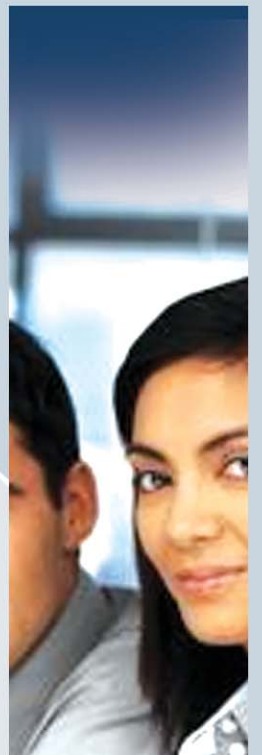
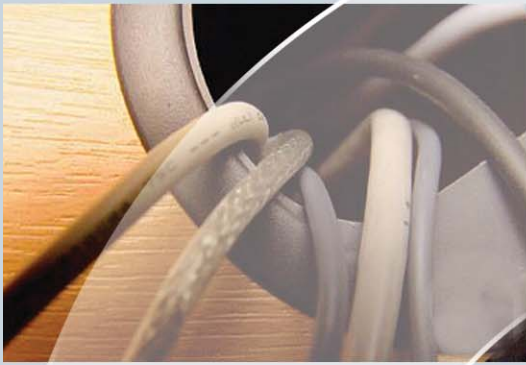
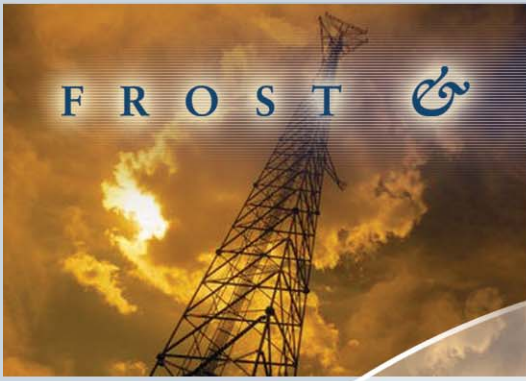


FROST &

SULLIVAN



Information Lifecycle Management (ILM):  
The Most Cost-Effective Way  
to Manage Data for Retention,  
Compliance, and Corporate  
Governance



A Frost & Sullivan White Paper

## **Information Lifecycle Management (ILM): The Most Cost-Effective Way to Manage Data for Retention, Compliance, and Corporate Governance**

*A Frost & Sullivan White Paper Sponsored by ARKIVIO and EMC  
Analyst and Author: Jarad Carleton, IT Consulting Analyst*

### **EXECUTIVE SUMMARY**

As companies around the globe move away from paper-based documents they are finding themselves awash in unstructured digital content that lacks the type of access, retention, and destruction controls of the paper-based world. Although the transition to digital content is a business necessity, the new reliance on digital content has led to a chaotic situation, where information has become scattered across multiple data repositories within the organization. Unfortunately, content chaos is the norm rather than the exception in many business environments, and leaves administrators in the untenable position of not being able to discern where unstructured information is stored or how many identical copies exist in the storage infrastructure.

This white paper examines the business challenges faced by organizations that must retain and archive large amounts of unstructured content while also restraining the growth of storage infrastructure expenditures. A business case is made for the necessity of a comprehensive Information Lifecycle Management (ILM) plan with an integrated hardware and software solution capable of proactively managing storage volumes, data types, retention periods, and data destruction cycles. Finally, two tightly integrated technologies offered by ARKIVIO and EMC are presented as a solution capable of meeting the demanding and complex ILM needs for businesses around the world.

## INTRODUCTION

For over a decade the information technology (IT) sector has been alluding to a time when digital information would overtake paper-based content as the preferred medium of sharing information. Although the transition is not yet complete, Frost & Sullivan estimates that 70 percent or more of all content used by companies today will be created, processed, shared, and stored digitally. In fact, the amount of digital content versus paper-based content is increasing with each passing year.

