Principled Technologies®

TEST REPORT SUMMARY
MARCH 2010

Dell PowerVault DL2100 powered by Symantec Backup Exec 2010: Source deduplication advantages in a database scenario

Executive summary

Companies with remote locations often back up remote workers' data via wide-area networks (WANs) to avoid the cost of local backups and local support staff. A large portion of the remote data remains constant from day to day, which means backup systems waste WAN and storage resources by transferring multiple copies of the same data to the central office. Source deduplication eliminates this inefficiency by transferring only unique chunks of data across the wire.

Symantec commissioned Principled Technologies (PT) to measure the source deduplication capabilities of the Dell™ PowerVault™ DL2100 powered by Symantec™ Backup Exec™ 2010 when backing up a SQL Server database. As Figure 1 shows, this solution delivered a 10.5 to 1 deduplication ratio. This indicates that by deduplicating the data at the source, the solution reduced the amount of data traveling over the WAN connection by a factor of more than 10, greatly reducing WAN traffic, media server storage capacity needs, and administrative overhead, which all translate to reduced costs.

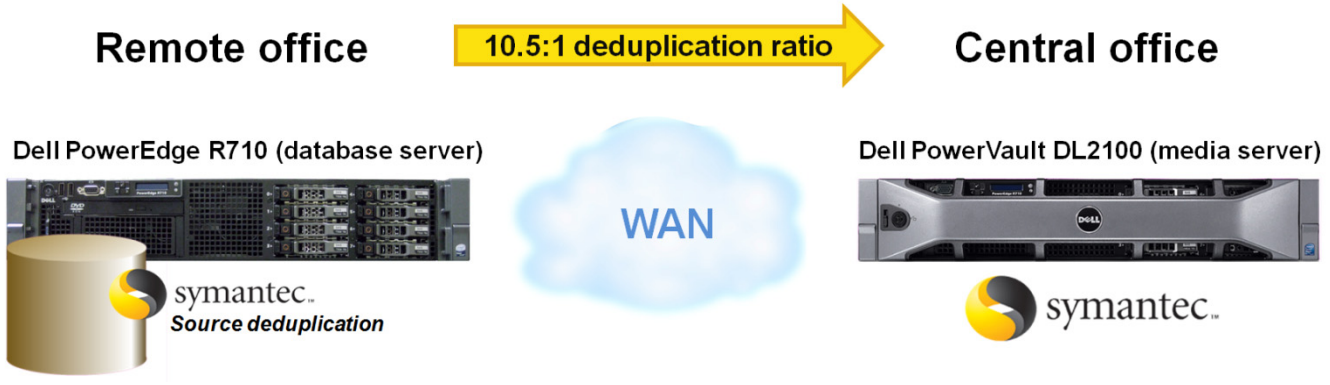

Figure 1: The source deduplication scenario we tested.

KEY FINDINGS

- The Dell PowerVault DL2100, powered by Symantec Backup Exec 2010, backed up the entire SQL Server database in an average of 27 minutes when performing source deduplication as opposed to 2.38 hours when not performing source deduplication.
- The Dell PowerVault DL2100, powered by Symantec Backup Exec 2010 performing source deduplication, achieved an average of 10.5:1 deduplication ratio on our database backup tests.

Symantec Backup Exec 2010 lets businesses easily eliminate duplicate backup data—in our case, SQL Server data. The backup technician uses the same familiar Backup Exec console to simply select the deduplication storage folder and select the option for remote access of the media server while configuring the backup job.

Source deduplication in Symantec Backup Exec 2010 eliminates redundant data directly on the source machine before that data makes its way to the media server. The remote agent compares the chunks of data on the remote server to the existing backup chunks on the media server, and if it finds a duplicate, only sends a marker to the media server. This process greatly reduces the amount of data that is transmitted to the media server and lowers the amount of data the media server must store. Source deduplication enables bandwidth optimization as part of the backup process. In our test case, Symantec Backup Exec 2010 reduced the backup window from 2.38 hours to an average of 27 minutes.

