



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.

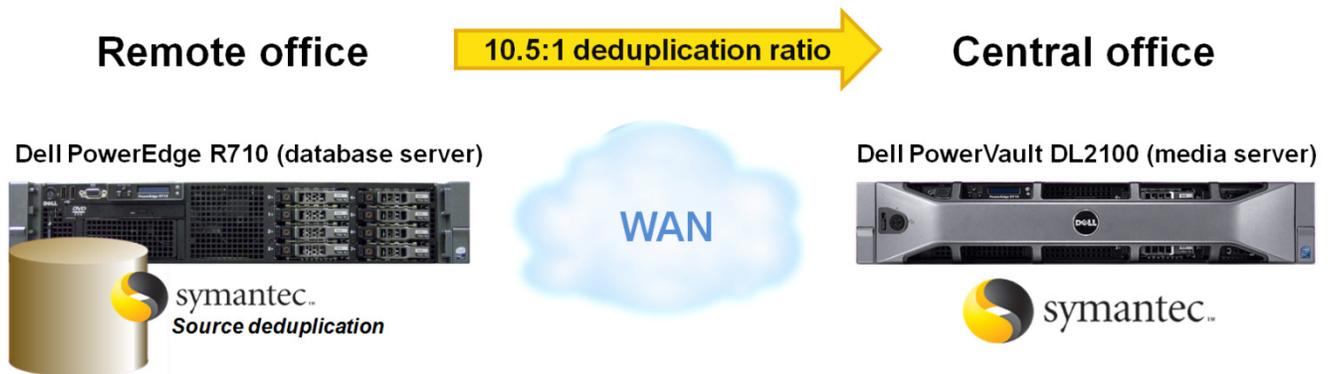


Figure 1: The source deduplication scenario we tested.

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.

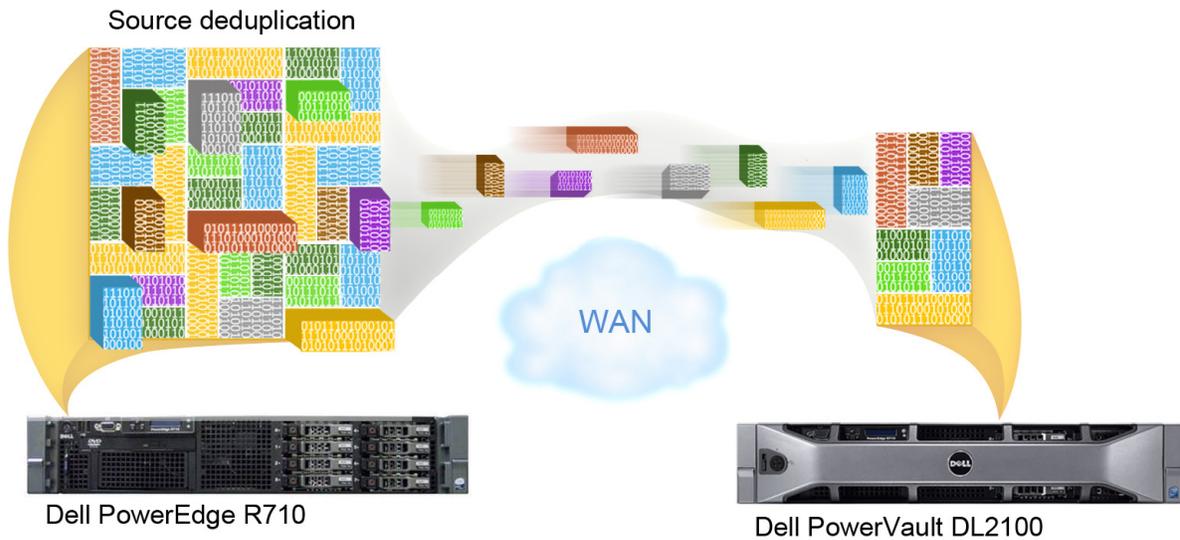


Figure 2. Source deduplication eliminates backup inefficiencies by transferring only unique chunks of data across the WAN.

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. For specific test results, see Appendix B.

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.

