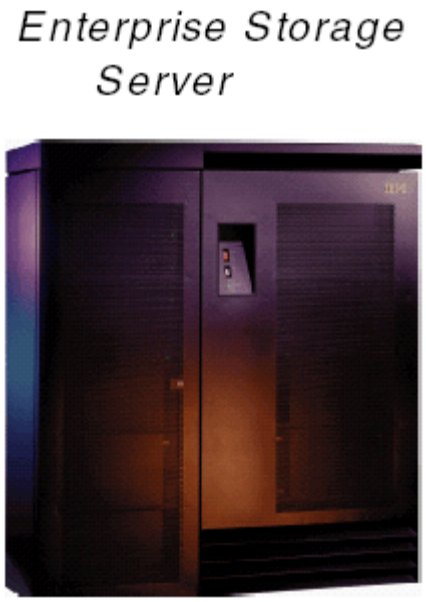
Storage Solutions for Oracle Database: Snapshot Backup and Recovery with IBM TotalStorage™ Enterprise Storage Server



DISCLAIMERS

The information in this publication is provided "AS IS" WITHOUT WARRANTY OF ANY KIND, implied or otherwise, including but not limited to any warranties of merchantability or fitness for a particular purpose. The responsibility for use of this information or implementation of any of these techniques lies with the customer and depends on the customer's ability to evaluate and integrate these techniques into their operating environment.

Accordingly, the recipient assumes all risks arising from the use of this information. In no event shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity. While IBM may have reviewed each item for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere.

The performance data contained in this document were obtained under a controlled, isolated environment. The results obtained in your operating environment may vary significantly. Accordingly, this does not constitute a performance guarantee or warranty.

Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Copyrights

The following terms are either registered trademarks or trademarks of the IBM Corporation in the United States or other countries or both:
Enterprise Storage Server, ESS, ESCON, FlashCopy, IBM, Netfinity, PPRC, Peer-to-Peer Remote Copy, Seascape, StorWatch, TotalStorage™, Tivoli, Tivoli Management Environment, TME,

The following terms are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries:

Microsoft, Windows, Windows NT, Windows 2000, BackOffice

Pentium, Pentium III Xeon, Pentium II Xeon and Intel are registered trademarks of the Intel Corporation.

Oracle is a trademark of Oracle Corporation in the United States, Other countries, or both.

McDATA is a registered trademark of McDATA Corporation.

QLogic is a registered trademark of QLogic Corporation.

Other product names mentioned herein may be trademarks of their respective companies.

Table of Contents

Executive Summary

As organizations move towards the need for constant availability of business-critical data, they need to have a fast, effective enterprise-class storage management solution to handle their business needs. One of the main objective for Storage Management is to maxize data availability and performance; IBM, using the latest snapshot technology, introduces high availability, zero downtime, and easy to use Oracle Snapshot Backup and Recovery solutions with IBM TotalStorage™ Enterprise Storage Server(ESS) to do the job.

This white paper describes how a customer can use the Enterprise Storage Server's Advanced Copy Services functions Peer-to-Peer Remote Copy (PPRC) and FlashCopy to perform snapshot backup for Oracle Databases. PPRC is a synchronous I/O mirroring technique between two peer ESS Storage Subsystems; it can help protect data from disaster and help to preserve data integrity. Flashcopy is a fast point-in-time copy technique designed to replicate all data on an entire ESS volume to another ESS volume within the same logical Subsystem(LSS). It can perform serverless backups of multi-terabyte databases in minutes. It can also improve uptime by providing the ability to recover in minutes, instead of hours or days. When used in combination, these two Copy Services have the ability to provide multiple, instant copies of the database (for serverless backup, data mining, application development or testing, and other usages) and can be used as part of overall plan to protect against disaster.

1 Introduction

In a critical computing environment, having the ability to quickly backup large databases without jeopardizing data integrity or performance of the online service is a very important aspect of today's technology trend. One of the recently developed and implemented technology, Snapshot Storage Systems, using IBM TotalStorage™ Enterprise Storage Server (ESS) and its Advanced Copy Services functions combined with Oracle's backup recovery features can fulfill this requirement.

This paper describes the use of the IBM TotalStorage™ Enterprise Storage Server (ESS) and its Advanced Copy Services PPRC and FlashCopy functions to perform snapshot backup and recovery for Oracle Database, using a set of documented scenarios. A snapshot is basically a copy image of storage devices or file systems on a secondary local or remote systems at a particular point in time. It duplicates data from standard storage devices by using the snapshot of mirror techniques or copy on write techniques. Snapshots can be accessed for read, while applications are running and modifying the original "live" data. The ESS creates the copy within its storage subsystems, using Advanced Copy Services designed specifically for this purpose. The techniques discussed in this paper leverage features of the ESS and Oracle Databases to allow rapid creation of copies of data, without compromising data integrity or performance of online operations.

2 Customer Requirements

2A Backup Requirements

As customers rapidly move to a 24x7 global computing environment, the backup window is shrinking just as rapidly. Customers can no longer afford the daily downtime on their production server to perform backups for a certain amount of times. This becomes especially true as databases become larger and larger. Online backups improve the availability of the database, but cost valuable host CPU, disk , and network resources. Even with the use of incremental and differential backups, the time necessary to perform a backup is significant.

2B Recovery Requirements

Should an error occur or disaster strike and recovery is needed, a restore often takes twice the amount of time as a backup (or more). During this time, the database will be unavailable. While backup techniques such as incremental and differential backups reduce the backup time significantly, they actually increase the amount of time required to restore.

Despite planned and unplanned outages, business needs often require that databases must be available within minutes. While disasters and hardware failures are rare, logical errors and software errors are more frequent. These errors require a time-consuming restore as well.
As databases become larger and contain more mission-critical data, they increasingly require higher availability. Conversely, the time required to backup and restore becomes longer.

2C System Cloning

Customers also require copies of a production system to do data mining and warehousing, application development or testing; this can be obtained by taking a system cloning using IBM Enterprise Storage Sever's Advanced Copy Services functions. System Cloning is a procedure for bringing up an existing system into operation on secondary server. If you want to create a copy of your production system in order to build a test system that contains data similar to the production system data, you can perform a system copy. There are two ways to do a system copy. One method is by using database-specific tools (tools provided by the database vendor). Some database vendors offer specific tools for copying a database. Using IBM Enterprise Storage Server's Copy Services functions, customer can make a system cloning. There are steps needs to follow after making the Flash Copy. For details steps, please reference whitepaper, "R/3 Homogenous System Copy - UNIX, Windows NT, OS/390 and OS400, release 4.6D".

3 Enterprise Storage Server Architecture and Overview

To help meet the challenges inherent in an enterprise-class database infrastructure, IBM introduced the Enterprise Storage Server, codenamed "Shark." Its rock-solid reliability, high performance, and Advanced Copy Services make it well equipped to meet the challenges.

The ESS provides enterprise-class storage for the applications across an entire corporation. Based on a third-generation Seascape architecture, the ESS is designed to deliver scalability, availability, performance, and connectivity to nearly every enterprise class operating system over Fibre Channel and SCSI. Availability and performance can be further improved by using its Advanced Copy Services features, including FlashCopy and Peer-to-Peer Remote Copy.

