Always On White Paper for Fujitsu ETERNUS Storage System

October 5, 2006

Fujitsu Limited

Contents

1. Introduction

This document describes a very efficient database environment that provides the combined benefits and advantages of the Microsoft SQL Server and Fujitsu ETERNUS storage systems.

1-1. Scope

Today, IT systems and their stable operation are becoming more and more critical to success of customer's businesses.   Among the IT technologies, successful deployment and sustained availability of database and the data on storage is one of the most important tasks for IT managers.   In this white paper, we discuss best practices of architecting our ETERNUS storage systems in SQL Server environment for its performance and availability.

1-2. Products mentioned in this white paper

Today's ETERNUS storage systems are Fujitsu's fourth generation of its disk array product series and meets all the requirements in mission-critical environment.   ETERNUS8000 is highly scalable in both vertical and horizontal dimensions, with high performance and large capacity disk array,; while, ETERNUS4000 is a flexible mid range disk array.   In all models, all the major components are redundantly equipped and are hot-swappable to ensure business continuity with high availability.

ETERNUS Multipath Driver is Fujitsu's software product that manages paths between servers and ETERNUS storage systems.   It provides path failover mechanism as well as load balancing function for availability and performance.

ETERNUS SF Storage Cruiser is Fujitsu's software product that provides the infrastructure management capability for SAN spanning from ETERNUS storage systems, fibre channel switches and servers/HBAs.

ETERNUS SF AdvancedCopy Manager is Fujitsu's software product that manages the advanced copy functions built in to the ETERNUS storage system and provides online, non-disruptive volume replication feature for point-in-time backup operation of SQL database.

2. Fujitsu's ETERNUS Storage System

2-1. High Availability

Fujitsu ETERNUS storage systems provide uncompromised data availability that maximizes the secure operation of databases in mission critical environments.

All major components are redundantly equipped. In the event of a component failure, ETERNUS automatically reallocates the rest of the active components for non-disruptive operation, and allows all the major components been swapped while operation in progress.

Disk drives are configured in RAID groups. In the event of a disk-drive failure, the data is quickly rebuilt to a hot-spare disk to maintain the disk redundancy.   In addition, ETERNUS storage systems constantly monitor each disk drive for pre-failure detection.   If one of the disk drives is likely to fail, ETERNUS proactively copies the data from the disk drive to a hot-spare disk drive.   This proactive copy provides superior data protection, because the redundancy is maintained while a disk drive is being replaced.

The interfaces between servers and ETERNUS storage systems also deliver a high availability mechanism. This mechanism is provided by ETERNUS Multipath Driver that manages multiple Fibre Channel paths.   Even if one path fails, operations of a database and other applications can continue by automatically rerouting the access through the rest of active paths.   This failover mechanism covers failures of HBA installed on server or failures of SAN switches, thus provides a high availability to guarantee continuous operation of database.

2-2. Performance Tuning

Fujitsu's ETERNUS storage system provides two features to manage the logical volumes.

Logical Volume Migration changes the RAID group where a logical volume located.   The entire logical volume is migrated from one RAID group to another in a manner that is transparent to servers and application software. One way is to migrate from one RAID group type to another RAID group type in place.   In other words, the disk drives used in the current RAID group will be used in the destination RAID group.   This feature is useful when reconfiguring a RAID group to a different RAID-level using the same set of disk drives, for example from RAID-5 to RAID-10.   Another feature is to prepare two separate RAID groups and migrate the logical volume(s) from one RAID group to another.   This is useful when logical volume need to be moved to a RAID group using different type of disk drives, for example, from a RAID-5 using 10,000 rpm disk drives to a RAID-1 or RAID-10 using 15,000 rpm disk drives.   In both features, the accesses to logical volumes are automatically switched within ETERNUS storage systems when the logical volume migration is complete, so these can be used to optimize database performance without interrupting production operations.

2-3. Point-in-time copy feature

Fujitsu ETERNUS storage systems provide point-in-time copy features that create a copy of the source volume data at a certain point-in-time within the storage system without disrupting the operations of a database or other applications.

With the strong demand for 24 x 7 x 365 non-stop operations, system administrators must make backups without disrupting production operations using the databases, without causing heavy loads on the servers. ETERNUS storage systems provide three types of replication features: Equivalent Copy (EC), Quick One Point Copy (QuickOPC), and Snap One Point Copy (SnapOPC) that create a copy of the source volume in variety ways without using processing power on the servers.

With EC, the source and target volumes are initially synchronized and then split (optionally on a temporary basis) when the replicated volume is required as a point-in-time copy.   The source volume and target volume can be defined in a same storage system for disk-to-disk backup, or may be defined in two physically separated storage systems.   The two storage systems can be separated by fibre channel distance (upto 200km) for additional data protection or by WAN (hundreds of miles) for disaster recovery.

With QuickOPC, a point-in-time copy can be instantly created on the target volume on demand, with the physical copy operation following.   SnapOPC has further enhanced the functionality of QuickOPC, where the size of the target volume is minimized by only copying the modified data on to the target volume.   These features can be chosen according to customers' business requirements.