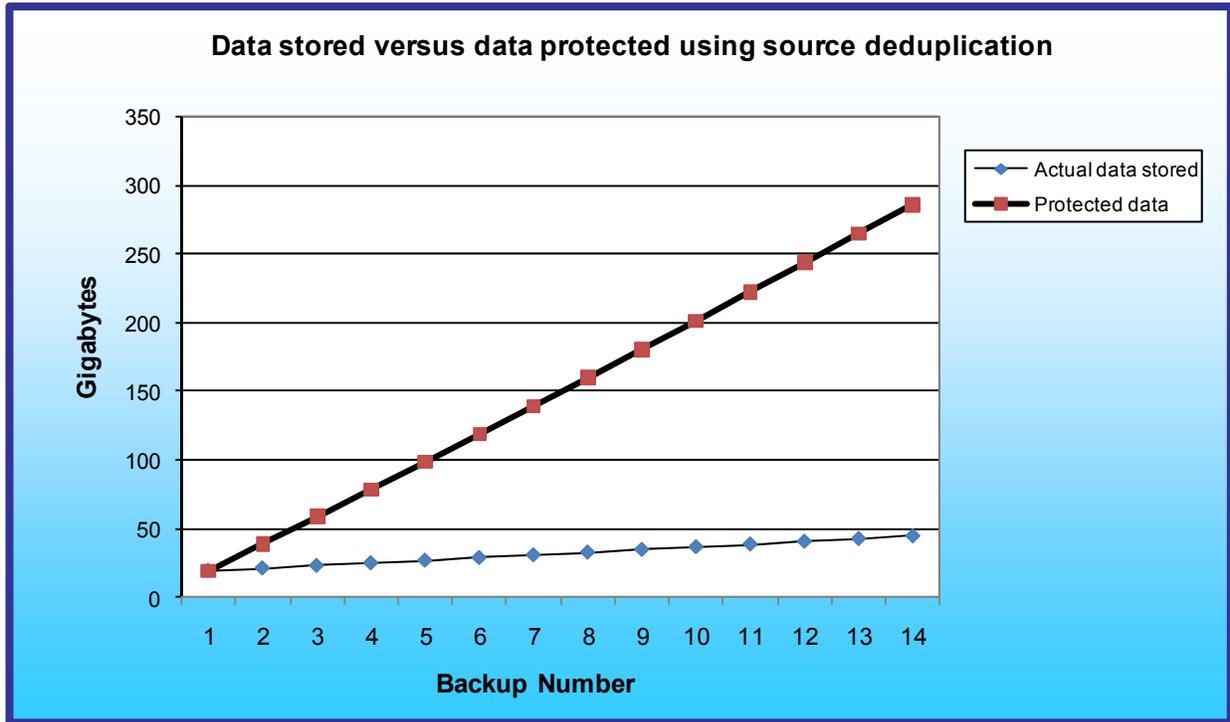


Figure 3. Comparison of protected data versus the actual data stored from our two-week database backup simulation.

Specifically, our scenario contained the following hardware:

- a Dell PowerEdge R710 database server at the remote office running SQL Server 2008 and the Symantec remote agent with source deduplication
- An active directory server and a client machine to simulate the remaining components of the remote office
- a Dell PowerVault DL2100 powered by Symantec Backup Exec 2010 at the central office for backup

Workload DVD Store

To build the workload, we used DVD Store Version 2 (DS2). DS2 is an open-source simulation of an online e-commerce DVD store, where customers log in, browse, and order products. DS2 has database components and Web server components, and includes driver programs that put heavy loads on these components. We used the included driver program to perform the workload only against the database component. The Dell PowerEdge R710 server under test ran a Microsoft SQL Server 2008 workload once per backup period, and we repeated this workload-plus-backup iteration fourteen times to simulate 14 days of activity. One copy of the DS2 database was on the server, with a 20GB database.

The DS2 benchmark’s main metric is orders per minute, or OPM. We used the OPM output by the benchmark to determine when we had reached at least 10,000 orders for the benchmark run. For the purposes of this testing

scenario, we analyzed the deduplication ratio and the elapsed time of each full backup source deduplication turned off, and then on.

We found we could achieve our target daily order rate of 10,000 orders by using specific DVD Store parameters. To mimic a business day where a company fulfilled 10,000 orders, we adjusted the DVD Store parameters to run with five execution threads, each executing with 0.1 seconds of think time. This mix of workload parameters, on average, achieved an orders-per-minute throughput of just over 2,000. Therefore, we ran this workload for at least 5 minutes to simulate the full business day, resulting in at least 10,000 orders. The test placed a moderate to heavy load on the server, modifying realistic amounts of data at the source for the Symantec Backup Exec 2010 software to deduplicate and then backup.

When DS2 is executing, simulated customers log in; browse movies by actor, title, or category; and purchase movies. Each DS2 order consists of a customer login, a number of searches for movies, and a purchase. Each search is by title, actor, or category. The title and actor searches use full-text search. The other customer actions, including adding new customers, exercise a wide range of database functions.

We ran the included driver application on client machines directly via its command-line interface. We used the default DS2 parameters and setup configuration, with the exceptions we note in the DVD Store setup section in the Test methodology section of this report. For more details about the DS2 tool, see <http://www.delltechcenter.com/page/DVD+Store>.

To measure the deduplication ratio and elapsed time from Symantec Backup Exec 2010, we created a backup job, targeted our specific 20GB database on the Dell PowerEdge R710 running SQL Server 2008, selected the deduplication storage folder as the destination, and selected the option to allow the remote agent to have direct access to the media server. We then executed this same backup job between workload iterations to gather the deduplication, job rate, and elapsed time statistics. As a comparison point, we also created an additional backup job, using a backup to disk destination instead with no deduplication. For more details on this process and the results, see the Methodology section and the Test results section.

Test configuration

Networking switch

- Dell PowerConnect™ 5448

Dell PowerVault DL2100 powered by Symantec Backup Exec 2010

- Dell PowerVault DL2100 with 2 x Intel® Xeon® Processor E5520s and 48GB RAM
- 4 x Broadcom BCM5709C Network Interface Cards
- One volume containing 2 x 160GB disks in RAID 1 configuration
- One volume containing 4 x 160GB disks in RAID 5 configuration

Database server

- Dell PowerEdge R710
- 2 x Intel Xeon Processor E5540s
- 4 x Broadcom BCM5709C Network Interface Cards
- 8 x 15K SAS 2.5" drives; 1 x 2-disk RAID 1 and 1 x 6-disk RAID 5
- Windows Server® 2008 R2
- SQL Server 2008

Active Directory server

- Dell PowerEdge 2650
- 2 x Intel Xeon Processor 2.80 GHz
- 2 x Broadcom NetXtreme Gigabit Ethernet NICs
- One volume containing 2 x 36GB SCSI disks in a RAID 1 configuration

DVD Store load-generating clients

- Two workstation-class machines
 - Windows Server 2003 R2 Enterprise
 - .NET Framework 3.5 SP1
 - DVD Store client

WAN emulator

- Dell PowerEdge T605
 - FreeBSD 8.0
 - Compiled with ipfw and dumynet
 - Use the following command line settings to restrict the bandwidth and add delay:

```
ipfw add 1000 pipe 1 ip from any to any
ipfw pipe 1 config delay 2ms bw 45813Kbit
```

Note: ipfw 2ms delay results in 15ms actual round-trip latency.