3A Enterprise Storage Server Architecture

IBM designed the ESS specifically for the enterprises requiring high reliability and availability. IBM's focus on reliability and availability clearly shows in the advanced architecture and features of the ESS:

Advanced Internal Architecture

Each ESS consists of two complete storage subsystems attached to fault-tolerant disks in a clustered configuration. Each of the two subsystems, or clusters, contains two buses to connect to the hosts, and four internal storage controllers to attach to the disks. The ESS also contains vast amounts of cache, dedicated nonvolatile storage (NVS), and intelligent load-balancing and parallelism. The result is a high performance, highly available storage server that is designed to avoid single points of failure. Figure 2 below shows the multiple paths to data available within the internal architecture of the ESS:

-

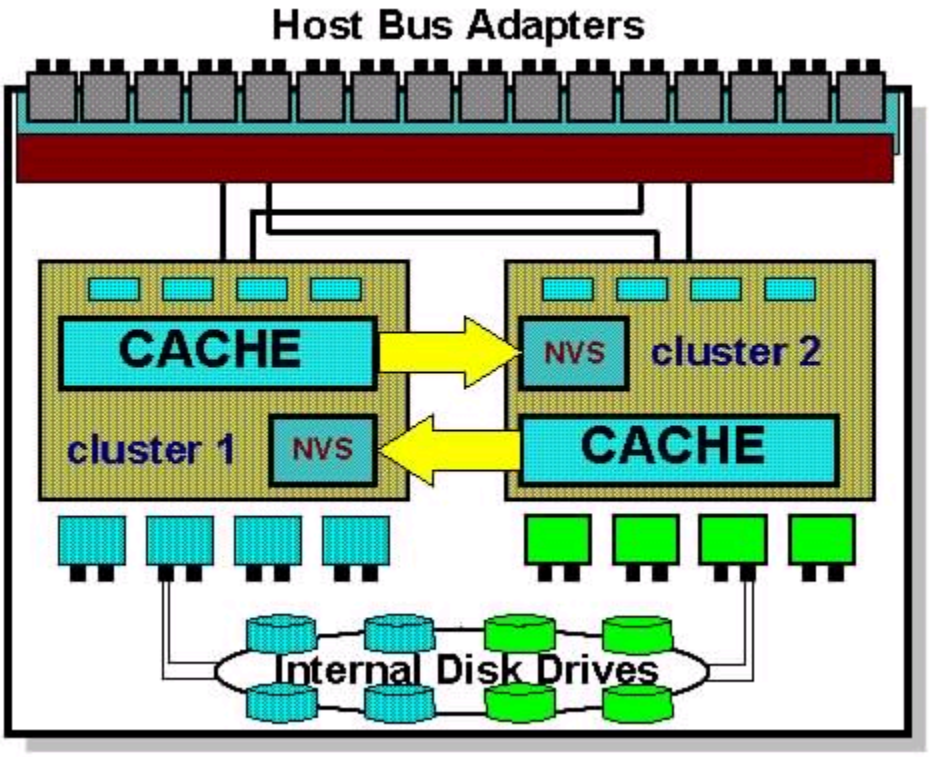


Figure 2 : The Enterprise Storage Servers's Internal Architecture

Advanced External Connectivity

The ESS can connect to multiple host systems using numerous connections. These connections can be multipathed for greater performance and redundancy using IBM's Subsystem Device Driver (SDD). The SDD software, installed on the host, automatically balances the I/O load across all available connections from the host to the ESS. In case of a connection loss, such as during a host bus adapter failure, SDD is designed to automatically failover to another path. These features support the host servers' ability to have fast, continuous access to data.

Other ESS Features

These features, along with support for non-disruptive upgrades and repairs, help provide the availability required in an enterprise-class solution. For more information about the ESS and other features of the ESS (such as SAN-capability, ease of management, and universal data consolidation), please visit the IBM storage web site at http://www.storage.ibm.com

3B Enterprise Storage Server Advanced Copy Services - FlashCopy

To further improve availability and performance, the IBM ESS features Advanced Copy Services. These serverless copy functions are implemented entirely within the storage server, with little or no impact on the host server. Figure 3 and Figure 4 below show how FlashCopy and PPRC works.

Flashcopy

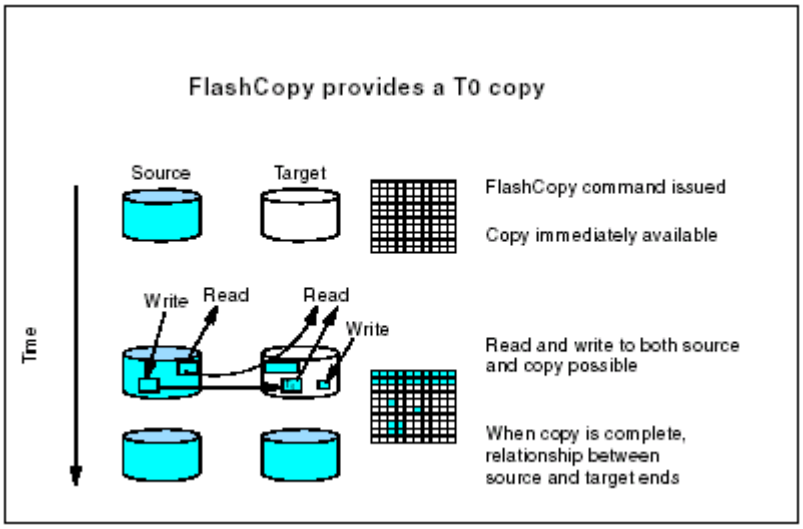


Figure 3 FlashCopy

FlashCopy provides an instant T0 copy of data. A T0 copy is a point in time copy of the storage volumes when the FlashCopy was invoked. The copy is available almost immediately after invoking the command. The copy process is transparent to the host server, as the ESS storage subsystems manage the copy internally. The point in time copy can now be accessed by a secondary server and be backed up to tape, typically with limited impact on the host server.

With FlashCopy, multi-terabyte databases can often be backed up and restored in minutes, instead of hours or days. Even for a tape subsystem capable of 100 GB/hr, a 1 TB database might take ten hours to backup. A restore from tape will take at least the same amount of time. Using the FlashCopy, the same database can be backed up almost instantaneously. Furthermore, the database often can be restored in minutes. Aside from backup and restore, FlashCopy can also provide a copy of "near-live" data clone for business intelligence or data mining, application testing and development.

When FlashCopy is invoked, the ESS creates another copy of the data by building a bitmap that records changed data. When the bitmap is complete (typically in seconds), the copy is logically complete. Both copies can immediately be used (read and write) separately without affecting the

other copy. If background copy option is selected, the ESS then begins physically copying the data to the target set of disks (at a rate much faster than a host server could copy). Whenever the host server writes to a block that has not yet been copied, the data is copied first, and then the write continues. In this way, a perfect T0 copy can be made instantly with little or no impact to the host server. When doing FlashCopying, NOCOPY option(Copy-on-Write) can also be used if you need the copy only for a short period of time. You must manually withdraw the pair to end the relationship and delete the target after you are done using the copy. In our validation environment, we used Copy-on-Write method to do our scenarios validation. Customers are recommended to use FlashCopy with background copy based on the requirements.

3C Enterprise Storage Server Advanced Copy Services - Peer-to-Peer Remote Copy

Peer-to-Peer remote copy

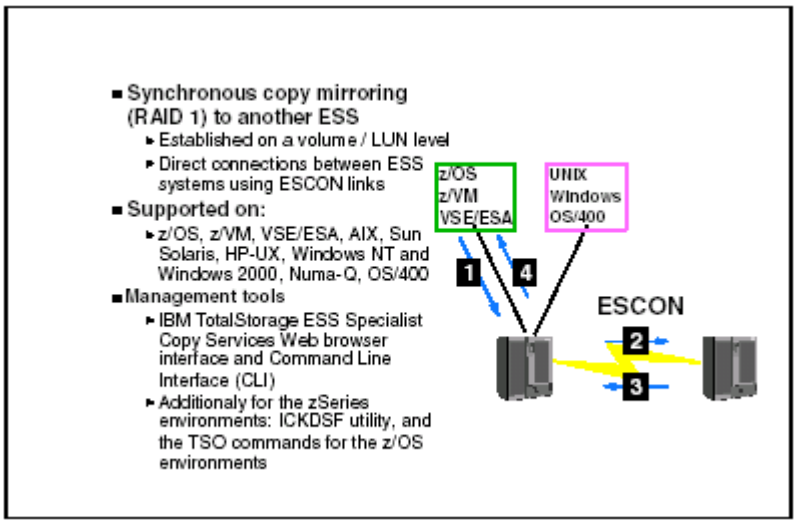


Figure 4 : Synchronous volumes copy - PPRC

PPRC is a remote mirroring technology (similar to RAID-1) that is designed to provides a synchronous copy of production data at a remote site. It is syncronous in that I/O is not completed until it was acknowledged from the remote site. Because the mirroring is done at the storage subsystem level, there is almost no impact on the host. PPRC is setup on a volume or LUN basis. Two or more ESSs are connected by ESCON links. The remote site can be up to 103 km away, or more with channel extenders. In case of a disaster at the production site, such as fire or earthquake, an up-to-the-second copy can be available at the remote site. The remote copy can be brought online using a standby server, usually with minimal business interruption.

Note: Before PPRC pairs can be established, logical paths must be defined between the logical control unit images. For more information about how to setup the PPRC, please refer to IBM Enterprise Storage Server, IBM Redbook, SG24-5465-00

4 Oracle Features and Requirements

In order to have a consistent image backups of the production database using Oracle's Offline(cold) backup and Online(hot) backup for the Snapshot Backup and Recovery solutions, we need to understand some of the key Oracle features. In validating the seven documented scenarios in our Split Mirror Backup Recovery environment, disk image backups play an important role in this backup and recovery procedures. Scenario1 and 2 involve with offline backup. Scenario 3 and 4 are online(hot) backup. We also developed three recovery scenarios.

