

## Symantec Backup Exec 2010: Source deduplication advantages in database server, file server, and mail server scenarios

### Executive summary

Symantec commissioned Principled Technologies (PT) to measure the source deduplication capabilities of Symantec™ Backup Exec™ 2010 when backing up a Microsoft® SQL Server® 2008 database server, a Microsoft Windows®-based file server, and a Microsoft Exchange Server 2007 mail server.

Companies with remote locations often back up remote workers' data via wide-area networks (WANs) to avoid the high costs of local backups and local support staff. With the source deduplication capabilities of Symantec Backup Exec 2010, companies can now make their remote backup scenario even more efficient and affordable. Because a large portion of the remote data remains constant from day to day, 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 from a remote office to the central office. A company using Symantec Backup Exec can reduce costs by reducing WAN traffic, lessening media server storage capacity needs, and curtailing administrative overhead.

### KEY FINDINGS

- Symantec Backup Exec 2010, performing source deduplication, achieved an average of 10.5 to 1 deduplication ratio on our database backup tests, an average of 24.6 to 1 deduplication ratio on our file server backup tests, and an average of 10.8 to 1 deduplication ratio on our mail server backup tests.
- Symantec Backup Exec 2010 reduced database backup time from over 2.38 hours to an average of only 27 minutes while performing source deduplication, reduced file server backup time from over 6 hours to an average of only 22 minutes while performing source deduplication, and reduced mail server backup time from over 4 hours to an average of only 2 hours 29 minutes while performing source deduplication.

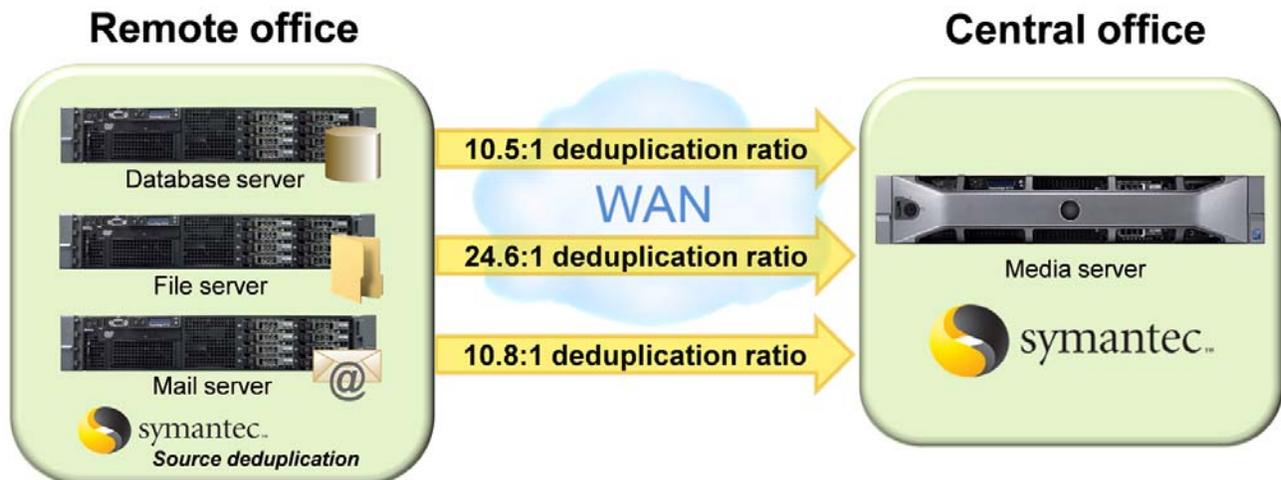


Figure 1: The source deduplication scenarios we tested.