2-4. Tiered Storage with MAID technology

Customers can choose and mix any type of disk drives in ETERNUS storage systems, ranging from 36GB to 146GB / 15,000 rpm disk drives, 300GB / 10,000 rpm disk drive and 500GB / 7,200 rpm disk drives. Although it is not recommended to create a RAID group using different type of disk drives, there is no additional hardware logic required to accommodate different types of disk drives within an ETERNUS storage system.   Combined with the variety of RAID levels 1, 10, and 5, customers can have RAID-10 using 146GB/15,000 rpm disk drives for high performance transaction operations and RAID-5 using 500GB/7,200 rpm disk drives for back up space.

ETERNUS Storage Systems are also equipped with MAID technology (Massive Array of Inactive Disks).   This feature manages the on/off state of the disk drives' spindle motors based on policy settings.   For example, customer can specify to spin up the motors of disk drives when these are used as target volumes for copy operation, and spin down the motors when the copy operation is completed.   By using this feature, customers can save the energy consumption upto 20%.

3. ETERNUS SF Storage Management software

3-1. ETERNUS SF Storage Cruiser

Fujitsu offers ETERNUS SF Storage Cruiser that provides the advanced storage infrastructure management capability based on the SAN (Storage Area Network) management features. With versatile management features for configuration, troubleshooting, performance and correlation, Storage Cruiser provides stable operation for any type of storage system, including SAN, DAS (Direct Attached Storage) and NAS (Network Attached Storage).

In SAN environments, it is not enough to be able to manage just the physical topology.   Without management of the logical topology, it is very difficult to exactly understand the impact to the entire system caused by a failure of a device and/or component.   ETERNUS SF Storage Cruiser not only graphically displays logical connections, but also shows component failures on the display.   This enables IT managers to easily and quickly react to failures in the system.

3-2. ETERNUS SF AdvancedCopy Manager

Fujitsu offers ETERNUS SF AdvancedCopy Manager, application software that provides high-speed backup/restore and replication capabilities in collaboration with ETERNUS storage systems. ETERNUS SF AdvancedCopy manager can perform online backup of SQL Server databases. This enables 24x7 continuity of the production environment.

By using AdvancedCopy Manager with SQL Server, customers can create a point-in-time copy set of the primary data file, secondary data file(s) and transaction log files.   This copy set can be used as a full backup of the SQL Server database or can be used for other non-production types of activities, such as testing new procedures or applications using the real database environment, etc.

4. Best Practices

4-1. Make sure to install System database and User database on different volumes

The AdvancedCopy Manager can only backup User database, and cannot backup System database (master, msdb, model and distribution).   The AdvancedCopy Manger creates entire copy of a volume, so the System database and User database should not be installed on a same volume.   The System database can be backed up by either SQL Server Management Studio or T-SQL.

4-2. Prepare backup volumes for both data file and log file

Backup is required for both data file and log file in case of SQL Server which is different from other databases such as Oracle where only table space needs to be backed up.
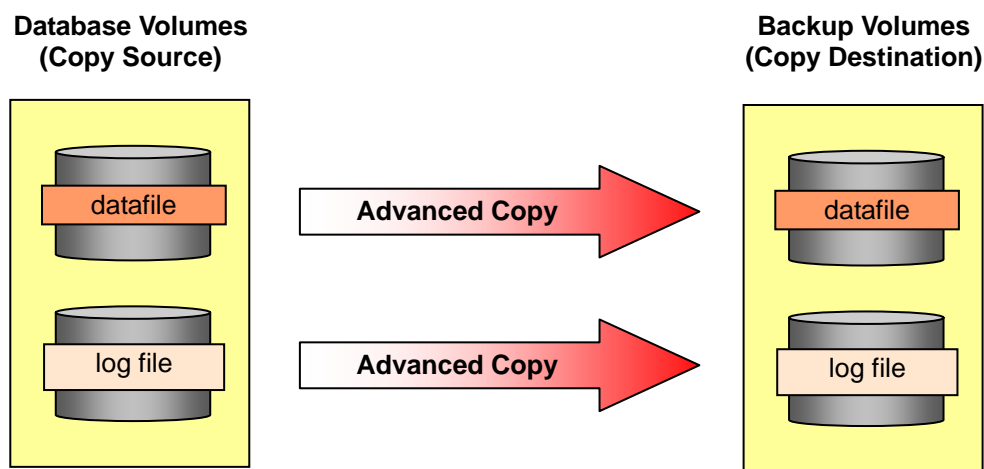
**Database Volumes
(Copy Source)**　　　　　　　　　　　　　　**Backup Volumes
(Copy Destination)**

| datafile | **Advanced Copy** → | datafile |
| log file | **Advanced Copy** → | log file |

Figure-1 Backup of data file and log file

4-3. Place only database files on a volume

AdvancedCopy Manager creates a copy of the entire volume (partition), so the volume for database files should not contain any other files.

a) If other files are located on the volume with the database, those files may be corrupted when database is restored, and could result in corruption of the entire file system.
b) Database files should not be located on a system drive or a volume that contains execution files and management files of SQL Server and AdvancedCopy Manager.

