Protecting Microsoft® SQL Server® with an Integrated Dell / CommVault Solution

Database Solutions Engineering

By Subhashini Prem and Leena Kushwaha
Dell Product Group
March 2009
THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

© 2009 Dell Inc. All rights reserved. Reproduction in any manner whatsoever without the express written permission of Dell, Inc. is strictly forbidden. For more information, contact Dell.

Dell, the DELL logo, PowerEdge, and PowerVault are trademarks of Dell Inc.; CommVault and Galaxy are registered trademarks of CommVault Systems, Inc.; Microsoft and SQL Server are either trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell Inc. disclaims any proprietary interest in trademarks and trade names other than its own.
Executive Summary

To alleviate the risk of downtime and quickly resume services after a disaster, every organization should have a reliable continuity plan that supports business operations and ensures databases are continuously available to customers. The Dell/CommVault integrated solution for Microsoft® SQL Server® provides a simplified platform that allows administrators to manage data protection and, in turn, maintain reliable business operations from a single interface.

Organizations must examine a few key considerations when selecting a data protection solution, such as how quickly their service level agreements require them to recover from a disaster and how much data they must ultimately recover. They also must consider how backup devices can be reused to maintain cost effectiveness and how their disaster recovery solutions can be most easily managed. A variety of different storage solutions address these concerns; however, many require a tradeoff between performance and cost. As this paper demonstrates, a solution that combines Dell hardware with CommVault software—specifically Dell™ PowerVault™ MD3000 and MD1000 disk arrays, PowerVault TL2000 tape libraries, and Dell PowerEdge™ R610 and R710 servers with CommVault® Galaxy® backup and recovery software—addresses these considerations in an efficient and cost-effective way.

CommVault Galaxy software offers online backup and recovery of SQL Server Databases. Galaxy includes a unified console for managing data protection operations. It allows administrators to protect databases at the instance, database, and file/file-group levels, making it easy to find and browse data and perform granular restore operations. The software utilizes the Microsoft Volume Shadow Copy Service (VSS) framework, which enables administrators to backup data when a database is online. Finally, in a multiple SQL Server environment where consolidated backup is an optimal way to utilize storage arrays, integrating Galaxy with both an MD3000 disk array and TL2000 tape library offers a variety of data protection strategies, such as disk-to-disk, disk-to-tape, and disk-to-disk-to-tape.

This paper begins by examining basic considerations for data protection and exploring various options for data backup and recovery. Then, it discusses different strategies that organizations can implement to protect SQL Server databases.
Table of Contents

Executive Summary ................................................................................................................................ iii
Introduction............................................................................................................................................. 2
Considerations for Selecting a Data Protection Solution ................................................................. 2
Dell / CommVault Solution for Data Protection ................................................................................. 3
Backup and Recovery Methods Using CommVault Galaxy ............................................................ 6
  Backup Methods for SQL Server ........................................................................................................ 7
  VSS-Enabled Backup .......................................................................................................................... 7
  Initiating a Backup Operation ............................................................................................................ 8
    Instance-Level Backup....................................................................................................................... 8
    Subclient-Level Backup .................................................................................................................... 9
    File/File-Group Level Backups ........................................................................................................ 11
  Consolidated Backups ....................................................................................................................... 12
  Recovery Options for SQL Server ..................................................................................................... 13
Dell / CommVault Backup Strategies ................................................................................................ 15
  Disk-to-Tape Backup Strategy ......................................................................................................... 15
  Disk-to-Disk Backup Strategy ......................................................................................................... 16
  Disk-to-Disk-to-Tape Backup Strategy ............................................................................................ 17
Conclusion ............................................................................................................................................. 19
Glossary: .............................................................................................................................................. 19
References ............................................................................................................................................. 20
Introduction
This paper articulates a data protection solution for SQL Server that integrates CommVault Galaxy software with PowerEdge R610, R710 servers, PowerVault MD3000 and MD1000 disk arrays, and PowerVault TL2000 tape libraries. It is part of a series of papers that focus on data protection and business continuity strategies for SQL Server. Visit www.dell.com/sql to obtain the other papers in this series.

Microsoft SQL Server databases form the backbone for many businesses, housing data for many critical applications. As a business grows, the amount of data it stores also grows, usually resulting in multiple databases spread across multiple servers. Because access to this data is critical for day-to-day business operations, Service Level Agreements (SLAs) demand that databases remain online and available—there is little tolerance for downtime. As the need to store more data grows, simply adding storage capacity to accommodate the growth without implementing data protection is not a sound business strategy. Factors like unplanned outages, hardware failures, viruses, and human errors can cause business units to cease operations for hours or even days, and other factors—like thefts and natural disasters—may trigger a complete shutdown of the business. A robust data protection strategy helps to mitigate these potential problems by ensuring that critical information remains available and well protected.

