hp storage solutions

october 2002

solutions implementation blueprint

hp StorageWorks datasafe for mySAP.com (Windows 2000/Oracle)

overview

When data security and availability are critical to the success of their businesses, mySAP.com customers require a computing solution that protects their information systems from disasters, such as power outages, earthquakes, fires, floods, or acts of vandalism. The effects of a disaster range from temporary loss of availability to outright physical destruction of a facility and its assets. In the event of such a disaster, the mySAP.com system must allow customers to shift their information processing activities to another site as quickly as possible. Procedures for disaster recovery must therefore be predictable, well defined, and immune to human error.

Site replication is a method of achieving disaster tolerance in a mySAP.com environment. Disaster tolerance (DT) is characterized by a short recovery time and avoidance of data loss. In a disaster-tolerant system based on this approach, redundant, active servers and client interconnects are located at geographically separated sites. As mySAP.com applications produce data, this data is copied by a replication system whose function is to maintain consistent replicas of the data at each site. Should the system at one site suffer a disaster, mySAP.com instances that were running at the now disabled site can be failed over to a surviving site that has the resources to support them. The process of failing over a mySAP.com application to the target node involves making the application's replicated data accessible, and starting instances on the target node to restore application availability.

executive summary

The HP StorageWorks Datasafe solution for mySAP.com enhances the Microsoft high-availability features with the disaster tolerant capabilities of HP StorageWorks Data Replication Manager (DRM), maintaining application performance dependent on the distance between the two sites. DRM over an Internet Protocol intersite link is a cost-optimized, high-performance solution for greater distances if the quality of service of this network is maintained.

Replicating the entire SAP database is a robust, managed solution for SAP customers. The solution provides a failover/recovery time at a remote computing site measured in minutes. This scenario allows existing HP customers a straightforward enhancement of their environment using DRM with careful planning.

Replicating only SAP database redo log information via DRM using the Oracle Standby Database mechanism also provides DT functionality up to the latest transactional update. In addition, this scenario requires less bandwidth and allows database changes to be propagated with a timed delay at the target site to protect the standby database from human error. The tradeoff for this scenario is the additional management effort required to maintain the standby database and the Oracle database expertise necessary in case of a site failover.

A non-clustered two-node configuration using DRM for replicating the operating system boot disk in addition to SAP database replication is a cost-effective, entry level DT solution with a minimum of necessary system management.

table of contents

datasafe solution overview

This document describes a method for configuring a disaster-tolerant system distributed over distant computer sites by combining the HP StorageWorks Data Replication Manager (DRM) Solution Kit with Microsoft Server technology in an SAP environment. In a stretched MSCS cluster using DRM, some member systems reside at one site and the others reside at a different site. A mySAP.com application can run the database server on the initiator site and the corresponding central instance or one dialog instance on the target site. All I/O occurs on the storage subsystem on the initiator site under non-disaster conditions. The DRM has exclusive access to storage at the target site, to which it replicates the I/O performed on the initiator site's storage. If a significant failure occurs at the initiator site, data processing can be resumed at the target site where the data is intact.

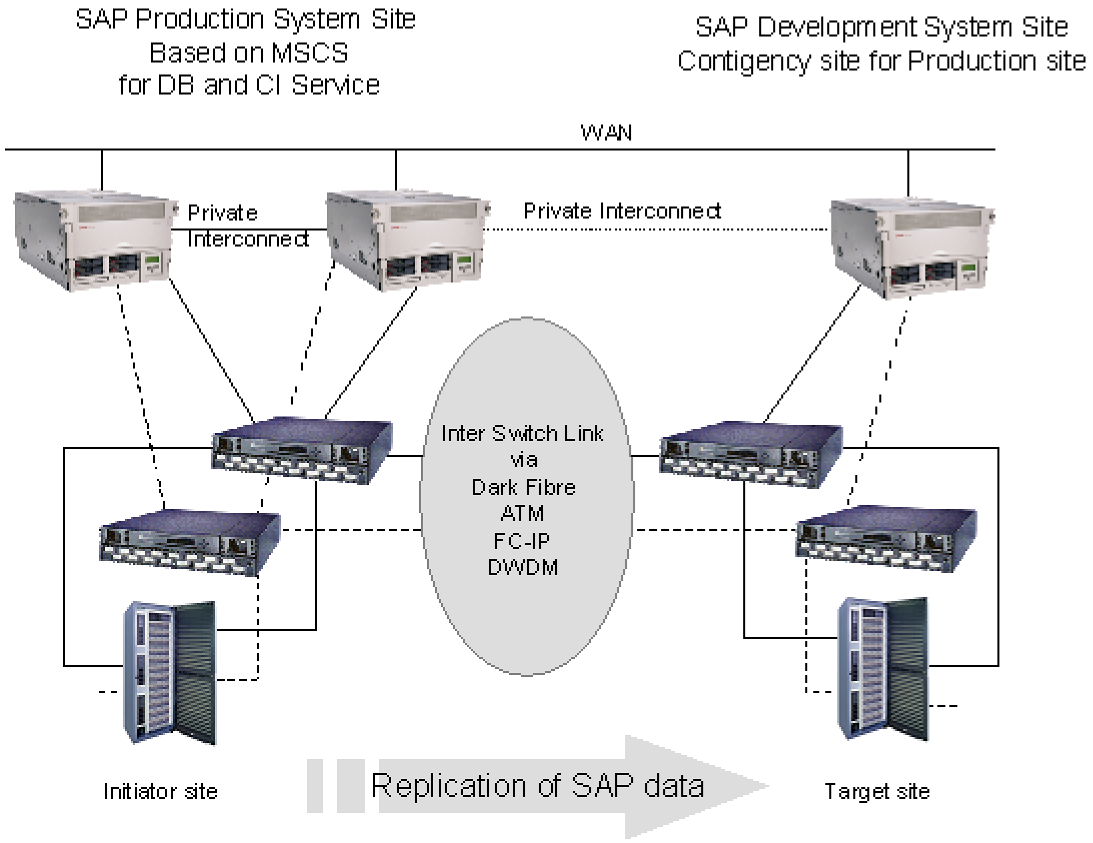


Figure 1

The Datasafe solution takes advantage of the best features of both the DRM Solution Kit and Microsoft Server technology. Cluster members can span distances across a commercial or college campus to a distance of up to 100 kilometers within a metropolitan area. Data replication hardware ensures correct and consistent mirroring across sites, while the HP management features for a Microsoft server environment allow you to manage all cluster members, regardless of whether they are at the local or remote site. These capabilities save time during normal system administration and recovery procedures. Although storage failover across sites is a manual process, cluster resources automatically restart mySAP.com applications at the target site when the systems are rebooted after a site failover is complete.

background

As customer applications and 24x7 access to data become business-critical, requirements for high-availability solutions with no single point of failure increase. Customers' ability to continue application processing and maintain data access in the event of a catastrophic disaster becomes critical to their business operations. Disaster-tolerant solutions provide high levels of availability with rapid data access recovery, no single point of failure, and continued data processing after the loss of one or more components of a configuration.

Please refer to the Glossary for terms and expressions used in the following sections.

data replication manager (DRM)

The Data Replication Manager Design Guide - Application Notes, describes DRM as a controller-based data replication software solution for disaster tolerance and data movement. DRM currently works with HSG80-based storage systems and allows all data to be mirrored between storage elements in two different storage arrays that can be in separate geographical locations as seen in Figure 1. Each I/O write access is sent to both storage locations, and reads occur only at the local storage location. DRM copies data online and in real time to remote locations via a local or extended storage area network (SAN).

DRM supports various options to connect the FC switches between the initiator and target site. Dark Fibre and FC-IP gateways are certified as well as ATM and DWDM. Replicating data between extended SANs over unlimited distances through a Fibre Channel-over-IP link is of special interest because in general it is more cost effective to provide bandwidth for a WAN instead of implementing a dedicated long distance Dark Fibre or WDM solution. For more information about DRM functionality, refer to the Features and Benefits of HSG80 Data Replication Manager - White Paper .

Currently most IP networks do not manage bandwidth to each individual connection. As traffic increases due to other demands on the network, bandwidth can be robbed from the DRM application. The following techniques can be used to minimize this effect:

- Create virtual private networks (VPNs) with Quality of Service (QoS) through premise routers for the DRM circuit.
- Create separate physical networks.
- Guarantee the bandwidth using a third-party router/QoS vendor.

The Application Notes – Data Replication Manager over an Internet Protocol Intersite Link cover the third-party Fibre Channel-Internet Protocol (FC-IP) gateway devices that are certified by HP for use in an HP StorageWorks Data Replication Manager FC-IP solution. In addition the application notes provide a case study and considerations on distance versus required bandwidth.

DRM and the Oracle Storage Compatibility Program

As part of Oracle's Storage Compatibility Program (OSCP), Oracle has created a test suite that tests remote mirroring technologies to ensure their compatibility with Oracle databases. The self-test suite is provided for qualified vendors. HP chose to implement these tests using HP StorageWorks Data Replication Manager. As a member of OSCP, HP has successfully completed all test requirements stated in Oracle's remote mirroring test suite. The results were submitted to Oracle for verification and approved for entry in the program.

replicating Oracle databases

The HP white paper Oracle Databases Replication and Solutions highlights various concepts concerning Oracle database replication scenarios and gives a comparison table for different replication scenarios.

Microsoft Cluster Service and Oracle Fail Safe in an SAP environment

The Microsoft Cluster Service (MSCS) currently provides high availability for services and resources in a two-node advanced server and up to four nodes in a data center configuration. MSCS allows every node in a cluster to be actively running. In case of a failure, the protected SAP database, the central instance or a dialog instance would be failed over to a surviving node that would have to assume the additional workload. The cluster server groups resources such as network names, IP addresses or disks and forms “virtual servers” with which clients communicate. The group or virtual server can run on any physical server at any point in time.

The Oracle Fail Safe product, integrated with MSCS, is responsible for failing over and restarting the SAP database on a surviving node in the solution configuration. The SAP database in an Oracle active-passive configuration with a single instance runs on one of the cluster members.

The HP White Paper ProLiant Cluster HA/F500 Enhanced DT Solution for Oracle Fail Safe describes in greater detail, how the combination of the products mentioned form a stretch cluster configuration.

replication scenarios for Oracle with DRM

In a DRM environment, there are two major configuration options for replicating the Oracle database synchronously to the target site with no potential data loss.

The Oracle Storage Compatibility Testing - Remote Mirroring White Paper describes a number of considerations for mirroring the entire Oracle database or 'shipping' only the database redo log information. As mySAP.com applications are based on the underlying database, these suggestions are also valid in an SAP environment. There are two possible scenarios.

replicating the entire Oracle database

In this configuration, all volumes that contain either Oracle data files, online redo log files or control files are configured equally at both sites and linked to each other via a remote copy set on the HSG80 CLI level. Although DRM supports asynchronous mirroring in a database environment all remote copy sets must be synchronous and treated as a single entity in the same association set.

- Depending on the customer's backup strategy, the Oracle archived redo log files do not need to be in the association set or replicated in this scenario at all, because archived redo log information is not necessary in the event of a disaster failover, when all database files are replicated.
- There is a maximum of 12 remote copy sets per HSG80 pair. Bearing in mind that a minimum of one remote copy set might be used for a system boot disk or one for the cluster quorum, 9 remote copy sets will be left for the Oracle SAP database. This can be a serious restriction.
- A real advantage of mirroring the entire database is that it is a much simpler solution to manage, because it does not require the maintenance of a second database at the standby site.
- A failover to the target site in case of a disaster (DRM unplanned failover) is faster when mirroring the entire database, because recovery is similar to a standard instance recovery for the database after a site failover.

replicating Oracle redo log information only

Here, the Oracle standby database mechanism is used only to replicate Oracle redo log information to the target site via DRM to achieve a disaster tolerant state for the SAP Oracle database. Using the Oracle standby database mechanism without DRM is a common approach at SAP customer sites today. These customers accept that the latest transactional updates in the Oracle database might get lost in the event of a disaster at the primary site. The setup of an Oracle standby database is integrated in the SAP BRBACKUP utility.