Figure 4 illustrates our Dell PowerVault DL2100 WAN test bed.

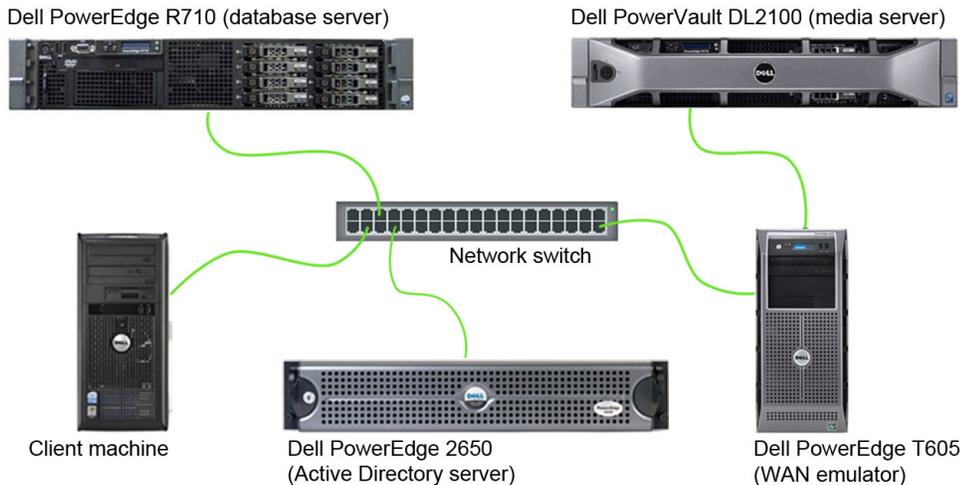


Figure 4: Dell PowerVault DL2100 WAN test bed.

Test methodology

Setting up and configuring the Active Directory server

We used a separate Active Directory server for our test bed.

Setting up the operating system

1. Boot the server, and insert the Windows Server 2003 R2 installation CD in the CD drive.
2. At the Welcome to Set up screen, press Enter to begin the installation.
3. At the Windows Licensing Agreement screen, press F8 to agree to the terms.
4. Press Enter to install in the default location.
5. Select Format the partition using the NTFS file system (Quick), and press Enter to continue.
6. At the Regional and Language Options screen, click Next.
7. Enter Name and Organization, and click Next. Note: Organization is an optional field, while username is required.
8. At the Your Product Key screen, enter your product key, and click Next.
9. At the Licensing Modes screen, click Next.
10. At the Computer Name and Administrator Password screen, type `Password1` as the new password in both fields, and click Next.
11. At the Date and Time Settings screen, change the Date & Time and Time Zone if appropriate, and click Next.
12. After the system restarts automatically, follow the prompt to log into the server.
13. At the Windows Setup screen, the installer will prompt you to insert the second installation CD. Insert the appropriate CD, and click OK.
14. At the Windows Server 2003 R2 Setup Wizard Welcome screen, click Next.
15. At the Windows Server 2003 R2 Setup Wizard Summary screen, click Next.
16. When the installer prompts you, click Finish to end the installation.

Promoting the active directory server using dcpromo

1. Click Start, click Run, type `cmd` in the prompt, and press Enter.
2. Type `dcpromo`, and press Enter.
3. Click Next.
4. Click Next.
5. Click Next.
6. Select Create a new domain in a new forest, and click Next.
7. Type `domainname.local` as the FQDN (Fully Qualified Domain Name), where `domainname` is the name of your domain, and click Next.
8. Set the forest functional level to Windows Server 2003, and click Next.
9. Set the domain functional level to Windows Server 2003, and click Next.
10. Leave defaults on the Additional Domain Controller Options screen, and click Next.
11. If a warning window appears asking about dynamically assigned IP addresses, click Yes.
12. Click Yes to continue.
13. Leave defaults at the Location for Database, Log Files, and SYSVOL screen, and click Next.
14. At the Directory Services Restore Mode Administrator Password screen, type `Password1`.
15. At the summary screen, click Next.
16. Click Finish.
17. Click Restart now.

Setting up and configuring the database server

Installing the operating system

1. Boot the Dell PowerEdge R710 server.
2. When prompted by the boot to enter the storage configuration, press Ctrl+R.
3. Highlight Controller 0 and press F2, and choose Clear Config.
4. Accept the warning of loss of data.
5. Highlight Controller 0 and press F2, and choose Create New VD.
6. Choose RAID 1, select the first two disks, leave other options at the default values, and press OK.

