



Quantum®



WHITE PAPER

BIG DATA: Managing Explosive Growth

The Importance of Tiered Storage



OIL & GAS

CONTENTS

- Introduction 3
- An Eye Toward Archiving 4
- Archiving Advantages 5
- The Right Choice 6
- Big Data Defined* 7
- Data Archiving or Data Backup? 7
- Accuracy Spells Success for Oil and Gas Companies 8

NOTICE

This White Paper may contain proprietary information protected by copyright. Information in this White Paper is subject to change without notice and does not represent a commitment on the part of Quantum. Although using sources deemed to be reliable, Quantum assumes no liability for any inaccuracies that may be contained in this White Paper. Quantum makes no commitment to update or keep current the information in this White Paper, and reserves the right to make changes to or discontinue this White Paper and/or products without notice. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or information storage and retrieval systems, for any person other than the purchaser’s personal use, without the express written permission of Quantum.

INTRODUCTION

It's no exaggeration. Organizations today are experiencing a data explosion. According to recent research from Enterprise Strategy Group¹, respondents to the firm's survey say they are experiencing 11 percent to 30 percent annual growth rates of data, with another 28 percent dealing with annual growth rates of 30 percent or more. With the amount of data created and replicated increasing exponentially, enterprise IT departments are grappling with how to manage, store and protect vast quantities of data.

The reasons organizations are collecting and storing more data than ever before is because their businesses depend on it. The trend toward leveraging Big Data (see "Big Data Defined", page 7) for competitive advantage and to help organizations achieve their goals means new and different types of information—website comments, pharmaceutical trial data, seismic exploration results, to name just a few—is now being collected and sifted through for insight and answers. For example, as part of its genome-sequencing process, the Swiss Institute of Bioinformatics creates huge amounts of data—a single experiment produces up to 743,000 files per run, with each run sized at an average of 2TB and performed every 3.5 days.

At the same time, organizations are capturing and converting more and more content into digital data. Data types such as video, which used to be considered nice to have on a company's website, have become essential marketing, training and communications tools. Advances such as special effects and high definition significantly increase the amount of data generated; it takes twice as much space to store 3-D video as 2-D video because the technique requires two cameras to shoot the same footage, putting a significant strain on storage resources.

Attempting to support such advanced technology can reveal significant gaps in IT architectures. When Brigham Young University-Hawaii (BYU-H) decided to transition its internally produced video to high definition, the insufficient storage capacity and transfer rates it experienced caused the institution to run out of storage space midproject, forcing staff to make on-the-spot decisions about what stored material could be erased to make room for new content. Consulting firm Coughlin Associates estimates that by 2015, close to 4 exabytes of storage capacity will be required for the creation of professional digital media content alone.

Also driving up the amount of data that companies must store today are the growing obligations that regulations and laws are placing on companies gathering and storing data about customers, partners and even employees.

As businesses, nonprofits and governments realize the importance of data—not only for day-to-day functions, but to their strategies and success going forward—IT departments face a rising challenge. According to a Gartner study² in November 2010, data growth is the biggest data center hardware infrastructure challenge for large enterprises, with 47 percent of survey respondents ranking it as one of their top three challenges. And 62 percent of those respondents say they plan to invest in data archiving or retirement by the end of 2011, to help them deal with the challenges created by data growth.

“While all the top data center hardware infrastructure challenges impact cost to some degree, data growth is particularly associated with increased costs relative to hardware, software, associated maintenance, administration and services,” said April Adams, research director at Gartner, in a prepared statement. “Given that cost containment remains a key focus for most organizations, positioning technologies to show that they are tightly linked to cost containment, in addition to their other benefits, is a promising approach.”

The challenges in developing strategies to support organizational dependence on data include:

- **Budget constraints.** In this uncertain economy, organizations want to make the most of what they already have, not to commit significant up-front investments to new technology.
- **Complexity sprawl.** With technology talent also strained, IT organizations are loath to add more complexity to their technology infrastructures and are instead looking for more ways to empower end users with easy-to-use tools.
- **Availability requirements.** Companies need to ensure their employees can get at their data easily, regardless of where it may be stored or how old it is.
- **Integrity concerns.** Companies expect the data they trust to storage will remain intact and unaltered. IT departments need to be able to promise such capabilities, and keep those promises.