- In this scenario all LUNs/storagesets that contain control files and online redo log files have to be in synchronous remote copy sets.
- The archived redo log files should still be replicated via DRM. In the event of a disaster, there is no guarantee that the latest archived redo log has been completely copied to the target site before the whole site is lost. As a result, it may be, that the DRM replicated online redo log files containing the latest transactional updates could not be applied.
- With HP StorageWorks Array Controller Software (ACS), no server access to the remote copy sets on the target site is allowed. Therefore the archived redo log information has to be copied over to the target site and has to be applied regularly. The most effective method to achieve this is to use Oracle functionality.
- One advantage of redo log shipping is that transactional updates can be applied to the target database with a timed delay. If the primary database information is destroyed through human error, the target standby database is protected from this kind of error being propagated immediately and a point in time recovery is possible.

- In general, an Oracle standby database can also run in a read-only mode, allowing the remote machine to be used as a query-only database for reporting and consistency checks. As mySAP.com applications tend to write to the database during startup, this feature has only limited value in this environment.
- Comparing the two replication scenarios, replicating only redo log information requires less bandwidth between the two sites. This is not that important in a campus environment where the customer is more flexible to increase bandwidth at a moderate cost compared to renting additional bandwidth from a telecommunications company.

The ORACLE8*i* Standby Database - Technical Report covers all details on log shipping as well as design and planning considerations on this topic.

managing DRM failover and failback

For various reasons it is necessary to use the service that DRM provides and fail over to the target site. Table 1 lists possible failover situations and the recommended actions in a specific situation. If a type of failure requires a site failover, it is important to verify that all components at the target site are operational before a failover is initiated. It might be preferable in some situations to fix a single component within an acceptable timeframe and continue processing, rather than performing a complete failover.

Table 1

Type of Failure	**Recommended action (Error_Mode = Failsafe)**
Total initiator site loss	Manual Intervention to fail over data processing to target site
Loss of initiator site fabric	Manual Intervention to fail over data processing to target site
Loss of initiator controller pair	Manual Intervention to fail over data processing to target site
Loss of all intersite links	Decide on which side should continue processing
Total target site loss	Manually continue processing at initiator site
Loss of target fabric	Manually continue at initiator site
Loss of target controller pair	Manually continue processing at initiator and target sites
Loss of single initiator controller	Failover not necessary
Loss of both initiator switches	Manual intervention to fail over to the target site and restart of processing at both sites
Loss of a single initiator switch	Failover not necessary
Extended power outage at the initiator site	Manual Intervention to fail over data processing to target site
Loss of a host bus adapter	Failover not necessary
Loss of single disk in redundant storage	Failover not necessary
Loss of single storageset	Failover not necessary
Loss of single host of cluster	Failover not necessary

An essential part of a DRM-based solution is the mechanism, for managing a planned/unplanned failover or failback operation in the event of a disaster or during maintenance operations. An HP-supported utility is the HSG Scripting Tool Kit (HSTK), providing automated failover and failback for DRM. The scripts require a system from which the CLI commands for DRM operations are sent to the HSG80 controller. Two communication configurations are possible, either out-of-band (maintenance port of the HSG80 controller) via terminal server or in-band with either Fibre Channel (HP StorageWorks Command Scripter) via (HP StorageWorks Command Console) LUN or an agent (Command Console). The system running the scripts can be a member of a production cluster or a dedicated server with at least one HBA in case of in-band communication.

datasafe solution verification

system configurations

To verify the functionality and performance of DRM in a mySAP.com Windows 2000 environment HP has set up two configurations running different workloads and two different Oracle replication scenarios. The different configurations were tested using direct Fibre connection as well as the Storage Edge Routers1000 from CNT as Intersite Links via FC over IP.

Please see the Appendix for setup details and software versions.

stretched cluster with DRM

This configuration used for the solution verification is an MSCS stretched cluster setup with DRM, the HA/F500-enhanced DT configuration is shown in Figure 2. Everything within the gray box is a standard MS cluster setup. At least one cluster member resides at the initiator storage site and one cluster member at the target storage site. The maximum number of nodes today is up to four in a Windows 2000 data center environment. Under normal operating conditions, the shared storage for the SAP database and the Central Instance runs at the initiator site. One SAP dialog instance has been set up on the cluster member at the target site. Storage devices (LUNs/Units) D and Q, containing the SAP database and the MSCS quorum disk are replicated via DRM to D' and Q' on the target site. Figure 2 only displays 2 units in order to simplify the graph. The detailed configuration is listed in the Appendix.

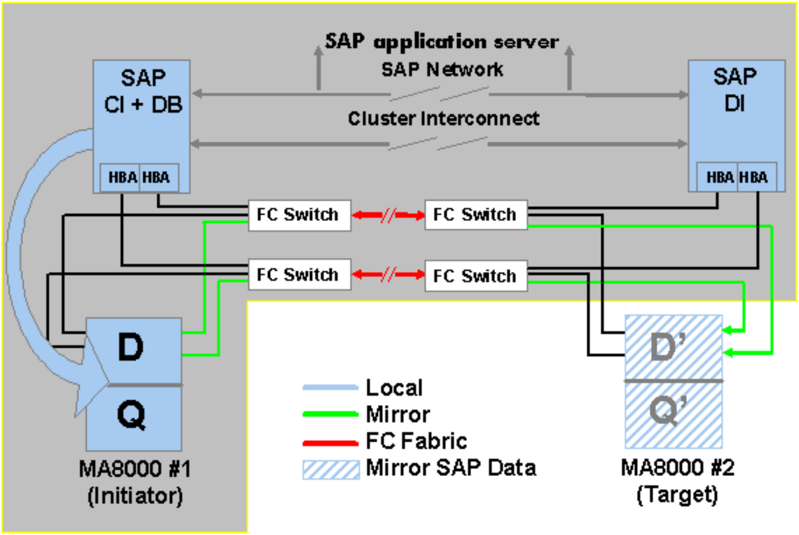


Figure 2

In the event of a node failure on the Initiator site, the SAP database service would be started automatically on the node at the target site, and access the shared storage on the initiator site while DRM replication continues. This could result in an overload of the intersite links (ISL) depending on the SAP load and the distance between the sites, because in this situation database access as well as DRM replication might use the same ISL. There are various configuration options for this situation:

- Have a second cluster member at the initiator site in the event of a node having down time but without being in a disaster situation.
- Have more than one ISL per fabric and use static routes on the FC switches or use the 2Gbit FC switch products and the licensed *trunking* feature.
- Fail over the entire site in the event of a node failure on the initiator site.

single system configuration with replicated system disk

This configuration consists of two non-clustered systems that are identical in terms of hardware. The production system runs the SAP database as well as the SAP central instance at the initiator site, while an SAP development system is configured to run with local storage on the target site as seen in Figure 3. The production system has its boot disk on the shared storage which is replicated to the target site as well as the SAP Oracle database. The configured page and swap space is local to both systems and not replicated to the target site.

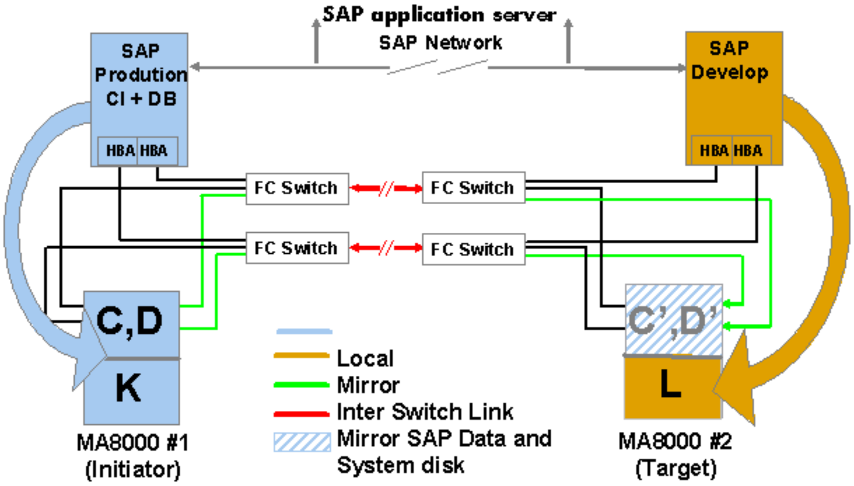


Figure 3

As all operating system and SAP data is replicated to the target site there is no need to change network configurations in order to make the SAP service visible. This is because the target site will boot with all production system properties in the event of a failover.

verification workloads

write intensive workload

The write-intensive workload is an ABAP program, started via transaction SE38 in the SAP front-end. This ABAP inserts a specified number of 200 Byte records into 5 Oracle tables containing unique indices into the USER1 table space of an R/3 standard database. This scenario simulates the behavior of a generic R/3 batch job. The tables are deleted and recreated after each run to ensure equal conditions for different runs. The size of the configured SAP/Oracle database has no direct impact on the workload.

To verify the solution in terms of functionality and DRM overhead, while not focusing on high-water benchmarking for a specific type of server hardware, the ABAP parameters have been adjusted as follows to make sure that neither the servers nor the network will become a bottleneck in the verification scenario:

The default workload specifies an insert/update of **1.5 million records** via 3 SAP D+W processes. A commit occurs every 1000 records. In this workload, **5 x 100MB** archived transaction log files are generated and replicated to the target site. The ABAP provides wall clock time for the whole run (transaction response time), as well as inserted records per second.

The ABAP program is completed in **66 seconds**, and provides the 100% write intensive baseline for a non-DRM scenario with all units/LUNs on one HSG80 with the exception of the Oracle achieved redo logs. The DRM overhead is calculated using this baseline in the two DRM scenarios.

On the SAN switches the portshowperf utility is used to monitor the switch port throughput.

The vtdpy display status utility is used to monitor the behavior of the HSG80 controller in terms of idle time.

standard SAP Sales and Distribution workload

The SAP Sales and Distribution (SD) benchmark has become a de-facto standard for SAP's platform partners and in the ERP (Enterprise Resource Planning) environment. In this benchmark, a defined set of business transactions must be run to simulate the behavior of interactive SAP users. For the Datasafe solution verification this benchmark has been set up to identify the relationship of the write intensive workload with the SD benchmark and the influence on a DRM environment. A detailed description of the SD benchmark is available at http://www.sap.com/benchmark/.

database replication verification scenarios

replicating the entire Oracle database

As discussed above, in this scenario the entire R/3 directory structure and all Oracle database files are replicated via DRM to the target site. All remote copy sets are SYNCHRONOUS and in FAILSAFE mode. All database-related remote copy sets belong to the same association set.

The ABAP program running the write intensive workload completes after 76 seconds, having generated 5 x 100MB achieved redo log information. After deleting all remote copy sets to run without DRM, the same workload is completed after 66 seconds. This means that the DRM overhead under a heavily write-intensive workload is within the range of 14% as compared to a non-DRM scenario as shown in Figure 4.