Hardware redundancy is one way to address business continuity. Databases can be protected from hardware problems—like power supply, network controller, and storage controller failures—by including redundant hardware components in servers and storage arrays. Similarly, disk drive failures can be avoided by storing SQL Server data on RAID volumes. The PowerEdge servers and PowerVault storage arrays featured in this paper provide these capabilities.

However, redundancy alone cannot protect an organization’s database deployment from every potential failure. That’s why it is critical to implement a data backup plan—so that crucial business information is protected both from human errors and natural disasters.

Considerations for Selecting a Data Protection Solution
Business environments differ widely, from small businesses running just a few servers to large data centers with complex server and storage architectures and everything in between. Each organization must choose a data protection solution that fits its unique business environment.

Following are four considerations to examine when choosing a business continuity solution:

- **What are the organization’s Recovery Point Objective (RPO) and Recovery Time Objective (RTO)?** RPO and RTO provide a baseline for how much data loss and how much downtime can be tolerated when recovering from an outage.

- **What are the organization’s backup retention policies?** If the backed up data exceeds the retention policies, it is possible to delete that data and reuse the media. Alternatively, data may be archived to different media and kept for long-term storage.
• **What are the organization’s overall goals?** Factors like performance, capacity, portability, and cost will determine the different storage media—such as disk, tape, and optical drives—to include in a data backup and recovery solution.

• **Will the organization use data protection software?** If a solution includes data protection software, it is important to understand the value it adds. For example, does the software schedule regular backup and restore jobs? Does it require little to no monitoring? And, most important, how easy is the software to manage?

**Dell / CommVault Solution for Data Protection**

Galaxy software integrated with PowerEdge servers and PowerVault storage provides a complete data protection solution for Microsoft SQL Server databases.

Figure 1 illustrates a Dell/CommVault solution that includes PowerEdge R610, R710 servers, MD3000 disk arrays, MD1000 storage arrays, and a TL2000 tape library along with Galaxy Software. Each production SQL Server host connected to a MD3000 storage array has a SQL Agent installed, which is responsible for backing up and restoring SQL Server databases. Backups from these hosts are consolidated by a backup agent installed on the backup server. The backup agent manages data transmission between production servers and backup storages (MD1000, TL2000). The management server communicates with all SQL agents and backup agents to coordinate operations such as backups, restores, and media management. The management server and backup server can be installed on the same physical server depending upon the hardware needs.
The PowerEdge R710 is a 2U rack mounted server designed around system management, usability, industrial design, power and thermals which address the needs of enterprise, datacenter as well as growing small businesses with 3 or more servers. With its huge internal capacity it can very well suited for production server.

The PowerEdge R610 is a 1U rack mounted server designed to address the needs of space-constrained customers that face the challenge of putting as much computing power in their data centers as possible; it fulfills the needs of customers who desire to pack the most CPU performance into the least amount of rack space. With the high speed Intel dual core and quad core processors with QuickPath Architecture it can manage heavy backup operations. It can be very well suited as a backup server.

The MD3000 is a SAS array that delivers high-performance, fault-tolerant direct attached storage (DAS) for a low cost. It is designed to provide high availability access for up to two servers running business
applications. The MD3000 is ideal for non-consolidated databases, because of its fully-redundant architecture that includes dual active-active RAID controllers, a mirrored cache, and I/O multipathing. It supports both SAS and SATA disk drives that can be scaled form 15 disks to 45 disks by adding additional MD1000 storage arrays—a flexible design that allows customers to add storage as the business grows and data increases.

The TL2000 is a tape library that helps automate backup processes to reduce the need for manual intervention, mitigating the risk of human error. It offers simplified operations, like importing and exporting tapes individually via a mail slot or twelve-at-a-time using one of two magazines. It also provides intuitive management, such as barcode-scanning technology that records the location of tape media, so you can be assured that your critical data is properly stored and ready if disaster strikes.

Galaxy software integrates with SQL Server applications to protect data in a way that allows administrators to easily understand and manage data. This application-centric and policy-based approach enables backup and recovery operations to be managed and scheduled from a unified console. Figure 2 shows the CommVault Galaxy unified console for managing all data protection operations. The left pane of this console lists all the production servers and their respective SQL Server instances. This pane is used to configure data protection operations, such as creating different policies and configuring storage devices that will be used for backup. The right pane lists subclients that have been configured for backup jobs. A subclient is used to configure a group of databases to use same backup policies. The “Job Controller” pane at the bottom of the console shows job status.
Backup and Recovery Methods Using CommVault Galaxy

CommVault Galaxy offers an application-integrated module-SQL Agent that provides a variety of methods to backup databases. It supports full, differential, and transaction log backups for SQL server at the instance, database, and file/file-group levels. Backed-up data can be restored in several ways,
including full, partial, and step restore. Note that the method used for backup and restore can affect the solution’s RTO and RPO.