For very large Oracle databases, the Snapshot backup capability is essential for the creation of reliable database backup copies without stopping the production system. This can be done with Oracle online backups. Online backups can be performed while the database is open. On the production system, Oracle database must be running in ARCHIVE mode because all the online redo logs and archive redo logs produce during Oracle database in online hot backup modes are needed for the recovery to produce a consistent point-in-time copy. By running Oracle Database in ARCHIVE mode, this provides 24 hours availability. Oracle features such as "ALTER TABLESPACE BEGIN BACKUP" and "ALTER TABLESPACE END BACKUP" for online(hot) backup capability are used to recreate backup copies of the primary database with little impact on the production OLTP system, database or user activities. Oracle Database needs to maintain datafile headers in a consistent state when they are being copied; this is the reason why we need to apply BEGIN BACKUP and END BACKUP commands. The ALTER TABLESPACE BEGIN BACKUP command will begin logging entire block images on the first change that Oracle encounters on each block owing to the DML activity (such as Insert, Update, Delete). When the Tablespaces are in backup mode, Oracle starts logging entire image before and after block images to the entire
Redo Logs. Oracle recommend to backup online tablespace serially because it minimizes the time between ALTER TABLESPACE ... BEGIN/END BACKUP statements. During online backups, more redo information is generated for the tablespace. Because FlashCopy performs a copy rapidly so it does not necessary to put tablespace in backup mode serially. The ALTER TABLESPACE END BACKUP command creates a redo log record containing the Oracle marker, BEGIN BACKUP checkpoint, also know as System Change Number (SCN). This SCN is also recorded in the header of the HOT BACKUP datafiles and ensures that all the redo generated during the backup has been applied to the datafiles. During an Online Backup, the Oracle database remain available. All transactions that are logged in the Redo log files during this backup period are required to be applied to the backup copy of the database to produce a consistent point-in-time copy.

There are two steps which are very important that need to be done after finish online backup all the datafiles. They are archiving the current online redo log files and backing up the current control files for the recovery. Archving the current online redo log files is very important because they contain a full before and after image of every block modified during the BACKUP period. This enables recovery from the Split Block phenomenon. They also contain the redo information required to rollforward datafiles as required to produce a consistent database. To have Oracle archive the current redo log files, Oracle feactures the command such as "ALTER SYSTEM ARCHIVE LOG CURRENT". This command causes Oracle to switch to a new log file. Oracle

then archive all the redo log files that have not yet been archived. This is same as issuing the command "ALTER SYSTEM SWITCH LOGFILE" and "ALTER SYSTEM ARCHIVE LOG ALL". After archived the current redo log files, the next step is to backup up all of the archive log files. Backup the current controlfile is also very important because it contains the SCN of the checkpoint completed after backup mode ended. When starting Oracle on the secondary, it is the mismatch between the SCN in the backup controfile and the SCN contained in each of the datafile headers that alerts Oracle to the need for recovery. In the normal method of recovery, you restore a backup to replace a damaged or destroyed datafile. Oracle compares the SCN in the datafile header with the SCN in the controlfile. Since the backup was made some time earlier, the SCN in the datafile is before the current SCN. Oracle knows transactions in archived redo log are required to bring that datafile up to the current SCN, so backing up the current controlfile is necessary. In Backup up the current controlfiles of the database, Oracle features "ALTER DATABASE BACKUP CONTROLFILE TO 'filename' REUSE". This command makes backup controlfile for starting up the database. TRACE option of the ALTER DATABASE BACKUP CONTROLFILE (TO TRACE)) statement can also be used if you want to change the SID, when the location or names of the datafiles need to be changed, or in order to change parameters established at database creation time.

Cold Backups are made with the database closed, which means that primary Oracle databases must be shutdown before doing the backups. Any file, be it datafile, redo log, control file, or archive log file is part of the backup. After the backup complete, start the instance. Database will then be up and running. Simple as they are to implement, offline backups do carry with them one big disadvantage: They require the database to be closed. In many corporations today, 24x7 availability is required, so shutting down the database for a backup is not an option. If you can afford the downtime to do an offline line backup, then that is probably the best way to do. Otherwise, you need to go with the approach of performing online backups.

It is strongly recommend that Hot Backups to be taken during periods of low IO or system activities. Hence the snapshot backup process, as demonstrated in this paper using the ESS's hardware assisted, near-instant copy (FlashCopy) and remote copy (Peer-To-Peer-Remote-Copy) functions, ensure that the backup process can be executed within a very short period of time, thus minimizing the impact to the production system.

Note: Oracle recommends to obey the Golden Rule of Backup and Recovery when developing backup and recovery strategy. The Golden Rule of Backup and Recovery is the set of disks or other media that contain the redundancy set should be separate from the disks that contain the datafiles, online redo logs, and control files. This helps prevent the failure of a disk that contains a datafile from cause the loss of the backups or redo logs needed to recover the datafile. For more information about how to follow Oracle's Golden Rule of backup and recovery, please reference Oracle 8i Backup and Recovery Guide.

5 Snapshot Infrastructure Setup and Configuration

To validate these methods of snapshot backup and recovery, we applied a set of test scenarios in the lab. We created an Snapshot Backup Recovery environment that was intended to closely resembles a typical customer environment. Within this section, we had included with very precise pictures of how we setup our SAN physical topology and what the split mirror infrastructure looks like.

5A Network Topology

These servers were connected into a Storage Area Network (SAN). The SAN had the following topology, shown in Figure 5 below:

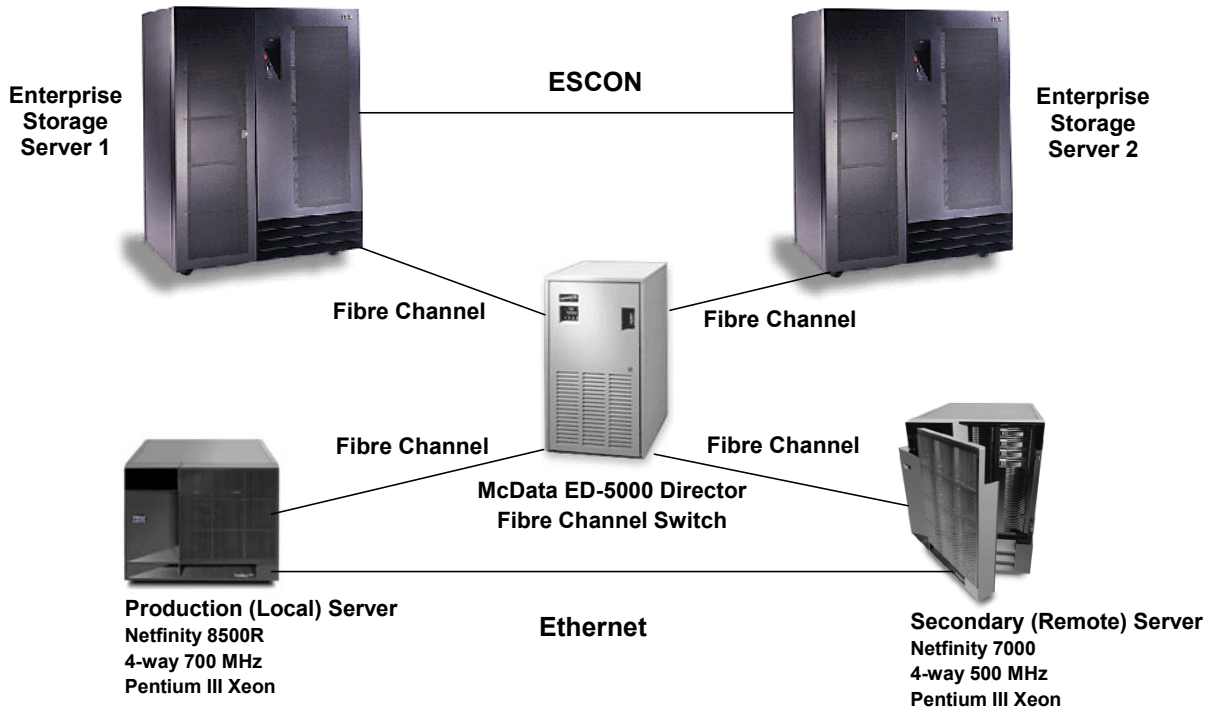


Figure 5 : The Storage Area Network Physical Topology

In this configuration, with multiple Fibre connection to McData switch via multiple ports cards and Subsystem Device Driver(SDD) supports fault tolerance. Even though the switch is an enterprise-class, fully redundant Director, the possibility of disaster and multiple failure still exists. The McData switch is meant to represent a SAN fabric, consisting of several redundant paths and switches from edge to edge.