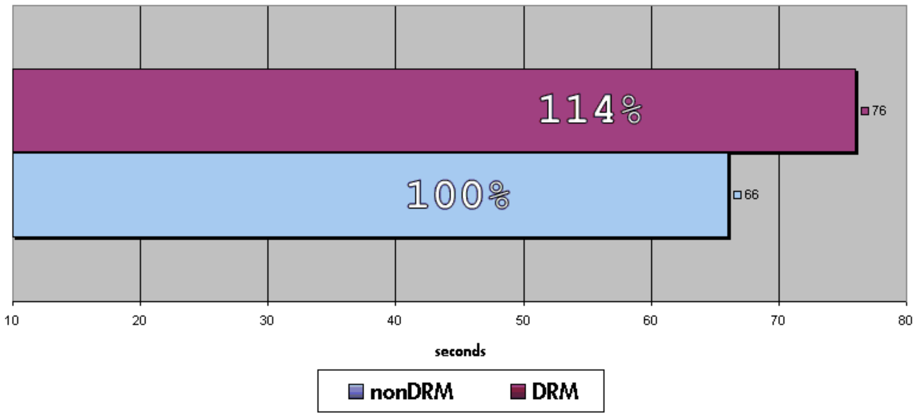


Figure 4

While running the job *without* DRM, the vtdpy performance utility of the HSG80 controller reported 35-40% idle time, which means that this workload puts a considerable load on the HSG80 controller as a controller is supposed to be saturated when around 25% idle time is reported via vtdpy. The perfmon utility on the database server reported a constant CPU idle time within the range of 50% and a total throughput of 25-28 MB/s. Considering a conservative average read/write ratio of 6:4, there is a DRM overhead within a range of less than 7%. When running the non-write I/O intensive SAP SD benchmark with a load of 200 simulated interactive users, a DRM overhead of less than 4% was measured in the verification environment.

replicating Oracle redo log information only

When replicating only the LUNs/storagesets containing Oracle redo log and control file information, there is less data to be transferred to the target site. This is reflected in the job completion time of the ABAP program shown in Figure 5.

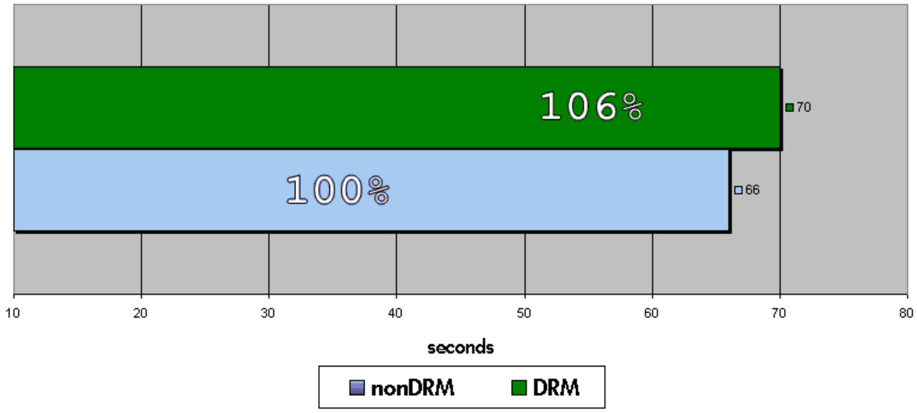


Figure 5

The write-intensive workload needs 6% more time to complete when the redo log information is replicated, compared to a situation with no DRM active. In a mixed read/write environment, the overhead caused by DRM in this scenario is within the range of 3%. Using the SAP SD benchmark load with 200 simulated users, there is a DRM overhead of less than 1% in the verification environment.

Please note that the measured DRM overhead in the lab environment is caused only by the additional I/Os for the controllers. The distance between the primary and target controller in this environment was 0 km and had practically no influence on the response time. Please see the HP StorageWorks Data Replication Manager Inter-site Link Performance Analyzer white paper on how to calculate the impact of distance for worst-case I/O scenarios for DRM.

intersite link via direct Fibre and FC-IP

Both verification configurations have been tested with direct Fibre and UltraNet Edge Storage Router 1000 FC-IP gateways as Intersite Links (ISL) between the initiator and target site.

Testing the FC-IP gateways, the FreeBSD dummynet utility was used to simulate IP traffic packet loss and route delays. Dummynet is a flexible tool for bandwidth management and for testing networking protocols. It works by intercepting packets in their way through the protocol stack, and passing them through one or more pipes which simulate the effects of bandwidth limitations, propagation delays, bounded-size queues, packet losses, etc.

Table 2 gives the job completion time in seconds of the write intensive workload while simulating various network delays using the UltraNet Edge Storage Router for DRM.

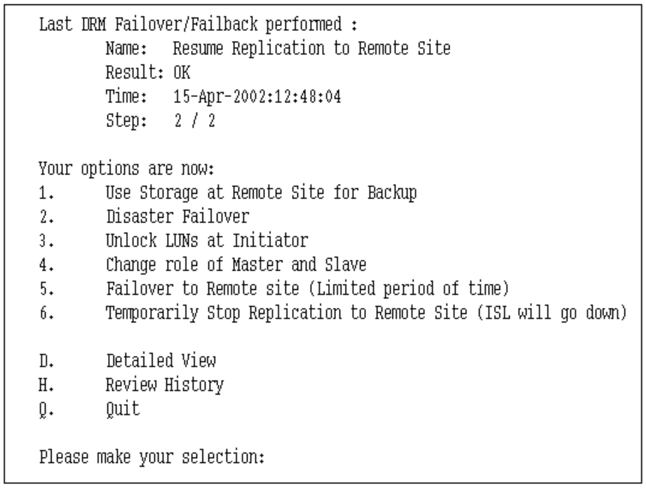
Table 2

Inter switch connection type	Route delay simulation in ms				
	0	10	20	50	100
Direct Fibre	76	n.a.	n.a.	n.a.	n.a.
100Mbit with Compression	80	159	262	625	1214
100Mbit with Compression and 1% packet loss	107	483	718	1676	7963
10Mbit with Compression	300	340	450	950	3100
10Mbit with Compression and 1% packet loss	351	584	885	2232	

Table 2 shows that having a zero latency network with a high quality of service the write intensive workload provides a similar response time as the direct Fibre connection. Once the quality of service of the IP network is reduced, the application response time increases.

failover and failback operations

In a DRM environment, after a site failover the data is already available at the target site. Failback moves data operations back to the initiator after the initiator site has been brought back online. The drmconsole script in Figure 6 shows the available options for managing a DRM site failover. Option 1 in drmconsole suspends remote mirroring for a volume and enables access to that volume on the target site, for example for a backup server. Using this functionality is beyond the scope of this paper. Please see the *Guide of Operations For Data Replication Manager and Clone and Snapshot scripts* or the DRM Scripting User Guide (HP DRM Scripting Kit) for further information on the supported Perl script toolkits.



```
Last DRM Failover/Failback performed :
        Name:   Resume Replication to Remote Site
        Result: OK
        Time:   15-Apr-2002:12:48:04
        Step:   2 / 2

Your options are now:
1.      Use Storage at Remote Site for Backup
2.      Disaster Failover
3.      Unlock LUNs at Initiator
4.      Change role of Master and Slave
5.      Failover to Remote site (Limited period of time)
6.      Temporarily Stop Replication to Remote Site (ISL will go down)

D.      Detailed View
H.      Review History
Q.      Quit

Please make your selection:
```

Figure 6

unplanned site failover (disaster)

In the event of a series of failures at the initiator site as described in Table 1, this might result in a total loss of access to the storage on this site. This is true for both the MSCS cluster configuration (Figure 7) and the non-clustered configuration with replicated system boot disk (Figure 8). This leads to an SAP production system halt situation at the initiator site. A human decision has to be made to initiate a site failover to the target site.

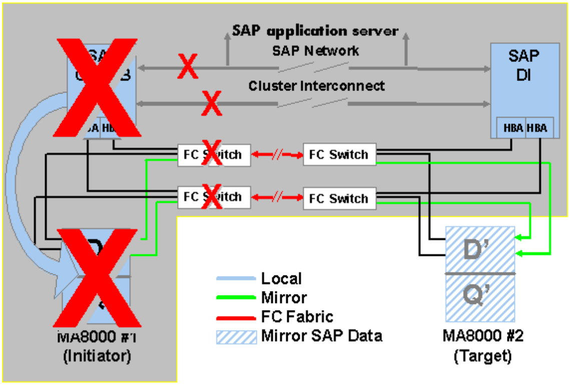
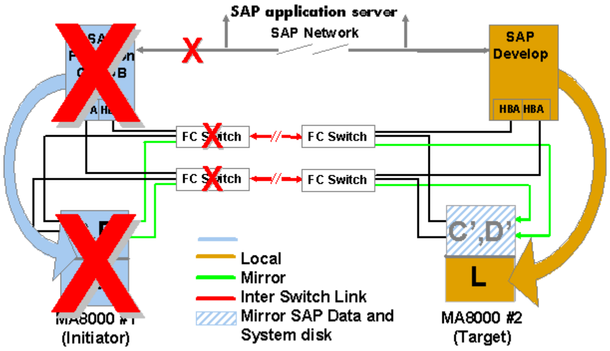


Figure 7

From the SAP point of view, it makes no difference whether the cluster shared storage or the database of a single server is not available anymore.

Figure 8



The following steps must be taken to complete a site failover for both configurations:

1. Run the drmconsole script on the management node (Figure 6) and choose option number 2 (disaster failover). The steps for the failover can be completed within a 10 to 12 minute time frame depending on the number of remote copy sets.
2. Reboot the systems at the target site.
3. SAP Oracle database recovery:

   a) If the entire database has been replicated, the defined MSCS SAP database service will automatically start the Oracle database instance. During database startup, Oracle will automatically perform an instance recovery. The time it takes until the database is available depends on the number of open transactions during the disaster situation.

   b) If only redo log information has been replicated the following steps must be taken:

   - Start up and recover the standby database in order to apply archived redo logs that were not successfully transferred via IP while the disaster occurred.
   - Shut down the standby database after all archived redo logs have been applied.
   - Apply the replicated online redo log information to the standby database using the latest replicated control file from the initiator site using the following options:

     ```
     SVRMGR> STARTUP MOUNT  <SID>;
     SVRMGR> RECOVER DATABASE;
     SVRMGR> ALTER DATABASE OPEN;
     ```
   - Start the SAP application

Figure 9

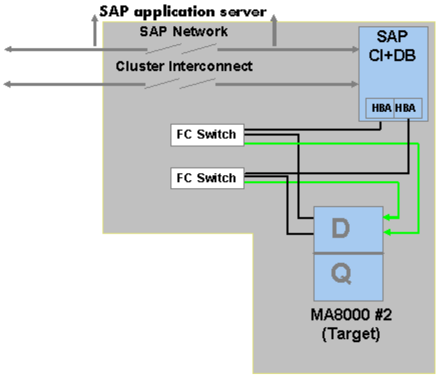
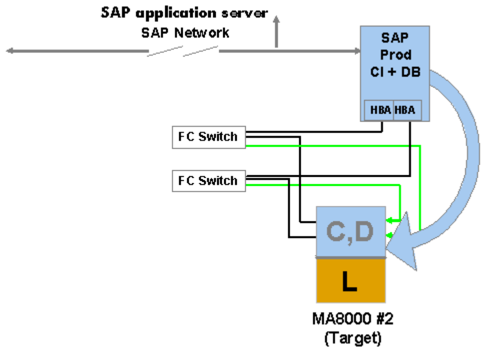


Figure 10



After the site failover has been completed, the SAP database and the central instance are running on the target site as shown in Figure 9 for the cluster environment and in Figure 10 for the single node configuration. At this point in time a clustered solution is very similar to a non-clustered configuration with the difference that once a node on the initiator site becomes available again, it can join the cluster. The development environment is not available for the non-clustered configuration at this point in time. The total failover time in the verification scenario is less than 15 minutes for replicating the entire SAP database and within the range of 19 minutes for the standby database scenario, depending on the amount of redo log information that has to be applied. In the verification scenario one 100MB archived redo log was applied within 30 seconds.

failback procedure

After a planned or an unplanned site failure has occurred, a DRM configuration still has no single point of failure (Disk, HBA, Cable, HSG80), but is no longer in a disaster tolerant state. To achieve this status again, the necessary actions depend on the customer's disaster plan and strategy, the type of disaster that had occurred, and the replication scenario that the customer is using.