7. Accept the warning regarding initialization.
8. Highlight the Virtual Disk, and press F2, Initialization, Start Init.
9. Highlight Controller 0, press F2, and choose Create New VD.
10. Choose RAID 5, select the remaining six disks, leave other options at the default values, and press OK.
11. Accept the warning regarding initialization.
12. Highlight the new Virtual Disk and press F2, Initialization, Start Init.
13. After initialization completes, press Esc to reboot the machine.
14. Insert the installation DVD for Windows Server 2008 R2 Enterprise x64 into the DVD drive.
15. Choose the language, time and currency, and keyboard input. Click Next.
16. Click Install Now.
17. Choose Windows Server 2008 R2 Enterprise (Full Installation), and click Next.
18. Accept the license terms, and click Next.
19. Click Custom.
20. Click the disk, and click Drive options (advanced).
21. Click New, Apply, Format, and click Next.
22. Let the installation process continue. The server will reboot several times.
23. After the installation completes, click OK to set the Administrator password.
24. Enter the administrator password twice, and click OK.
25. Click Start, Control Panel, and double-click System.
26. Click Change Settings.
27. Click Change.
28. Enter the new computer name, and click OK.
29. Click OK to restart, click Close, and click Restart Now.

Installing Windows updates on the database server

1. Click Start, type `update`.
2. Click View Advanced Options.
3. Click Never Check for Updates.
4. Click OK.
5. Click Check for Updates.
6. When the application prompts you to install an update for Windows Update, click Install Now.
7. Click View available updates to ensure you install the necessary updates.
8. Click Install.
9. Agree to the license terms, and click Finish.
10. When the application requests you do so, reboot the machine.
11. After reboot, check Windows Update again for further updates. Click Start, Windows Update.
12. Click Check for updates.
13. If Windows lists no important updates, close the updates window.

Joining the domain

1. Click Start, right-click Network, and choose Properties.
2. Click Change adapter settings.
3. Right-click the relevant network adapter, and choose Properties.
4. Click Internet Protocol Version 4 (TCP/IPv4).
5. Click properties.
6. Click Use the following IP address.
7. Enter a valid IP address, Subnet mask.
8. Click Use the following DNS server addresses.
9. Enter the IP address of the Active Directory Server.
10. Click OK.
11. Click Close.
12. Click Start, right-click Computer, and choose Properties.
13. Click change settings.
14. Click change.
15. Choose Domain, and enter the domain name.

16. Enter valid credentials, and click OK.
17. When prompted by Windows to restart, restart the machine.

Installing SQL Server 2008

1. Insert the DVD into the DVD drive.
2. Click Run SETUP.EXE on the AutoPlay menu.
3. If the application prompts you to install the .NET Framework, click OK.
4. Select the Accept radio button, and click Install.
5. When the installation finishes, click Exit.
6. If any warnings appear about compatibility, accept them as installing SP1 in the next section should solve them.
7. If the application prompts you, click OK to install a hotfix for Windows (KB942288).
8. Click Restart now to restart the server upon completion.
9. After rebooting, log into the server.
10. Click Start, and click Computer.
11. Double-click the CD/DVD drive.
12. Click Installation.
13. Click New SQL Server stand-alone installation.
14. On the Setup support rules screen, click OK.
15. Specify the Enterprise Evaluation free edition. Click Next.
16. Click the checkbox to accept the license terms. Click Next.
17. Click Install to install the setup support files.
18. You may see a Windows Firewall warning. For now, ignore this, and click Next.
19. On the Feature Selection screen, select Database Engine Services, Full-Text Search, Client Tools Connectivity, Client Tools Backward Compatibility, Management Tools –Basic, and Management Tools – Complete. Click Next.
20. On the Instance configuration screen, leave the default selection of default instance, and click Next.
21. On the Disk space requirements screen, click Next.
22. On the Server configuration screen, the domain credentials for the SQL Server Agent, and enter the domain credentials for the SQL Server Database Engine. Click Next.
23. On the Database Engine Configuration screen, select Mixed Mode.
24. Enter a password for the system administrator (sa) account.
25. Click Add Current user. This may take several seconds.
26. Click Next.
27. On the Error and usage-reporting screen, click Next.
28. On the Installation rules screen, click Next.
29. On the Ready to Install screen, click Next.
30. After installation completes, click Start, All Programs, Microsoft SQL Server 2008, Configuration Tools, SQL Server Configuration Manager.
31. Expand the Network Configuration node, and select the relevant instance.
32. In the right pane, right-click TCP/IP, and choose Enable.
33. In the left pane, click SQL Server Services.
34. In the right pane, right-click the SQL Server service, and select Restart.
35. Create a SQL Server login for the ds2user (see the DVD Store setup section below for the specific script to use).
36. Place the DVD Store backup file you prepared in the default location for SQL Server: C:\Program Files\Microsoft SQL Server\MSSQL10.MSSQLSERVER\MSSQL\Backup.

Installing SQL Server 2008 SP1

Service Pack 1 is available for download from

<http://www.microsoft.com/downloads/details.aspx?FamilyID=66ab3dbb-bf3e-4f46-9559-ccc6a4f9dc19&displaylang=en>.

1. Run SQLServer2008SP1-KB968369-x64-ENU.exe.
2. On the Welcome screen, click Next
3. On the License terms screen, accept the license terms, and click Next.
4. On the Features screen, select available features, and click Next.

5. On the Check files in use screen, click Next.
6. On the Update screen, click Next.
7. When the update completes, click Next, and click Close.