5B Snapshot Infrastructure Setup

To attain higher availability, we devised a two-ESS and two-server configuration. The production server and ESS 1 are located at the production site, of course. The secondary server and ESS 2 are located at a remote site. The following figure is a logical view of the environment configuration:

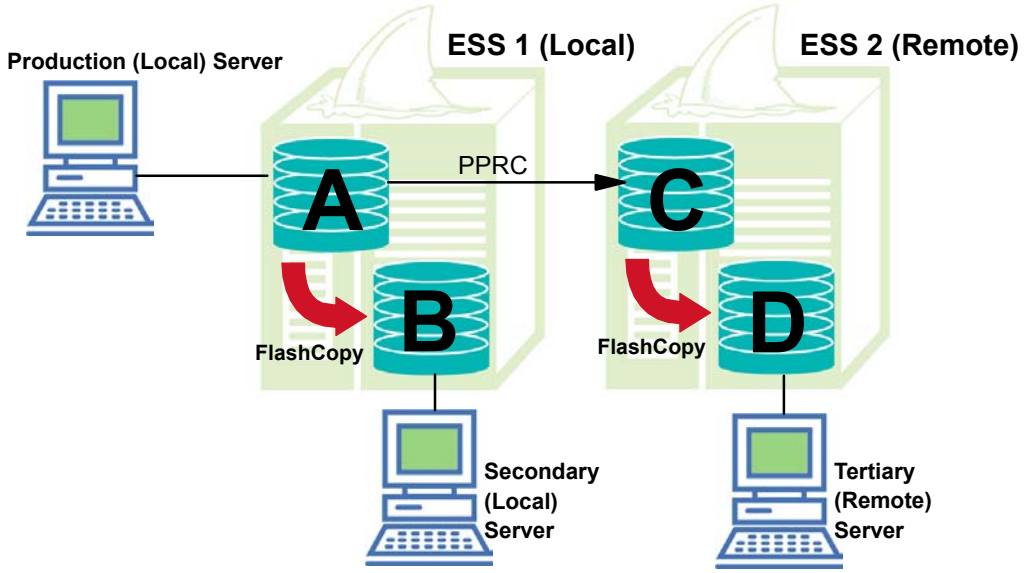


Figure 6 : Split Mirror Infrastructure

At the production site, the ESS contains two copies of the production database. The first copy (A) is the production database. Periodically (perhaps every night), the database is quiesced and FlashCopy (B) is created. The FlashCopy (B) is then sent to tape using the resources of a secondary server (not the production server). After that, the FlashCopy (B) is maintained in case a rapid restore is required. At the remote site, the second ESS keeps a synchronous copy (C) of the production data. Should the primary server fail, the backup server at the remote site can take over, typical with little or minimal business interruption. It is also possible to take a FlashCopy of that remote copy to produce a fourth copy (D) that can be used for additional protection, data mining or business intelligence, application development, etc.

Customers may choose to have a secondary server at the local site to send the backup to tape, and a tertiary server at the remote site to take over in case of disaster. In our test environment, our second server served as both the secondary and tertiary.

5C Hardware and Software Setup

The test environment contained two Intel-based IBM Netfinity Servers and two Enterprise Storage Servers all connected through an enterprise-class McData ED-5000 Fibre Channel Director. Using Fibre Channel, the Storage Area Network (SAN) provided a fast path to data without adding congestion to the IP network. Each of the Netfinity Servers connected to the SAN via two Fibre Channel host bus adapters (HBAs). Each of the ESSs connected to the SAN using four Fibre Channel adapters, and to each other over 8 ESCON links. In total, there were eight redundant logical paths from each host to each ESS. Each server was configured to use four of the eight available paths. The paths were logically combined through the use of IBM's multipathing software, Subsystem Device Driver (SDD). They appeared to Windows as a single logical path (i.e., one instance of each volume). Lists of detail Hardware and software configurations are shown in Appendix C:

5D Storage Configuration

We set up the Oracle database to spread the data and logs across all of the available disks in the ESS for best performance. We also spread the load across both clusters, all host bus adapters, and all internal logical subsystems (LSSs) available within the ESS. This helps avoid "hot spots" within the storage subsystem that receive most of the load while other parts of the subsystem are idle.

Furthermore, we also setup in a way that all of the online redo and archive redo logs were multiplexed on a separate set of disks. Oracle control file was multiplexed. The ESS is design to protect against single failures (such as a disk, adapter, CPU, or memory module failure), and many double and triple failures, with its fully redundant, highly available architecture. The possibility exists, however remote, that a combination of hardware failures can occur that would cause data loss (usually during a disaster). Having a mirror of the log is extra insurance against such a scenario. Additionally, in some scenarios, we have a mirror at a remote site to protect against
disasters, such as an earthquake or fire.

Oracle has recommended Strip and Mirror Everything **(S.A.M.E.)** method for the storage configuration. The idea of this configuration is to make extensive use of striping across large sets of disks in a storage configuration.

For more information on how to implement the methodology **S.A.M.E** please reference "Optimal Storage Configuration Made Easy" by Juan Loaiza, Oracle Corporation.

6 Snapshot Backup Process

This section discusses various situations in which customers might use FlashCopy and PPRC to perform backups. Restore scenarios are discussed in Section 7, "Recovery". Using FlashCopy and PPRC, it is possible to make dozens of copies in different configurations. We devised four likely scenarios a customer might use to perform a database backup

Backup Scenario 1 : Customer requires serverless fast backup and recovery
Solution : Offline Local FlashCopy backup
As with each of these scenarios, Scenario 1 provides an instant "consistent" copy of the database to perform a backup. For customers who require the ability to recover as quickly as possible, this scenario allows for a fast FlashCopy restore.

Online option is also available for scenario 1. Detail steps for online variant are described in subsequent sections.

Backup Scenario 2 : Customer requires disaster recovery + Scenario 1 requirements
Solution : Offline remote FlashCopy backup
With PPRC and FlashCopy, this scenario provides an instant copy on a second ESS. This scenario adds further protection and reduces or eliminates any impact on the primary storage subsystem. With the second ESS, the database can typically survive a total site failure.

Backup Scenario 3 : Customer requires 24x7 availability + Scenario 2 requirements
Solution : Online remote FlashCopy backup
Customers that require 24x7 availability can use this solution and achieve the disaster protection benefits of Scenario 2 without having to shut down the database.

Backup Scenario 4 : Customer requires 24x7 availability with extra copy + Scenario 3 requirements
Solution : Online local FlashCopy with additional remote FlashCopy backup
As with Scenario 3, this scenario provides 24x7 availability and the protection of a remote copy. Additionally, this scenario provides an extra copy of the database that can be used for functions such as data mining and warehousing, application development or testing, etc.

Note: Scenario 3 and scenario 4 are online hot backup. For more information about how to do online hot backup in Oracle Database, please reference Oracle 8i Backup and Recovery Guide for more detail.

While these backup scenarios performed as expected under significant load in the lab, we recommend that the backup be performed at times of minimal load (such as during the night). Each of our scenarios included error detection mechanisms. We recommend that customers use rigorous error checking in their implementation of these scenarios.

Warning: Do not start Oracle Database on the secondary system if the secondary database will be used for backup and restore. Instead, use an application such as NTBackup or Tivoli Storage

Manager to make a flat file-level backup. Restarting Oracle Database can create multiple parallel redo log timeline so that it can no longer be used for roll-forward recovery of the production database. The steps for starting the database on the secondary should only be used for system cloning.

Each of these scenarios are described in further detail in subsequent sections.