To failback the verification configuration to the original primary site, the following steps are taken after an unplanned site failover when no hardware replacement at the initiator site is required:

In the scenario replicating the entire database

- Start the drmconsole script on the management node and choose the appropriate option. The script is aware of a preceding site failover and offers only valid actions
- In the *first step* of the script, the controller at the initiator site is restarted and prepared for the renormalization of the remote copy sets for the Oracle database and the cluster LUNs. This step can run while SAP services are up and running
- In the *second step* of the script, the renormalization process is started. This will impact the performance of the running SAP environment as described in Figure 2
- *Before the third step* of the script can start, the SAP service and the cluster needs to be shutdown, because this step will disable access to the units at the target site and failback the remote copy sets to the initiator site
- *After the third step* of the script is completed the cluster can be rebooted and is in the same state as before the disaster had occurred

In the scenario for replicating redo log information

- Establish a standby database on the node at the primary site on the primary site storage. The Oracle Storage Compatibility Testing - Remote Mirroring White Paper suggests several ways to achieve this:
  - Reverse role via database copy and via restoring backup
  - Reverse role via recovery
  - Direct fallback via DB Copy
  - Direct fallback via restoring backup

- Run the drmconsole script on the management node to failback the remote copy sets for this scenario as described above. In the *second step* of the script the performance of the SAP service will hardly be affected, as only the LUN containing redo log information is normalized.

path failure and DRM normalization

At both sites, a DRM configuration has no single point of failure in the I/O path from the server to the data on disk. There are at least two paths in two distinct fabrics to ensure that an unplanned site failover (disaster) can only happen if a series of failures occur. To test this functionality in a mySAP.com environment, a path failure was simulated by powering off one FC switch while the write intensive workload was running. The path failure was acknowledged after 30 seconds and the running job completed within the range of 130 seconds compared to the 70 seconds in which the job is completed without error conditions.

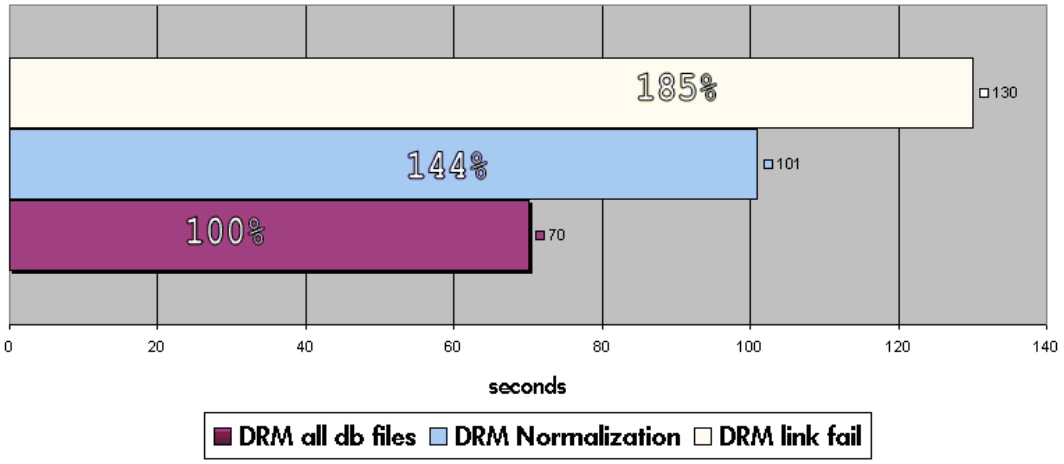


Figure 11

The normalization process in a DRM configuration is the full copy of a LUN between the initiator and the target site controllers with a size of 64KB I/O. This must happen when a remote copy set is created or once the two sites are out of synchronization. In the verification configuration with direct Fibre ISL, the portperfshow utility for the FC switches reported 8 MB/s for a two-member mirror set and 15 MB/s for a six-member raid 0+1 set. Normalizing five remote copy sets in parallel on one HSG80 controller did not exceed 30MB/s.

Running the write-intensive ABAP program while a full normalization was in progress, the job completion time increased by about 44% to 101 seconds (Figure 11). This means that it might not be acceptable in terms of application response time for a customer to renormalize after a disaster to prepare a failback while the SAP service is available.

When comparing the performance between a direct Fibre connection and various FC-IP gateway configuration options (Figure 12) during the normalization process in a zero latency network, one can see that switching on compression nearly doubles the throughput in the verification scenario. And the throughput of the IP gateways within the range of 21 MB per second demonstrates good performance in a network with no packet transmission delays and no packet loss.

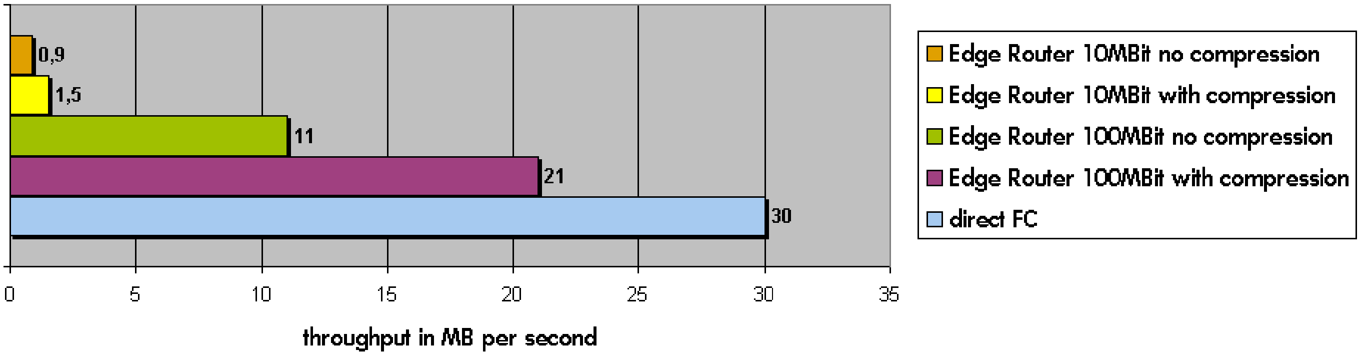


Figure 12

summary

The HP StorageWorks Datasafe for mySAP.com solution enhances the Microsoft high-availability features with the disaster tolerant capabilities of HP StorageWorks Data Replication Manager, maintaining application performance dependant on the distance between the two sites. The DRM overhead in a 100% write-intensive SAP-specific workload is 14% compared to the same workload without DRM in a zero latency SAN. The average DRM overhead in this environment running a mixed workload is less than 6%.

DRM over an Internet protocol intersite link is a cost-optimized, high-performance solution for greater distances if the quality of service of this network is maintained. Compared to a Direct Fibre connection the use of FC-IP gateways in a zero latency 100Mbit network, a write-intensive workload showed an overhead compared to direct Fibre connection in the range of 3 % for a full database replication scenario. Once the quality of service for the IP network is reduced through bandwidth limitations and route delays, the DRM overhead increases dramatically.

Replicating the entire SAP database is a robust, managed solution for SAP customers. The solution provides a failover/recovery time at a remote computing site measured in minutes. A medium-sized SAP configuration placed on six RAIDsets can fail over to a recovery site within less than 15 minutes. This scenario allows existing HP customers a straightforward enhancement of their environment using DRM with careful planning.

Replicating only SAP database redo log information via DRM using the Oracle standby database mechanism also provides DT functionality up to the latest transactional update. In addition, this scenario requires less bandwidth and allows database changes to be propagated with a timed delay at the target site to protect the standby database from human error. The tradeoff for this scenario is the additional management effort required to maintain the standby database and the Oracle database expertise necessary in the event of a site failover, creating a longer failover time. For greater distances and SAN integration in existing networks, however, the combination of Oracle log shipping with DRM and FC-IP gateways maintains the best level of price/performance.

An important consideration in a SAP customer's DT plan is the necessary time it takes to be in a disaster tolerant state again following a disaster and subsequent failover. Resynchronizing of RAIDsets in a zero latency SAN for a medium-sized SAP database is in the range of 30MB/s per HSG80 with direct Fiber ISLs and 21 MB/s using FC-IP gateways. As distance between the sites increases and the quality of service for the IP network decreases, the throughput is to be seen in direct relation to these criteria and can be reduced to half a MB per second or less. This could cause a full resynchronization of an SAP database to take days or even weeks. In this case the Oracle standby database scenario makes it possible to restore a backup from tape on the target site and to roll forward using archived redo log information.

A non-clustered, two-node configuration using DRM for replicating the operating system boot disk in addition to SAP database replication is a cost-effective entry level DT solution with a minimum of necessary system management.

One success story demonstrating a productive implementation at a customer site can be found at
ftp://ftp.compaq.com/pub/solutions/enterprise/ha/sucessstories/Robert-Bosch-SS.pdf

appendix

Description and setup of the verification configuration for the clustered as well for the non-clustered configuration.

A - Description of hardware for the solution

Table 3

	Initiator Site	#	Target Site	#
Server	PL8500	2	PL8500	1
	CPU	8	CPU	8
	4 GB memory		4 GB memory	
	KGPSA-BC	2	KGPSA-BC	2
Storage	HSG80 pair	1	HSG80 pair	1
	10K RPM disk drives	24	10K RPM disk drives	24
FC Infrastructure	SAN Switch/16	1	SAN Switch/16	1
	SAN Switch/8	1	SAN Switch/8	1
FC IP Gateways	CNT StorEdge Router 1000			2
SAN mgmt	HP OpenView Storage Management Appliance			1
Client (SAP)	ML370			1
Network Infrastructure	The servers and the client are connected via a 10/100 NICs			

B - Description of software for the solution

Table 4

Software	Version
Windows 2000 Advanced Server and Data Center	SP3
HP StorageWorks Secure Path	4.0
HP StorageWorks Array Controller Software	8.7 P-0
Fabric OS	2.6c
SAP R/3	4.6D
Oracle	8.1.7
HP StorageWorks Command Scripter	1.0A
DRM Failover Scripts	1.6

C - HSG80 disk storage map

Table 5

	SCSI 1	SCSI 2	SCSI 3	SCSI 4	SCSI 5	SCSI 6
Position 0	RCS1		RCS2			
	D1 SAPBIN		D2 QUORUM		D99 DRMLOG	
	DISK10000 18.2GB/1 10K	DISK20000 18.2GB/1 10K	DISK3000 18.2GB/1 10K	DISK40000 18.2GB/1 10K	DISK50000 18.2GB/1 10K	DISK60000 18.2GB/1 10K
Position 1	RCS6		RCS5			
	D6 ORABIN		D5 SAPARLOG			
	DISK10100 18.2GB/1 10K	DISK20100 18.2GB/1 10K	DISK30100 18.2GB/1 10K	DISK40100 18.2GB/1 10K	DISK50100 18.2GB/1 10K	DISK60100 18.2GB/1 10K
Position 2	RCS4					
	D4					
	SAPLLOG					
	SAPMIRL1		SAPMIRL2		SAPMIRL3	
	DISK10200 18.2GB/1 10K	DISK20200 18.2GB/1 10K	DISK30200 18.2GB/1 10K	DISK40200 18.2GB/1 10K	DISK50200 18.2GB/1 10K	DISK60200 18.2GB/1 10K
Position 3	RCS3					
	D3					
	SAPDATA1					
	SAPMIR1		SAPMIR2		SAPMIR3	
	DISK10300 18.2GB/1 10K	DISK20300 18.2GB/1 10K	DISK30300 18.2GB/1 10K	DISK40300 18.2GB/1 10K	DISK50300 18.2GB/1 10K	DISK60300 18.2GB/1 10K