Configuring additional drives in Disk Management

1. Click Start, Administrative Tools, Computer Management.
2. Click Disk Management.
3. Right-click the uninitialized disks (if any), and choose Initialize Disk.
4. Choose MBR.
5. Right-click the volume meant for SQL server data, and choose New Simple Volume. Click Next.
6. Accept the default volume size. Click Next.
7. Choose E: from the drop-down menu for the drive letter. Click Next.
8. Type `sqldata` for the volume name. Click Next.
9. Choose NTFS, 64K allocation unit size, and check the Perform a quick format checkbox. Click Next.
10. Click Finish.

Setting up and configuring the Dell PowerVault DL2100 powered by Symantec Backup Exec 2010

Setting up the operating system

1. Power on the Dell PowerVault DL2100 appliance.
2. At the license agreement screen, accept license agreements, and click Next.
3. At the Set up Windows screen, click Start.
4. At the default password warning, click OK.
5. At the welcome screen, click Next.
6. At the locate Symantec licenses screen, click Next.
7. At the Storage connections screen, click Next.
8. At the Change default local administrator password screen, enter the new password. Enter the password again for confirmation, and click Next.
9. Select the network interface you wish to configure, and click Next.
10. Enter network configuration information, and click Next.
11. Enter a new hostname. Join the domain if your configuration requires. Click Next.
12. At the SNMP settings screen, click Next.
13. At the Disk configuration policy screen, choose Automatic with confirmation (default), and click Next.
14. Under the deduplication scenario, choose Internal storage controller (default), and click Next. Under the Backup to Disk scenario, choose to not configure deduplication at this point.
15. Click Finish. After the configuration completes, click Reboot. The system will reboot.
16. After reboot, press Ctrl+Alt+Delete to login.
17. Click Switch User, and click Other User.
18. Log in as a domain user with administrative credentials.
19. Wait a few minutes for the Configuration Wizard to continue.
20. At the Symantec Backup Exec 2010 welcome screen, click Next.
21. Accept license agreements, click Next.
22. At the license key screen, enter license keys, and click Next.
23. On the components screen, choose desired components, and click Next.
24. On the user credentials screen, enter the password, and click Next.
25. On the summary screen, click Install.
26. After installation completes, click Finish.
27. After configuration of the appliance completes, click Exit. Do not shut down the appliance until automatic disk provisioning is complete. You can check the status of this operation on the Appliance Console.
28. On the PowerVault DL Backup To Disk Appliance Console, click Launch Symantec Backup Exec.
29. Click OK when the application prompts you regarding the disk configuration changes.
30. On the Device Warning screen, click OK when the application informs you of the configuration of the deduplication folder.
31. Open Backup Exec 2010, click Tools, Backup Exec Services, and click Restart All Services.
32. Click OK upon completion of the services being restarted.

33. Click Tools, Options, and click Device and Media.
34. Select the Deduplication Storage Folder 1 storage device as the default storage location.
35. Open Backup Exec 2010, and click Tools, LiveUpdate to download the latest updates for the Backup Exec 2010.

Installing the Symantec Remote Agent on the database server

1. Log into the database server.
2. Map a drive to the C:\ root share of the Dell PowerVault DL2100 appliance. Use command line text as follows:

```
net use x: \\dl2100hostname\c$
```

3. While in the context of the mapped drive, navigate to X:\Program Files\Symantec\Backup Exec\Agents\RAWXS64 using the following command line text:

```
cd X:\Program Files\Symantec\Backup Exec\Agents\RAWXS64
```

4. Execute the setupaax64.cmd file to install the remote agent.

Setting up the Remote Agent Direct Access

1. Open Symantec Backup Exec 2010.
2. Click the Devices tab.
3. Click the Configure devices assistant.
4. Click Remote Agent with Direct Access.
5. Provide the fully qualified server name of the server running the agent you wish to provide direct access to.
6. Click OK.

Setting up the Backup to Disk folder (only when using the backup-to-disk scenario)

1. Open Symantec Backup Exec 2010.
2. Click the Devices tab.
3. Click the Configure devices assistant.
4. Click Backup to Disk folder.
5. If the application asks about configuring additional storage devices, click Yes.
6. Click Next.
7. Name the backup to disk folder.
8. Provide the path to the folder, and click OK.
9. Click Next.
10. On the Maximum Size screen, choose No, and click Next.
11. Type 200 for the maximum size (GB) for the backup-to-disk file. This is adjustable.
12. Click Next.
13. Type 250 for the maximum number of backup sets per backup-to-disk file.
14. On the concurrent jobs screen, click Next.
15. On the low disk space threshold screen, enter the low space threshold for your environment, and click Next.
16. Click Next to finish the configuration.

Configuring the backup job

1. Open Symantec Backup Exec 2010.
2. Click the Job Setup tab.
3. Click New Job.
4. Browse Domains→Active Directory Domains→Domainname→Servername, and select Microsoft SQL Server.
5. Highlight the desired database.
6. Under Destination, click Device and Media.

7. When running a deduplication scenario, select Deduplication Storage Folder. When running a backup to disk scenario, select the Backup To Disk device.
8. When running a deduplication scenario, ensure the Allow this job to have direct access to the device checkbox is checked. This enables source deduplication.
9. Under Settings, click General.
10. Uncheck the Verify after backup completes checkbox.
11. Under Settings, click Microsoft SQL Server.
12. Under Consistency check before backup, choose None.
13. Click Run Now.