6A Backup Scenario 1 - Offline Local FlashCopy Backup

This scenario requires one ESS at the production site. While this scenario provides very fast backups and restores, it does not offer any protection against disasters, such as fire or earthquake. Logically, the data is being copied from copy (A) to copy (B) in the following Figure:

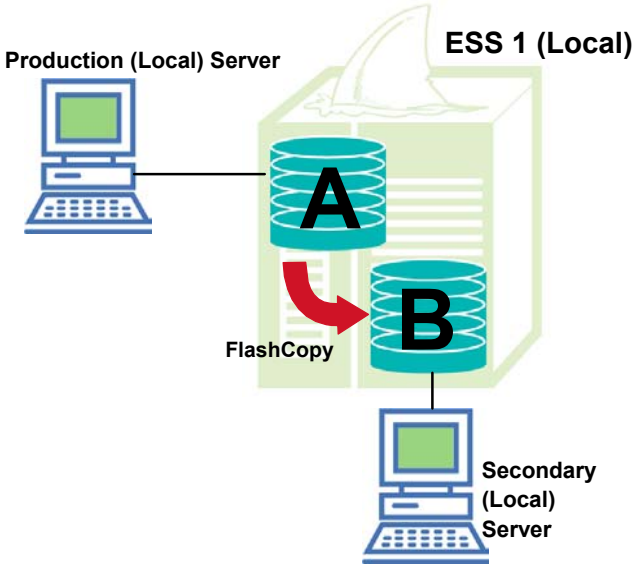


Figure 7: Backup Scenario 1 - Local FlashCopy Backup

Offline Variant
Customers that require a known, consistent state and can tolerate a brief outage should use an offline FlashCopy backup. The amount of time to require to perform FlashCopy is greatly reduce to perform Oracle Backup. The steps for performing the backup are shown as follows:

1. Stop all applications and Oracle Database on production server
2. Stop production Application services and Oracle services
3. Flush file system buffers on remote server
4. Perform FlashCopy Backup from copy (A) to copy (B)
5. Restart Application services and Oracle services on production server
6. Restart Oracle Database and Application on production server
7. Refresh secondary filesystem (i.e. make secondary server aware of new disks)

8. Restart all the Application and Oracle Services
9. Restart Oracle Database and Application at secondary server

Online Variant

If customers that require 24x7 availability can use a Scenario 1 variant to perform an online backup instead of an offline backup. While the backup is not necessarily in a known state, it can be made consistent using the redo generate by the online backup. In general, an online hot backup is a backup taken while a database is open. In this case, we use oracle's features " Alter tablespace begin backup" and " Alter Tablespace end backup" for online hot backup. The following are the steps to perform Local Safety Oracle online hot backup

1. Switch the redo logs and begin backup mode for all the tablespaces on production server.
   Alter System Archive Log Current
   Alter Tablespace Begin Backup
2. Flush file system buffers on remote server
3. Perform FlashCopy from copy (A) to copy (B) (only on the Datafile Volumes)
4. END backup mode for all the tablespaces and switch the redo logs on production server
   Alter Tablespace End Backup
   Alter System Archive Log Current
5. Make a backup controlfile in a directory that will be FlashCopied along with the Archive Redo Logs in stop 6 below
   Alter Database Backup Controlfile to (safe directory for the current controlfile) Reuse
6. Perform FlashCopy from copy (A) to copy (B) (Archive Logfile, Redo Logs Volumes and current controlfiles)
7. Refresh remote filesystem (i.e. make remote server aware of new disks)
8. Copy back all the current controlfiles to its directories for recovery on remote server. These controlfiles are the ones flashcopied along with the Archive RedoLogs in step 6. This controlfile needs to be copied to the location(s) listed in the initialization parameter file (init.ora) or it needs to be updated to contain the location of the current backup controlfile. This is the point where offline verification can be performed (using DBVerify) and a backup that is suitable for recovery can be made.
9. Restart Oracle Services and Application services on remote server
10. Recover Oracle Database on remote server
    Recover Database Using Backup Contolfile Until Cancel
    Alter Database Open Resetlogs
11. Restart Application on remote server - system (clone of the primary system) is up and running

Note: Switching the redo logs before putting the Tablaspace in Backup mode in not necessary, but it is recommended. The purpose of doing it is to save the recovery time.

6B Backup Scenario 2 - Remote FlashCopy Backup

To add protection from disasters (such as fire, earthquake, etc.), the three remaining scenarios require second ESS at the production site and one ESS at the remote site for disaster recovery. For both Scenarios 2 & 3, the FlashCopy is done from the remote synchronous mirror (C) to the remote copy (D), as shown in Figure below:

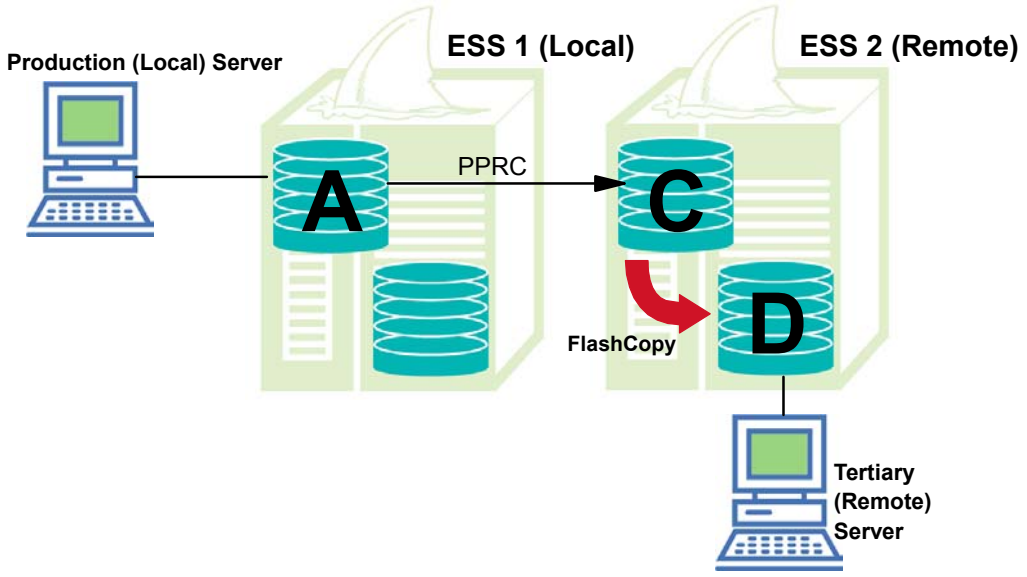


Figure 8 : Backup Scenarios 2 and 3 - Remote Flashcopy Backup

This setup provides the rapid backup functionality of Scenario 1 and adds disaster recovery protection. Furthermore, the FlashCopy (D) of the database at the remote ESS (ESS 2) can be used to send the backup to tape, eliminating any overhead on the production ESS and production server. At the beginning of the scenario, the database volumes should already be in a PPRC relationship and need to be syncronized, indicated by volumes in the "duplex" state. The following steps are involved in the offline remote FlashCopy backup:

1. Stop all applications and Oracle Database on production server
2. Stop production Application services and Oracle Services
3. Flush file system buffers on remote server
4. Perform FlashCopy Backup at remote ESS 2
5. Restart Oracle services and Application services on production server
6. Restart Oracle Database and Application on production server - system is back in production
7. Refresh remote filesystem (i.e. make remote server aware of new disks)
8. Restart Oracle and Application services on remote server
9. Restart Oracle Database and Application on remote server

6C Backup Scenario 3 - Online Remote FlashCopy Backup

This scenario is an online version of Scenario 2. The setup is as described in Figure 8 above. This scenario would be used by customers requiring 24x7 availability. Instead of shutting down Application and Oracle Database, we temporarily set Oracle Database in online Hot backup mode. The steps involved in performing the online remote FlashCopy backup are as follows:

1. Switch the redo logs and begin backup mode for all the tablespaces on production server
   Alter System Archive Log Current
   Alter Tablespace Begin Backup
2. Flush file system buffers on remote server
3. Perform FlashCopy at the remote ESS2 (only on the Datafile Volumes)
4. End backup mode for all the tablespaces and switch the redo logs on production server
   Alter Tablespace End Backup
   Alter System Archive Log Current
5. Copy the current controlfiles to a specific directory for recover on primary server
   Alter Database Backup Controlfile to (safe directory for the current controlfile) Reuse
6. Perform FlashCopy at the remote ESS2 (Archive Logfile, Redo Logs and the current controlfiles)
7. Refresh remote filesystem (i.e. make remote server aware of new disks)
8. Copy back all the current controlfiles to its directories for recovery on remote server. These controlfiles are the ones FlashCopied along with the Archive Redo Logs in step 6. These controlfiles need to be copied to the location(s) listed in the initialization parameter file (init.ora) or the pfile needs to be updated to contain the location of the current backup contolfile. This is the point where offline verification can be performed (using DBVerify) and a backup that is suitable for recovery can be made.
9. Start Oracle Services and Application services on remote server
10. Recover Oracle Database on remote server
    Recover Database Using Backup Contolfile Until Cancel
    Alter Database Open Resetlogs
11. Start Application on remote server - System (clone of the production system) is up and running

Faster Backup Variant for Very Large Database Customers

For large, multi-terabyte databases, FlashCopy may take minutes to complete. Some applications may not tolerate I/O being suspended for that length of time. This time can be greatly reduced by using the suspend feature of PPRC. In this situation, the I/O is only frozen for the time it takes to suspend the PPRC link (typically a few seconds). Then, the remote mirror is in a consistent state, and I/O resumes on the production server. Now, the FlashCopy backup can take minutes without impacting the production server. As soon as the FlashCopy is logically complete, the PPRC link is resynched (the transactions that occurred while PPRC was suspended is written to the remote mirror), and the mirrors are in sync again.