4-4. Minimize the number of databases for each backup operation

It is possible to install more than one database on a same volume, or to share a volume with multiple log files, however from the ease of operation point of view, it is recommended that the number of databases on a volume is minimized.
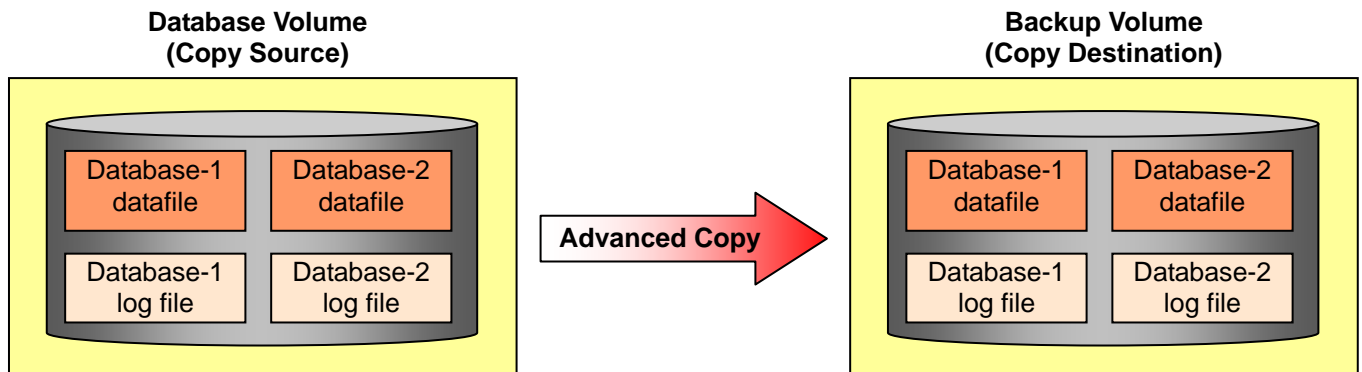
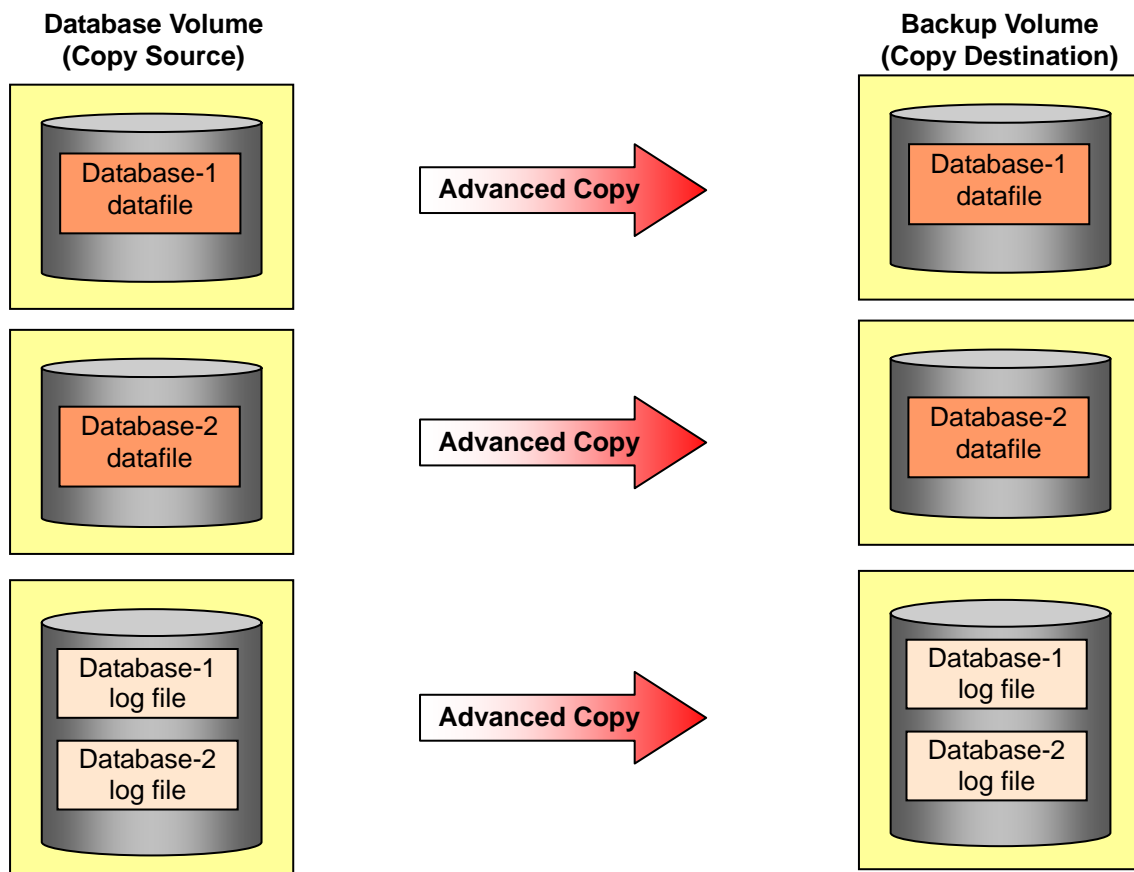Figure-2 Two databases sharing a volume (possible but not recommended)

Figure-3 Two databases sharing a volume for log file (possible but not recommended)

AdvancedCopy Manager creates a copy of the entire volume (partition).   If a volume contains more than one database file, then all the databases on the volume should be backed up at the same time, and it complicates the backup and restore operations.