Setting up and configuring the load-generating client

For the DVD Store scripts, we used a workstation client machine to simulate a number of users putting a load on the server. For our client, we installed the .NET 3.5 framework on the client, as the DVD Store test executable requires this. We created a performance counter log on the client machine to track the number of orders per minute (OPM) the database server returned. We followed this process for the installation:

1. Install Microsoft Windows Server 2003 R2 Enterprise x86 Edition Service Pack 2 on the client.
2. Assign a computer name of `Client1` for the database client.
3. For the licensing mode, use the default setting of five concurrent connections.
4. Enter a password for the administrator log on.
5. Select Eastern Time Zone.
6. Use typical settings for the Network installation.
7. Type `Workgroup` for the workgroup.
8. Install Windows Updates, .NET 3.5 framework, and copy the DVD Store client executable onto the machine.

Setting up DVD Store

Data generation overview

We built the database schema using the scripts in the DS2 distribution package, though we made a few minor modifications. The DS2 stress tool provides options to generate 10MB, 1GB, or 100GB datasets. To get the tool to generate the 20 GB of user data we used in this test, we had to make a few straightforward changes to the source code and to the DVD Store application's scripts. Note: We created our test data on a Linux system to take advantage of the larger RAND MAX.

Editing the `ds2_create_orders.c` module

The module `ds2_create_orders.c` defines constants that define the maximum values for the customer ID and the product ID. The constants for the 20GB database size did not exist. We added the constants for this size.

On the command line for the `ds2_create_orders.c` module, we specified the size. The available options were S (small), M (medium), and L (large). We added the case W for the 20GB database. In the switch statement that sets the values for the variables `max_cust_id` and `max_prod_id`, we added cases that assigned them the proper values for the 20GB database size.

We recompiled the `ds2_create_orders.c` module on Linux, following the instructions in the header comments. We used the following command line:

```
gcc -o ds2_create_orders ds2_create_orders.c -lm
```

Editing the `ds2_create_cust.c` module

We had to make the same changes to the `ds2_create_cust.c` module that we made to the `ds2_create_orders.c` module. On the command line for the `ds2_create_cust.c` module, we specified the size. The available options were S (small), M (medium), and L (large). We added the case W for the 20GB database. In the switch statement that sets the values for the variables `max_cust_id` and `max_prod_id`, we added cases that assigned them the proper values for the 20GB database size.

We recompiled the `ds2_create_cust.c` module on Linux, following the instructions in the header comments. We used the following command line:

```
gcc -o ds2_create_cust ds2_create_cust.c -lm
```

Generating the data for the 20GB database

We used shell scripts to run all four of the executables that generate the data. The distribution did not include shell scripts for the 20GB size. We wrote shell scripts based on the `ds2_create_cust_large.sh` and `ds2_create_orders_large.sh` scripts. The `ds2_create_prod` and `ds2_create_inv` executables did not ship with associated shell scripts, so we created shell scripts using the instructions in the readme files. We ran the shell scripts in the following order to generate the data for the 20GB database:

1. `ds2_create_orders_20gb.sh`
2. `ds2_create_inv_20gb.sh`
3. `ds2_create_prod_20gb.sh`
4. `ds2_create_cust_20gb.sh`

We waited until the processes finished before we moved onto the next step.

Creating the database

We modified the database creation SQL Server scripts in the DVD Store distribution package to build the database schema, which includes the file structure, tables, indices, stored procedures, triggers, and so on. We built a master copy of the 20GB database. We used that master backup file to initiate our Symantec Backup Exec backup simulations. We stored the backup file on the C: drive of the database server for quick access. Subsequent backups via Symantec Backup Exec 2010 were stored on the Dell PowerVault DL2100 appliance.

We followed these steps to create the database:

1. We created the database and file structure using database creation scripts in the DS2 download. We made size modifications specific to our 20GB database and the appropriate changes to drive letters.
2. We created database tables, stored procedures, and objects.
3. We set the database recovery model to bulk-logged to prevent excess logging.
4. We loaded the data we generated into the database. For data loading, we used the import wizard in SQL Server Management Studio. Where necessary, we retained options from the original scripts, such as Enable Identity Insert.
5. We created indices, full-text catalogs, primary keys, and foreign keys using the database-creation scripts.
6. We updated statistics on each table according to database-creation scripts, which sample 18 percent of the table data.
7. On each SQL Server instance, we created a `ds2user` SQL Server login using the following Transact SQL (TSQL) script:

```
USE [master]
GO
CREATE LOGIN [ds2user] WITH PASSWORD=N'',
DEFAULT_DATABASE=[master],
DEFAULT_LANGUAGE=[us_english],
CHECK_EXPIRATION=OFF,
CHECK_POLICY=OFF
GO
```

8. We set the database recovery model back to full.

We made the following changes in the build scripts:

- Because we varied the size of the datasets, we sized the files in our scripts to reflect the database size and the number of files per filegroup. We allowed for approximately 40 percent free space in our database files to ensure that filegrowth activity did not occur during the testing.
- We followed Microsoft's recommendation of having 0.25 to 1 file per filegroup per logical processor. We used eight files per filegroup on the test server.

- We did not use the DBCC PINTABLE command for the CATEGORIES and PRODUCTS tables, both because Microsoft recommends against this practice and because the commands do nothing in SQL Server 2008.
- We added the FORCESEEK hint in the BROWSE_BY_ACTOR stored procedure because in initial testing we found that SQL Server 2008 was choosing an index scan instead of the preferred index seek, adding unnecessary overhead. We created a SQL Server login we called `ds2user` and mapped a database user to this login. We made this user a member of the `db_owner` fixed database role.
- Using the DVD Store scripts as a reference, we created the full-text catalog and index on the PRODUCTS table manually in SQL Server Management Studio.