AN EYE TOWARD ARCHIVING

Setting strategies to archive data in an automated, reliable, cost-effective way is emerging as the right approach for organizations in a wide variety of industries. Archiving older data so that it’s automatically moved to less expensive storage media—such as cheaper disk or tape—frees up costly primary storage. That means as companies collect more data they can prioritize it based on how old it is and how often it is accessed. When done properly, archived data is inexpensively but reliably stored and easily accessed.

However, in today’s complex world of storage and data management, archiving is not always given adequate consideration—or worse yet, not done at all. Archiving data is often confused with backing up data (see “Data Archiving or Data Backup?” page 7), as companies assume they can do one or the other to cover their bases. However, companies that simply back up data are wasting expensive storage equipment and tying up precious IT resources by keeping all data on a primary storage tier, regardless of the age of the data or how frequently it is accessed. What’s more, these companies are missing the big picture by failing to think about their long-term needs, including which type of storage is best suited for their data over the years.

“Most companies perform backups, but not as many companies today are truly archiving their data,” says Eric Bassier, director of Quantum’s tape automation product line. “Industries like media and entertainment, which are dealing with such massive amounts of digital data that they just can’t keep it all on disk, are ahead of the broader market.”

ARCHIVING ADVANTAGES

Well-planned archiving strategies provide organizations with a number of benefits:

- **Cost savings.** Archiving systems enable the movement of data that hasn't been accessed in a certain amount of time—usually defined by the IT department—from expensive disk-based primary storage to less expensive second-tier or third-tier storage, either disk-based or tape. This frees up primary storage space for the ever-growing amount of data organizations collect on a daily basis.
- **Reduced management.** Unlike backup strategies that require users to request restores from the IT department, comprehensive archiving strategies allow end users to find needed files themselves. Archiving solutions that include file management software make this task simpler by giving end users a Windows-like file management system (i.e., a “file system”) that's easily navigated, so they can find and access their files in their original format. Solutions that offer automated data management take care of the task of moving a file to second- and third-tier storage once it has gone a certain number of days—typically 30, 60 or 90, depending on the settings chosen by the IT department—without being accessed. This automation saves IT from having to constantly get involved in the data management process by making sure data is tiered to the right level at the right time.
- **The right storage media for the right data.** Organizations need fast access to data that is called up often. To suit these performance needs, primary, disk-based storage is the right fit. However, as data gets older and is accessed less often, companies can reap significant cost advantages by moving older data to lower-end disk or tape. While these solutions don't provide the fast access times that high-speed disk does, such performance isn't required for older data that isn't accessed for long periods of time. By moving older data to these types of media, companies reduce the amount of expensive disk required.
- **Assurance of data availability.** Archived data is often kept for years, so an important element of a comprehensive archiving strategy is a solution that enables information stored for the long term to be available through the years. Solutions that regularly check the health of storage media provide assurance that data stored for the long term remains intact.

These features are what set comprehensive archiving solutions apart from less full-featured offerings that promise to safeguard information, but can't perform when that information needs to be accessed.

“It's not really a good archiving solution if, when you need to find a file, you first have to find a backup snapshot from years ago, rebuild the server, then search through and find that file,” Bassier says. “Archives should be searchable and easily accessible.”

Companies that implement comprehensive archiving solutions and also back up these systems just as they would their primary storage systems find they can more easily respond to business needs, compliance pressures and the changing nature of corporate data. Most important, they can do this without having to overhaul their IT architectures or purchase additional primary storage, while freeing up IT staff to perform more strategic tasks.