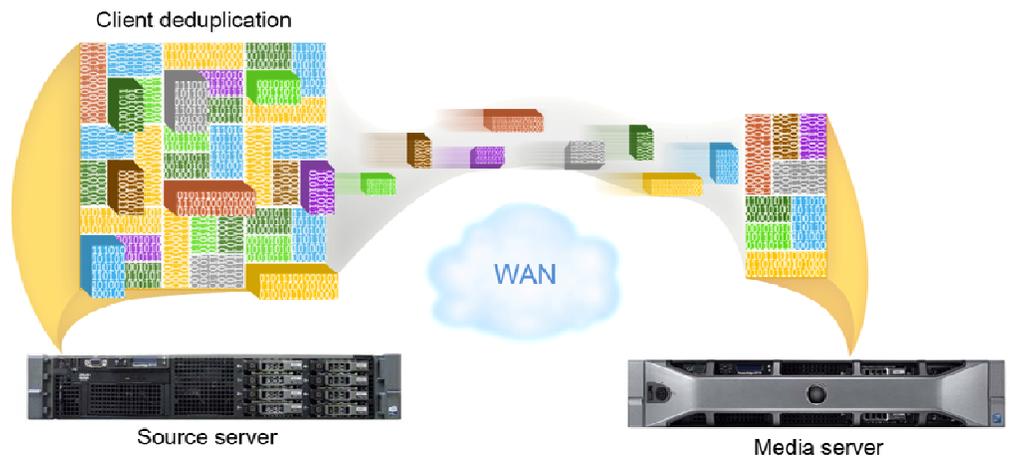
As Figure 1 shows, Symantec Backup Exec 2010 delivered a 10.5 to 1 deduplication ratio when backing up a SQL Server database server, delivered a 24.6 to 1 deduplication ratio when backing up a Windows-based file server, and delivered a 10.8 to 1 deduplication ratio when backing up an Exchange 2007 mail server. By deduplicating the data at the source, the deduplication solution reduced the amount of data traveling over the WAN by a factor of more than 10 when backing up a database server, a factor of more than 24 when backing up a file server, and a factor of more than 10 when backing up a mail server. In our test cases, Symantec Backup

Exec 2010 reduced the database backup window from 2.38 hours to an average of 27 minutes, reduced the file server backup window from over 6 hours to an average of 22 minutes, and reduced the mail server backup window from over 4 hours to an average of 2 hours 29 minutes.

Symantec Backup Exec 2010 lets businesses easily eliminate duplicate backup data, including SQL Server data, files stored on a Windows-based file server, and Exchange 2007 data on a mail 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 at the central office. This process greatly reduces the amount of data 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 a straightforward process, the backup technician uses the same familiar Symantec 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.

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

In our database scenario, 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, simulating real-world data set changes in a database server. Based on our



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

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.

In our file server scenario, we used a file server running Microsoft Windows 2008 R2 and a custom workload to simulate data change rates. Our custom workload browsed through the specified directory on the file server, which contained 50 GB of data, updating, deleting, and adding files, simulating real-world data set changes in a file server. Our custom workload changed 3 percent of Microsoft PowerPoint®, Excel®, and Word® files, deleted 1 percent of all files, moved 2 percent of all files, and added between 160 and 200 files per iteration.

In our mail server scenario, we used a mail server running Microsoft Windows Server® 2008 x64 and Microsoft Exchange Server 2007, along with a client running Microsoft Exchange Load Generator (LoadGen) to simulate data change rates. We configured our Microsoft LoadGen workload with an average load use case and 300 users to simulate a medium-sized business's mail server activity.

The three scenarios included an Active Directory® server, a client machine on the "remote office" side of the WAN, and the application-specific server, which was either the database server, file server, or mail server. We set our WAN emulator to use a T3 bandwidth with 15ms latency. We installed a media server powered by Symantec

Backup Exec 2010 on the “central office” side of the WAN. On all three servers, we performed a full backup of the relevant data, and then we executed the workload and backup operations, using full backups each time.

In all three scenarios, we allowed the media server to auto-provision the storage for deduplication, and configured the remote agent on the database server, file server, or mail server to have direct access to the media server, for the purposes of source deduplication.

## Test results

### Database server backup scenario

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 Symantec Backup Exec 2010. We then proceeded to simulate 13 additional days of activity, running the DVD Store workload and then completing a full SQL Server database backup each time.

