

WHITE PAPER

An Effective Strategy for Storing Unstructured Data

A MAJORITY OF IT PROFESSIONALS SURVEYED SAY IT'S CHALLENGING KEEPING UP WITH FULL BACKUPS DUE TO THE INCREASING VOLUME OF UNSTRUCTURED DATA THAT'S STORED. WHAT'S THE SOLUTION?

The rapid growth of unstructured data is putting greater pressure on organizations' storage resources, and on the IT executives responsible for managing the systems and information so critical to business operations.

Unstructured data now accounts for half of all the data stored by organizations today, according to a survey of IT professionals conducted by IDG Research Services in September 2013. The IT pros surveyed are involved in the purchase of long-term data storage and archive solutions at their organizations.

Unstructured data can include video content, images, email, Web pages, audio files, PowerPoint presentations, data generated on social-media sites, as well as file data that is difficult to reduce or compress.

In fact, 33 percent of the IDG Research survey respondents said they anticipate the proportion of unstructured data to increase over the next 12 months.

Business executives are increasing demands on IT leadership to get more value out of this data. According to the IDG Research survey, a majority of respondents agree that IT is under pressure to store more data and make it accessible around the clock.

» Holding Onto Data

In the past, organizations focused on protecting and storing data to manage risk, both in terms of security and regulatory compliance. Now, however, businesses are shifting their focus from risk management to saving unstructured data in a cost-effective manner so it can be analyzed and potentially repurposed later.

For example, a media and entertainment company that shoots footage for a movie or television show might decide to store all the video data in case it is needed later for a special program. This could also be the case with everything from employees' PowerPoint files to social-media marketing campaigns. If data has any potential use down the road, the general thinking is to store it.



The problem with this strategy is that most companies do not or cannot store this unstructured data in a cost-effective way. They are storing the data on primary storage resources such as disks and primary network-attached storage.

The sheer volume of unstructured data—and the fact that this data is growing at such a rapid rate—makes this approach a poor choice for organizations looking to control costs. Consider these findings from the IDG Research survey:

- » More than 60 percent of survey respondents said it is challenging to keep up with full backups due to the increasing volume of unstructured data that is stored.
- » Nearly 40 percent said the volume of unstructured data at their organizations is straining current storage resources and staff.

Clearly, organizations need new approaches to store and manage this growing volume of unstructured data more effectively and within their IT budgets.

What Is Unstructured Data?



UNSTRUCTURED DATA is commonly referred to as information that does not have a predefined data model or is not organized in a predefined way. It might contain elements such as dates and numbers, resulting in irregularities that make this type of data more difficult to understand compared with data stored in a fielded form in databases.

Industry research shows that unstructured data accounts for at least half of all data stored by organizations, and with the growth of social-media sites and the use of online video and images for business applications, the volume of unstructured data is likely to grow significantly.

Unstructured data has become so prominent that it has helped give rise to the term Big Data, which refers to extremely large volumes of data that are difficult to analyze with traditional data analytics tools.

For many organizations, unstructured data represents a potential treasure trove of information to help enhance business decisions, tap new sources of revenue and provide better customer service. The ability to analyze this data effectively to glean actionable knowledge from it is therefore a critical capability for enterprises.

» Tiered Storage Architecture

Many companies would consider new and less expensive ways to store unstructured data. More than 70 percent of respondents in the IDG Research survey said they are open to evaluating more cost-efficient methods for storing unstructured data.

One effective approach to consider is tiered storage, which enables organizations to essentially match storage costs to data value over time. With tiered storage, companies can assign various types of data to different types of storage media, thus reducing total storage cost.

The various data categories can be based on how often the data is needed and how much protection it requires; assigning data to specific storage media can be an ongoing activity. Choosing from different media means companies do not need to store unstructured data in costly, primary storage resources. The result is greatly reduced storage costs for unstructured data.