D - SANswitch port allocation

Table 6

	Port	0	1	2	3	4	5	6	7
Initiator	TopFabric	PRI TOP1 (HSG80)	PRITOP2 (HSG80)		PL85-INI-HBA1				ISL
	BottomFabric	PRI BOT1 (HSG80)	PRIBOT2 (HSG80)		PL85-INI-HBA2		SWMA		ISL
Target	TopFabric	TARTOP1 (HSG80)	TARTOP2 (HSG80)		PL85TAR-HBA1		ML37HBA1		ISL
	BottomFabric	TARBOT1 (HSG80)	TARBOT2 (HSG80)		PL85TAR-HBA2				ISL

E - Oracle database layout

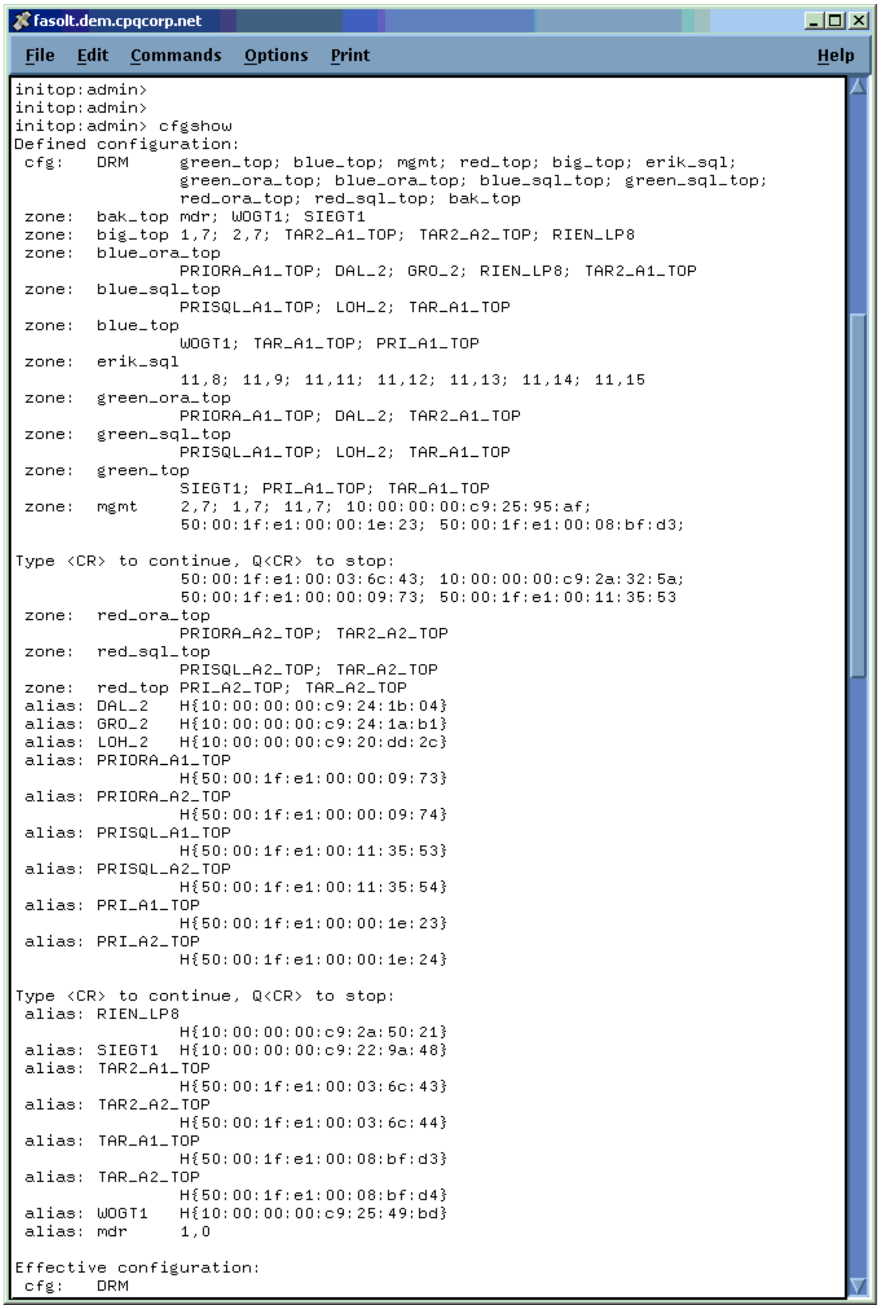
The physical layout of the SAP Oracle database was adjusted to the workload of the solution verification scenario. The size of the configured SAP R/3 standard table spaces has no direct impact on the test scenario.

Table 7

Name	Drives	RAID	LUN	Description
SapArch	2 x 18.2-GB	1	D5	Archived redo logs
OrigLogA OrigLogB	6 x 18.2-GB	1+0	D4	Redo log set A Redo log set B
SapData1	6 x 18.2-GB	1+0	D3	DB data area 1
Oracle SapReorg	2 x 18.2-GB	1	D2	Oracle exec Reorg, backup, check, stat and trace
SAPBIN	2 x 18.2-GB	1	D1	SAP binaries

F - Set up zoning in a DRM configuration

The *Data Replication Manager Configuration Guide* explains the way zoning on the FC switches has to be set up in a DRM environment. Figure 13 shows the overall zoning of the top fabric of the verification SAN through the telnet interface of the initiator top fabric FC switch. The relevant zones for this solution are the members of the green_ora_top, blue_ora_top and the red_ora_top zones.

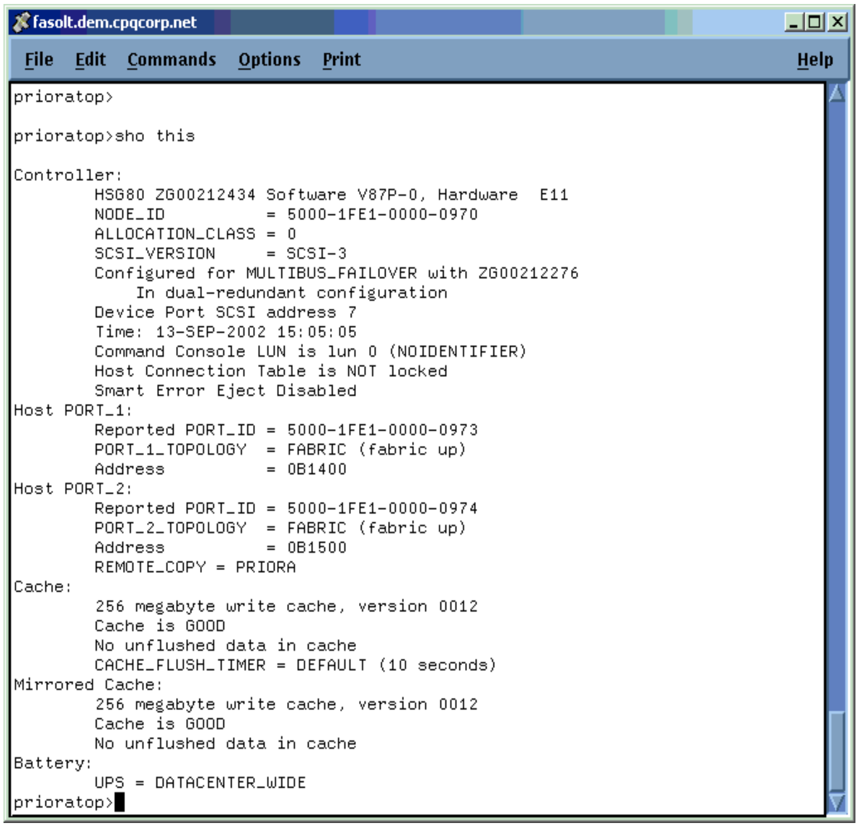


```
initop:admin>
initop:admin>
initop:admin> cfgshow
Defined configuration:
 cfg:   DRM     green_top; blue_top; mgmt; red_top; big_top; erik_sql;
                green_ora_top; blue_ora_top; blue_sql_top; green_sql_top;
                red_ora_top; red_sql_top; bak_top
 zone:  bak_top mdr; WOGT1; SIEGT1
 zone:  big_top 1,7; 2,7; TAR2_A1_TOP; TAR2_A2_TOP; RIEN_LP8
 zone:  blue_ora_top
                PRIORA_A1_TOP; DAL_2; GRO_2; RIEN_LP8; TAR2_A1_TOP
 zone:  blue_sql_top
                PRISQL_A1_TOP; LOH_2; TAR_A1_TOP
 zone:  blue_top
                WOGT1; TAR_A1_TOP; PRI_A1_TOP
 zone:  erik_sql
                11,8; 11,9; 11,11; 11,12; 11,13; 11,14; 11,15
 zone:  green_ora_top
                PRIORA_A1_TOP; DAL_2; TAR2_A1_TOP
 zone:  green_sql_top
                PRISQL_A1_TOP; LOH_2; TAR_A1_TOP
 zone:  green_top
                SIEGT1; PRI_A1_TOP; TAR_A1_TOP
 zone:  mgmt    2,7; 1,7; 11,7; 10:00:00:00:c9:25:95:af;
                50:00:1f:e1:00:00:1e:23; 50:00:1f:e1:00:08:bf:d3;

Type <CR> to continue, Q<CR> to stop:
                50:00:1f:e1:00:03:6c:43; 10:00:00:00:c9:2a:32:5a;
                50:00:1f:e1:00:00:09:73; 50:00:1f:e1:00:11:35:53
 zone:  red_ora_top
                PRIORA_A2_TOP; TAR2_A2_TOP
 zone:  red_sql_top
                PRISQL_A2_TOP; TAR_A2_TOP
 zone:  red_top PRI_A2_TOP; TAR_A2_TOP
 alias: DAL_2   H{10:00:00:00:c9:24:1b:04}
 alias: GRO_2   H{10:00:00:00:c9:24:1a:b1}
 alias: LOH_2   H{10:00:00:00:c9:20:dd:2c}
 alias: PRIORA_A1_TOP
                H{50:00:1f:e1:00:00:09:73}
 alias: PRIORA_A2_TOP
                H{50:00:1f:e1:00:00:09:74}
 alias: PRISQL_A1_TOP
                H{50:00:1f:e1:00:11:35:53}
 alias: PRISQL_A2_TOP
                H{50:00:1f:e1:00:11:35:54}
 alias: PRI_A1_TOP
                H{50:00:1f:e1:00:00:1e:23}
 alias: PRI_A2_TOP
                H{50:00:1f:e1:00:00:1e:24}

Type <CR> to continue, Q<CR> to stop:
 alias: RIEN_LP8
                H{10:00:00:00:c9:2a:50:21}
 alias: SIEGT1  H{10:00:00:00:c9:22:9a:48}
 alias: TAR2_A1_TOP
                H{50:00:1f:e1:00:03:6c:43}
 alias: TAR2_A2_TOP
                H{50:00:1f:e1:00:03:6c:44}
 alias: TAR_A1_TOP
                H{50:00:1f:e1:00:08:bf:d3}
 alias: TAR_A2_TOP
                H{50:00:1f:e1:00:08:bf:d4}
 alias: WOGT1   H{10:00:00:00:c9:25:49:bd}
 alias: mdr     1,0

Effective configuration:
 cfg:   DRM
```

Figure 13

G - Configuring storage

Install and cable initiator and target site storage, FC switches and target site storage as described in *Data Replication Manager Configuration Guide*. Set up initiator site and target site HSG80 storage the same way at the CLI level. Figure 14 shows the configuration characteristics of the HSG80 at the Initiator site.