Unfortunately, the transition away from paper-based content has led to digital content chaos, a scattering of digital content across multiple data repositories in many organizations. As a result, administrators are unable to discern where unstructured information is stored or how many identical copies exist in the storage infrastructure. Content chaos is directly linked to rising storage infrastructure costs, versioning and access control problems, and ensuring strict data retention and expiration schedules. All of these problems have a negative affect on IT initiatives that focus on cost containment, productivity enhancement, and financial performance improvement. Because the transition to digital content is an unquestionable business necessity, business has begun to focus on ways to rein in storage infrastructure growth and bring the digital content chaos under control.

Analysis of the problems associated with digital content has led many IT professionals to realize that current storage systems are both inadequate and inefficient in dealing with corporate governance and regulatory compliance. In order to address the problems with digital content growth, information lifecycle management (ILM) initiatives are being implemented across all industries with a need to manage file system and email content from creation to expiration.

The concept of ILM is straightforward and involves a system wherein data types are analyzed and moved through different tiers of storage depending upon business needs such as value of the data to the business, secure online data permanence, and a standardized retention and data expiration policy across the organization. While the inherent advantages of ILM are improved regulatory compliance and business continuity, corporate governance improvement and cost savings are the most important benefits an organization will achieve with an ILM initiative.

### Corporate Governance vs. Corporate Compliance

Corporate governance consists of company policies that address a company's assets and employees. These policies are often associated with operational efficiency, including lowering operational costs, increasing worker productivity, and improving financial performance.

Corporate compliance occurs when a company implements policies in order to obey established laws and regulations in the nation or state where it conducts business. Quite often, the only difference between the two is the organization requiring adherence - the Compa(governance) or the Government (compliance).

## ILM BUSINESS CHALLENGES

### Data Restoration from Tape

Tape storage has been a preferred method for data archiving because of the low cost associated with the systems and the tapes themselves. Unfortunately, what earlier corporate operational efficiency efforts failed to reveal is that the total cost of ownership (TCO) of a tape storage system is much higher than the cost of the system, tapes, and offsite tape storage. Restoring data from tape is not only time-consuming, but it can cost over \$1 million for large-scale recovery efforts.

The reason data recovery costs from tape storage are so high is because the company must use extensive manpower for each job. Employees must search tapes for pertinent data, rent equipment to read tapes that are several years old, and in some cases work overtime to accelerate the data restoration process due to time-sensitive litigation or regulatory audits. Although data restoration requests resulting from litigation or compliance events are random, internal requests for archived tape data can occur for a number of business reasons. Some examples might include finding new uses for pharmaceuticals, using prior research data to create new compounds in the petrochemical industry, or referencing old engineering designs in a variety of industries.

### Duplicate File Storage

Many organizations rely on traditional hierarchical storage management (HSM), a procedure that moves data from hard drives to optical disk or to tape. This data migration procedure was established as a way to archive old or less valuable data on slower and less expensive storage media than magnetic disks. As prices for hard drives fell over the past decade, administrators have shown a preference for adding SCSI or ATA hard drives when additional storage is needed rather than following old corporate HSM practices. Providing additional online storage has been well received among business users, but it has failed to rein in out-of-control storage infrastructure costs related to storing multiple copies of the same files.

Storing multiple copies of the same file is a problem that all organizations face and that most have failed to effectively manage. This is because it is difficult to determine if different copies are different versions, if the file is stored in multiple repositories, or if an identical file has been stored repeatedly with different names in different directories. Frost & Sullivan has observed that unstructured data is typically saved between four to six times and that this number can increase exponentially when the file is forwarded multiple times as an email attachment.

The file duplication problem has become so severe that it is attracting the attention of both corporate governance and corporate compliance officers. For example, administrators are unable to assert control over unstructured data of a proprietary and sensitive nature, or data that is subject to corporate- or government-mandated retention and disposition periods. Furthermore, as more content is created in digital format, the duplicate file storage problem will get worse and demand for additional storage capacity will continue unabated.