Editing the workload script - `ds2xdriver.cs` module

To use the 20GB database we created earlier, we had to change the following constants:

- In the routine `Controller()`, we changed the string sizes. We added the `W` option for the 20GB database size. DS2 uses the sizes string to interpret the `db_size_str` option.
- In the class `Controller`, we changed the arrays `MAX_CUSTOMER` and `MAX_PRODUCT`. To each, we added values specifying the bounds for the customer and product IDs. The `Controller()` routine uses these arrays.
- We added a command line parameter for the database name:

```
-database_name
```

Editing the workload script - `ds2sqlserverfns.cs` module

We changed the connection string to increase the number of available connections, to not use the default administrator (“sa”) account, and to include a parameter for the database name. We raised the available connections limit from the default of 100 to 200 to allow room for experimentation. We created a user account we called `ds2User` and used that account.

The `ds2connect` routine in the `ds2sqlserverfns.cs` module defines `sConnectionString`. We used the following string; the changes we made appear in bold:

```
string sConnectionString = "User ID=ds2User;Initial Catalog=+" + dbname + "";Max  
Pool Size=200;Connection Timeout=120;Data Source=" + Controller.target;
```

Recompiling the `ds2sqlserverdriver.exe` executable

We recompiled the `ds2xdriver.cs` and `ds2sqlserverfns.cs` module in Windows by following the instructions in the header comments. Because the DS2 instructions were for compiling from the command line, we used the following steps:

1. Open a command prompt.
2. Use the `cd` command to change to the directory containing our sources.
3. Run the batch file `C:\Program Files\Microsoft Visual Studio 9.0\Common7\Tools\vsvars32.bat`. This sets up the environment variables for us.
4. Execute the following command:

```
csc /out:ds2sqlserverdriver.exe ds2xdriver.cs ds2sqlserverfns.cs  
/d:USE_WIN32_TIMER /d:GEN_PERF_CTRS
```

Performing the test

To perform the test, we created a backup job in the Symantec Backup Exec 2010 console, then used a combination of batch files, the `psexec` utility, and the `bemcmd.exe` utility to execute our workload and backup jobs from our client machine.

The testing procedure consisted of the following general steps:

1. In SQL Server Management Studio, restore the initial backup of the SQL Server DVD Store database.

2. Create a backup job in Symantec Backup 2010 console, choosing the deduplication storage folder as the destination, and also selecting the option to allow the remote agent to have direct access to the Dell PowerVault DL2100.
3. Run the backup job to initially backup the database to the media server.
4. Execute the DVD Store workload to simulate one business day of activity, using 5 threads, 0.1 think time, and run for 5 minutes.
5. Rerun the backup job in Symantec Backup Exec 2010.
6. Execute the DVD Store workload to simulate one business day of activity, using 5 threads, 0.1 think time, and run for 5 minutes.
7. Repeat steps 5 and 6 through the end of testing.
8. Export the job details results data.

For the backup to disk scenario, we performed similar tasks as above, but targeted the backup destination to the backup to disk device, as opposed to the deduplication storage folder device.

Appendix A – Test server information

Figure 5 presents detailed information for the test servers we used in this report.

Servers	Dell PowerEdge R710	Dell PowerVault DL2100
General dimension information		
Height (inches)	3.50	3.50
Width (inches)	17.50	17.50
Depth (inches)	27.00	27.00
U size in server rack (U)	2	2
General processor setup		
Number of processor packages	2	2
Number of cores per processor package	4	4
Number of hardware threads per core	2	2
CPU		
Vendor	Intel	Intel
Name	Xeon E5540	Xeon E5520
Stepping	D0	D0
Socket type	LGA1366	LGA1366
Core frequency (GHz)	2.53	2.26
L1 cache	32 KB + 32 KB (per core)	32 KB + 32 KB (per core)
L2 cache	1 MB (4 x 256 KB)	1 MB (4 x 256 KB)
L3 cache (MB)	8	8
Platform		
Vendor and model number	Dell PowerEdge R710	Dell PowerVault DL2100
Motherboard model number	00W9X3	0M233H
Motherboard revision number	A03	X02
BIOS name and version	Dell 1.3.6	Dell 1.3.6
Memory modules		
Total RAM in system (GB)	24	48
Vendor and model number	Micron MT36JSZF51272PY-1G4D1AB	Samsung M393B1K70BH1-CF8
Type	PC3-10600 DDR3	PC3-8500 DDR3
Speed (MHz)	1,333	1,067
Speed in the system currently running @ (MHz)	1,067	1,067
Timing/latency (tCL-tRCD-iRP-tRASmin)	7-7-7-20	7-7-7-20
Size (GB)	24	48
Number of RAM modules	6 x 4 GB	6 x 8 GB
Chip organization	Double-sided	Double-sided

Servers	Dell PowerEdge R710	Dell PowerVault DL2100
Hard disk		
Vendor and model number	Seagate ST973451SS	Western Digital WD1602ABKS
Number of disks in system	8	6
Size (GB)	73	160
Buffer size (MB)	16	8
RPM	15,000	7,200
Type	SAS	SATA
Controller	Dell PERC 6/i Integrated RAID Controller	Dell PERC 6/i Integrated RAID Controller
Driver version	LSI 4.5.0.64 (02/06/2009)	Dell 2.24.0.64 (01/19/2009)
Operating system		
Name	Windows Server 2008 R2 Enterprise	Windows Server 2008 Standard
Build number	7600	6002
File system	NTFS	NTFS
Language	English	English
Network card/subsystem		
Vendor and model number	Broadcom BCM5709C NetXtreme II	Broadcom BCM5709C NetXtreme II
Type	Integrated	Integrated
Driver version	Microsoft 4.8.4.1 (12/23/2008)	Broadcom 5.0.15.0 (10/2/2009)
Optical drive		
Vendor and model number	TEAC DV28SV DVD-ROM	PLDS DVD+-RW DS-8A3S
USB ports		
Number	4	4
Type	2.0	2.0

Figure 5: Detailed system configuration information for the two test servers.