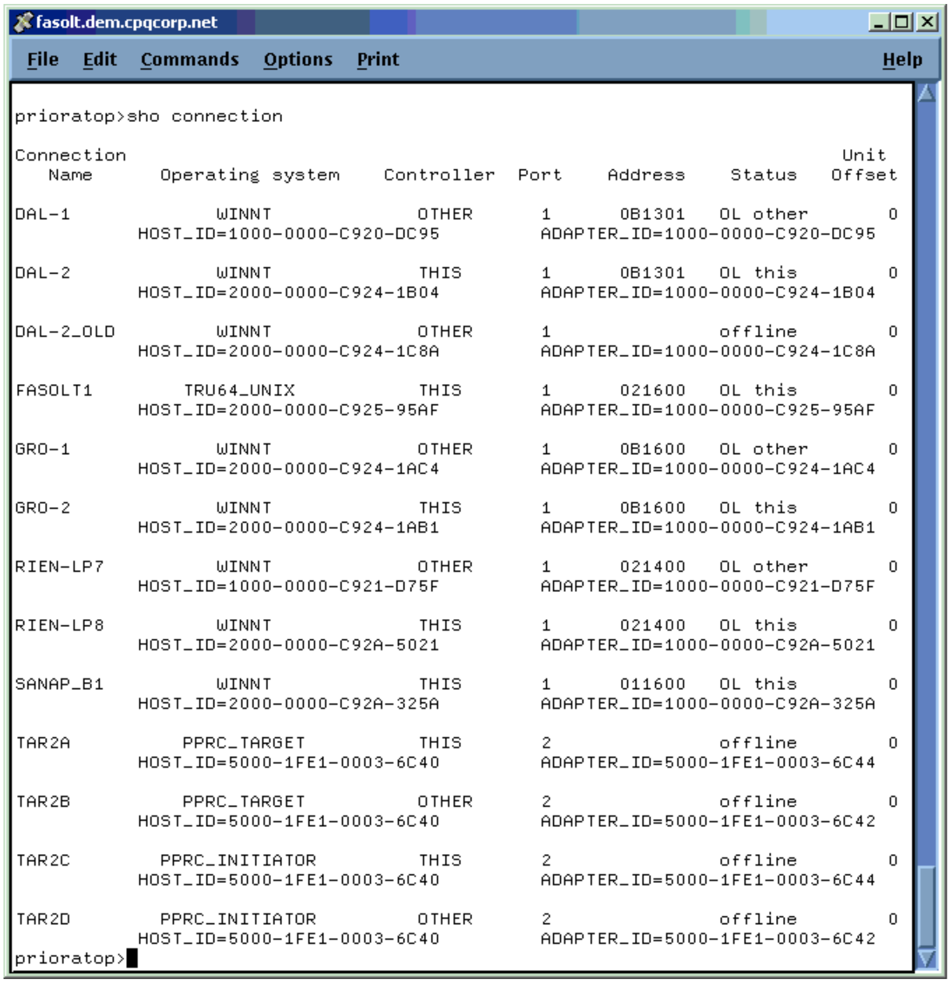


```
prioratop>

prioratop>sho this

Controller:
        HSG80 ZG00212434 Software V87P-0, Hardware  E11
        NODE_ID            = 5000-1FE1-0000-0970
        ALLOCATION_CLASS = 0
        SCSI_VERSION       = SCSI-3
        Configured for MULTIBUS_FAILOVER with ZG00212276
            In dual-redundant configuration
        Device Port SCSI address 7
        Time: 13-SEP-2002 15:05:05
        Command Console LUN is lun 0 (NOIDENTIFIER)
        Host Connection Table is NOT locked
        Smart Error Eject Disabled
Host PORT_1:
        Reported PORT_ID = 5000-1FE1-0000-0973
        PORT_1_TOPOLOGY  = FABRIC (fabric up)
        Address            = 0B1400
Host PORT_2:
        Reported PORT_ID = 5000-1FE1-0000-0974
        PORT_2_TOPOLOGY  = FABRIC (fabric up)
        Address            = 0B1500
        REMOTE_COPY = PRIORA
Cache:
        256 megabyte write cache, version 0012
        Cache is GOOD
        No unflushed data in cache
        CACHE_FLUSH_TIMER = DEFAULT (10 seconds)
Mirrored Cache:
        256 megabyte write cache, version 0012
        Cache is GOOD
        No unflushed data in cache
Battery:
        UPS = DATACENTER_WIDE
prioratop>
```

Figure 14

H - Managing the HSG80 connection table

The HSG80 connection table should be maintained with meaningful connection names and the correct operating system types as shown in Figure 15.

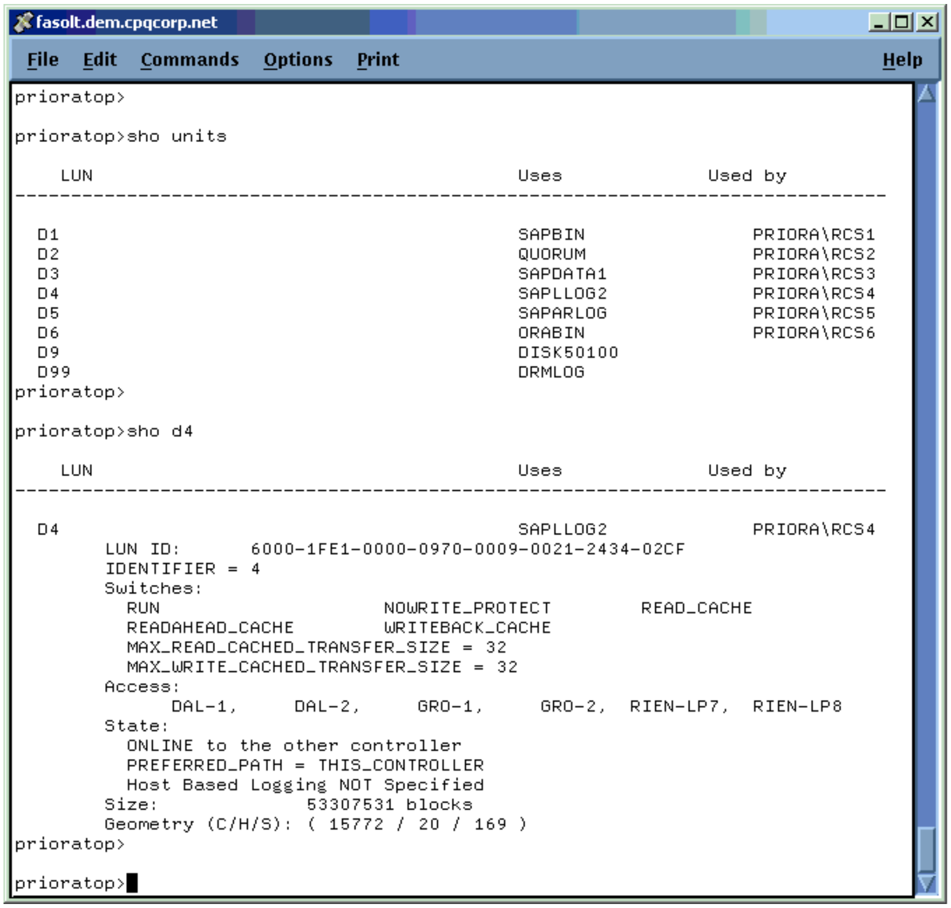


```
prioratop>sho connection

Connection                                                                Unit
   Name      Operating system    Controller  Port    Address    Status   Offset

DAL-1             WINNT             OTHER      1      0B1301   OL other       0
           HOST_ID=1000-0000-C920-DC95         ADAPTER_ID=1000-0000-C920-DC95

DAL-2             WINNT             THIS       1      0B1301   OL this        0
           HOST_ID=2000-0000-C924-1B04         ADAPTER_ID=1000-0000-C924-1B04

DAL-2_OLD         WINNT             OTHER      1               offline        0
           HOST_ID=2000-0000-C924-1C8A         ADAPTER_ID=1000-0000-C924-1C8A

FASOLT1        TRU64_UNIX           THIS       1      021600   OL this        0
           HOST_ID=2000-0000-C925-95AF         ADAPTER_ID=1000-0000-C925-95AF

GRO-1             WINNT             OTHER      1      0B1600   OL other       0
           HOST_ID=2000-0000-C924-1AC4         ADAPTER_ID=1000-0000-C924-1AC4

GRO-2             WINNT             THIS       1      0B1600   OL this        0
           HOST_ID=2000-0000-C924-1AB1         ADAPTER_ID=1000-0000-C924-1AB1

RIEN-LP7          WINNT             OTHER      1      021400   OL other       0
           HOST_ID=1000-0000-C921-D75F         ADAPTER_ID=1000-0000-C921-D75F

RIEN-LP8          WINNT             THIS       1      021400   OL this        0
           HOST_ID=2000-0000-C92A-5021         ADAPTER_ID=1000-0000-C92A-5021

SANAP_B1          WINNT             THIS       1      011600   OL this        0
           HOST_ID=2000-0000-C92A-325A         ADAPTER_ID=1000-0000-C92A-325A

TAR2A          PPRC_TARGET          THIS       2               offline        0
           HOST_ID=5000-1FE1-0003-6C40         ADAPTER_ID=5000-1FE1-0003-6C44

TAR2B          PPRC_TARGET          OTHER      2               offline        0
           HOST_ID=5000-1FE1-0003-6C40         ADAPTER_ID=5000-1FE1-0003-6C42

TAR2C        PPRC_INITIATOR         THIS       2               offline        0
           HOST_ID=5000-1FE1-0003-6C40         ADAPTER_ID=5000-1FE1-0003-6C44

TAR2D        PPRC_INITIATOR         OTHER      2               offline        0
           HOST_ID=5000-1FE1-0003-6C40         ADAPTER_ID=5000-1FE1-0003-6C42
prioratop>
```

Figure 15

I - Set up the initiator units

Figure 16 shows how the initiator site units have been set up. HBA access is not allowed for the corresponding units at the target site. Only the right-hand ports of the HSG80 controller have access to the units on the target site under normal operating conditions.