Note: In this variant of Scenario 3, the mirrors are out of sync for the few minutes it takes to perform the FlashCopy. Should a disaster strike at the production site, the transactions that occurred during FlashCopy backup may be lost. Therefore, this variant is not appropriate for all customers.

Here are the steps in performing this type of backup:
1. Switch the redo logs and begin backup mode for all the tablespaces on production server.
2. Flush file system on remote server
3. Suspend PPRC to remote ESS 2
4. Perform FlashCopy (only the Datafile volumes) at remote ESS 2
5. Resync PPRC to remote ESS 2
6. End backup mode for all the tablespaces and switch the logfiles on production server
7. Copy the current controlfiles to a specific directory for recover on primary server
8. Suspend PPRC to remote ESS 2
9. Perform FlashCopy(only the Archive Logfiles, Redo Logs and the current controlfiles) at remote ESS 2
10. Resync PPRC to remote ESS 2
11. Refresh remote filesystem (i.e. make remote server aware of new disks)
12. Recover Oracle Database using the current controlfiles
13. Start Oracle Database at remote server
14. Start Application Services on remote server

6D Backup Scenario 4 - Online Local FlashCopy with Extra Remote Copy

In this comprehensive solution, which uses the technologies from the three previous scenarios, there are four copies of the database at any time. Copy (A) is the production copy; (C) is its synchronous mirror for disaster recovery. Copy (B) is a local FlashCopy that can be used for extremely rapid restores, while (D) is a copy that can be used to backup to tape, data mining, and other functions. This scenario would be used by customers requiring 24x7 availability, extremely rapid restores, and disaster recovery capability. As with the previous two scenarios, Scenario 4 requires one ESS at the production site and one ESS at the remote site for disaster recovery. The logical view of this scenario is described in figure 9 below:

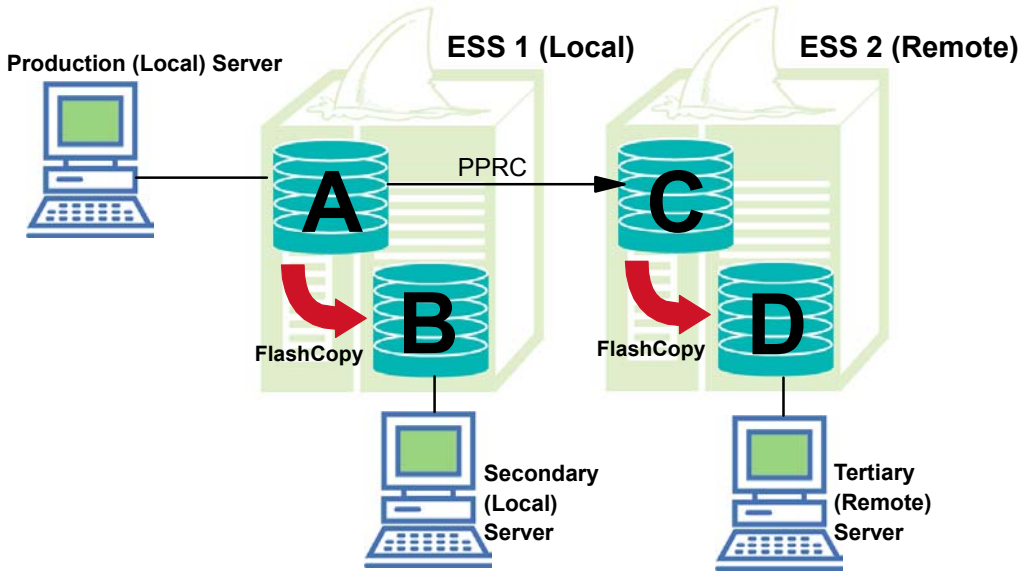


Figure 9 : Backup Scenario 4 - Online Local FlashCopy with Extra Remote Copy

Oracle Online(hot) Backup

As in Scenario 3, we set Oracle Database in Online Hot Backup Mode. While the FlashCopy tasks on the production (Step 3) and remote ESSs (Step 6) are usually done at the same time, these copies can also be individually scheduled at different intervals to allow the customer to recover to different points in time. This allows for the maximum level of flexibility for the customer.

1. Switch the redo logfiles and begin backup mode for all the tablespaces on production server
   Alter System Archive Log Current
   Alter Tablespace Begin Backup
2. Flush file system buffers on remote server
3. Perform FlashCopy at the remote ESS2 (only on the Datafile Volumes)
4. Perform Safety FlashCopy at ESS1(Only on the Datafile Volumes)
5. End backup mode for all the tablespaces and switch the redo logs on production server
   Alter Tablespace End Backup
   Alter System Archive Log Current
6. Copy the current controlfiles to a specific directory for recovery on production server
   Alter Database Backup Controlfile to (safe directory for the current controlfile) Reuse
7. Perform FlashCopy at the remote ESS2 (Archive Logfiles, Redo Logs and the current controlfile)
8. Perform FlashCopy at the ESS1 (Archive Logfiles, Redo Logs and the current controlfile)
9. Refresh remote filesystem (i.e. make remote server aware of new disks)
10. Copy back all the current controlfiles to its directories for recovery on remote server. These controlfiles are the ones flashcopied along with the Archive RedoLogs in step 6. These controlfiles need to be copied to the location(s) listed in the initialization parameter file

(init.ora) or the pfile needs to be updated to contain the location of the current backup contolfile. This is the point where offline verification can be performed (using DBVerify) and a backup that is suitable for recovery can be made.

11. Restart Oracle Services and Application services on remote server
12. Recover Oracle Database using the current controlfiles on remote server

    Recover Database Using Backup Contolfile Until Cancel

    Alter Database Open Resetlogs
13. Start Application on the remote server - System (clone of the production system) is up and running
14. Optional - bring up the Oracle and Application on the Safety Copy on ESS1 copy (B)

Warning: Do not bring up safety copy on ESS1 copy(B) if copy(B) will be used for restore. Otherwise, it will create parallel redo log timelines so that it can no longer be used for rollforward recovery.

7 Recovery Process

In addition to the four backup scenarios, we devised three recovery scenarios (Scenario 1, 2, 3) that can be used to recover from most errors or disasters.

Recovery Scenario 1 : Customer requires disaster recovery
Solution : PPRC recovery
In the situation where the production database is corrupted or the production ESS lost (due to natural disasters, etc.) we can use the backup taken in Backup Scenario 3 to restore from the remote copy on ESS 2. While the synchronous nature of PPRC protects against disasters, it does not protect against logical errors; the errors are propagated to the PPRC Target on the remote ESS as well. This scenario would designed to protect against logical errors. It can also help protect against rolling disasters. The FlashCopy taken at the remote site can be used to restore the production database much faster than possible from tape. Optionally, if the log files are not damaged and contain no logical errors, they can be used to roll forward to the point of failure.

Recovery Scenario 2 : Customer requires point-in-time recovery
Solution : FlashCopy simple recovery
In the scenario where the production database is corrupted or contains logical errors, the customer would need to restore the database to a previous point-in-time from the FlashCopy backup. The FlashCopy backup taken in Scenario 1 contains a consistent backup, and can be used to restore the database.

Recovery Scenario 3 : Customer requires recovery to point-of-failure
Solution : FlashCopy full recovery
This scenario would be used by customers who can not afford to lose any data in case of a system failure. As long as control file and archive redo logs are intact, this scenario can be used to recover the database up to the point of failure.

Before attempting any recovery, we recommend making a copy of the online redo logs. we also recommend backing up the data files. Even if the data files are not startable, they may be repairable. While each of these scenarios have been fully validated in the lab, the backup provides yet another safety net for mission-critical data.