If multiple databases are installed on one volume, it will cause the following disadvantages.
a) If multiple databases are backed up at the same time, all the databases are locked until all the backup operations completes.   This will unnecessarily prolong the duration of backup, and may result in an error in the backup.

5

b) Restore operation of the AdvancedCopy Manager cannot be used for restoring one particular database, since other database(s) on the same volume will also restored at the same time.   If this is the case, the targeted database can be restored by using either Copy command or ftp.

4-5. Execute Database and Backup on separate servers

AdvancedCopy Manager has the flexibility to run in a configuration such that Database and Backup are both running on the same server.   However, it is recommended to run the Database on one server, and to run the Backup on another server.   By separating these servers, Database performance is not affected by the backup activity, and the system is able to provide consistent and predictable performance.
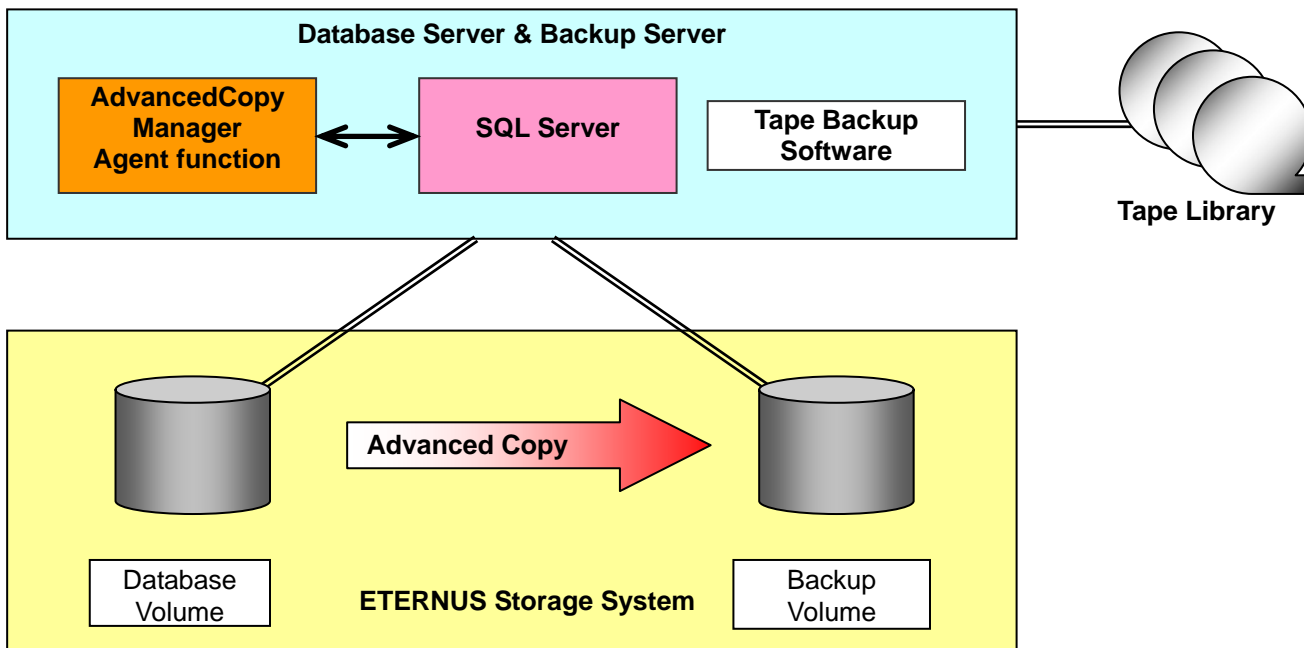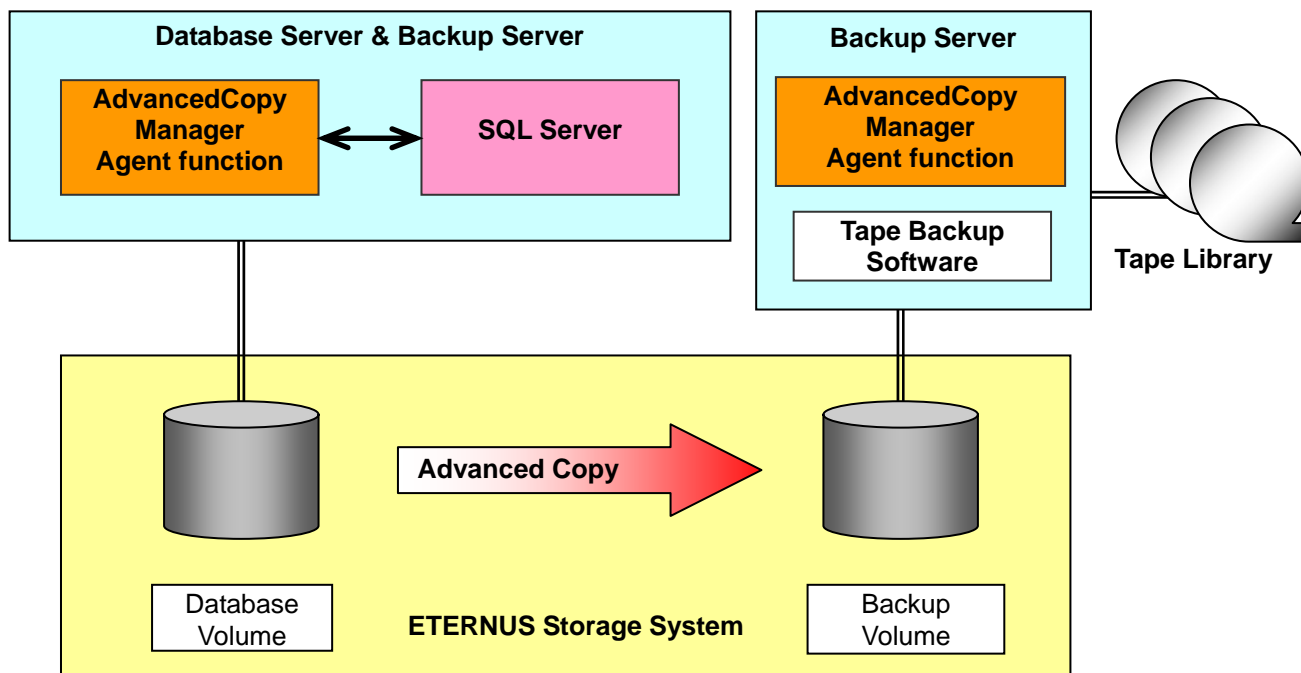
Figure-4 Database and Backup running on a same server