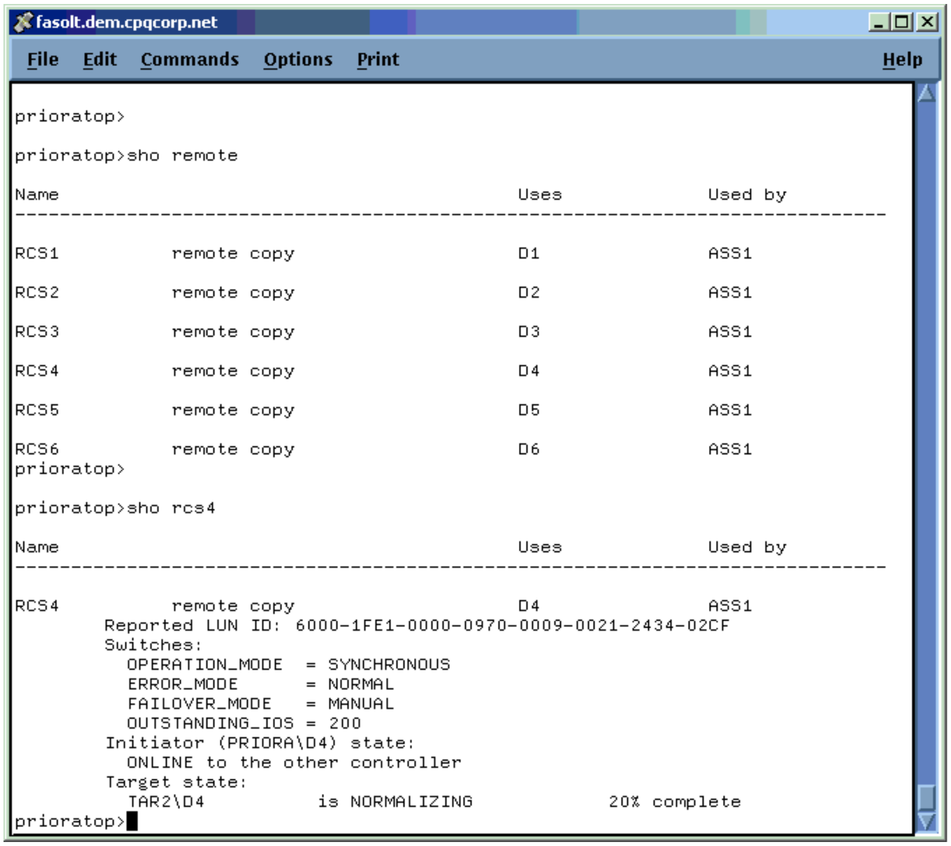


```
prioratop>

prioratop>sho units

    LUN                                          Uses             Used by
------------------------------------------------------------------------------

  D1                                             SAPBIN               PRIORA\RCS1
  D2                                             QUORUM               PRIORA\RCS2
  D3                                             SAPDATA1             PRIORA\RCS3
  D4                                             SAPLLOG2             PRIORA\RCS4
  D5                                             SAPARLOG             PRIORA\RCS5
  D6                                             ORABIN               PRIORA\RCS6
  D9                                             DISK50100
  D99                                            DRMLOG
prioratop>

prioratop>sho d4

    LUN                                          Uses             Used by
------------------------------------------------------------------------------

  D4                                             SAPLLOG2             PRIORA\RCS4
        LUN ID:       6000-1FE1-0000-0970-0009-0021-2434-02CF
        IDENTIFIER = 4
        Switches:
          RUN                      NOWRITE_PROTECT        READ_CACHE
          READAHEAD_CACHE          WRITEBACK_CACHE
          MAX_READ_CACHED_TRANSFER_SIZE = 32
          MAX_WRITE_CACHED_TRANSFER_SIZE = 32
        Access:
                DAL-1,     DAL-2,     GRO-1,     GRO-2,  RIEN-LP7,  RIEN-LP8
        State:
          ONLINE to the other controller
          PREFERRED_PATH = THIS_CONTROLLER
          Host Based Logging NOT Specified
        Size:             53307531 blocks
        Geometry (C/H/S): ( 15772 / 20 / 169 )
prioratop>

prioratop>
```

Figure 16

J - Remote copy set configuration

Figure 17 shows the setup for remote copy sets in the verification scenario. The OPERATION_MODE has to be SYNCHRONOUS in a database environment. The figure shows that RCS4 is in the process of NORMALIZATION. This is the initial full copy of D4 on the initiator site to D4 on the target site.

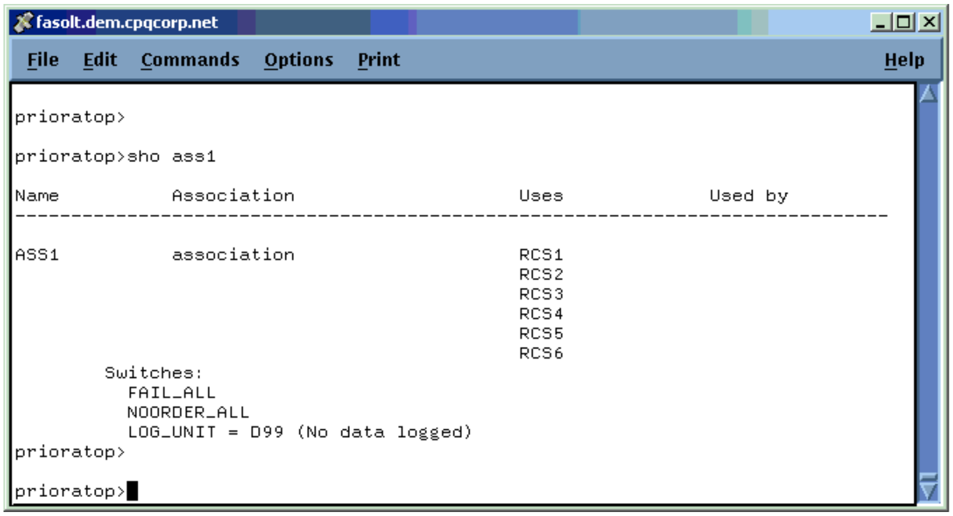


```
prioratop>

prioratop>sho remote

Name                                         Uses             Used by
------------------------------------------------------------------------------

RCS1          remote copy                    D1               ASS1

RCS2          remote copy                    D2               ASS1

RCS3          remote copy                    D3               ASS1

RCS4          remote copy                    D4               ASS1

RCS5          remote copy                    D5               ASS1

RCS6          remote copy                    D6               ASS1
prioratop>

prioratop>sho rcs4

Name                                         Uses             Used by
------------------------------------------------------------------------------

RCS4          remote copy                    D4               ASS1
        Reported LUN ID: 6000-1FE1-0000-0970-0009-0021-2434-02CF
        Switches:
          OPERATION_MODE   = SYNCHRONOUS
          ERROR_MODE       = NORMAL
          FAILOVER_MODE    = MANUAL
          OUTSTANDING_IOS = 200
        Initiator (PRIORA\D4) state:
          ONLINE to the other controller
        Target state:
          TAR2\D4          is NORMALIZING            20% complete
prioratop>
```

Figure 17

K - Association set of the SAP Oracle database

Figure 18 shows the setup for the association set. The verification scenario setup has unit D99 assigned as DRM LOG_UNIT.

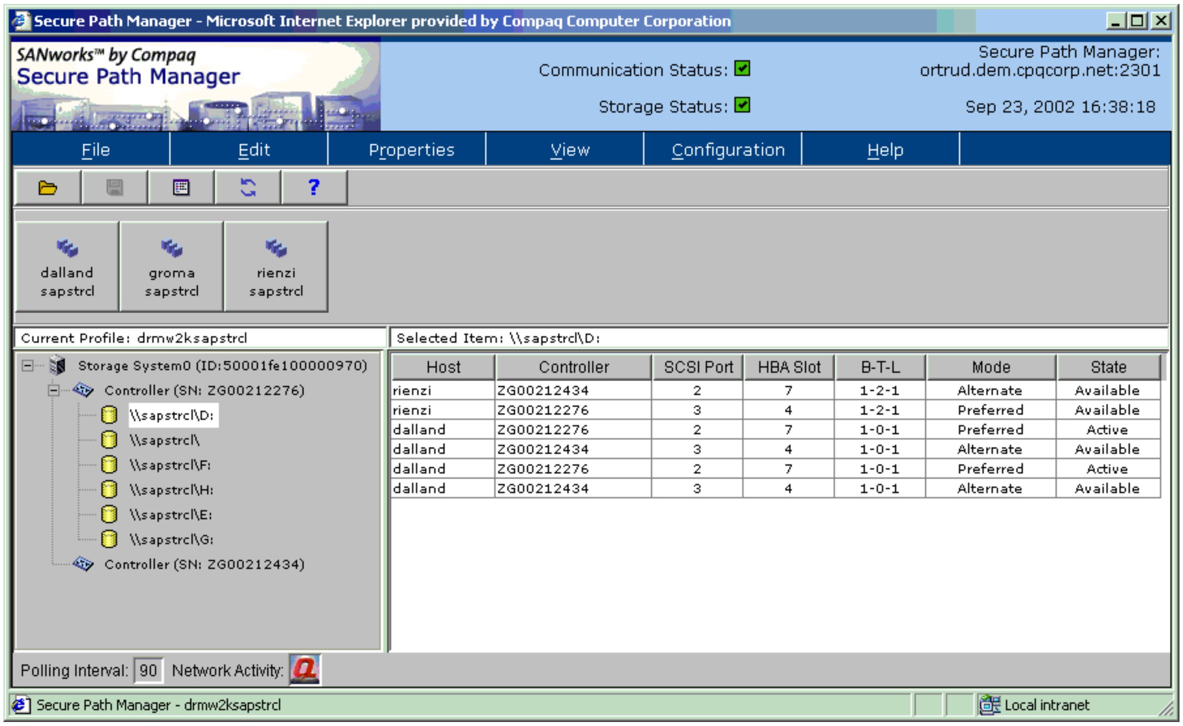


```
prioratop>

prioratop>sho ass1

Name          Association                    Uses             Used by
------------------------------------------------------------------------------

ASS1          association                    RCS1
                                             RCS2
                                             RCS3
                                             RCS4
                                             RCS5
                                             RCS6

        Switches:
          FAIL_ALL
          NOORDER_ALL
          LOG_UNIT = D99 (No data logged)
prioratop>

prioratop>
```

Figure 18

L - Set up secure path

As discussed in the Data Replication Manager Design Guide - Application Notes, the maintenance of two separate fabrics is a prerequisite for DRM. This is because only one fabric would represent a single point of failure. To maintain two distinct paths to the storage, HP StorageWorks Secure Path software provides the functionality for switching between paths in case one path has a problem with completing application I/Os. Figure 19 shows the administration tool of Secure Path V 4.0 running on an HP OpenView Storage Management Appliance.



Communication Status: ☑

Storage Status: ☑

Secure Path Manager: ortrud.dem.cpqcorp.net:2301

Sep 23, 2002 16:38:18

Current Profile: drmw2ksapstrcl

Selected Item: \\sapstrcl\D:

```
⊟ Storage System0 (ID:50001fe100000970)
    ⊟ Controller (SN: ZG00212276)
        \\sapstrcl\D:
        \\sapstrcl\
        \\sapstrcl\F:
        \\sapstrcl\H:
        \\sapstrcl\E:
        \\sapstrcl\G:
    Controller (SN: ZG00212434)
```

Host	Controller	SCSI Port	HBA Slot	B-T-L	Mode	State
rienzi	ZG00212434	2	7	1-2-1	Alternate	Available
rienzi	ZG00212276	3	4	1-2-1	Preferred	Available
dalland	ZG00212276	2	7	1-0-1	Preferred	Active
dalland	ZG00212434	3	4	1-0-1	Alternate	Available
dalland	ZG00212276	2	7	1-0-1	Preferred	Active
dalland	ZG00212434	3	4	1-0-1	Alternate	Available

Polling Interval: 90 Network Activity:

Secure Path Manager - drmw2ksapstrcl

Figure 19

M - Set up the HSG Scripting Tool Kit (HSTK)

Install Command Scripter V1.0 and upgrade to V1.0A. Verify that Command Scripter can communicate with the HSG80 where C: is a device/LUN on the HSG80.

```
# cmdscript -fC:
```

Install the HSTK

```
# perl install.pl
```

Set the environment variable CLONE_HOME

```
>set CLONE_HOME=C:\%CLONE_HOME%\hstk\scripts
```

Create the configuration file of the target site controller

```
> generate_cfg.sh com=cs tar C:
```

Create the configuration file of the initiator site controller

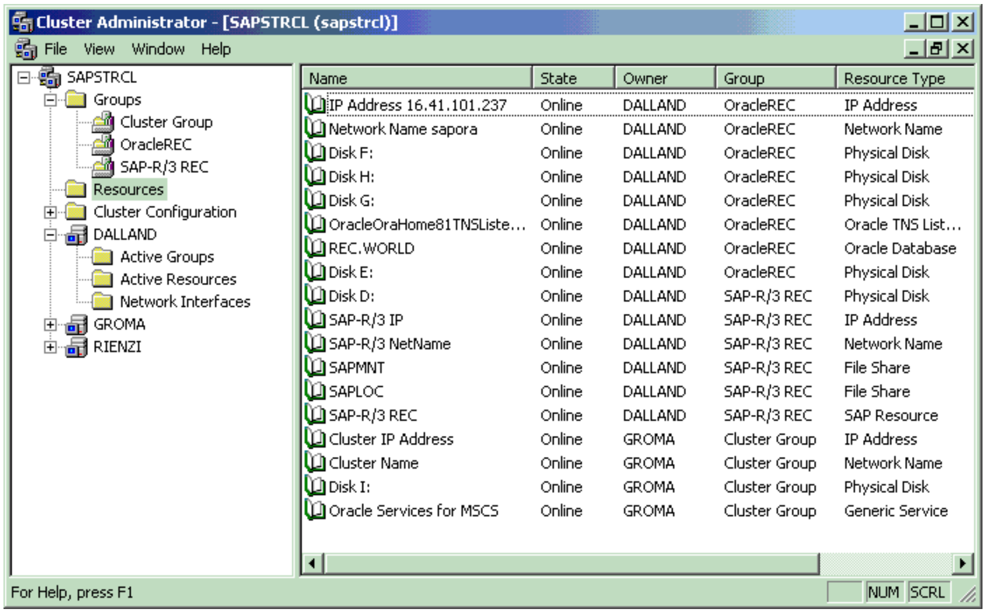
```
> generate_cfg.sh com=cs pri C:
```

Adjust the access rights in target site configuration file in preparation for a site failover

Run `drmconsole` to verify the HSTK works.

