

## I D C T E C H N O L O G Y S P O T L I G H T

---

# Cloud Storage: The Efficient Way to Augment Storage Infrastructure

April 2009

Adapted from *Finding Ways to Efficiently Manage Fast-Growing File-Based Storage* by Noemi Greyzdorf, IDC #216131

Sponsored by Iron Mountain

---

### Introduction: The Need to Control Storage Costs

The economic downturn is not expected to slow the rate of data creation. In some cases, recent events are forcing organizations to retain more data for longer periods of time while keeping it available. The challenge is no longer two-dimensional — more data needs more storage. The challenge now is how to store more data for longer periods of time without experiencing a linear increase in storage costs, management overhead, and energy and facilities consumption.

In 2008, for the first time, file-based capacity shipped was greater than block-based capacity shipped. By 2012, IDC expects 78% of all storage capacity will be file based. This is not just for primary systems but also for systems that store copies of data for data protection, disaster recovery, test and development, archiving, and collaboration.

The amount of data is growing, more data is retained, and the data that is retained is retained for a longer period of time. This is happening while IT budgets are expected to remain flat in 2009 and grow a mere 2.3% in 2010. The theme that has emerged in this environment is greater efficiency, or do more with less.

In an effort to extract more value from existing assets, IT managers are looking at processes, technologies, and services that can help them reach their objective. This Technology Spotlight explores both the processes and the approaches to managing data and its underlying storage assets more efficiently and effectively. The paper also discusses Iron Mountain's cloud-based Virtual File Store infrastructure.

### Definitions

To level the discussion, IDC defines the terms that are used throughout the document:

- **Capacity optimization** refers to the act of reducing the data footprint and gaining more usable capacity as a result of compression, single instancing, or deduplication. Compression is the elimination of white space, single instancing refers to deduplication data at the file or object level, and deduplication assumes the elimination of redundancies at the subfile level.
- **Storage management** software includes products that monitor and report on the allocation, performance, utilization, and consumption of storage resources. This includes storage resource management, storage system management, and SAN management software.

- **Cloud storage** refers to the inherently scalable, flexible, and redundant infrastructure that is the foundation of any storage services, such as online backups, online archiving, and storage capacity. It assumes that storage services built on top of a cloud infrastructure are priced and delivered as a utility, charging for the resources utilized.

## Three Steps to Curbing File-Based Storage Sprawl

1. The first challenge most enterprises experience with curbing file-based storage sprawl is knowing what data they have; what value it represents to the organization; and as a result, how it needs to be stored, protected, and disposed of. It is surprising how quickly a few unrelated files can perpetuate significant storage consumption. As an example, a typical .jpg file is 2MB, assuming decent but not great resolution. This may not seem like much when dealing with terabytes or even petabytes of data, but if there are many 2MB files and each is perpetuated throughout the online, data protection, and disaster recovery systems, then a significant amount of storage becomes consumed by data that may pose no value to the enterprise. Personal data doesn't have to be the only culprit in consuming increasing storage resources; there are many files that exist without owners, files that are old and haven't been accessed in over a year, and files that may no longer possess value to the organization that consume capacity. Understanding the realities of your situation will help you formulate better policies around file-based storage management.
2. In an effort to increase productivity and collaboration, more data is being created in digital form. Additionally, advances in technology have resulted in the generation of ever larger files. Whether it is instrumentation data, a presentation, or a legal document, files are consuming more storage capacity. Assuming that only files that represent value to the organization remain on the file-based storage systems, the next step in reducing storage sprawl is to leverage capacity optimization technologies. Capacity optimization refers to the process by which the capacity consumed by a file or a data set is reduced. This can be achieved through compression, single instancing, or deduplication. The approach you select will depend on how active your data is and what type of data you store. For online file-based data that is active, inline compression may be the best solution. It will reduce the data footprint inline, allowing the life of the storage asset to be extended. In an archive situation or in a backup, single instancing can help reduce the number of times a unique file is stored. Deduplication at the subfile level can reduce the data footprint of static content in the online storage system or in a backup system. Not all deduplication approaches are the same; some address the needs of a backup and restore system, while others make sense only in an online storage system. In the end, whether any of these approaches make sense and can help you reduce your file-based storage capacity consumption depends on the payback you realize with the deployment of these systems.
3. Data sprawl is having an effect on the datacenter as a result of the increasing storage system footprint. Some of the storage resources deployed are not fully utilized, not only because too much capacity had been acquired up front but also because some of this capacity is used only periodically in support of various projects or seasonal activity. Examples of such projects and activity include marketing campaigns, ediscovery and litigation support, development collaboration, and reporting. In these instances, organizations could benefit from leveraging storage-as-a-service offerings. Storage as a service allows capacity to be provisioned and consumed on as-needed basis without any upfront infrastructure investment or management overhead. "Outsourcing" for storage capacity can dramatically reduce the overall cost of storage.