### Data Restoration from Tape

Restoring data from tape is not only time-consuming, but it can cost over \$1 million for large-scale recovery efforts.

### Duplicate File Storage

Frost & Sullivan has observed that unstructured data is typically saved between four to six times and that this number can increase exponentially when the file is forwarded multiple times as an email attachment..

### Structured vs. Unstructured Data

Structured data is data that resides in fixed fields within a relational database.

Unstructured data is file-based data that does not reside in a fixed location format. A word processing document, presentation, web content, and image are all typical examples of unstructured data.

### Standardized Record Retention and Disposition Policy

Regardless of an organization's size, a standardized record retention and disposition policy is a business necessity from a corporate governance and corporate compliance perspective. Approximately 50 percent of the digital content in an organization consists of unstructured or file-based data and another 30 percent is semi-structured or email data. Since the amount of digital content in the organization grows each day, IT departments are overwhelmed and unable to identify how many times unstructured data is stored, where it is stored, and how many versions of the data are stored in different repositories across the company.

The difficulty in understanding the nature of unstructured data and the lack of classification technology makes it almost impossible for organizations to assign meaningful classifications that meet the needs of the business. As a result, most organizations have been unable to implement an ILM strategy that includes the ability to differentiate between different data types and treat them according to their business value. This results in higher storage infrastructure costs because the organization continues to buy additional high-cost, primary storage rather than manage the information and push less critical data to lower cost secondary storage. A poorly implemented ILM plan can also negatively impact productivity and operational costs. For example, without a standardized and uniform data retention policy, valuable data can be inadvertently destroyed while useless data such as music files or personal pictures are routinely retained. Ensuring data permanence for valuable business data, and data destruction for content with no business value is critical for corporate governance as well as compliance initiatives.

## **FILE ARCHIVING AND CORPORATE GOVERNANCE BEST PRACTICES**

As a result of the business challenges facing every industry, corporate governance efforts have begun to focus on ILM improvements that also help to restrain soaring storage infrastructure costs. In examining these same issues, Frost & Sullivan has found that first, it is important to determine the type and purpose of the data. Then the data should be categorized based on its value to each department or to the organization as a whole. Classifying data in this manner enables administrators to identify what does not need to reside on primary storage.

In most companies, up to 80 percent of data residing on high-performance primary storage should be moved to a lower-cost storage tier. This migration is important because it improves the performance of the primary storage tier, while enabling the company to purchase storage capacity that matches the data's value.

### **Record Retention**

Fifty percent of the digital content in an organization consists of unstructured or file-based data; another 30 percent is semi-structured or email data.

### **Best Practices**

In most companies, up to 80 percent of data residing on high-performance primary storage should be moved to a lower-cost storage tier.

## **FILE ARCHIVING AND CORPORATE GOVERNANCE BEST PRACTICES**

Finally, companies should consider an online storage archive capable of ensuring data permanence, data authenticity, and single-instance storage capabilities. Only with single-instance storage technology and matching the right storage to data based on the business value of the information stored will a company be able to slow the growth of storage expenditures and develop a comprehensive ILM plan. If an IT department insists on using offline storage for long-term data archival, it is likely that the decision will cost more than what it is capable of saving the company in cheap optical or tape media.

Furthermore, optical and tape media should not be haphazardly replaced with serial ATA or SCSI drives. These types of low-cost alternatives cannot provide information on when data was archived, accessed, or altered. Only an ILM system with online storage is capable of providing non-repudiable audit trails that are necessary for legal discovery and regulatory compliance.

## **TECHNOLOGIES CAPABLE OF MEETING CORPORATE GOVERNANCE BEST PRACTICES**

### ARKIVIO® auto-stor

Technology advances over the past five years have transformed long-term file archiving from an uncertain and inefficient process to one that is able to meet the demands of today's business environment. ARKIVIO has developed an integrated solution capable of performing a deep-level scan of an organization's storage volumes to give administrators the information they need to develop an all-encompassing ILM plan with automated data management tools to facilitate implementation.