**Backup Methods for SQL Server**

SQL Server databases can be backed up using *full, differential,* and *transaction log* backups. A full backup makes a complete copy of all data pages. For very large databases, performing a full backup takes a considerable amount of time to complete, and it consumes a lot of storage space. Therefore, it is a good idea to limit full database backups to a weekly or monthly schedule, and schedule regular differential backups to copy only those data pages that have changed since the last full backup. Differential backups are considerably smaller than full backups; however, remember that a differential backup does not copy the pages changed since the last differential backup—rather, it copies all pages changed since the last full backup. So, if a full backup is performed on Sunday, and then differential backups are run during the week, Monday’s differential backup will likely be smaller than the subsequent differential backups performed throughout the remainder of the week. When restoring from a set of backups that use this scheme, the full backup should be restored first, followed by the differential backups.

A transaction log is a record of all operations that have been performed on the database. Transaction log backups store copies of the transaction log on a backup device, with these backups, a database can be recovered to a specific point in time. These backups take significantly less time and resources to create than full or differential backups, and they are considerably smaller. As a result, transaction log backups can be taken more frequently than full and differential backups. To fully restore a database using transaction log backups requires a full backup, the last differential backup (if available), and all transaction log backups taken since the last full or differential backup; the restore operations must be performed in sequential order.

Note: Differential and Transaction log backups can only be created on databases that use the Full Recovery model in SQL Server.

A typical backup strategy combines all three backup types to help reduce the RPO. For example with a weekly full backup, you’ll achieve an RPO of one week, which you can reduce down to a day by adding daily differential backups. This can be further reduced to an hour by taking transaction log backups every hour.

**VSS-Enabled Backup**

Most backup operations take the production server offline until the backup is complete, prompting many organizations to perform backups during non-business hours. The problem with this approach is that failing to perform backups during business hours can result in losing an entire day’s data if the database fails before the next backup is performed.

Galaxy uses VSS to perform backups while a Production server is online. Write operations can continue while open files are copied and backed up. This works by pausing the database and queuing transactions while a snapshot is created. This process happens quickly, and write operations can
continue once the snapshot is created, which eliminates locked-file conditions and maintains the integrity of the database.

**Initiating a Backup Operation**

Galaxy offers flexibility in managing how databases, data files, and file groups are backed up; backups can be initiated from the instance, subclient, or file/file-group level.

**Instance-Level Backup**

When a backup is initiated at the instance level, all databases in that instance that are configured for backup are backed up together, which simplifies administration. As shown in Figure 3, the instance named `SQL-Server1\INSTANCE2` hosts `database1` and `database2`. The backup is initiated at the instance level by selecting **Backup All Subclients**, which backs up all the databases as shown in the **Job Controller** pane.
Subclient-Level Backup

Instead of always backing up a complete instance, administrators can also select a subset of databases within an instance to back up with a single operation—a method called "subclient level backup." This is useful in situations where different tiers of databases are housed on the same database instance. All these databases will be backed up in parallel. However if more than one database share the same storage volume then the backup job will run sequentially, thus diminishing backup operation performance. To reduce the time required to back up multiple databases that share the same storage volume for backup, Galaxy provides a streaming feature that allows multiple database backup jobs to run in parallel.
Data files and log files of the database can set to use different backup devices for backup. In such a case, increasing streaming count per storage volume further reduces backup time. For example, database1 contains data file data1.mdf and log file log1.ldf; similarly, database2 contains data file data2.mdf and log file log2.ldf. Data files data1.mdf and data2.mdf are backed up on “backup-volume1” and log files log1.ldf and log2.ldf are backed up on “backup-volume2” -- then increasing the stream count to two per storage device will create a total of four threads. Each thread runs in parallel, which allows the backup operations to complete much more quickly.

Figure 4 shows this type of configuration, in which different storage devices are used for backing up data files and log files. In addition, the stream count for each backup volume has been increased. This was accomplished by configuring a “Data Storage Policy” and a “Log Storage Policy.” Each policy has an associated storage device and stream count.
Figure 4: Configuring streams for Data and Log Backups

File/File-Group Level Backups

File and file-group level backups allow you to back up selected portions of a database. SQL Server databases have three different kinds of files: primary files, secondary files and log files. Each database can have one or more primary files and secondary files and log files. By using File/File-Group level backup, a single primary/secondary file or a group of primary/secondary files can be backed up as part of the same operation.
This functionality can be critically important for very large databases in which complete database backups are incredibly time intensive. File and file-group level backups support only full and differential backups.

