ADDRESSING THE CHALLENGES OF DATA GROWTH WITH SYMANTEC DEDUPLICATION AND THE DELL POWERVAULT DL2100

By Charles Butler

Combining the Symantec™ Backup Exec™ 2010 Deduplication Option with the Dell™ PowerVault™ DL2100 – Powered by Symantec Backup Exec can help dramatically reduce bandwidth and storage requirements for backup infrastructures while providing versatile, enterprise-class data protection.

Although data growth is not new, the pace of that growth has become more rapid, the locations of data more dispersed, and the links between data sets more complex than ever—leaving enterprises constantly seeking ways to overcome their data protection challenges. Data deduplication offers the opportunity to dramatically reduce the amount of bandwidth and storage required for backups and to centralize backup data in support of disaster recovery strategies. Although this technology has existed for several years, many organizations have yet to take advantage of the enhanced efficiencies it can provide.

Dell and Symantec have worked together to address the explosion of enterprise data by creating the Dell PowerVault DL2100 – Powered by Symantec Backup Exec. The combination of this integrated backup-to-disk solution and the Backup Exec 2010 Deduplication Option enables organizations to flexibly deduplicate data across their backup environments through either client deduplication (also called *source deduplication*, which occurs at the remote client servers using Backup Exec agents) or media server deduplication (also called *target deduplication*, which occurs in-line when a Backup Exec media server writes data to disk).

To help simplify deployment and scalability, the Backup Exec Deduplication Option does not use capacity-based licensing; instead, it is delivered through a single license key that enables all deduplication functions within Backup Exec, including client deduplication, media server deduplication, and appliance deduplication through integration with Symantec OpenStorage–enabled intelligent disk devices (see the "Intelligent backups through Symantec OpenStorage" sidebar in this article). Incorporating deduplication into the Backup Exec platform can provide a variety of advantages in enterprise backup environments, including helping organizations to dramatically reduce bandwidth and storage capacity requirements, centralize and manage backup operations, and reduce rotations and use of tape for disaster recovery.

DEDUPLICATION AND COMPRESSION

Deduplication and compression technologies can both help reduce the size of a backup, but the two approaches differ in key ways. Compression can be applied to files, directories, or even volumes, but typically lacks awareness of the underlying data and therefore cannot recognize that identical files exist in different directories. In addition, a typical compression algorithm—unlike the deduplication process—can neither recognize changed data nor capture these unique blocks at a subfile level.

When administrators back up data on a remote server and choose to compress that data before backing it up, the compression would generally result
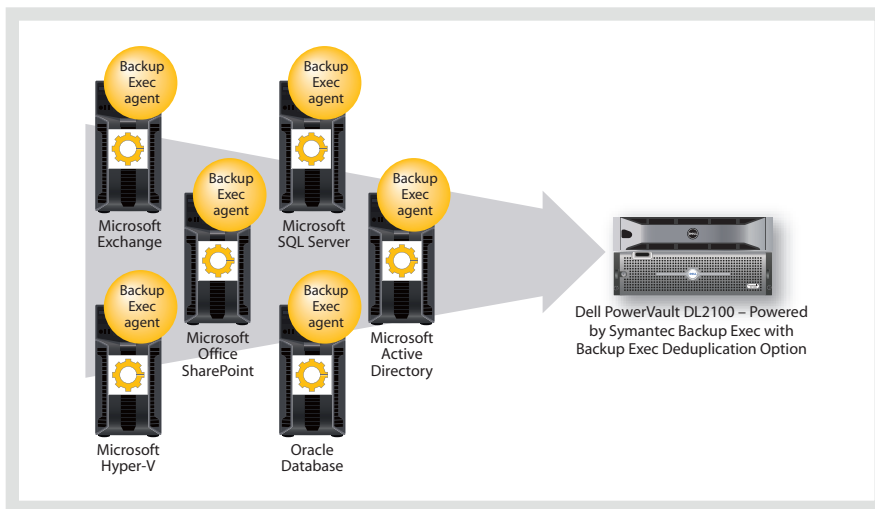
Figure 1. *Client deduplication performed on the source servers is designed to reduce the amount of data sent over the network, helping optimize both bandwidth and storage usage*

in reduced sizes for files sent to the local backup application or across the network to a centralized backup application. When using Symantec Backup Exec deduplication, in contrast, files on the remote server might be recognized as identical to files already stored, before compression occurs—in which case only a very small amount of data would need to be transmitted and stored. Other systems sending data to the same Backup Exec media server would also be aware of similar data that may have already been transmitted and stored, helping reduce backup sizes across the environment.

BANDWIDTH AND STORAGE OPTIMIZATION

Administrators can deploy Symantec Backup Exec 2010 deduplication to help optimize both bandwidth and storage using client deduplication, or to help optimize only storage using media server deduplication. The degree of storage optimization can depend on factors such as the type of data, data change rate, and retention period.

Backup Exec client deduplication is supported on Microsoft® Windows® OS–based systems, and requires a Backup Exec client agent on the system to be protected (see Figure 1). This agent enables the data to be deduplicated at this source

client before it is transmitted to the backup target, helping significantly reduce the amount of data that must be sent over the network. In addition, Backup Exec agents are designed to leverage the deduplication of data across all systems under protection in the environment, helping achieve higher levels of deduplication than would otherwise typically be possible.