Appendix B – Detailed test results

Figures 6 and 7 present detailed results of our tests using source deduplication and our test using a backup-to-disk folder.

Dell PowerVault DL2100 powered by Symantec Backup Exec 2010 <i>with source deduplication</i>				
Backup number	Job time (HH:MM:SS)	Job rate (MB/min)	Deduplication ratio	Total backup size (GB)
1	03:10:11	104	1.0:1	19.25
2	00:21:19	1,014	11.1:1	19.49
3	00:22:20	974	10.6:1	19.71
4	00:39:52	534	10.5:1	19.91
5	00:22:50	970	10.5:1	20.09
6	00:41:22	523	10.3:1	20.26
7	00:41:33	525	10.3:1	20.42
8	00:23:35	961	10.3:1	20.58
9	00:23:22	977	10.5:1	20.72
10	00:23:52	985	10.5:1	20.85
11	00:23:32	983	10.5:1	20.98
12	00:23:47	977	10.5:1	21.10
13	00:23:49	981	10.5:1	21.21
14	00:23:42	992	10.7:1	21.32

Figure 6. Backup metrics when using source deduplication.

Dell PowerVault DL2100 powered by Symantec Backup Exec 2010 <i>without source deduplication</i>		
Job time (HH:MM:SS)	Job rate (MB/min)	Total backup size (GB)
02:22:59	138	19.25

Figure 7: Backup metrics when using no source deduplication.

About Principled Technologies

We provide industry-leading technology assessment and fact-based marketing services. We bring to every assignment extensive experience with and expertise in all aspects of technology testing and analysis, from researching new technologies, to developing new methodologies, to testing with existing and new tools.

When the assessment is complete, we know how to present the results to a broad range of target audiences. We provide our clients with the materials they need, from market-focused data to use in their own collateral to custom sales aids, such as test reports, performance assessments, and white papers. Every document reflects the results of our trusted independent analysis.

We provide customized services that focus on our clients' individual requirements. Whether the technology involves hardware, software, Web sites, or services, we offer the experience, expertise, and tools to help you assess how it will fare against its competition, its performance, whether it's ready to go to market, and its quality and reliability.

Our founders, Mark L. Van Name and Bill Catchings, have worked together in technology assessment for over 20 years. As journalists, they published over a thousand articles on a wide array of technology subjects. They created and led the Ziff-Davis Benchmark Operation, which developed such industry-standard benchmarks as Ziff Davis Media's Winstone and WebBench. They founded and led eTesting Labs, and after the acquisition of that company by Lionbridge Technologies were the head and CTO of VeriTest.



Principled Technologies, Inc.
1007 Slater Rd., Suite 250
Durham, NC 27703
www.principledtechnologies.com
info@principledtechnologies.com

Principled Technologies is a registered trademark of Principled Technologies, Inc.
All other product names are the trademarks of their respective owners.

Disclaimer of Warranties; Limitation of Liability:

PRINCIPLED TECHNOLOGIES, INC. HAS MADE REASONABLE EFFORTS TO ENSURE THE ACCURACY AND VALIDITY OF ITS TESTING, HOWEVER, PRINCIPLED TECHNOLOGIES, INC. SPECIFICALLY DISCLAIMS ANY WARRANTY, EXPRESSED OR IMPLIED, RELATING TO THE TEST RESULTS AND ANALYSIS, THEIR ACCURACY, COMPLETENESS OR QUALITY, INCLUDING ANY IMPLIED WARRANTY OF FITNESS FOR ANY PARTICULAR PURPOSE. ALL PERSONS OR ENTITIES RELYING ON THE RESULTS OF ANY TESTING DO SO AT THEIR OWN RISK, AND AGREE THAT PRINCIPLED TECHNOLOGIES, INC., ITS EMPLOYEES AND ITS SUBCONTRACTORS SHALL HAVE NO LIABILITY WHATSOEVER FROM ANY CLAIM OF LOSS OR DAMAGE ON ACCOUNT OF ANY ALLEGED ERROR OR DEFECT IN ANY TESTING PROCEDURE OR RESULT.

IN NO EVENT SHALL PRINCIPLED TECHNOLOGIES, INC. BE LIABLE FOR INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH ITS TESTING, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. IN NO EVENT SHALL PRINCIPLED TECHNOLOGIES, INC.'S LIABILITY, INCLUDING FOR DIRECT DAMAGES, EXCEED THE AMOUNTS PAID IN CONNECTION WITH PRINCIPLED TECHNOLOGIES, INC.'S TESTING. CUSTOMER'S SOLE AND EXCLUSIVE REMEDIES ARE AS SET FORTH HEREIN.