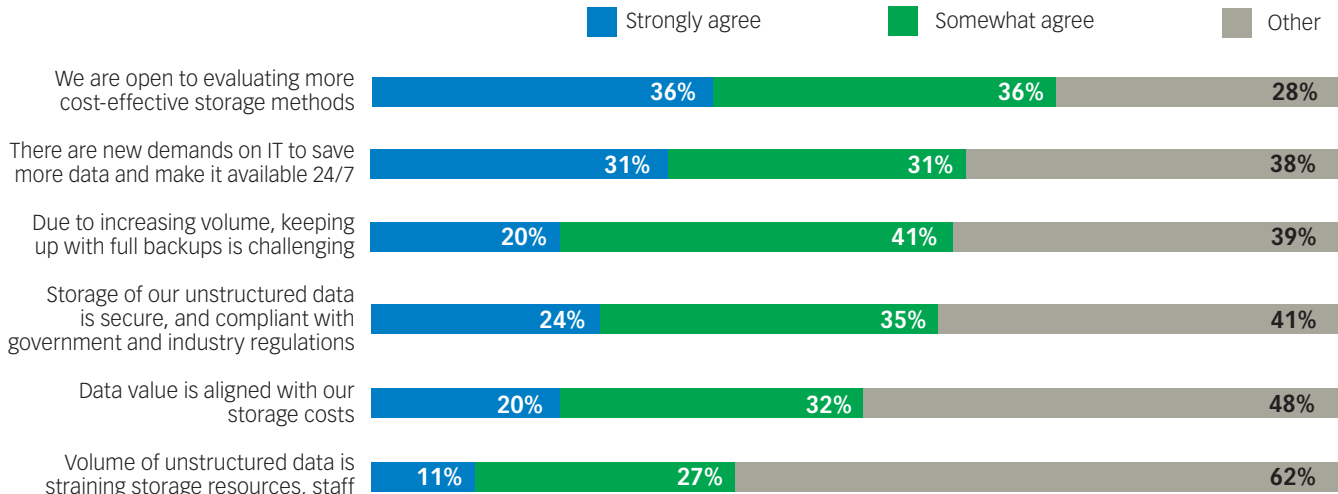
However, a tiered storage strategy can lead to a new set of challenges:

- » Nearly half of the survey respondents (48 percent) cited data protection as a concern, followed by concerns about integrating multiple solutions with existing infrastructure (44 percent) and uncertainties about the best way to classify and/or store different types of data (42 percent).
- » Traditional data reduction techniques such as deduplication and compression are not always effective for unstructured data. Multimedia data such as JPEG or video files cannot be compressed all that much using these techniques.



UNSTRUCTURED DATA

Rate your agreement with the following statements regarding the storage of unstructured data at your organization



SOURCE: IDG Research Services, 2013

An organization backing up 100 TB of mostly static unstructured data on a regular basis will put a huge strain on its storage infrastructure.

» Because companies are backing up the same static data day after day, they are putting more pressure on their servers, networks and storage systems. An organization backing up 100 terabytes of mostly static unstructured data on a regular basis will put a huge strain on its storage infrastructure.

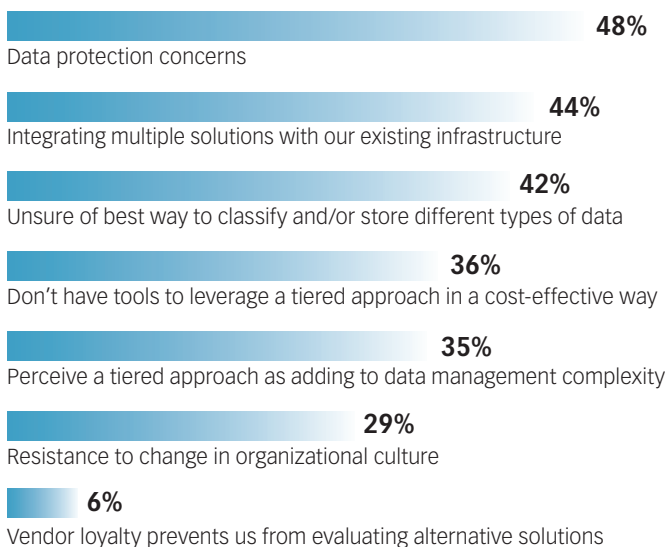
» **New Technologies Address Challenges**

The good news for organizations considering a tiered storage architecture for their unstructured data is that new storage technologies can help address these challenges.

To address the issue of data protection in a tiered storage environment, these new solutions offer self-protection capabilities that keep data safe. They feature advanced, variable-block deduplication to deliver high performance and efficiency and also enable optimized and affordable replication for disaster recovery.

1010 010 TIERED APPROACH

Data protection, solution integration and data classification top the list of concerns regarding a tiered approach to storing unstructured data



SOURCE: IDG Research Services, 2013

Biola University Deploys Effective Storage for Growing Video Library



BIOLA UNIVERSITY, A PRIVATE

Christian college serving more than 6,000 students in Southern California, has been aiming to reach an even broader audience through extended courses and distance learning via online programs.