Our testing scenario represents a company using the Dell PowerVault DL2100 appliance to automate and simplify the remote office backup process to ensure consistent and reliable data backup. Because the company is sensitive to the high cost of WAN connections and backup storage, the solution uses source-side deduplication. Figure 2 presents a visual depiction of source deduplication, which eliminates backup inefficiencies by transferring only unique chunks of data across the WAN.

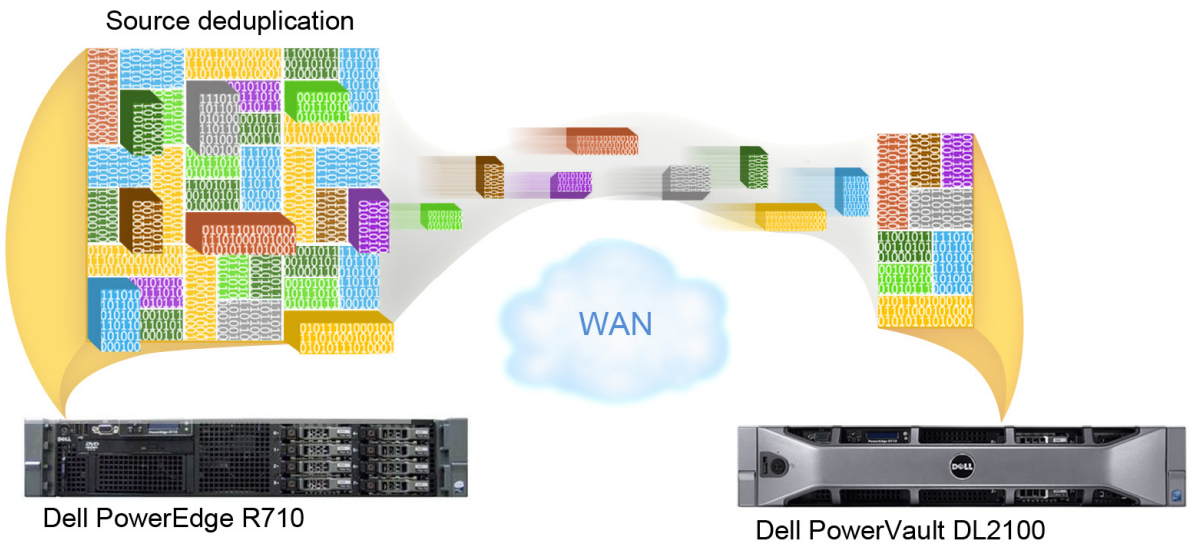

Figure 2. Source deduplication eliminates backup inefficiencies by transferring only unique chunks of data across the WAN.

We used Microsoft® SQL Server® 2008 and the DVD Store benchmark tool to simulate data change rates. The DVD Store tool is an open-source simulation of an online e-commerce DVD store, where customers log in, browse, and order products, thus simulating real-world data set changes in a database server.

Our scenario includes an Active Directory server, a database server running SQL Server 2008, and a client machine on the "remote office" side of the WAN. We set our WAN emulator to use a T3 bandwidth with 15ms latency. We installed a Dell PowerVault DL2100 powered by Symantec Backup Exec 2010 on the "central office" side of the WAN. Based on our assumption that the company would process at least 10,000 orders per day, we ran a DVD Store benchmark workload that produced a minimum of 10,000 orders. We then repeated the backup operation, using full backups each time.

Test results

We used the DVD Store benchmark and SQL Server 2008 to simulate a real-world database backup scenario of a business receiving approximately 10,000 orders per day. We performed an initial full backup of our 20GB database using the Dell PowerVault DL2100 and Symantec Backup Exec 2010. We then proceeded to simulate a total of 14 days of activity, running the workload, then a full SQL Server database backup each time.

For the backup device, we allowed the Dell PowerVault DL2100 to auto-provision the storage for deduplication, and configured the remote agent on the Dell PowerEdge™ R710 database server to have direct access to the Dell PowerVault DL2100 appliance, for the purposes of source deduplication.