Using ARKIVIO technology, an organization can perform an agent-less ILM assessment of online storage volumes that provides an in-depth analysis of how much data is stored, where it is stored, which departments own it, the value of that information to the company, and how many duplicate files are stored by the organization. In most organizations, ILM assessments reveal that up to 80 percent of data on primary storage can be migrated to second-tier storage, thus freeing up terabytes of space on high-performance storage systems.

Following the assessment, the IT department uses the ARKIVIO® auto-stor software to easily group and classify all unstructured data across the organization and begin creating migration and retention policies for different data types, departments, and work groups. Administrators only need to apply the policies once, since they are automatically updated as the organization adds additional storage capacity and new data that fit existing storage classes.

**TECHNOLOGIES  
CAPABLE OF  
MEETING  
CORPORATE  
GOVERNANCE BEST  
PRACTICES**

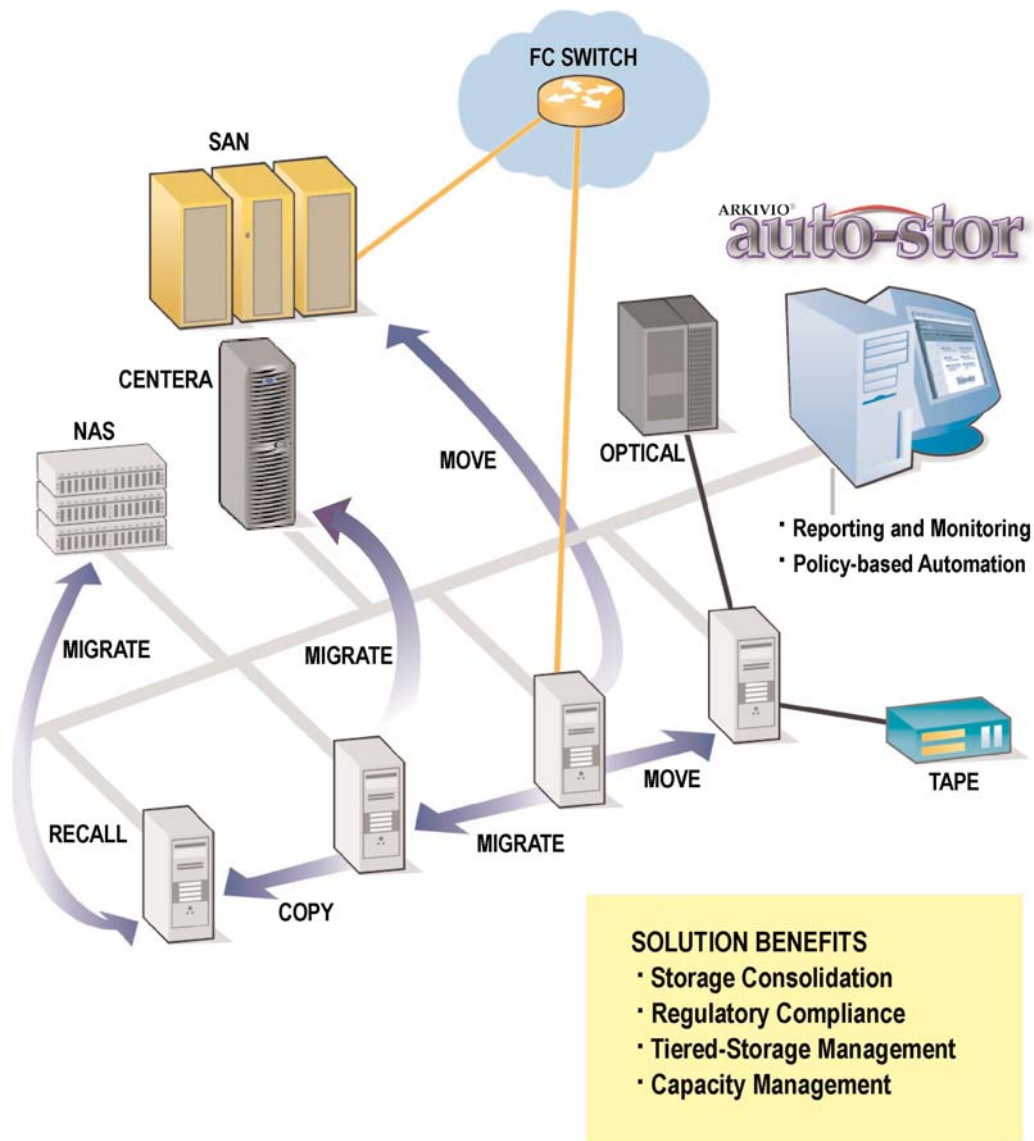
ARKIVIO® auto-stor also enables administrators to organize storage into pools that can be grouped from anywhere on the network and prioritized based on function or value. A storage group is not bound by a physical device or a single vendor's hardware; it can be as small as a single volume or it can contain volumes from multiple, heterogeneous platforms. Storage groups can also be created around retention periods (seven-year repository, fifteen-year repository, etc.). After data and storage classification are complete, migration of data from primary storage to second-tier storage can begin.

