



By K. E. H. Polanski

EXTENDING THE ADVANTAGES OF DATA CENTER CONSOLIDATION WITH DE-DUPLICATION TECHNOLOGY

Hardware consolidation has become a key way for enterprises to increase efficiency and reduce ongoing data center costs. By combining this approach with de-duplication technologies such as those in CommVault® Simpana™ 7.0 software, which are designed to remove unnecessary copies of data, they can reclaim storage space and extend consolidation down to the data itself.

Enterprises must constantly contend with the pressures of limited floor space, energy, and cooling systems. One widespread approach to coping with these pressures is server consolidation through virtualization, which helps reduce the number of physical servers required to support key systems. To complement this type of approach and help further increase return on investment, many organizations are also considering implementing de-duplication technologies, such as those in CommVault Simpana 7.0 software. These technologies are designed to eliminate duplicate copies of data, reclaim storage space, and extend consolidation from the physical infrastructure down to the data itself.

UNDERSTANDING DIFFERENT APPROACHES TO DE-DUPLICATION

De-duplication technologies encompass a variety of approaches. One common type is hardware-based de-duplication, in which the de-duplication functionality is integrated into the storage system itself. This model can work well for individual storage systems, allowing them to store more logical data with less storage space than would be possible without de-duplication; however, it is limited to data within that specific device. Software de-duplication, in contrast, can work with many different types of storage across

a data center environment, providing more flexibility than hardware de-duplication.

De-duplication can also work on both the file level and the block level. File-based de-duplication uses hash algorithms to find and remove duplicate files, while block-based de-duplication uses hash algorithms to find and remove duplicates of the blocks that make up the files. By providing reference pointers to a single copy of a file or block, rather than storing multiple copies in different locations, de-duplication helps reduce storage requirements and increase efficiency. By working along two different dimensions, file- and block-based approaches can serve as complements to one another, with the file-based approach removing “horizontal” copies of files and the block-based approach working “vertically” within these files to remove duplicate blocks.

When evaluating de-duplication approaches, organizations should keep in mind that they are not implemented in isolation, but as part of a comprehensive system of data management used in a particular environment. That environment typically includes some combination of backup and recovery, archive, snapshot, and replication technologies in addition to de-duplication. It may also require encryption or compression when data is transmitted across wide area networks or stored on tape for long-term

Related Categories:

CommVault

Dell PowerVault storage

Server consolidation

Storage

Storage software

Visit DELL.COM/PowerSolutions for the complete category index.

“De-duplication technologies are designed to eliminate duplicate copies of data, reclaim storage space, and extend consolidation from the physical infrastructure down to the data itself.”

off-site retention, and may include service-level agreements for levels of data recovery that the organization must support and enforce.

When taking these factors into account, the differences between file- and block-based de-duplication can be significant. For example, when file-based de-duplication removes copies of files, the image of the file is preserved. When block-based de-duplication removes identical blocks, in contrast, the image is not preserved, and must be rebuilt before files can be restored—an important consideration for organizations where rapid recovery is a high priority. In addition, while file-based de-duplication can integrate easily with encryption and compression processes, doing so with block-based de-duplication typically causes a net expansion of the data footprint.

Both file- and block-based de-duplication can be hardware independent, but file-based approaches are typically software implementations, while block-based

approaches are typically hardware implementations. In these forms, file- and block-based de-duplication can work together, with the software enabling de-duplication of files stored on virtually any disk while the hardware provides complementary block-based functionality for specific devices.

USING DE-DUPLICATION WITH BACKUP AND ARCHIVE SYSTEMS

Single-instance store (SIS) capabilities, such as those provided by CommVault Simpana 7.0, are a key method of file-based de-duplication. By comparing the contents of files to locate duplicates regardless of differences in file names, properties, or attributes, SIS helps eliminate duplicate copies of files and attachments found in backups and archives stored on disk. This method helps ensure that exact digital duplicates can be matched and eliminated across data sets, applications, clients, and OS platforms to help maximize compression.

As new backup and archive copies are made, this compression ratio can continuously increase as the number of eliminated duplicate files increases. As shown in Figure 1, SIS is particularly well suited for eliminating duplicate files across backup cycles, eliminating duplicate files across backup and archive copies, and consolidating tape copies of data to SIS disk copies for search and discovery. SIS can apply to both primary backup and archive copies of data as well as to secondary copies created by storage policies. As logical copies of data age and are gradually retired, reference links to the single physical copy can be eliminated until none remain, after which the physical copy can also be removed. Figure 2 lists some key features and advantages of the SIS capabilities of CommVault Simpana 7.0.

IMPLEMENTING DE-DUPLICATION IN THE DATA CENTER

Tape backup, the traditional method of backup and archiving, is still useful and appropriate for many uses, particularly off-site rotation and long-term data preservation. However, combining de-duplication with Serial Attached SCSI (SAS) and Serial ATA (SATA) storage systems enables organizations of all sizes to maximize the advantages of simple, cost-effective disk-based data management.