For our backup job, we browsed to the database server, selected our database, directed the backup to the deduplication storage folder, and allowed the remote agent direct access. From that point, our backup job used source deduplication. The initial complete backup took just over 3 hours. Subsequent full database backups took an average of just over 25 minutes, achieving a deduplication ratio of at least 10:1 using source deduplication. For

comparison, we ran a single backup using a backup-to-disk storage device on the Dell PowerVault DL2100 without source deduplication. This took roughly 2 hours and 23 minutes. Because this method does not deduplicate the data, it requires greater capacity requirements on the media server and the WAN.

As Figure 3 shows, the difference between the amount of data protected versus the actual amount of data stored becomes greater over time, benefiting the business by requiring significantly less long-term storage capacity.

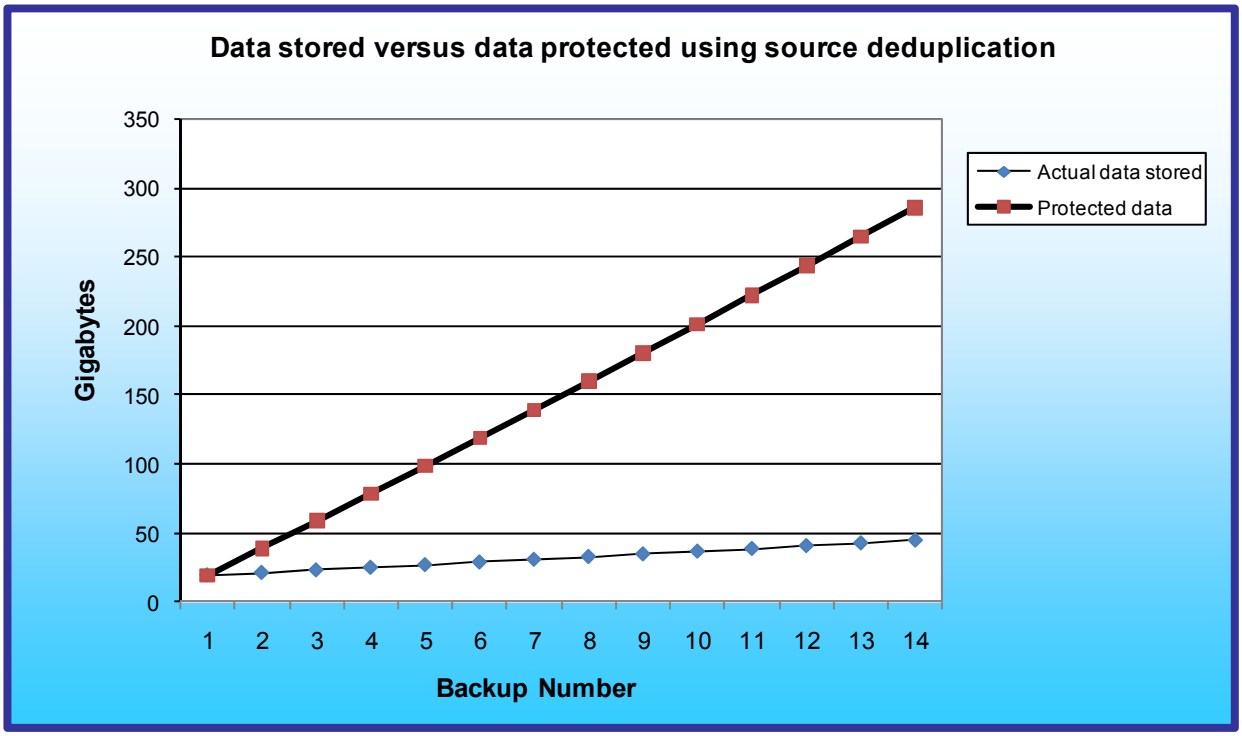

Figure 3. Comparison of protected data versus the actual data stored from our two-week database backup simulation.

In our tests, the Dell PowerVault DL2100 powered by Symantec Backup Exec 2010's ability to eliminate duplicate data in our SQL Server backups greatly reduced the time and overhead costs associated with moving and storing the backup data. We provide more details on the results and system configuration from our test in our full report, which you can read at http://www.principledtechnologies.com/clients/reports/Symantec/db_source_dedupe.pdf.