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 database 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 media server without source deduplication, which took roughly 2 hours and 23 minutes.

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

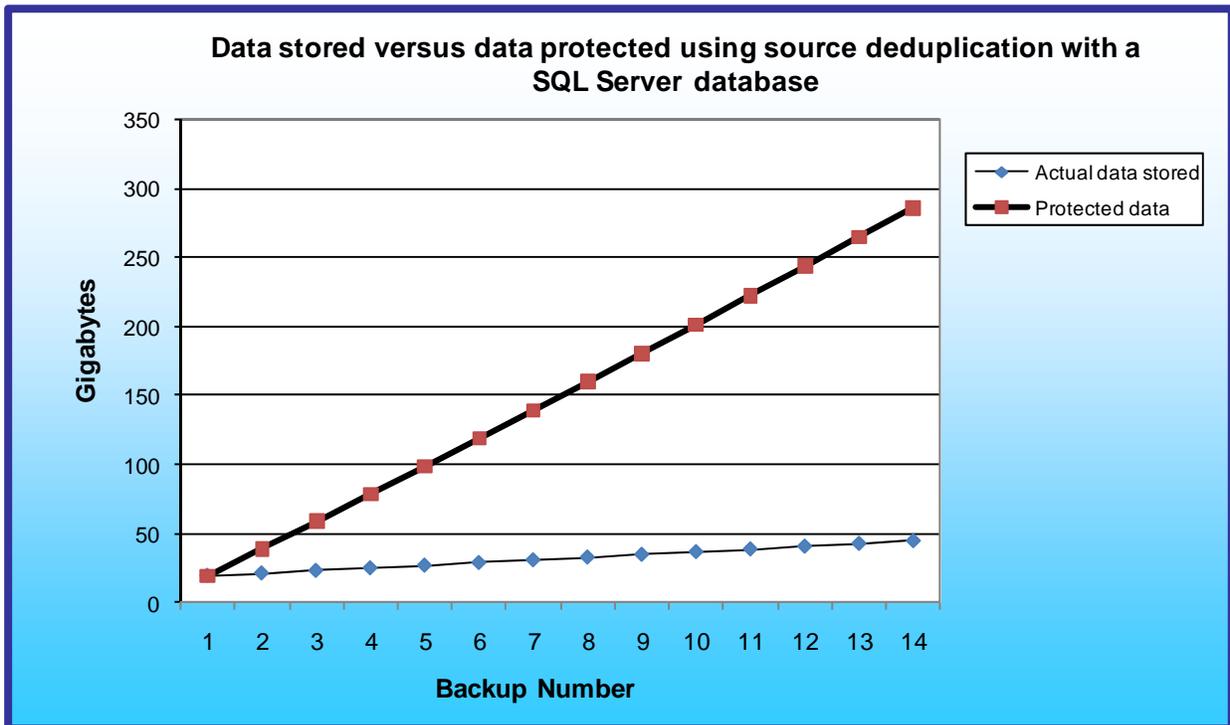


Figure 3: Comparison of protected data versus the actual data stored from our 2-week database backup simulation of a SQL Server database.

### File server backup scenario

We used a custom file server workload to simulate a real-world file server backup scenario of a business modifying approximately 3 to 5 percent of their files per day. We performed an initial full backup of our 50GB file

share using Symantec Backup Exec 2010. We then proceeded to simulate 13 additional days of activity, running the workload, and then executing a full file server backup, each time.

For our backup job, we browsed to the file server, selected the relevant folder, 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 6 hours. Subsequent full file server backups took an average of just over 22 minutes, achieving a deduplication ratio of at least 24.6:1 using source deduplication. For comparison, we ran a single backup using a backup-to-disk storage device on the media server without source deduplication, which took roughly 6 hours and 5 minutes.

As Figure 4 shows, the difference between the amount of file server data protected versus the actual amount of data stored becomes greater over time, benefiting the business by requiring significantly less long-term storage capacity.