Figure-5 Database and Backup running on separate servers (recommended)

4-6. Performance Tuning

Database operations sometimes require performance tuning after the production operations have started. ETERNUS storage systems provide a Logical Volume migration feature that enables relocation of logical volumes from one RAID group to another in a manner that is totally transparent to the Windows Operating System and SQL Server.

When a Logical Volume Migration is invoked, the data of the targeted logical volume is copied from current RAID group, such as RAID-5 with 10,000 rpm disk drives, to another RAID group, such as RAID10 with 15,000 rpm disk drives.   All the writes are applied to source and destination while copy is in progress, and when the copy is completed, then I/Os from servers are automatically switched from the old logical volume to the new logical volume.

5. Deployment of SQL Server on ETERNUS with online backup feature

Here is a typical sequence that customers would use to prepare the environment for SQL Server and start operating it.

The figure below depicts the block diagram of this example.
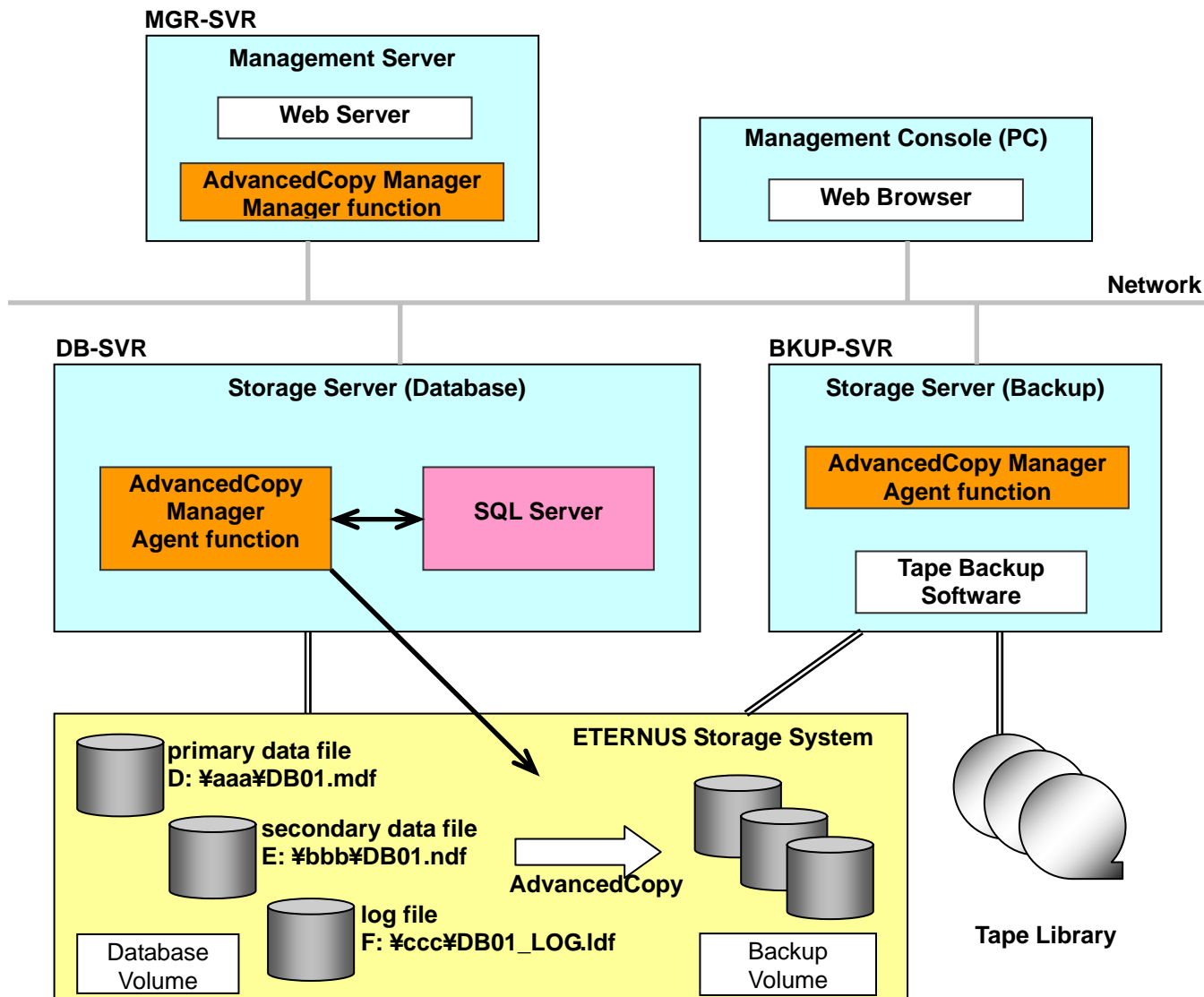
Figure-6 Block diagram of Example configuration