Prior to data migration, administrators can run a simulation based on policies they establish for data and storage types. The simulation provides best- and worst-case scenarios of how much space will be freed on primary storage, and how much secondary storage will be needed after duplicate files have been eliminated if an EMC Centera is chosen as the second-tier repository. The simulation can also assess data that should be removed and create deletion policies to eliminate unwanted data. This process frees space on primary storage and eliminates the need to back up or restore unwanted data in a disaster recovery situation. Additional information provided from the simulation includes a summary of files moved and an in-depth view of individual files that will be moved on a server-by-server basis.

When the administrator is satisfied that the policies are working as intended, data migration to secondary storage can be initiated immediately or during a scheduled maintenance window. As data is migrated from primary to secondary online storage, original files on primary storage are replaced with small, 3-kilobyte stub files or links that look and act like actual files for business users. When a business user attempts to access a file on primary storage that has been migrated, the stub or link immediately accesses the data directly from secondary online storage, eliminating the need to restore the data to primary storage. In a Network Attached Storage (NAS) environment, the storage administrator can choose to have the data recalled to primary storage rather than being directly read from the second tier. Furthermore, any changes to migrated data, whether recalled or accessed directly from secondary storage, will be stored as a separate file to ensure the permanence of the original data. For compliance or governance purposes, policies can also be created that move data to the repository, completely eliminating it from primary storage and traditional end-user access. Data can also be copied to the repository after it is created rather than waiting for it to age. This allows a repository like Centera to ensure the retention and immutability of the data for the entire length of time the data needs to be retained by the organization.

Lastly, ARKIVIO® auto-stor makes data backups more efficient for administrators by cleaning up the storage environment so that full backups are only done for data that requires it. Ongoing reports allow administrators to monitor the effectiveness of their policies so they can fine-tune or create new policies as their data environment changes over time. This information also prepares them to acquire the right type of primary and secondary storage at just the right time.