## Considering Iron Mountain

Virtual File Store (VFS) is a service based on Iron Mountain's scalable, redundant cloud-based infrastructure. VFS enables organizations to store their files without having to design, build, and support internal infrastructure. End users may subscribe to VFS for any amount of capacity they require and grow and shrink that capacity based on business demands. The cost of the service depends on how much capacity is consumed and for what period of time.

VFS is a service that doesn't require end users to make any investments in hardware, software, or people. Once end users have subscribed, data can be moved to Iron Mountain's cloud using the NAS appliance deployed within the customer's datacenter. The NAS appliance supports all standard network protocols such as CIFS and NFS and communicates directly with the cloud. Following are three ways data can be moved to the VFS appliance:

- Users can write directly to the NFS or CIFS mounts. Once data is written to the appliance, users can read, open, and modify files as though they were working off a local NAS system.
- Using scripts, file virtualization, or HSM, data can be moved to the VFS appliance based on predefined parameters. These parameters are determined by the business unit or administrators and typically include age of file, last time file was modified, type of file, or even file owner. The migration of the file to VFS is transparent to users, who still see the file icon in their directory.
- Using specialized software such as archiving or data protection software, data can be moved to the VFS appliance and thus the cloud. An archiving application would write a file to a storage device locally the same way it would write a file to VFS NAS.

Often, users already have existing archives that may be relatively large, making the migration to appliance and the cloud over the Internet impractical. Iron Mountain offers an alternative way to move data to the VFS cloud and to retrieve it. This service is called the Data Shuttle. Data is copied to a local device that is then securely shipped to Iron Mountain. Data is uploaded locally, eliminating the time otherwise required to move data over the Internet. The same service is available if a large data set needs to be retrieved from the archive. The data would be copied to a local device and shipped securely to the customer, making the retrieval simple and fast.

Once data has been stored in the VFS cloud, users may apply additional services available from Iron Mountain. These services include discovery, litigation support, compliance archiving, and disaster recovery.

### **Challenges**

Storage as a service, or cloud storage, is an emerging market that is not always well understood. Organizations have become familiar with online backups and even email archiving services, but they are often confused by the term "cloud." Cloud is a way to describe an infrastructure designed to deliver storage as a service. Iron Mountain is expected to face the following common challenges associated with either a new technology or a new paradigm:

- Iron Mountain needs to explain to end users the concept of a service. Users are used to buying boxes and typically want to know what is in the back end. It is a challenge to have a conversation about storage capacity without getting bogged down in a discussion of the infrastructure.
- End users are still concerned about the security and availability of their data in the cloud. Though Iron Mountain has been securely storing 93% of the Fortune 1000's data for decades, the concept of the cloud is unfamiliar to many. Iron Mountain must take the time to explain why

transporting data via the VFS appliance is safe as well as assure users that once the data is in Iron Mountain's possession, it is safe. Safety concerns not only are related to the underlying hardware but also are associated with safety from viruses, administrative errors, and physical intrusions.

- For some, Iron Mountain is still the provider of long-term storage of paper documents and tapes. Making the market aware of its capabilities in this space will be critical in establishing a market position.
- During challenging economic times, organizations seek to conserve cash, and a service such as VFS is well positioned to help them do that. In less dire economic times, organizations tend to buy based on a budget and a project. Providing customers with financing options should allow Iron Mountain to service a broader base of customers.

## Conclusion

In today's data-intensive market, balancing the need to effectively store information in support of business and the associated costs of storage is a true challenge. More data is being created and needs to be stored, but this doesn't mean there aren't ways managers can contain data sprawl as a result. Managers can contain their file-based storage sprawl in three ways:

- Manage what gets stored and where
- Reduce the data footprint
- Augment existing storage assets with outsourced storage services where appropriate

In response to the emerging challenges weighing on storage managers, Iron Mountain has architected an infrastructure that allows storage capacity to be delivered seamlessly to organizations seeking some relief.

---

### ABOUT THIS PUBLICATION

This publication was produced by IDC Go-to-Market Services. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Go-to-Market Services makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

### COPYRIGHT AND RESTRICTIONS

Any IDC information or reference to IDC that is to be used in advertising, press releases, or promotional materials requires prior written approval from IDC. For permission requests, contact the GMS information line at 508-988-7610 or [gms@idc.com](mailto:gms@idc.com). Translation and/or localization of this document requires an additional license from IDC.

For more information on IDC, visit [www.idc.com](http://www.idc.com). For more information on IDC GMS, visit [www.idc.com/gms](http://www.idc.com/gms).

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 [www.idc.com](http://www.idc.com)