In each recovery scenario, it is assumed that Oracle Database and Application are already installed and all pertinent patches applied. Each of these scenarios are described in further detail in subsequent sections.

7A Recovery Scenario 1 - PPRC Recovery

This scenario requires two ESSs and assumes the database setup of Backup Scenario 3. Backup Scenario 3 makes periodic (perhaps daily) FlashCopy backups of the PPRC target on the remote ESS, and this backup is used to restore the production database on the primary ESS. Figure 10 shows the logical view of the scenario:

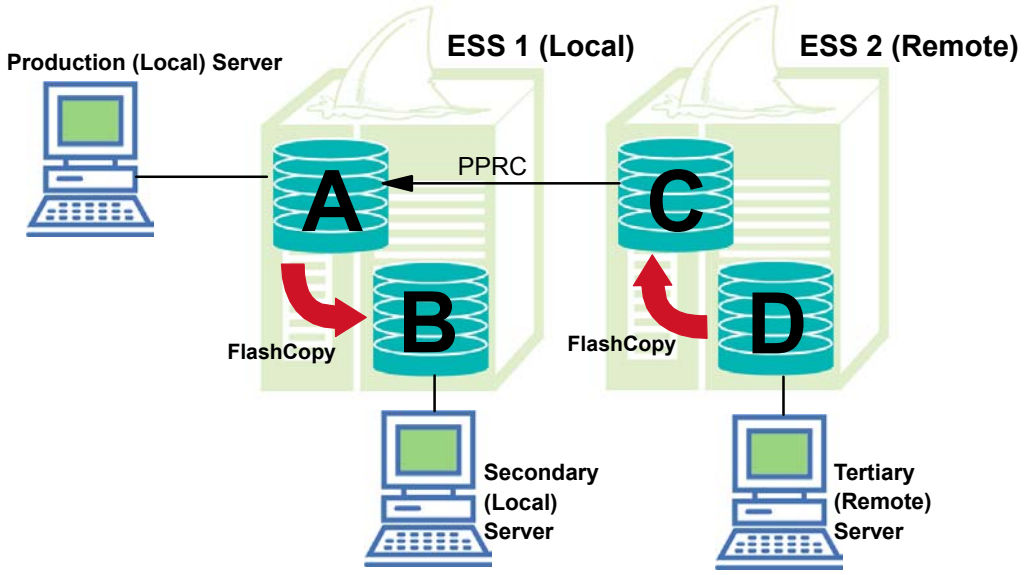


Figure 10 : Recovery Scenario 5 - Disaster Recovery

This scenario assumes that there is no local copy available. The steps in detail are:

1. Terminate PPRC relationship from production ESS 1 to remote ESS 2
2. Flush file system buffers on production system
3. Perform reverse FlashCopy at remote ESS 2 from copy (D) to copy (C)
4. Establish PPRC relationship from remote ESS 2 copy (C) to production ESS 1 copy (A)
5. Wait for PPRC to complete using ESS CLI tools that monitors the tasks
6. Terminate PPRC relationship from remote ESS 2 copy (C) to production ESS 2 copy (A)
7. Refresh production filesystem (i.e. make production system aware of new disks)
8. Perform Safety FlashCopy on ESS 1 from copy (A) to copy (B)
9. Start Application and Oracle services on production server
10. Start Oracle Database and Application on production server - system is back in production
11. Perform Withdraw Flashcopy on ESS 2 from copy (D) to copy (C)
12. Re-establish PPRC relationship from production ESS 1 copy (A) to remote ESS 2 copy (C)
13. Optional - Start Oracle and Application on the remote server and the Safety copy on ESS 1 copy (B)

Once the database is restored successfully (after step 10), we take a safety FlashCopy (B) to be used for rapid restores in case the production database becomes corrupt again. Then, at the conclusion of this scenario, we restore the ESS configurations to the production state, with the production database (A) being mirrored on the remote site (C) through PPRC.

Note: If Copy D was not brought up on the secondary server, then media recovery would be required on Copy A in order to have a consistent backup to the primary. If this recovery

mechanism is chosen, a new "Redo Log Timeline " is created, and any archive redo logs created after the backup was made can not be applied to this database. If brought up on secondary already have new redo timeline, it can be used to restore up to the point of end backup but can not be used for rollforward recovery.

7B Recovery Scenario 2 - FlashCopy Simple Recovery

While Recovery Scenario 1 provides protection against disasters, its method of restore can take several hours for very large databases. Though PPRC is usually faster than multiple tape drives, FlashCopy can provide even faster restores. With this scenario, the customer is able to restore to the time of the last backup within a very short period of time. In each case, as well as in the case of Recovery Scenario 3 (full roll forward recovery), the logical setup is as shown in Figure below:

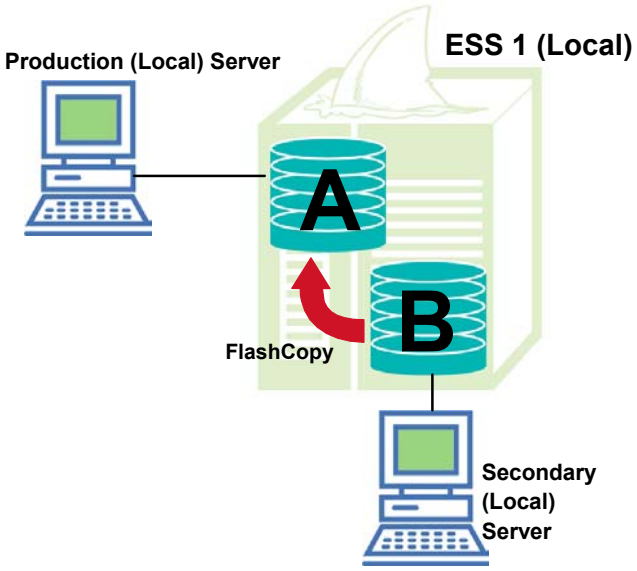


Figure 11 : Recovery Scenario 6 and 7 - Local FlashCopy Restore

This scenario is useful for recovering from logical errors. If a user accidentally drops a table, for example, a FlashCopy can return the database to a consistent, known prior state (the state at the time of the last backup). Any changes made after that last backup need to be redone (except the ones that caused logical errors, of course)

Oracle Database Simple Recovery

The Oracle database must not be in use during the recovery process. Scripts performed the following steps:

1. Stop all the applications and Oracle Database on remote server
2. Flush file system buffers on production system.
4. Perform a FlashCopy from copy (B) to copy (A)

5. Refresh production filesystem (i.e. make production system aware of new disks)
8. Restart Oracle and Application services on production server
9. Restart Oracle and Application on the production server - system is back in production

Note: this assumes that copy(B) was backed up with offline option.

7C Recovery Scenario 3 - FlashCopy Full (Roll Forward) Recovery

While Recovery Scenario 2's ability to recover to the time of the last backup is useful, many customers require the ability to recover to the moment of failure (or any arbitrary point in time). Recovery Scenario 3 provides this capability, while maintaining the fast restores.

Logically, the FlashCopy is from copy (B) to copy (A), shown in Figure 11 above. The details steps to perform Flashcopy Rollforward Recovery are shown below:

1. Use SSQJ to create some workload on production server
2. Stop SSQJ on Application console on production server
3. Stop all the applications and Oracle database on production server
4. Stop production Oracle services and Application services
5. Copy controlfile to a specific directory for later rollforward recovery on production server
6. Flush file system buffers on production system (Only Datafile volumes)
7. Perform a FlashCopy from copy (B) to copy (A) (only Datafile volumes)
8. Refresh production filesystem (i.e. make production system aware of new disks)
9. Copy back all the controlfiles to its directories for rollforward recovery
10. Restart Oracle services and Application services
11. Perform the rollforward recovery (Please reference Oracle 8i Backup and Recovery for more detail about rollforward recovery)
    Recover Database
    Alter Database Open
12. Start Application on production system - system is now available on production
13. Restart Oracle and Application on remote server

After applying the online redo logs, database will be recovered up to the point of failure. For more information about restoring Oracle Database, please reference *Oracle 8i Backup and Recovery.*

8 Process Automation and Solution Integration

In the lab, we developed various applications and utilities to fully integrate Oracle Database and ESS features to create a complete backup and recovery solution. We also created scripts to fully automate the Snapshot Backup Recovery scenarios described in previous sections. The scripts execute a customized set of snapshot routines set up for a particular customer environment. The customer can use these scripts, applications and utilities as a one-stop backup and restore solution for even the most complex storage configurations and customer requirements. The solutions developed are focused on providing rapid backups with little or no impact on the production environment, while also protecting against disasters using the remote mirroring technology of the ESS.