It's not really a good archiving solution if, when you need to find a file, you first have to find a backup snapshot from years ago, rebuild the server, then search through and find that file. ▶▶

—Eric Bassier, Quantum

THE RIGHT CHOICE

Because archiving has become a strategic initiative to support business goals, IT departments must be sure to consider a number of variables and choose carefully among the solutions on the market today. Quantum's family of archiving solutions, which includes its StorNext® data management software, the Scalar® i6000 tape library and the StorNext AEL Archive, give organizations the features they need to safely and easily implement a comprehensive archiving strategy:

- Quantum's StorNext end-to-end data management software enables organizations to build infrastructures that consolidate resources, speeding workflow and lowering operations costs. StorNext provides a high-performance file system that is great at dealing with large files, and also offers data sharing and retention in a single solution, so companies don't have to piece together multiple products that may not integrate well. StorNext works with any operating system or hardware platform to make all data easily accessible to all hosts, and automates the process of moving data as it ages. Enterprise data management features include online tiering and archiving, deduplication, distributed data movement and partial file retrieval.
- Quantum's intelligent Scalar i6000 tape library is designed to significantly improve the security and manageability of enterprise backup, disaster recovery and archiving processes. The library features Extended Data Life Management (EDLM) to provide policy-based data integrity checking for tapes stored in the library. EDLM scans every tape in the library periodically to make sure the data is still readable and to ensure there is no degradation in the tape. The feature can be set to automatically alert an administrator or the StorNext software when damaged files are found, so that they can be moved to new tape. The i6000 also includes an Active Vault feature that provides a cost-effective, safe and secure alternative to vaulting archived tapes.
- Quantum's StorNext AEL Archive provides StorNext software and the Scalar i6000 tape library in an integrated solution, comprising everything an organization needs to build out its archiving strategy. The preconfigured, high-performance, scalable appliance delivers lower operational costs than solutions sold individually, but is an open system that integrates with third-party applications. The StorNext AEL Archive provides the industry's most highly reliable, long-term archiving solution. This unique appliance tightly integrates Quantum hardware and software to deliver a reliable, self-monitoring and self-healing, high-capacity archive intended for the long-term storage of near-line content.

Organizations that have had firsthand experience dealing with the consequences of data explosion know that comprehensive storage solutions are the right answer. To solve the data storage and management challenges unearthed by its move to HD video, BYU-H deployed StorNext without disruption to the production process; users only noticed they were working from a different drive once StorNext was implemented. StorNext now acts as a metadata controller for BYU-H's storage-area network (SAN) and enables multiple professionals from different fields to work on a file simultaneously. And data-transfer rates have been boosted from 800 MB/sec to 8 GB/sec, leading to a significant efficiency and productivity hike.

"StorNext allows our graphics, editing and audio people to all access data at the same time. It is a wonderful solution since we can now ingest audio and video while others are editing the same file. Our efficiency in editing is up by at least twofold," says Russell T. Merrill, director of Instructional Media & Production with BYU-H.

Whether it's the onslaught of Big Data, the prevalence of new types of digital data in corporate environments or burdensome compliance requirements, organizations of all sizes are pressed to deal with data explosion. Strategic, comprehensive archiving solutions present affordable options that save IT resources, reduce strain on IT personnel and can scale with a company's needs.

BIG DATA DEFINED

Big Data has become the center of attention in the IT world, and the term has taken on a few meanings. Generally speaking, Big Data refers to relatively new data types—video, imaging, audio, etc.—that produce large files.

The term also refers to large collections of small data—social networking website comments, underwater photos of the ocean floor, feeds from traffic cameras—that when combined create meaning. Typically Big Data implies fast growth, so even modest data stores of certain types of information are expected to quickly grow and become Big Data. Organizations require storage capabilities that can handle big files and rapidly growing data stores in a single location.

DATA ARCHIVING OR DATA BACKUP?

The difference between backing data up and archiving data boils down to this: Backing up data makes a copy of the information and stores it somewhere else; archiving data moves the data to secondary storage. Other differences include:

BACKUP

- For disaster recovery
- Makes a secondary copy of data that may never be accessed
- Does not reduce disk capacity or the costs of storing data
- Protects data at a volume level

ARCHIVE

- For long-term data storage
- Is the primary copy of data that will be accessed
- Reduces primary-disk capacity and storage costs because it moves data to another tier
- Protects data at a file level, making it easier to find a file for restoration

It has enabled us to dramatically increase our seismic data processing efficiency.