These programs use video (such as recorded courses, seminars and other events), and the amount of video continues to grow every semester. The university has steadily increased the number of recorded events, and to enhance presentations it has adopted multiple camera angles, high-definition video content and advanced editing techniques, all of which add significant data volume.

The university produces at least five to 10 terabytes of raw video footage each semester; edited files add even more terabytes on top of that. To repurpose content easily, the marketing group wanted to indefinitely store all raw video footage as well as edited final copies.

The university's marketing and IT departments worked to centralize the video content on a single server, but costs increased and external drives began to stack up. The university needed a solution that fit well into a large IT environment, allowing the marketing group to leverage tape libraries in the data center. The Scalar Linear Tape File System (LTFS) appliance from Quantum met that requirement.

The university now uses the Quantum Scalar LTFS appliance as an integral part of its storage process for video production. As camera footage is ingested, a copy is immediately transferred to the LTFS file share at the data center, ensuring that the original footage is safe. The marketing department then begins the editing process, and at the end of the semester a second copy of the raw source and completed project is again put on the LTFS for archival purposes, ensuring that all the material remains together.

The ability of the appliance to present tape storage as just another file share is critical for both the IT group and other departments at Biola, significantly simplifying the storage environment. The IT group does not have to stage data, no special client software is needed and individual groups can manage their own storage needs.

The solution also allows the university to deliver these benefits while leveraging its existing IT infrastructure. The LTFS appliance sits in front of an existing Quantum Scalar i500 tape library, and enables individual drives to be assigned to remote LTFS access or traditional data backup.

Unlike traditional data protection tools that back up VMs as if they were physical servers, new solutions are easier to deploy and decrease overall costs.

These new products combine deduplication and advanced virtual machine (VM) optimization utilities, offering data reduction and reduced recovery time for both physical and virtual data protection. Storage appliances also feature direct tape creation, providing integrated tiered storage capabilities to further reduce costs for long-term retention.

Unlike traditional data protection tools that back up VMs as if they were physical servers—which can strain server and network resources with agents and uncoordinated I/O—these new solutions are easier to deploy and manage and decrease overall costs.

The new tools provide an automated, more efficient approach to backing up and archiving VMs. Users can restore or boot VMs in seconds without a backup application, providing immediate access to all files in all backups without complicated restore applications. Users can minimize the impact on virtual servers and networks by optimizing VM image contents before transmitting those images to backup storage, thus significantly reducing the cost of using traditional backup applications to protect virtual environments.

An effective unstructured data storage solution should have certain key characteristics:

- » Tiered storage options for any cost profile and data access profile
- » Options to move data automatically between tiers, such as primary, disk, object and tape
- » Self-protection without the need for a daily backup over the network

Today's newest solutions have these characteristics, enabling organizations to leverage tiered storage for their unstructured data without jeopardizing data protection or putting too much strain on the existing infrastructure.

To learn more about the latest solutions for effective tiered storage environments, visit www.quantum.com/tieredstorage

The Ark Opts for Object Storage to Handle Unstructured Data



THE ARK POST PRODUCTION is one of London's leading post-production facilities, with two sites that manage the duplication, conversion, encoding and editing of digital media.

The company was faced with tightening production deadlines and increasing amounts of media content to manage in various formats. It needed a solution capable of handling a multi-petabyte archive of unstructured data while providing self-protection and fast retrieval times.

In addition, content creators and video editors needed to seamlessly and immediately access and share content across various editing and VFX applications, in which each toolset is supported by different operating systems.

The Ark deployed Quantum's Lattus-M object storage and StorNext appliances to manage, share and preserve its volumes of digital assets and provide a seamless and efficient end-to-end media workflow.

Lattus-M integrates next-generation object storage technology with StorNext Storage Manager in a highly scalable and efficient disk-based archive, and enables policy-based movement of content and rapid retrieval of archives at nearline speeds. It offers much greater durability than traditional RAID offerings, including true self-healing and self-protection capabilities. The result is a highly resilient, disaster-tolerant archival repository that does not need to be backed up.

StorNext delivers the high performance and flexibility of the StorNext File System in a pre-configured metadata appliance that enables collaboration across the entire workflow.