These solutions are designed to work in many different environments, and are compatible with various customer-preferred enterprise solutions management consoles like TME, OpenView or BMC. In establishing this snapshot backup and recovery solution for Oracle Database and ESS, IBM has created end-to-end procedures that are intended to enable this solution to seamlessly integrate into customer environments.

Appendix

A) Operating System Consideration

Snapshot disk-copying technologies have only recently gained popularity in the marketplace. Most operating systems predate these technologies, and as such, these operating systems were not designed to accommodate them. Therefore, until these operating systems fully integrate these advanced technologies, there are some aspects to consider when using FlashCopy and PPRC to ensure a smooth backup and restore. These aspects are related to volume management in Windows only, and not to Oracle Database. For more information on using FlashCopy and PPRC in a Windows environment, please consult the IBM Redbook, Implementing ESS Copy Services on UNIX and Windows NT/2000, available at http://www.redbooks.ibm.com. A few of the operational considerations affecting Windows are briefly outlined below:

Windows NT considerations

Both PPRC and FlashCopy are supported when using simple disks and fault-tolerant disks (such as volume sets). Consider these four tips when planning for Copy Services on Windows NT:

- For fault-tolerant disks, essential configuration information is stored in the Windows Registry (not on the actual disk). Therefore, when initially defining volume sets, (since FlashCopy and PPRC does not copy the data stored in the Windows Registry) a utility from IBM automatically performs this registry copy. Alternatively, a special procedure can be used, as outlined below:

  1. Define and format on primary system
  2. Document the Drive Signature, Partition Number, and Order Selected
  3. Perform FlashCopy
  4. Run FTEdit (available on the NT resource kit) on secondary and define volume set
  5. Repeat whenever volume set is changed

- After Service Pack 6, it is possible to have the FlashCopy source and target volumes accessible by the same server. In this case, use Disk Administrator to write a different disk signature on the target volume and assign a drive letter. Prior to Service Pack 6, the FlashCopy source and target volumes must be attached to different servers.

- To avoid rebooting after a FlashCopy, define an identical set of disks on the target machine. Then, use Disk Administrator to unassign the drive letter, perform the FlashCopy, reassign the drive letter, and the drive is immediately available.

- To flush file system buffers in Windows NT and 2000, use Disk Administrator to unmount (unassign) the drive letter. To fully automate the entire solution, you can develop a command-line utility to perform the file system flush without unassigning the drive letter, as we did in the lab.

Windows 2000 considerations

Windows 2000 supports two types of disks: basic disks and dynamic disks. Basic disks are the same as Windows NT disks with the same restrictions. For dynamic disks, Windows 2000 incorporates licensed technology from Veritas, called the Logical Disk Manager (LDM). The LDM can create five types of dynamic volumes: simple, spanned, mirrored, striped, and RAID-5.

On Windows NT the information relating to the disks is stored in the Windows NT registry. With Windows 2000, this information is stored on the disk drive itself in a partition called the LDM database, which is kept on the last few tracks of the disk. Each volume has its own 128 Bit Globally Unique Identifier (GUID). As the LDM database is stored on the physical drive itself, with Windows 2000 it is possible to move disk drives between different computers.

Having the drive information stored on the disk itself imposes some limitations when using Copy Services functionality on a Windows 2000 system with dynamic disks:

- The source and target volumes must be of the same physical size. Normally the target volume can be bigger than the source volume. With Windows 2000 this is not the case, for two reasons:

1. The LDM database holds information relating to the size of the volume. As this is copied from the source to the target, if the target volume is a different size from the source, then the database information will be incorrect, and the host system will return an exception.
2. The LDM database is stored at the end of the volume. The copy process is a track-by-track copy, and unless the target is an identical size to the source the LDM database will not be at the end of the target volume.

- It is not possible to have the source and target FlashCopy volume on the same Windows 2000 system. Each dynamic volume has its own 128 Bit Globally Unique Identifier (GUID). As its name implies, the GUID is unique to one system. When performing a FlashCopy, the GUID is copied as well, so this means that if you try to mount the source and target volume on the same host system, you would have two volumes with exactly the same GUID. This is not allowed and you will not be able to mount the target volume.

- Each disk contains information about every other dynamic disk on the system. Therefore, after a FlashCopy or PPRC, the information on the other disks may be rendered inaccurate. Windows only checks the information in the LDM database on bootup. While the disks can be used without rebooting, as in Windows NT, the drives will continue to work as expected only until the first reboot. The first reboot after a FlashCopy will require the use of Disk Management to "import" the volume. Therefore, we recommend rebooting the system and performing an import before using the disk.

B) Hardware and Software Configuration

Production Server
IBM Netfinity 8500R 8RY
4/8 700 MHz Pentium III Xeon processors w/2MB L2 cache each
2.0 GB memory
2 9.1 GB 7200 RPM internal SCSI disk drive
1 Qlogic 2200 Fibre Channel adapter (1 Gbit)
1 IBM Netfinity 10/100 Ethernet adapter
Windows NT Enterprise Edition, Service Pack6a
Windows NT Resource Kit
Oracle 8.0.6 32bit, Service Pack 8.0.6.1.0
Windows 2000 Advanced Server, Service Pack2
Windows 2000 Resource Kit
Oracle 8.1.7 32bit
Application Load Generator
SDD
CLI 1.3.3.27 for ESS with JDK 1.1.8.6

Secondary Server
IBM Netfinity 7000 M10
4-way 500 MHz Pentium II processors
1 GB memory
2 9.1 GB 7200 RPM internal SCSI disk drive
1 Qlogic 2300 Fibre Channel adapter (1 Gbit)
1 IBM Netfinity 10/100 Ethernet adapter
Windows NT Enterprise Edition, Service Pack 6a
Windows NT Resource Kit
Oracle 8.0.6 32bit, Service Pack 8.0.6.1.0
Windows 2000 Advanced Server, Service Pack 2
Windows 2000 Resource Kit
Oracle 8.1.7 32bit
Application Load Generator
SSQJ Software
SDD (Subsystem Device Driver)
CLI 1.3.3.27 for ESS with JDK 1.1.8.6

Production Storage Server
IBM Enterprise Storage Server F20
16 GB cache
64 18.2 GB 10000rpm HDDs
12 Fibre Channel adapters
4 dual port ESCON adapters

Backup Storage Server
IBM Enterprise Storage Server F20
16 GB cache
64 18.2 GB 10000rpm HDDs

12 Fibre Channel adapters
4 dual port ESCON adapters
Fibre Channel Switch
McData ED-5000 Fibre Channel Director with 32 ports

References

1. *IBM Enterprise Storage Server* -- IBM Redbook , Document Number SG24-5465-00, July 1999

2. *Oracle8i DBA Bible* - Jonathan Gennick, Carol McCullough-Dieter, and Gerrit-Jan Linker, 2000 IDG Books World Wide, Inc.

3. *IBM Enterprise Storage Server Performance Monitoring and Tuning Guide* -- IBM Redbook, Document Number SG24-5656-00, March 2000

4. "Storage Solutions for Microsoft SQL Server Snapshot Backup and Recovery with IBM's Enterprise Storage Server" -- White Paper, http://www.storage.ibm.com/hardsoft/diskdrls/technology.htm

5. "SAP R/3 Storage Management Split Mirror Backup & Recovery on IBM's Enterprise Storage Server on DB2 OS390" -- White Paper, http://www.storage.ibm.com/hardsoft/diskdrls/technology.htm

6. "Storage Management for SAP and Oracle Split Mirror Backup / Recovery with IBM's Enterprise Storage Server (ESS)" -- White paper, http://www.storage.ibm.com/hardsoft/diskdrls/technology.htm

7. "Backup and Restore Concepts for mySAP.com System Landscapes, Version 1.4" -- White Paper, http://service.sap.com/atg

8. *Implementing ESS Copy Services on UNIX and Windows NT/2000* -- IBM Redbook, Document Number SG24-5757-00, March 2001

9. "R/3 Homogenous System Copy - UNIX, Windows NT, OS/390 and OS400, release 4.6D" -- White Paper

10. *Oracle 8i Backup and Recovery Guide, Release 2 (8.1.6),* Parts Number A76993-01 - Connie Dialeris, Joyce Fee, Lance Ashdown, 1999