The bandwidth reductions possible when using Backup Exec client deduplication make this approach well suited for protecting distributed servers or virtual machines with bandwidth (and I/O) constraints. Traditional full backups send all of the data across the network to the backup application, which then directs the data to a media source where it can be compressed or deduplicated; traditional incremental backups, similarly, transmit entire files during the backup process even though only a small portion of a given file may have changed. Deduplicating this data at the source enables Backup Exec to perform full backups that transmit only data that does not already exist on the backup target, and to perform incremental backups that send only changed segments of each modified file rather than the entire file.

In many data centers, bandwidth constraints may not be a concern. In this type of environment, administrators may prefer to use Backup Exec media server deduplication, which performs the deduplication at the storage target after the data has already been sent over the network (see

INTELLIGENT BACKUPS THROUGH SYMANTEC OPENSTORAGE

The Symantec OpenStorage initiative is a disk-based innovation that allows Symantec Backup Exec to take advantage of the advances of intelligent disk storage devices—including storage reduction, backup image duplication, synthetic backups, replication, and energy efficiency. The tight integration with third-party intelligent disk storage devices provides enhanced management of backup images and additional functionality while helping avoid the limitations of tape emulation.

Intelligent disk storage vendors write plug-ins for the OpenStorage application programming interface (API) that integrate with Backup Exec. These plug-ins provide Backup Exec with visibility into the properties and capabilities of the storage devices as well as control over the backup images stored on them. Backup Exec can then treat them as disk devices rather than tape devices, as in the case of virtual tape libraries.

Using OpenStorage plug-ins and the OpenStorage API, Backup Exec can control when backup images are created, duplicated, and deleted, while the intelligent disk storage devices control how the images are actually stored in and replicated between devices. In addition, intelligent disk storage manufacturers in the OpenStorage program can add value to an overall solution through specialized innovations such as backup image deduplication, wide area network (WAN)–optimized backup image replication for disaster recovery, power management, and more.

Figure 2). This approach is typically easier to incorporate into existing backup architectures than client deduplication.[1]

VERSATILE PLATFORM AND APPLICATION PROTECTION

The Backup Exec 2010 Deduplication Option supports data protection for a variety of platforms and applications. Administrators can enable either client or media server deduplication to help protect Microsoft platforms and applications such as Hyper-V™, Exchange, SQL Server®, SharePoint®, and Active Directory® software; VMware® virtual machines (with an agent installed in the guest OS); Symantec Enterprise Vault™, SAP®, IBM® Lotus® Domino, and IBM DB2 software; and Oracle® databases on Microsoft Windows operating systems. It additionally supports media server deduplication for VMware virtual machines using vStorage application programming interfaces (APIs); Linux®, UNIX®, and Novell® NetWare® operating systems; and Oracle databases on Linux operating systems.

EFFICIENT BACKUPS AND DATA PROTECTION

For IT environments where bandwidth is at a premium, combining the Dell

The Dell PowerVault DL2100 – Powered by Symantec Backup Exec provides an integrated backup-to-disk solution that supports flexible deduplication

PowerVault DL2100 – Powered by Symantec Backup Exec with client deduplication through the Backup Exec 2010 Deduplication Option enables bandwidth-efficient, storage-optimized data protection for remote systems and virtualized environments. For environments where bandwidth is not a concern, Backup Exec media server deduplication enables efficient use of storage resources and can help simplify deployment in existing backup infrastructures. Although specific results can vary depending on type of data, change

rate, retention period, and other factors, Symantec Backup Exec 2010 deduplication technology in the Dell PowerVault DL2100 can help organizations create a flexible backup infrastructure to help maximize management and storage cost efficiencies.

Charles Butler is a technical director in the Data Protection Group at Symantec. He has a B.S. in Electrical and Computer Engineering from the University of Colorado at Boulder and an M.B.A. from St. Edward's University.
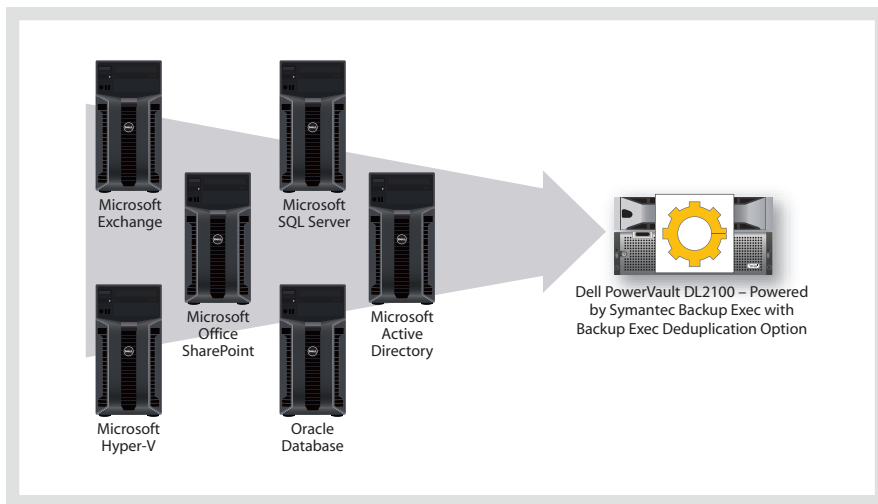

Figure 2. Media server deduplication performed on the Dell PowerVault DL2100 helps optimize storage while simplifying deployment in existing backup infrastructures

---

[1] For more information on types of deduplication and the advantages and disadvantages of different approaches, see "Demystifying Deduplication," by Joe Colucci and Kay Benaroch, in *Dell Power Solutions*, 2010 Issue 1, DELL.COM/Downloads/Global/Power/ps1q10-20100235-Colucci.pdf.