—Tiengang Zhang, Chief Engineer,
EDRI of Daqing Oil Field

ACCURACY SPELLS SUCCESS FOR OIL AND GAS COMPANIES

In order to better understand the earth's natural resources that are the core of their business, oil and gas companies must constantly gather and piece together disparate data. Collecting this data helps these companies to discover natural resources, assess research output and characterize reservoirs for exploration. Before deciding whether or not to begin extracting resources, however, they must also determine if the right geological elements are in place to ensure that the oil and gas have accumulated in sufficient quantities in a given location.

Such determinations are based on geologists' assessment of the earth's surface via satellite images and geophysicists' interpretation of seismic data that helps them comprehend what's going on beneath the earth's surface. Pulling all of the information together to paint an accurate picture is essential to the business viability of these companies. Making the wrong decision has expensive consequences, potentially costing these companies millions of dollars and setting their projects back by months.

To help oil and gas companies achieve successful operations, high-performance, reliable and scalable storage solutions are needed to manage data and image acquisition from a variety of sources, ranging from tens of megabytes to multiple terabytes per day. Specialists then work together to combine this information from various sources into an accurate view of Earth's subsurface in the form of 3-D and 4-D models that help companies determine oil field production potential.

Petroleum research giant Daqing Oil Field Petroleum Exploration and Development Research Institute (EDRI) in China has more than 1,800 petroleum technical and engineering personnel working to increase the pace of oil and gas exploration. In order to support this goal, the organization's Geophysics Service Center (GSC) needed to boost its capacity for seismic data processing. Yet with this increase in computation and disk space came performance bottlenecks, as well as a single point of failure for its file system architecture.

GSC looked to a combination of Quantum StorNext File System, StorNext Distributed LAN Client (DLC), StorNext Storage Manager archival system and a Quantum Scalar i2000 tape library to provide better equipment optimization, improved resource utilization and greater overall efficiency. In particular, the deployment of shared file systems and rational usage of disk storage space helped the institute meet its needs for joint 3-D seismic data processing.

"Our implementation of StorNext has eliminated the performance bottleneck and removed the single point of failure of our previous file system architecture. It has enabled us to dramatically increase our seismic data processing efficiency," says Tiengang Zhang, chief engineer with EDRI of Daqing Oil Field. The StorNext File

System operates in conjunction with the shared storage pool available in GSC's SAN. This permits seismic operations to take place through many different I/O nodes to meet the massive demands of shared file system space for seismic prestack time/depth migrating processing. In addition, StorNext performs seismic data archival, retrieval, data protection and vaulting through the Scalar i2000 tape library.

With the right storage solutions at hand, oil and gas companies can manage the information required to make successful business decisions in a timely, cost-effective manner.

Our implementation of StorNext has eliminated the performance bottleneck and removed the single point of failure of our previous file system architecture. ▶▶

—Tiengang Zhang, Chief Engineer,
EDRI of Daqing Oil Field

1 Enterprise Strategy Group video commissioned by Quantum in 2010: Deduplication and Quantum's DXi portfolio

2 <http://www.gartner.com/it/page.jsp?id=1460213>

This page is intentionally left blank.

ABOUT QUANTUM

Quantum Corp. (NYSE:QTM) is the leading global specialist in backup, recovery, and archive. From small businesses to multinational enterprises, more than 50,000 customers trust Quantum to solve their data protection, retention and management challenges. Quantum's best-of-breed, open systems solutions provide significant storage efficiencies and cost savings while minimizing risk and protecting prior investments. They include three market-leading, highly scalable platforms: DXi®-Series disk-based deduplication and replication systems for fast backup and restore, Scalar® tape automation products for disaster recovery and long-term data retention, and StorNext® data management software for high-performance file sharing and archiving. Quantum Corp., 1650 Technology Drive, Suite 800, San Jose, CA 95110, (408) 944-4000, www.quantum.com.

www.quantum.com • 1.866.809.5230

Quantum®

Preserving the World's Most Important Data. Yours.™

©2012 Quantum Corporation. All rights reserved. Quantum, the Quantum logo, DXi and StorNext are registered trademarks of Quantum Corporation and its affiliates in the United States and/or other countries. All other trademarks are the property of their respective owners.

WP00167A-v01 Feb 2012