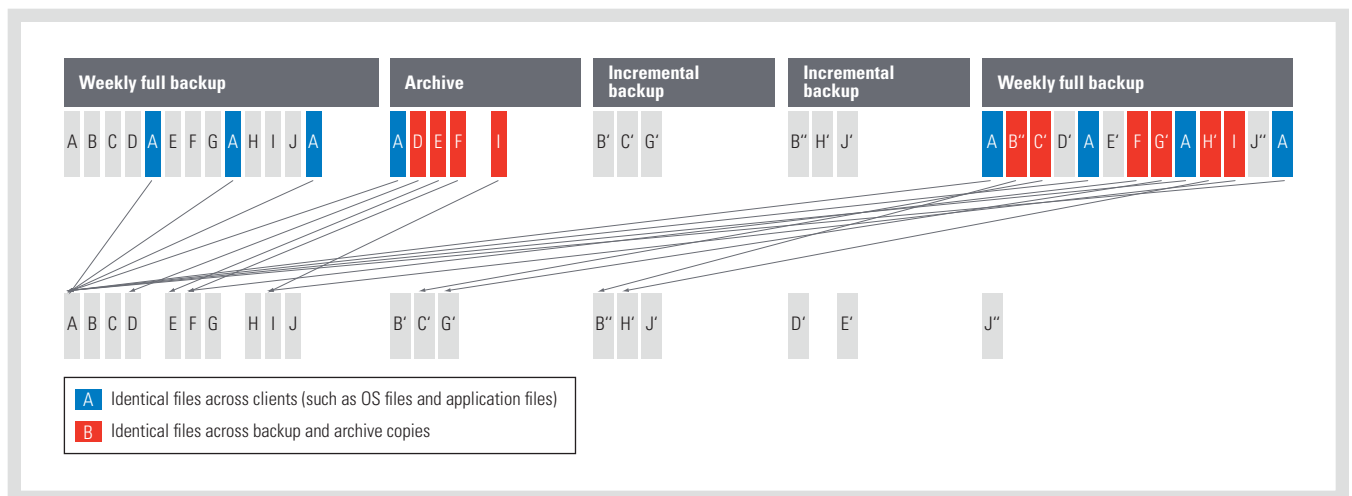


Figure 1. Elimination of duplicate files across backups and archives through single-instance store de-duplication

Storage systems such as the Dell™ PowerVault™ MD3000 and PowerVault MD3000i, for example, combine disk-based storage with tape and are well suited for SIS backup and archiving for small organizations. Using SIS with these systems can help organizations keep more data on disk for longer periods than they could without this functionality, and can help accelerate recovery while also preserving data on tape. Organizations can also use the same approach with Dell/EMC CX series arrays, perhaps in combination with virtual or physical tape libraries, to help accommodate data growth while still taking advantage of de-duplication features.

In addition, organizations can combine SIS with other technologies to help further increase storage efficiency. For example, backup systems that enable aggressive “incremental forever” backup cycles without compromising data recovery can dramatically reduce the amount of data duplicated as part of the normal backup process, helping increase the cost-effectiveness of disk-based data retention even without de-duplication. Data archiving systems can also help reduce

“As enterprises strive to reduce storage costs, particularly those related to backup and archiving, they can take advantage of de-duplication technologies such as those in CommVault Simpana 7.0 to help reclaim storage space, accelerate data recovery, and maximize return on investment.”

the amount of data that organizations must back up; by moving old data to a secondary storage location for long-term retention, organizations can avoid the need to repeatedly protect that data across multiple full backup cycles. This type of holistic approach to data management can help maximize compression ratios and reduce the amount of data that must be stored on either disk or tape.

INCREASING STORAGE EFFICIENCY

De-duplication can be a key part of efforts to consolidate hardware and increase

efficiency. As enterprises strive to reduce storage costs, particularly those related to backup and archiving, they can take advantage of de-duplication technologies such as those in CommVault Simpana 7.0 to help reclaim storage space, accelerate data recovery, and maximize return on investment. [↪](#)

K. E. H. Polanski is a partner at the KEHP Group, a marketing and public relations firm specializing in storage and data management. She was previously the director of product marketing at CommVault, and has worked in product marketing and management, business development, and channel marketing at EMC, Legato, the Qualix Group, and Octopus Technologies. She has a degree in Computer Science from Augustana College.

Feature	Advantages
Removal of duplicate files from backup and archive copies	By reducing the physical disk space required to store logical data, organizations can reduce storage requirements, increase the amount of data they can hold in short-term retention on disk, and enable rapid recovery and search access.
Support for existing disk hardware	Organizations can typically implement the software without changing their physical infrastructure, easily migrate SIS copies to new disk arrays, and use the software with different types of storage from different vendors.
Support for encryption and compression	Unlike block-based de-duplication, organizations can combine SIS with data compression and encryption to secure and further compress data sent over networks and written to disk.
On-demand use	Administrators can run de-duplication when the initial backup or archive operation occurs or apply it later to existing data copies, including those made with previous versions of CommVault software for searching or staging purposes.
Rapid restore	Unlike block-based de-duplication, SIS does not need to rebuild file images before restoring files, helping accelerate data recovery.
Support for block-based de-duplication	Organizations can write SIS data copies to de-duplication appliances, enabling them to combine SIS file duplicate removal and hardware-based block duplicate removal to help further reduce storage requirements.

Figure 2. Key advantages of the single-instance store features in CommVault Simpana 7.0

MORE
ONLINE
DELL.COM/PowerSolutions

QUICK LINKS

CommVault:
www.commvault.com

Dell and CommVault:
DELL.COM/CommVault