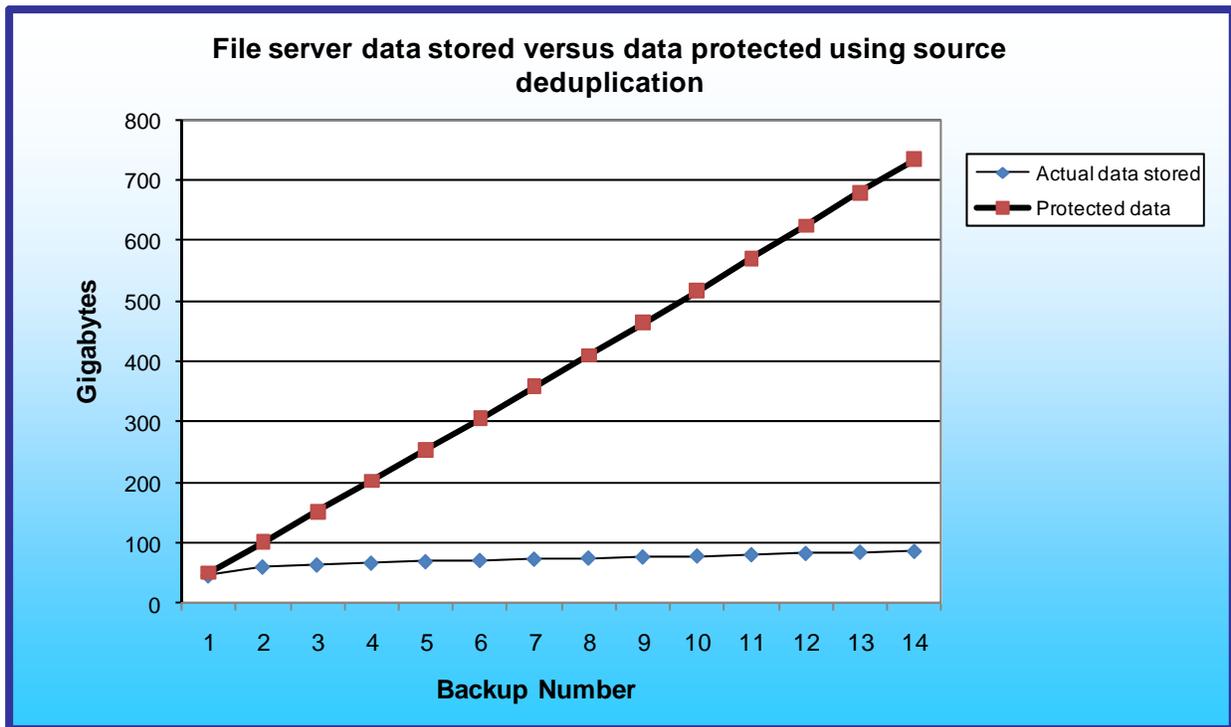


Figure 4: Comparison of protected file server data versus the actual data stored from our 2-week file server backup simulation.

### Mail server backup scenario

We used Microsoft LoadGen to simulate a real-world Exchange 2007 server backup scenario of a business email configuration, with 300 users modifying their Exchange data at an average daily usage level. We performed an initial full backup of our Exchange storage group using Symantec Backup Exec 2010. We then proceeded to simulate 13 additional days of activity, running the workload and then performing a full storage group backup each time.

For our backup job, we browsed to the mail server, selected the Exchange storage group, directed the backup to the deduplication storage folder, allowed the remote agent direct access, and left the default usage of Granular Recovery Technology (GRT) enabled. From that point, our backup job used source deduplication. The initial complete backup took just under 9 hours. Subsequent full mail-storage-group backups took an average of just over 2 hours, achieving a deduplication ratio of at least 10.8:1 using source deduplication. For comparison, we ran a single backup using a backup-to-disk storage device on the media server without source deduplication, which took roughly 4 hours and 19 minutes.

As Figure 5 shows, the difference between the amount of mail server data protected and the amount of actual data stored increases over time, benefiting the business by requiring significantly less long-term storage capacity.