Note: A full file group backups may not be sufficient to restore the entire database. To restore the database completely, transaction log backups need to be run after file/file group backups. Galaxy provides an option to automatically run transaction log backups after successful completion of a Full or Differential backups of the file/file-group level backup.

**Consolidated Backups**

Databases from multiple production servers can be backed up using a single backup server. As shown in Figure 5, databases DB1, DB2, and DB3 on production servers Production Server1, Production Server2, and Production Server3 are backed up and managed from one backup server. This backup server controls all of the backup storage devices.

![Consolidated Backups Diagram](image)

**Figure 5: Consolidated Backups**

In Galaxy, SQL Server databases from different production servers can be backed up using a single, unified console. These databases cannot be simultaneously initiated for backup—instead, they are
initiated sequentially. If the sequentially initiated jobs use different storage volumes, then the jobs will complete in parallel; however, if they use the same storage volume, then jobs remain in a pending status until resources are available. You can avoid pushing jobs into pending status by increasing the streaming count at the storage device level, creating multiple streams to run backup jobs in parallel.

Backup data is transferred from the production servers to the backup server through the LAN, which can substantially increase the network load. To reduce the amount of data that must be transferred, you can compress the database on the production server before sending it over the LAN. This will also reduce the space that is required for storing backup copies. Another option would be to increase your available bandwidth by teaming NICs or replacing your 1GbE NIC with a 10GbE NIC. Remember, however, that using software to compress the database is processor intensive. So, if you need to reduce backup storage consumption more than you need to reduce the network load, you should compress the database on the backup server before sending it to the backup storage media. Any time you transfer data over a network; there is a possibility that it will be compromised. To help alleviate this concern, increase data security by encrypting data at the production server before transporting it to the backup server.

A consolidated backup environment with multiple Production Servers housing multiple databases requires a significant amount of backup storage space. To maintain a cost-efficient architecture, administrators must decide what data can be deleted and what data must be stored longer term according to data retention policies. The data aging feature in Galaxy identifies and deletes data that exceeds specified data retention rules; it then recycles that media for reuse.

**Recovery Options for SQL Server**

Galaxy provides several restore options to recover databases at as granular level as possible. Databases can be restored completely, partially, or at the file level—each of these restore options varies in its RTO. The type of backup operation will also influence how the data can be restored.

Complete restore has a longer RTO than a partial restore, because it renders the database unavailable until the recovery operation is complete. Alternatively, a partial restore restores the primary file group first, and then takes the database back online; remaining secondary file groups then restore sequentially.

Individual files or file groups can also be recovered using database backups or file/file-group backups. These restores are commonly used for situations in which only individual data files on a database are corrupted, since a complete database restore has a high RTO. Remember that file/file-group level restores cannot run without transaction logs; so, if your plan is to execute these restores, you must back up transaction logs regularly.

A SQL Server database can be recovered to its original location, or it can be restored to a new location. If your data problem is simple—for example, some data in the database is damaged or a few rows or tables are mistakenly deleted—you’ll likely want to recover the data to its original location using backup copies. On the other hand, if the production server or storage device is no longer available due to
hardware problems or natural disaster, you can also restore the database to another production server. This is possible only if the backup server is not co-located with the production server.

Remember that selected databases from a single instance can be recovered in a single restore operation. As shown in Figure 6, *Database1* and *Database2* from *SQL-Server2\instance1* are selected to recover at the same time.

![Figure 6: Recovering Multiple Databases in a Single Recovery Operation](image-url)
Dell / CommVault Backup Strategies

Galaxy software integrated with PowerVault storage offers a variety of backup strategies including disk-to-disk, disk-to-tape, and disk-to-disk-to-tape. This flexibility helps organizations to iteratively implement their data protection plan.

Disk-to-Tape Backup Strategy

Disk-to-tape backup is a standard approach that uses tape for data backup and restore. Figure 7 illustrates the configuration of a disk-to-tape backup that uses Dell/CommVault technologies. The PowerEdge R710 server hosting the SQL Server database is connected to the production storage (MD3000), which stores database files. After receiving the backup command from Management Server, the databases are backed up from the production server to the backup server through the LAN, and then stored in the tape library (TL2000), which is connected to a backup server.

