Big Data;

- *StorNext Q-Series* **disk arrays** that enable predictable and scalable performance and capacity for high-speed access to archived data;
- *StorNext AEL Archives* for cost-effective, automated near-line secondary storage and off-line data retention; and
- *StorNext Storage Manager* enabling cost savings and management simplicity with policy-based tiered storage and archiving.

Each of these will be discussed.

Quantum's StorNext Platforms

Quantum's StorNext Family has been put together and updated specifically to address the massive growth in data center storage requirements in a wide variety of industries with Big Data technology. This technology has been designed to protect the enterprise investment in existing server and storage products and to ingest data as quickly as possible in order to distribute that data across the storage network, via SAN or LAN, via Fibre Channel (FC) or NAS, to the most "value-appropriate" tier of storage, using the best infrastructure available.

Since 2001, Quantum has had over 6,000 customer installations of StorNext, with over 60,000 file system clients deployed. StorNext has a shared file system that is operating system independent, supporting a wide variety of clients: *Windows*, *Linux*, *UNIX*, and *Mac*. StorNext is compatible with a number of third-party disk vendors for maximum flexibility and investment protection. In addition to Quantum's own hardware, StorNext also supports a wide assortment of disk arrays and tape libraries from Dell, HP, IBM, and Oracle (Sun/STK). StorNext is both server- and disk-vendor agnostic. The StorNext Family consists of an infrastructure made up of appliances, arrays, libraries, and software, as described below.

StorNext M330 Metadata Appliance

The StorNext M330 Metadata Appliance provides any size enterprise or department with concurrent, high-speed, sustained, reliable access to a shared storage pool, which is especially important for rich media environments. Unfortunately, the typical smaller workgroup may not have the IT budget or the expertise to meet the operational challenge of Big Data. This is where the M330 shines. In support of all

workgroups, the M330 Metadata Appliance has been pre-integrated, with software installed at the factory for simplified deployments and immediate reliability. It provides advanced graphical reporting, displaying the health of the system, which enables optimal performance with minimal support.

The M330 comes in a 6U, rack-mountable chassis to provide concurrent, shared high-speed access to a virtualized, consolidated storage pool. It is configured with two metadata controllers, duplicated for failover, with a dedicated metadata array with 300GB, 15K SAS drives. The base M330 includes a ten-client file system SAN license (for ten servers) and a high-availability license option. Designed for reliability, in addition to the metadata controllers, the M330 has redundant fans and power supplies, automated SAS I/O failover, a hot spare in the in the metadata array, and 4GB of mirrored RAID battery-backed cache.

StorNext G300 Gateways

The StorNext Gateway Appliances, the G301 and G302, are purpose-built for StorNext environments[3]. They are both compatible with the M330 Metadata Appliance. They provide flexible, affordable high-speed access to Big Data by leveraging high-speed networking technologies to access data stored in StorNext for rapid content delivery and on-demand access. The G301 can provide 1GbE Ethernet access, while the G302 provides 10GbE access; both have 8GB dual FC SAN connectivity. Both are pre-integrated with StorNext software at the factory, with the integrated monitoring tool providing graphical views of performance metrics for all network activity across the gateway(s).

Either gateway is ideal for data centers with server farms requiring high-speed, on-demand access to shared data, with Ethernet access up to 1.6GB per second for a single G302.[4] Both also enable the IT staff to add new clients dynamically, with one StorNext SAN client license included, to accommodate changes in the workflow or the overall configured StorNext environment.

The StorNext G301 and G302 are configured with quad-core Intel E5620 processors running at 2.4GHz, and 24GB of memory.

---

[3] A conversion from the G301 to the G302 is available.

[4] Throughput scales when multiple G300s are combined in a single cluster.

They both have a SAS RAID adapter for internal disks, with two 500GB SATA drives, configured in a 1+1 RAID-1 set, and redundant power supplies and fans. In addition, each has two built-in 1GbE ports[5] and two 8Gb FC ports. The G302 also comes with two built-in 10GbE ports.

StorNext Q-Series Storage

The StorNext tiered storage architecture also includes the StorNext Q-Series, a family of primary disk systems that are optimized to deliver high-performance in a Big Data environment, enabling data centers to store and manipulate data at very high rates. The Q-Series family is composed of four models: *QM1200*, *QS1200*, *QS2400*, and *QD6000*. All have been designed to deliver a balanced blend of performance, scalability, and density, while significantly lowering the total cost of ownership (TCO) when compared with traditional storage solutions.

The Q-Series supports up to 3.6GB/second for sequential writes and 3.2GB/second for sequential reads. All models have dual-active controllers with automated I/O path failover for high availability, with redundant, hot-swappable components for reliability, and a mirrored data cache with battery backup and de-stage to flash SSD. Used in a StorNext policy-driven tiered environment, these arrays enable data centers to have virtually unlimited space for Big Data.

StorNext QM1200 Storage

The QM1200 is a 2U, modestly-sized high-performance rack-mounted storage array, optimized for StorNext Metadata storage, with from 12 to 48 15K 600GB SAS drives and up to 28.8TB of storage, with RAID 1 and RAID 10 capability. It supports an 8Gbps FC interface with eight ports. It can support up to three 2U expansion drawers for each base system, each with 12 more drives.

StorNext QS1200 Storage

The QS1200 is a 2U, high-performance, larger capacity rack-mounted storage array with from 12 to 96 15K 600GB SAS drives and up to 57.6TB of storage, with RAID 6 capability. It supports an 8Gbps FC interface with 8 or 16 ports. It can support up to seven 2U expansion drawers for each base system, each holding 12 more drives.

---

[5] The G301 also supports eight 1GbE NIC ports.