5-1. Server Configuration and Database configuration

In the above block diagram, there are two servers; Production Server and Backup Server. In addition to these servers, a Management Server and a console (PC) are required.

(1) Production Server
This is the server where SQL Server 2005 actually runs. In order to execute online backup, AdvancedCopy Manager Agent function needs to be installed on this server.

(2) Backup Server
This is the server to perform Backup operation from the copy volume in the ETERNUS to tape library. This server also requires AdvancedCopy Manager Agent function to be installed.

(3) Management Server and console
  This is the server to manage all the functions of AdvancedCopy Manager and requires AdvancedCopy Manager
  Manager function be installed.   This management function does not need to be executed on a separate server,
  and may be installed on the Production Server or a Backup Server.   For the command inputs and monitoring a PC
  running a browser is required.

There are 3 database volumes; Primary data file, secondary data file and Log file that make SQL Server
database and all these volumes need to be located on ETERNUS Storage System.   The ETERNUS SF
AdvancedCopy Manager copies volumes (partitions), so it is recommended that no other file is located on the
volumes of SQL database.   It is also recommended that a separate volume is allocated for Primary, secondary
and log file.

The following tables show the actual configuration used in the examples described in 5-2, 5-3 and 5-4.

<Server Configuration>

Server type	Server name	Comment
Management server	MGR-SVR	
Storage server	DB-SVR	Production server running on MSCS AdvancedCopy Manager's logical node name is "nodeAGT" SQL Server virtual server name is "VSERVER"
Storage server	BKUP-SVR	Backup server

<Database configuration>

Database name	Database file name	Database file type
DB01	D:¥SQLSVR¥DB01.mdf	primary data file
	E:¥SQLSVR¥DB01.ndf	secondary data file
	F:¥SQLSVR¥DB01.ldf	transaction log file
DB02	D:¥SQLSVR¥DB02.mdf	primary data file
	E:¥SQLSVR¥DB02.ndf	secondary data file
	F:¥SQLSVR¥DB02.ldf	transaction log file
DB03	D:¥SQLSVR¥DB03.mdf	primary data file
	E:¥SQLSVR¥DB03.ndf	secondary data file
	F:¥SQLSVR¥DB03.ldf	transaction log file

<Database volume allocation>

Drive letter	AdvancedCopy Manager device name
D:	g1d1p1@DB-SVR
E:	g1d2p1@DB-SVR
F:	g1d3p1@DB-SVR

(*)An AdvancedCopy Manager device name has the format: g?d?p? where "g", "d" and "p" indicates a disk
array system number, a Logical Unit Number, and a partition number respectively.

5-2. General Preparation

There are a few preparatory work items to be done to execute online back up of SQL Server.

(1) Preparation for Log application
If Log needs to be applied after restore operation, then the "recovery model" of the database needs to be set to
either "Full" or "Bulk logged."

(2) Setup of AdvancedCopy Manager
Access the url described below then register the servers, volumes and setup policies.
http://"the address of the Management server(:port address)"/swstorage/index.html

(3) Preparation of device map file
AdvancedCopy Manager has a capability to automatically allocate the backup volume from the pre-registered

pool.   However, if it is required to explicitly specify the backup volume, then a device mapping file needs to be created. This file defines the correlation between Database volumes and Backup volumes, and SQL Server backup command uses this file to find the appropriate backup volume.

The following is an example of the device mapping file.

```
# Device map file for ETERNUS SF AdvancedCopy Manager SQL Server Support Function

# database volume names         backup volume names
# (original volume names)       (replica volume names)

# Database DB01, DB02, DB03
g1d1p1                          g1d11p1@BKUP-SVR      # D: primary data file
g1d2p1                          g1d12p1@BKUP-SVR      # E: secondary data file
g1d3p1                          g1d13p1@BKUP-SVR      # F: log file

# Database DB04, DB05, DB06
g1d4p1                          g1d14p1@BKUP-SVR      # G: primary data file
g1d5p1                          g1d15p1@BKUP-SVR      # H: log file
:
```

(4) Recording the output path of metadata file
At the execution of backup, SQL Server creates a metadata file that contains path name of the database data file, the size, file type, etc.   AdvancedCopy Manager keeps this file in the backup volume.   This metadata file is put on database volume at the time or recovery.   This metadata file is required for the SQL Server restore command, so the absolute path name needs to be recorded.