## ARKIVIO® auto-stor Managing ILM in a Network



Source: ARKIVIO

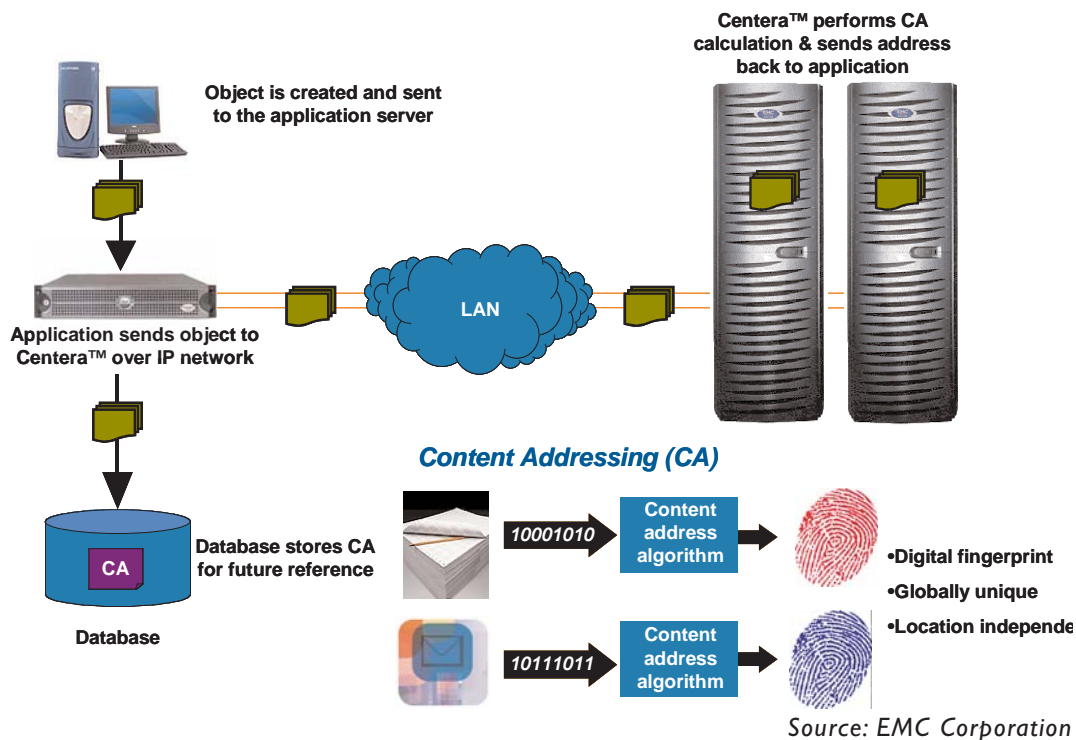
## EMC Centera™ Content Addressed Storage (CAS) System

Although ARKIVIO® auto-stor manages ILM in the organization, the problems of file storage duplication and data permanence require more than ILM management software. Therefore, a storage infrastructure capable of meeting corporate governance needs should also be capable of single-instance storage, archive data availability, security, authenticity, and non-repudiation as well as file retention and disposition in any business environment.

Frost & Sullivan Whitepaper  
Sponsored by: ARKIVIO and EMC

EMC Centera™ with its Content Addressed Storage (CAS) system provides an advanced storage technology capable of delivering tamper-proof audit trails as well as single-instance storage capabilities. Single-instance storage enables an organization to decrease the amount of additional storage infrastructure needed each year. Redundant file storage is problematic in storage infrastructures, but is not an insurmountable problem when using Centera™. CAS technology makes single-instance storage possible by addressing stored data based on a Content Address (CA) derived from the file itself rather than on the information's physical or logical placement in the storage infrastructure. The addressing and encryption functions are similar to a public key infrastructure (PKI), which can be thought of as digital fingerprints.

### How Centera™ Works



The unique Content Address of each piece of data ensures security, authenticity, and nonrepudiation, all of which are critical from a regulatory perspective. Since government mandates insist on a WORM (write once, read many)-based storage infrastructure for insurance and mortgage banking, Centera™ was designed to enforce application-based retention periods within its micro code. Centera™ only permits authorized Records Managers to lengthen retention periods, but prohibits shortening them. This feature turns Centera™ into a virtual locked filing cabinet that enables Records Managers to assert control over unstructured and semi-structured content in ways that are not possible with optical and tape storage. Because ARKIVIO is Centera™ certified, the combined solution enables high levels of data control that are made possible via a Centera™ API (application program interface) that ARKIVIO® auto-stor integrates with one or more Centera™ systems.