Figure 7: Disk-to-Tape Backup

Tapes are a common choice for long-term data retention because they are typically very reliable. They are also portable, so they can be used to easily transfer data offsite to safeguard from natural disasters. Since tape is a sequential access device, backup time and RTO for a single database itself takes more...
time when compared to the disk storage. This will be more time consuming if multiple databases are backing up together and restored together as well. To reduce the backup time, data can be compressed on the backup server before sending it to the tape library; this can not only boost the performance of data protection operations, but it also boosts the virtual capacity of the tape.

Galaxy provides an AuxCopy feature that is used for creating multiple secondary-backup copies. Secondary copies are saved on different tape media than primary copies, providing extra protection to the primary backup copies. If the primary backup copy is unavailable due to media failure, then secondary copies can be used to recover the data. Secondary copies can also be promoted to primary copy for continuing backup operations.

**Disk-to-Disk Backup Strategy**

Disk-to-disk backup uses reliable and higher-performance disk media to store backup files. Since disk storage is randomly accessible, recovery time (RTO) is less than that of tape storage. Another advantage of using disk is that multiple hosts can access the disk simultaneously.

Figure 8 illustrates a configuration of disk-to-disk backup that uses Dell/CommVault technologies. A PowerEdge R710 server hosting the Production database is connected to production storage (MD3000), which stores database files. After receiving the backup command from Management Server, the databases are backed up from the production server to a backup server through the LAN and stored in backup storage (MD1000), which is connected to backup server.
By using Aux Copy feature multiple backup copies can be stored on the backup storage which provides extra protection to the primary backup copies. If the primary backup copy is unavailable due to media failure, then secondary copies can be used to recover the data and they can also be promoted to primary copy for continuing backup operations.

**Disk-to-Disk-to-Tape Backup Strategy**

Disk-to-disk-to-tape backup allows administrators to backup the database first to disk (for faster retrieval) and then to tape (for long-term retention).

As shown in Figure 9, a PowerEdge R710 server hosting Production database is connected to production storage (MD3000). On receiving the backup command, databases are backed up from the production
server to the backup server through the LAN and stored in primary backup storage (MD1000-1). By using Aux Copy feature the backup copies can be created on multiple backup storage devices (MD1000-1, MD1000-2). In a Disk to Disk to Tape configuration multiple backup copies can be created on different disk storages simultaneously. On a primary storage failure the secondary storage can be used for continuing the backup operation. On the other way the secondary copies can be used to copy the backup copies to tapes for archival without intervening the production server. These copies not only provide extra protection, but also help to comply with governmental regulations for data retention.

Figure 9: Disk-to-Disk-to-Tape Backup
Conclusion

The Dell/CommVault integrated solution provides a comprehensive solution for protecting SQL Server running on Dell hardware. This flexible data protection strategy simplifies data administration by combining the data backup functionality in Galaxy software with MD3000 and MD1000 disk-based storage and TL2000 tape-based storage to offer a variety of data backup and recovery methods including disk-to-tape, disk-to-disk, and disk-to-disk-to-tape at the file/file-group, database, and instance levels. Customers can choose from several disk-based data protection techniques and backup methods—including full, differential, and transaction log—depending on their business requirements. By using MD3000 disks for the storage array and MD1000 for primary backup, customers get a scalable infrastructure that allows them to add additional storage to their growing business environment as needed.

Glossary:

**Aux Copy:** An auxiliary copy operation that creates secondary copies of data associated with data protection operations, independent of the original copy

**Data Aging:** A policy that determines the data retention period; after data exceeds its retention period, it is automatically deleted to free up media space for reuse

**Data Compression:** A technique for compressing data before sending it to storage media to reduce the quantity of data sent to storage; often doubles the effective capacity of the media depending on the nature of the data

**Data Encryption:** Data security technique that protects data for transmission over non-secure networks

**RTO:** Recovery Time Objective (RTO) indicates how quickly a business can recover from failure

**RPO:** Recovery Point Objective (RPO) indicates how much data is lost when recovery is performed

**Streams:** A logical channel that connects a database to storage media; multiple streams are parallel channels through which data can flow, thus improving the rate at which data can be written to the storage media
References

The Changing Dynamics of Backup and Recovery in the Small and Medium Business (SMB) Market:
http://0301.netclime.net/1_5/M/U/S/ESGResearch_SMBDataP.pdf

CommVault Solutions for SQL Server:

A CommVault White paper: CommVault Galaxy Backup & Recovery:

CommVault Products: Advanced Features:

Files and Filegroups implementation in SQL Server:

Understanding SQL Server backup types:
http://www.teratrax.com/articles/sql_server_backup_types.html

Transaction Log Backups