

## Behind Data Dedupe – Data Reduction

Having engulfed the IT team, the data avalanche is set to hit the wider business. Organizations now have to impose restrictions on email inboxes and local storage facilities because of the amount of data being produced by users across the network.

The problem may feel insurmountable; however a key way of reducing data is to examine how much information is simply replicated by multiple users. With standardized operating systems and applications, come thousands of identical files on legions of computers. Add to that identical attachments stored in multiple recipients' inboxes and it is easy to see how much duplicate documents add to an organization's storage requirements.

Unless organizations want to risk facing an all-engulfing data avalanche, the amount of data being stored has to be reduced, or managed more efficiently. Vendors have been quick to address this critical pain point; however it is unclear as to whether these technologies have the capacity to cope with new developments in data such as bigger file sizes, multimedia formats and distributed data. Unless the situation is evaluated now, companies may find themselves left with a quick-fix solution that could quickly leave them in the same position as before.

### **DATA DEDUPLICATION (DATA DEDUPE)**

Data Deduplication has quickly risen to the top of the IT agenda as a method to help reduce storage and power costs through streamlining the amount of information needing to be backed up. It also helps to address issues such as business continuity, e-discovery and compliance requests.

Deduplication technologies can take a myriad of forms, but there are several fundamental methodologies:

- Elimination of identical duplicate files across the network
- Incremental backups – finding the differences between today's and yesterday's files and only saving the changes
- File compression – further reducing the volume of data stored

These techniques are highly effective at stripping out a huge amount of backed up data that simply is not required. The technology can also work across distributed data centers, ensuring one centralized version of a document is backed up, rather than several different versions held on different devices.

However, deduplication really only tackles the initial symptoms of the data mountain and will not be able to match the growing average size of files as video and media files become increasingly popular. Compression is already implemented within these file formats, which will mean a reduction rate at the transmission stage. Effectively storage could get worse, rather than better.

## **DATA REDUCTION**

Data reduction takes deduplication one step further, moving it from a reactive to a proactive approach to data management. The technique automates data movement and deletion from the desktop, which reduces the physical volume of data moving around the organization.

Policy driven, the technique ‘tags’ files that are deemed no longer required – this is established through a rules-based system that can be set up by administrators or IT managers. These files can then be extracted from their current position and either moved to the archive or deleted securely.

Data reduction should technically reduce the requirement to educate users about how to manage their own data storage effectively. Moving data management to an automated, policy-driven mechanism removes the need for workers to worry about where and when their data is backed up.

However, ensuring users understand why data reduction policies are in place and how they can help remove any blockages to the backup pipeline will always help an organization’s long-term data strategies succeed. Common practices, such as using email inboxes as a secondary storage system for large documents such as PowerPoint presentations, will always continue. At the same time, IT managers should still encourage users to take a robust and rigorous approach to their individual storage habits.

One thing in storage remains constant – the amount of data we produce on a daily basis will continue to grow. Data reduction policies need to be conceived and executed now to ensure employees are not brought to a halt by a data avalanche.

## **PATENTED DATA REDUCTION TECHNOLOGIES — CONNECTED® BACKUP FOR PC**

The Connected Backup for PC product is a proven solution that automatically protects and restores user or server data files, operating systems, and applications using a single central Data Center while reducing IT costs and email storage demands.

Connected Backup for PC is based on patented technologies that minimize bandwidth requirements and ensure reliable, consistent and secure backups even over a dial-up connection.

These strategies together minimize network traffic, reduce data storage requirements by up to 85 percent, and enable users to use the service over slower connections.

The Connected Backup for PC consists of the following patented data dedupe technologies:

- **SendOnce®** – backs up common files to a SendOnce pool, preventing extra copies of shared files from being stored on the data center server.
- **Delta Block®** – allows only changed data – at the block level – to be saved. When a document changes, only the changes are backed up. This minimizes network traffic, reduces backup time, enables backup even over dial-up connections and cuts the overall size of stored files.
- **Connected EmailOptimizer™** – reduces the storage space for email files by 30 percent. This optional service recognizes duplicate email attachments and stores them only once, significantly reducing storage requirements.

#### **DATA REDUCTION TECHNIQUES - LIVEVAULT®**

LiveVault is an automated, fully managed server backup and recovery solution. It automatically and continuously backs up servers to a secure off-site vault, from which it can be quickly and easily restored in the event of human error, virus attack, system failure or disaster.

LiveVault uses delta-block techniques to avoid transmitting more data than necessary. This technology is proven to produce up to a 1000:1 ratio of the data under protection compared to the amount of data transmitted to perform the backups. LiveVault's delta backup techniques are more effective at saving bandwidth by providing fast and efficient backup and restores by only transmitting the data that changes, using the amount of bandwidth specified by the client.

© 2009 Iron Mountain Incorporated. All rights reserved. Iron Mountain, the design of the mountain, Connected, LiveVault, SendOnce and Delta Block are registered trademarks and Iron Mountain Digital and EmailOptimizer are trademarks of Iron Mountain Incorporated. All other trademarks and registered trademarks are the property of their respective owners.

 **IRON MOUNTAIN** DIGITAL™  
120 Turnpike Road  
Southborough, Massachusetts 01772  
(800) 899-4766

Iron Mountain Digital, the world's leading provider of data backup/recovery and archiving software and storage as a service (SaaS), offers a comprehensive suite of data protection and e-records management software and services to thousands of companies around the world. For more information, visit our Web site at [www.ironmountain.com/digital](http://www.ironmountain.com/digital).