StorNext QS2400 Storage

The QS2400 is a 2U, mid-performance, very large capacity rack-mounted storage array with from 24 to 192 10K 900GB SAS drives and up to 172.8TB of storage, with RAID 6 capability, 24 drives in the base system and 24 in each expansion drawer. It supports an 8Gbps F.C. interface with eight ports. It can support up to seven 2U expansion drawers for each base system.

StorNext QD6000 Storage

The QD6000 is a slower, massive, 4U, rack-mounted storage array with from 20 to 360 900GB 10K SAS drives[6], or 3TB 7.2K NLSAS drives and up to 1.08PB of storage, with RAID 6 capability, a minimum of 20 drives in the base and up to 60 drives in the base and each expansion drawer. It supports an 8Gbps FC interface with eight ports. It can support up to five 4U expansion drawers for each base unit.

StorNext AEL Archives

If a vendor is going to deliver a full, *long-term* storage solution for Big Data, it must include an economical long-term archiving capability. Unless the enterprise has an unlimited supply of power *and* an unlimited budget, that solution must include tape[7].

Quantum's StorNext AEL500 is as an archive solution for a StorNext file system that gives a customer 41 or 87 slots of LTO-5 storage, three drives for data movement, and an innovative approach to protecting data on tape. Data in a StorNext AEL Archive is automatically scanned to verify the tape – at three user-configurable levels of thoroughness – and if a tape is suspect or contains an error, data is automatically migrated off the suspect tape to a new tape. StorNext AEL500 Archives can grow to 409 slots. Quantum also offers the *AEL6000* for even greater scalability. These libraries represent a family of high-capacity, cost-effective archives with automated, policy-based storage tiering to lower the TCO of the infrastructure, and automatic media integrity validation to deliver a reliable, self-monitoring, self-healing, and high-capacity archive.

AEL Archives combine data management policies with cost-effective tape storage to de-

liver PBs of data accessible to applications through a simple file system interface, thereby enhancing the StorNext File System with an AEL archive, the data center can store frequently accessed data on high-performance disk, while rarely accessed data can migrate to inexpensive tape media. AEL Archives also includes the *StorNext Storage Manager* and Quantum's *Extended Data Life Management (EDLM)* features to test and verify the media in the library and provide health reports for all cartridges.

The StorNext AEL500 Archive includes a Quantum *Scalar i500* tape library[8] with over 400 slots for *LTO-5*[9] cartridges and from 3-to-18 LTO-5 tape drives, for a total capacity of over 600TB. The StorNext AEL6000 Archive includes a Quantum *Scalar i6000* tape library[10] and scales to over 5,000 slots for LTO-5 cartridges and from 4-to-96 LTO-5 tape drives, for a total capacity of over 8PBs. It should be noted that as each new generation of LTO tape arrives, the capacity of each tape cartridge increases. With the arrival of LTO-6 expected in Q412, the total capacity of the libraries will continue to grow without increasing the number of slots[11], protecting the investment in StorNext architecture, without an increase to license fees.

StorNext Storage Manager

The StorNext Storage Manager is an optional software component of StorNext that provides advanced data preservation services. Storage Manager is responsible for automated policy-based transparent data movement as well as capacity management of storage devices. This is done through an integrated policy engine, which the IT staff easily can customize to meet specific needs.

Policies are assigned at a directory level, enabling the IT staff to create different tiered storage strategies based on unique requirements

---

[6] Drives are added in 10-drive increments.

[7] 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.

[8] See **The Clipper Group Navigator** dated November 17, 2005, entitled *ADIC Continues Tape Leadership Role - Delivering Automation Solution to Mid-Range*, available at http://www.clipper.com/research/TCG2005072.pdf.

[9] See **The Clipper Group Navigator** dated January 31, 2010, entitled *LTO Program Announces Next Gen Tape – LTO-5 Raises the Bar for Tier-3 Storage*, available at http://www.clipper.com/research/TCG2010002.pdf.

[10] See **The Clipper Group Navigator** dated April 24, 2010, entitled *Need a Long-Term Archive Solution? Quantum Adds LTO-5 to Scalar Tape Libraries*, available at http://www.clipper.com/research/TCG2010021.pdf.

[11] Assuming that data on older cartridges is consolidated onto LTO-6 cartridges.

within a single namespace. For instance, certain files could be automatically copied to tape for data protection but never removed from primary disk because of their need for continuous use. Other files might be automatically copied to SATA or SAS disk and then removed from primary disk, so only the copy remains after 30 days of inactivity, thus creating a tiered storage environment. Other files never may be copied and would be cleared out by staff, as needed. The Storage Manager's flexible features can be used to move files to tiered storage for cost reduction, data protection, and long-term archiving.

Conclusion

The continued growth of data – Big Data – in the enterprise data center is inevitable. It is being created and ingested via policy-driven automation for a myriad of reasons and it needs to be managed, preserved, and archived in the most appropriate, most performant, yet economical, fashion possible. Whether the data center is recovering a single file or a single block, whether it is doing a disaster recovery or an operational restore, it must be efficient in delivering whatever data is being retrieved for analysis.

In order to keep the TCO of the IT infrastructure as low as possible, the data center must deploy a multi-tiered architecture. The IT staff of most major enterprises is currently fully capable of acquiring and assembling the necessary infrastructure components from any number of storage vendors on their own – but is this what they want to do? Should they?

Quantum has assembled a complete archiving solution to fit the needs of the SMB or larger enterprise. Their solution includes metadata servers, gateways, disk arrays, and tape libraries – all of the components required to deploy a multi-tiered architecture. If infrastructure matters to you, StorNext may well be the Big Data archiving solution that you seek for your enterprise.

About The Clipper Group, Inc.

The Clipper Group, Inc., now in its twentieth year, is an independent publishing and 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

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.