(5) Automatic backup operation
Backup operation of AdvancedCopy Manager can be automated using Server Management Studio.

5-3. Execution of Backup operation

The backup of SQL Server is executed in three major steps.
-    Start Synchronization (if synchronized backup is used)
-    Backup the database
-    Backup the transaction log

(1) Start Synchronization
If Synchronized Backup operation is necessary then synchronization of database volume and backup volume needs to be completed before backup.   The following is an example of this operation.

```
C:¥>set SWSTGNODE=nodeAGT
C:¥>C:¥Win32App¥AdvancedCopyManager¥bin¥swsrpstartsync g1d1p1 g1d11p1@BKUP-SVR
FROM=g1d1p1@DB-SVR,TO=g1d11p1@BKUP-SVR swsrpstartsync completed
C:¥>C:¥Win32App¥AdvancedCopyManager¥bin¥swsrpstartsync g1d2p1 g1d12p1@BKUP-SVR
FROM=g1d2p1@DB-SVR,TO=g1d12p1@BKUP-SVR swsrpstartsync completed
C:¥>C:¥Win32App¥AdvancedCopyManager¥bin¥swsrpstartsync g1d3p1 g1d13p1@BKUP-SVR
FROM=g1d3p1@DB-SVR,TO=g1d13p1@BKUP-SVR swsrpstartsync completed
C:¥>
```

(2) Backup the database
Backup of the database is performed using SQL Server backup command (swsrpbackup_sql) through the steps described below.

a. By sending "BACKUP DATABASE WITH SNAPSHOT" statement,   AdvancedCopy   Manager instructs   SQL Server to begin a backup operation.

b. AdvancedCopy Manager receives the metadata information from SQL Server and then writes them to the metadata files.

c. All the I/Os are frozen by SQL Server while the backup operation is active.   The updates are all stored in SQL Server's buffer memory so the database files (.mdf, .ndf, and .ldf) do not change.

d. AdvancedCopy Manager either starts OPC or suspends EC to copy the database files into backup volumes.

e. SQL Server thaws the I/O freeze and all the updates in buffer memory as well as subsequent changes are applied to the database files.

The following is an example of taking backup of three databases; DB01, DB02 and DB03.   The device mapping table is G:¥SQLADM¥devmap.txt in this example.

```
C:¥>set SWSTGNODE=nodeAGT
C:¥>C:¥Win32App¥AdvancedCopyManager¥bin¥swsrpbackup_sql –Xserver VSERVER
–Xdevmap G:¥SQLADM¥devmap.txt DB01 DB02 DB03
DB01 DB02 DB03 swsrpbackup_sql completed
C:¥>
```

(3) Backup the transaction file
Backup of the transaction log is executed either by Management Studio or T-SQL.

5-4. Execution of Restore operation

The recovery of SQL Server is executed in three major steps.
-   Backup of the transaction log (if recovery to the most recent status is required)
-   Restoration of database
-   Recovery of database

(1) Backup of the transaction log
If the database needs to be recovered to the most recent status, then the transaction log needs to be backed up first.   The current active transaction log is backed up by executing "Backup Log" statement with "NO_TRUNCATE" option.   If this step is not executed, or if the transaction log cannot be backed up because of some failure, then the database cannot be recovered to the most recent status.

(2) Restoration of database
The restoration of database requires two steps; File restoration and database restoration.

a. File restoration
All the data files and log files need to be restored as follows.
-   Delete the database using Enterprise Manager.   If log is going to be applied later, the backup of the database and recovery history should not be deleted.
-   Copy the data on the backup volume to database volume using replication command "swsrpmake."

Note1: It is not possible to restore only log file.
Note2: Metadata will be restored at the same time as the database file.
Note3: If the backup volume has been stored in tape library using AdvancedCopy Manager, then data can be restored directly from tape to the database volume.

Below is an example of using Snapshot copy to restore the files.

```
C:¥>set SWSTGNODE=nodeAGT
C:¥>C:¥Win32App¥AdvancedCopyManager¥bin¥swsrpmake g1d11p1@BKUP-SVR g1d1p1
FROM=g1d11p1@BKUP-SVR,TO=g1d1p1@DB-SVR swsrpmake completed
C:¥>C:¥Win32App¥AdvancedCopyManager¥bin¥swsrpmake g1d12p1@BKUP-SVR g1d2p1
FROM=g1d12p1@BKUP-SVR,TO=g1d2p1@DB-SVR swsrpmake completed
C:¥>C:¥Win32App¥AdvancedCopyManager¥bin¥swsrpmake g1d13p1@BKUP-SVR g1d3p1
FROM=g1d13p1@BKUP-SVR,TO=g1d3p1@DB-SVR swsrpmake completed
C:¥>
```

b.   Database restoration
The database can be restored by SQL Server restore command "swsrprestore_sql".
Note1: The user who actually restored the database will automatically become the owner of the database.   If
the database owner needs to be changed, then use a stored procedure "sp_changedbowner".