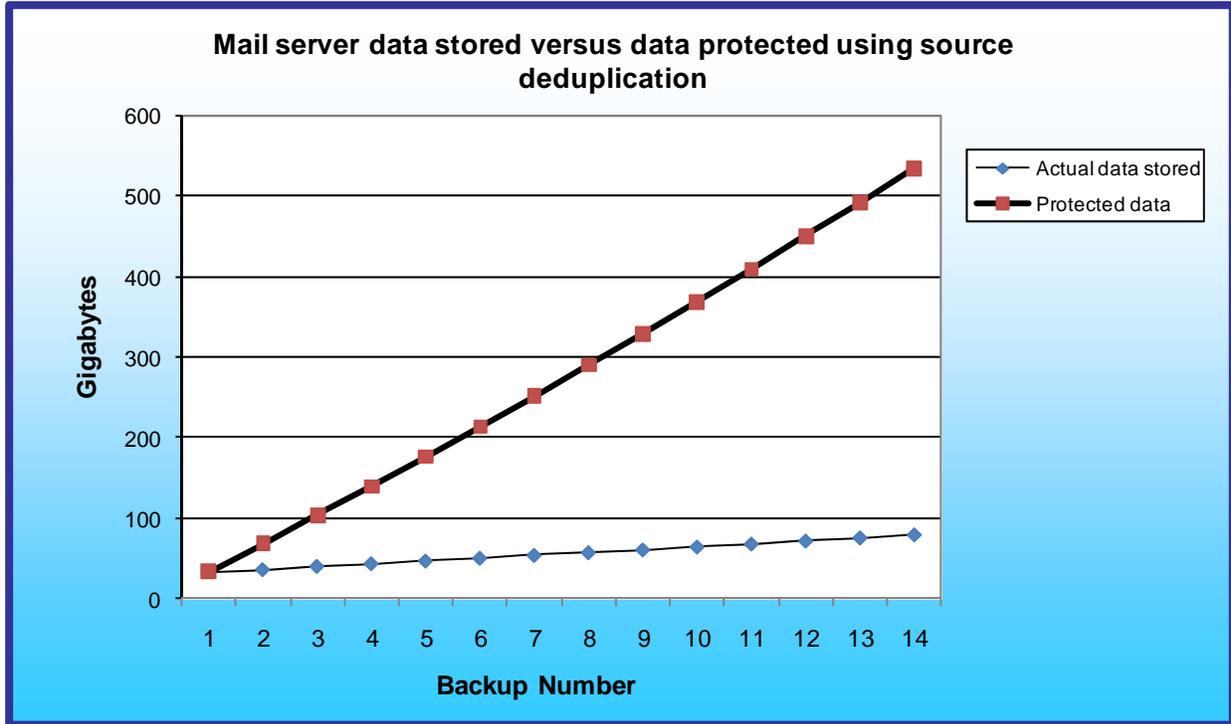


Figure 5: Comparison of protected mail server data versus the actual data stored from our 2-week Exchange 2007 backup simulation.

In our tests, Symantec Backup Exec 2010's ability to eliminate duplicate data in our SQL Server database server, Windows-based file server, and Exchange 2007 mail 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 tests in our full reports, which you can read at

[http://www.pricipledtechnologies.com/clients/reports/Symantec/db\\_source\\_dedupe.pdf](http://www.pricipledtechnologies.com/clients/reports/Symantec/db_source_dedupe.pdf),  
[http://www.pricipledtechnologies.com/clients/reports/Symantec/fs\\_source\\_dedupe.pdf](http://www.pricipledtechnologies.com/clients/reports/Symantec/fs_source_dedupe.pdf), and  
[http://www.pricipledtechnologies.com/clients/reports/Symantec/email\\_source\\_dedupe.pdf](http://www.pricipledtechnologies.com/clients/reports/Symantec/email_source_dedupe.pdf).

Principled Technologies, Inc. 1007 Slater Road, Suite 300, Durham, NC 27703 [www.pricipledtechnologies.com](http://www.pricipledtechnologies.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.