For business continuance purposes as well as for compliance needs, Centera™ stores content and safeguards it using content parity protection within the same storage infrastructure. Thus, in the event of a storage node failure, Centera™ is capable of healing itself by detecting the fault and generating new copies of the content objects onto a healthy node. As this process takes place, the failing storage node is isolated from the rest of the system and can be hot swapped at any time without disruption. Hot swapping without application disruption is possible since applications lack the knowledge of a physical or logical placement of fixed content within Centera™.

Finally, although data retention is important, the information lifecycle management (ILM) process isn't complete without a way to automatically implement data end-of-life policies immediately after retention periods expire. ARKIVIO® auto-stor works through the Centera™ API to provide IT departments this capability utilizing U.S. Department of Defense 5015 standards for permanent data deletion. Automating this process to coincide with corporate policy and regulatory requirements frees up IT personnel and easily applies a standard data destruction policy that helps eliminate potential liabilities.

## **CONCLUSION**

Data classification is a critical function that must be addressed in order to standardize and implement a comprehensive ILM plan for any industry. This is important not only for the purpose of controlling storage infrastructure costs, but also to ensure that data is retained and disposed of efficiently and in accordance with corporate and regulatory guidelines. In today's competitive global business environment, making sure that the right data gets to the right place at the right time is often the difference between success and failure.

ARKIVIO and EMC offer an efficient and cost-effective method to address corporate governance challenges with technologies that greatly improve ILM while reducing operational costs and storage expenditures. By automating the ILM and data migration process and integrating it with secure online secondary storage, it is easier for companies to improve corporate governance and reduce liability risk. It is the opinion of Frost & Sullivan that those organizations wanting to maintain a competitive edge while restraining costs should seriously consider the combined ARKIVIO® auto-stor/EMC Centera™ solution.

## CONTACT US

Palo Alto

New York

San Antonio

Toronto

Buenos Aires

Sao Paulo

London

Oxford

Frankfurt

Paris

Israel

Beijing

Chennai

Kuala Lumpur

Mumbai

Shanghai

Singapore

Sydney

Tokyo

**Silicon Valley**  
2400 Geng Road, Suite 201  
Palo Alto, CA 94303  
Tel 650.475.4500  
Fax 650.475.1570

**San Antonio**  
7550 West Interstate 10, Suite 400,  
San Antonio, Texas 78229-5616  
Tel 210.348.1000  
Fax 210.348.1003

**London**  
4, Grosvenor Gardens,  
London SW1W 0DH, UK  
Tel 44(0)20 7730 3438  
Fax 44(0)20 7730 3343

**877.GoFrost**  
[myfrost@frost.com](mailto:myfrost@frost.com)  
<http://www.frost.com>

### ABOUT FROST & SULLIVAN

Based in Palo Alto, California, Frost & Sullivan is a global growth consulting company. This white paper is part of Frost & Sullivan's ongoing strategic research into the Information & Communication Technologies industries. Frost & Sullivan regularly publishes strategic analyses of the major markets for products that encompass storage, management, and security of data. Frost & Sullivan also provides custom growth consulting to a variety of national and international companies.

The information presented in this publication is based primarily on interviews and therefore is subject to fluctuation. Frost & Sullivan takes no responsibility for any incorrect information supplied to us by manufacturers or end users.

This publication may not be downloaded, displayed, printed, or reproduced other than for non-commercial individual reference or private use within your organization, and thereafter it may not be recopied, reproduced or otherwise redistributed. All copyright and other proprietary notices must be retained. No license to publish, communicate, modify, commercialize or alter this document is granted. For reproduction or use of this publication beyond this limited license, permission must be sought from the publisher.

For information regarding permission, write:  
Frost & Sullivan  
2400 Geng Rd., Suite 201  
Palo Alto, CA 94303-3331, USA