Below is an example of the database restoration

```
C:¥>C:¥Win32App¥AdvancedCopyManager¥bin¥swsrprestore_sql –Xserver VSERVER –Xmeta
D:¥SQLSVR¥DB01.swsrp-dmp DB01
DB01 swsrprestore_sql completed
C:¥>C:¥Win32App¥AdvancedCopyManager¥bin¥swsrprestore_sql –Xserver VSERVER –Xmeta
D:¥SQLSVR¥DB02.swsrp-dmp DB02
DB02 swsrprestore_sql completed
C:¥>C:¥Win32App¥AdvancedCopyManager¥bin¥swsrprestore_sql –Xserver VSERVER –Xmeta
D:¥SQLSVR¥DB03.swsrp-dmp DB03
DB03 swsrprestore_sql completed
C:¥>
```

(3) Recovery of database
If the database needs to be recovered to the most recent point-in-time or to a particular time in the past, then
the log needs to be applied using Enterprise Manager.

6. ETERNUS Conformance to the Core Requirements

All models of Fujitsu ETERNUS3000, ETERNUS4000, ETERNUS6000 and ETERNUS8000 complies with the Core Requirements defined in the Microsoft Always On (v1.02) program as described below.

Core 1.00 Windows Logo Certification

All the ETERNUS storage systems are Microsoft Logo certified.   The record can be found at the URL listed below by entering "ETERNUS" in the search field and clicking Go.

http://www.windowsservercatalog.com/

Core 1.01 Core Windows API Support

All the ETERNUS storage systems support Windows API used by SQL Server.

Core 1.02 Stable Media

All the ETERNUS storage systems comply with the requirements of Write-Ahead Logging (WAL) to maintain ACID (Atomicity, Consistency, Integrity and Durability) properties of database and to ensure data integrity.

Core 1.03 Forced Unit Access (FUA) and Write Through

All the ETERNUS storage systems are equipped with mirrored cache write mechanism where any write data is stored in two physically separated cache memories in the storage system.   This means that even if one of the caches fails before the data in the cache is written back to the stable media (in this case, disk drive), the data is available in another cache, and will be written to stable media from there.

In addition to this protection mechanism, all the ETERNUS storage systems are equipped with battery backup mechanisms.   In case of power outage, ETERNUS4000, ETERNUS6000 and ETERNUS8000 use battery to save all the dirty data in cache (data written in the cache but not yet written on disk drive) to the designated system space on multiple disk drives before it powers itself down.   When power comes back, all the saved data is brought back to cache and is written to the stable media during the start-up process.   ETERNUS3000, on the other hand, is equipped with battery backed up cache memory.   In case of power outage, the battery unit in ETERNUS3000 keeps cache memory alive.

By having the above mentioned mechanisms, all the ETERNUS storage systems honor the write-to-stable media intent.

Core 1.04 Asynchronous Capabilities

All the ETERNUS storage systems support asynchronous I/O capabilities.   It does not convert asynchronous I/Os into synchronous operations in the storage system.

Core 1.05 Write Ordering

All the ETERNUS storage systems preserve the write ordering on each logical volume.   Even in the case of a database spanning more than one logical volume, it is not necessary to define any special grouping on these logical volumes since ETERNUS always honors the order of write commands.

In case of remote mirroring, if a database spans more than one logical volume in one ETERNUS storage system and these multiple logical volumes are copied in Asynchronous mode over the distance, then a

"Consistency Mode" of the Asynchronous mode needs to be enabled to preserve the write ordering among all the member volumes on the copy operation over the distance.

Core 1.06 Torn I/O Protection

All the ETERNUS storage systems store data in 512 byte per sector format on the disk drives. For RAID-5 and RAID-10 Groups, the stripe depth is 64KiB (65,536 bytes), ensuring that logical blocks (1K, 2K, 4K, 8K, 16K, 32K, and 64K) written to LBAs that are an exact multiple of the logical block size, will not be torn between different disk drives.

Core 1.07 NTFS Support

All the ETERNUS storage systems support NTFS capabilities.

Core 1.06 Torn I/O Protection

7. Technical references

The specification and features of all the ETERNUS storage systems are published on Fujitsu's website at the url listed below.

http://www.fujitsu.com/global/services/computing/storage/system/

About This White Paper

This white paper is devoted to providing technical information and an overview of the usage of Microsoft SQL Server 2005 & 2007 and Fujitsu ETERNUS storage systems.

The contents of this document may be modified without any prior notice. Please contact FUJITSU LIMITED if you find any error in the descriptions.   FUJITSU LIMITED is not responsible for any damage or loss that might be caused by the contents of this document.

FUJITSU LIMITED

Global Marketing
Platform Business Group
4-1-1, Kamikodanaka, Nakahara-ku, Kawasaki 211-8588 Japan

E-MAIL: storage-system@fujitsu.com
"Fujitsu Storage System Web Site" http://www.fujitsu.com/storage/