N - MSCS SAP cluster setup

After the installation of Oracle and Oracle FailSafe, the configuration of SAP in a MSCS environment is configured as shown in Figure 20.



- SAPSTRCL
  - Groups
    - Cluster Group
    - OracleREC
    - SAP-R/3 REC
  - Resources
  - Cluster Configuration
  - DALLAND
    - Active Groups
    - Active Resources
    - Network Interfaces
  - GROMA
  - RIENZI

Name	State	Owner	Group	Resource Type
IP Address 16.41.101.237	Online	DALLAND	OracleREC	IP Address
Network Name sapora	Online	DALLAND	OracleREC	Network Name
Disk F:	Online	DALLAND	OracleREC	Physical Disk
Disk H:	Online	DALLAND	OracleREC	Physical Disk
Disk G:	Online	DALLAND	OracleREC	Physical Disk
OracleOraHome81TNSListe...	Online	DALLAND	OracleREC	Oracle TNS List...
REC.WORLD	Online	DALLAND	OracleREC	Oracle Database
Disk E:	Online	DALLAND	OracleREC	Physical Disk
Disk D:	Online	DALLAND	SAP-R/3 REC	Physical Disk
SAP-R/3 IP	Online	DALLAND	SAP-R/3 REC	IP Address
SAP-R/3 NetName	Online	DALLAND	SAP-R/3 REC	Network Name
SAPMNT	Online	DALLAND	SAP-R/3 REC	File Share
SAPLOC	Online	DALLAND	SAP-R/3 REC	File Share
SAP-R/3 REC	Online	DALLAND	SAP-R/3 REC	SAP Resource
Cluster IP Address	Online	GROMA	Cluster Group	IP Address
Cluster Name	Online	GROMA	Cluster Group	Network Name
Disk I:	Online	GROMA	Cluster Group	Physical Disk
Oracle Services for MSCS	Online	GROMA	Cluster Group	Generic Service

Figure 20

The *R\3 Installation on Windows: Oracle Database* guide in SAP Library provides a step-by-step description for Oracle and MSCS integration within a mySAP.com environment.

O – Set up Fibre Channel boot

The following steps need to be taken in order to prepare a Windows 2000 installation on shared storage in order to replicate a systems boot disk to a target site.

- We recommend that you enable SCSI-3 mode on the HSG80.
- Prepare a LUN D1 on the shared storage and provide access to the HBA's on the system on which you intend to install Windows 2000.
- Change the order of the I/O controllers so that the FC HBA is in position 1 and can be seen before any local smart array controller.
- During system initialization, choose to enter the Emulex Bios <ALT E > and specify FABRIC POINT TO POINT as adapter attribute and select the created LUN as the primary boot device when prompted in the menu.
- Perform a standard SMART Start installation on ProLiant systems and hit F6 during the initial Windows 2000 boot phase to apply the KGPSA driver. This allows you to choose the prepared unit on the shared storage as a Windows 2000 installation device afterwards.
- Change the Windows 2000 page file to be on local storage.

A detailed setup is described in the Application Notes – Booting Windows from a Storage Area Network . In addition Microsoft has published a technical article with additional recommendations related to booting from a SAN at http://support.microsoft.com/default.aspx?scid=kb;EN-US;q305547.

P - Set up the Oracle standby database

The logical steps for creating a standby database for Oracle on a target site are:

- restore datafiles

- restore standby controlfiles after created on the primary database with

```
SVRMGR> ALTER DATABASE CREATE STANDBY CONTROLFILES AS <filename> ;
```

- modify init.ora files (if applicable)
- set up primary database tnsnames.ora file and test connection (if applicable)
- set up listener on the standby side (if applicable)
- mount the standby database using the standby controlfile

```
SVRMGR> STARTUP NOMOUNT ;
SVRMGR> ALTER DATABASE MOUNT STANDBY DATABASE ;
SVRMGR> RECOVER STANDBY DATABASE ;
```

The detailed steps for creating the initial standby database can be found in the *Major Preparation* section of the ORACLE8*i* Standby Database - Technical Report

glossary

array controller software (ACS): Software that is contained on a removable PCMCIA program card that provides the operating environment for the array controller.

association set: A group of remote copy sets that share common attributes. Members of an association set can be configured to transition to the same state at the same time. An association set shares the same log unit, has its host access removed from all members when one member fails , keeps I/O order across all members

asynchronous mode: A mode of operation of the remote copy set whereby the write operation provides command completion to the host after the data is safe on the initiating controller, and prior to the completion of the target command. Asynchronous mode can provide greater performance and faster response time, but the data on all members at any time cannot be assumed to be identical. *See also* synchronous mode.

availability: The percentage of time a functional unit is operational during a given interval of time. The interval of time and "operational" are defined by the requirements of the user.

block: A stream of data stored on disk or tape media that is transferred and error-checked as a unit. In a disk drive, a block is also called a *sector* (the smallest collection of consecutive bytes addressable on a disk drive). In integrated storage elements, a block contains 512 bytes of data, error codes, flags, and the block address header.

business continuity: This is the broadest term that covers all aspects of keeping your business *in* business including recovery, planning, information technology, environmental, and crisis situations. The concept of business continuity is gaining wide acceptance throughout the business world.

cache: A fast, temporary storage buffer in a controller or computer

cache memory: A portion of high-speed memory used as an intermediary between a data user and a larger amount of storage. The objective of designing cache into a system is to improve performance by placing the most frequently used data in the highest performance memory.

cascaded switch: As applied to the Data Replication Manager, a cascaded switch is one where its output is connected to the input of another switch, which then may in turn be connected to another switch or host or controller.

CLI: Command Line Interface. The CLI is the configuration interface that operates the controller software.

clone: A utility that physically duplicates data on any unpartitioned single-disk unit, stripeset, mirrorset, or striped mirrorset.

connection: As applied to the Data Replication Manager, a connection between two end Fibre Channel ports. An example is the connection between a Host Bus Adapter (by way of the Fibre Channel Switches) and the HSG80 controller. CLI commands available are ADD CONNECTIONS, SET *connection name*. *See also* link.

controller: A hardware device that uses software to facilitate communications between a host and one or more storage devices organized in an array. The HS-series StorageWorks family of controllers is all array controllers.

controller failover: The process that takes place when one controller in a dual-redundant configuration assumes the workload of a failed companion controller. Failover continues until the failed controller is repaired or replaced. The CLI command is SITE_FAILOVER. *See also* failback, dual-redundant configuration, *and* planned failover.

data integrity: The assurance that the data you receive is exactly what was sent you and that it stays that way until deliberately modified. Hardware problems, power failures, disk crashes, or software errors can threaten data integrity.

disaster tolerance: 1. As applied to high availability, the ability to maintain data integrity in operations following a catastrophic event to a computing site. 2. As applied to DRM, disaster tolerance provides the ability for rapid recovery of user data from a remote location when a significant event or a disaster occurs at the primary computing site. *See also* remote copy sets.

dual-redundant configuration: A storage subsystem configuration consisting of two active controllers operating as a single controller. If one controller fails, the other controller assumes control of the failing controller's devices. *See also* controller failover, site failover *and* failback.

fabric: A network of Fibre Channel switches or hubs and other devices.

failback: The process of restoring data access to the newly-restored controller in a dual-redundant controller configuration. The failback method (full copy or fast-failback) is determined by the enabling of the Logging or Failsafe switches, the selected mode of operation (synchronous or asynchronous), and whether the failover is planned or unplanned. *See also* controller failover, site failover, *and* dual-redundant configuration.

failover: *See* controller failover *and* site failover.

failsafe locked: The failsafe error mode can be enabled by the user to fail any write I/O whenever the target is inaccessible or the initiator unit fails. When either of these conditions occurs, the remote copy set goes into the inoperative (offline) state and the failsafe error mode is "failsafe locked." The CLI command SET *remote-copy-set-name* ERROR_MODE=FAILSAFE enables this error mode.

fast-failback: The synchronization of the initiator site with the target during a planned failover of the initiator subsystem. Write operations are logged to the target site write history log and, during the fast-failback, the initiator site is updated from the write history log. *See also* mini-merge, unplanned failover, planned failover, *and* write history logging.

Fibre Channel: An ANSI standard name given to a low-level protocol for a type of serial transmission. The Fibre Channel specifications define the physical link, the low level protocol, and all other pertinent characteristics.

frame: The basic unit of communication using Fibre Channel protocol. Each frame consists of a payload encapsulated in control information. The initiator breaks up the exchange into one or more sequences, which in turn are broken into one or more frames. The responder recombines the frames into sequences and exchanges. *See also* initiator.

ISL: Intersite link or Inter switch link. The abbreviation is context sensitive. *See also* multiple intersite links.

initiator: 1. A SCSI device that requests an I/O process to be performed by another SCSI device, namely, the SCSI target. The controller is the initiator on the device bus. 2. For subsystems using the disaster tolerance Data Replication Manager solution, the initiator is the site that is the primary source of information. In the event of a system outage, the database would be recovered from the target system. *See also* target.

latency: The amount of time required for a transmission to reach its destination.

link: A connection between two adjacent Fibre Channel ports, consisting of a transmit fiber and a receive fiber. An example is the connection between the Fibre Channel switch port and the HSG80 controller. *See* also connection.

logical unit: A physical or virtual device addressable through a target ID number. The logical unit numbers (LUNs) use their target's bus connection to communicate on the SCSI bus. *See* also LUN.

Logical Unit Number: *See* LUN.

LOG_UNIT: A CLI command switch that (when enabled) assigns a single, dedicated log unit for a specific association set. The association set members must all be in the NORMAL error mode (not failsafe).

LUN: Logical Unit Number. A value that identifies a specific logical unit belonging to a SCSI target ID number. A number associated with a physical device unit during a task's I/O operations. Each task in the system must establish its own correspondence between logical unit numbers and physical devices.

Mean Time Between Failure (MTBF): a statistically derived length of time a user may reasonably expect a component, device, or system to work between two incapacitating failures.

mini-merge: As applied to the Data Replication Manager, the data transfers to be made whenever a target becomes inaccessible. This occurs when both links or both target controllers have gone down. The transfers that would have been made are instead logged into the association set's assigned log unit to wait until the remote copy set subsystem comes back online. *See* fast-failback, write history logging.

mirroring: The act of creating an exact copy or image of data.

mirrorset: 1. A group of storage devices organized as duplicate copies of each other. Mirrorsets provide the highest level of data availability at the highest cost. Another name for RAID 1. Also called *mirrored units* or *mirrored virtual disks*.

2. Two or more physical disks configured to present one highly reliable virtual unit to the host.

3. A virtual disk drive consisting of multiple physical disk drives, each of which contains a complete and independent copy of the entire virtual disk's data.

multiple intersite links

Each intersite link (ILS) is a fiber link between two switches. As applied to Data Replication Manager, increasing bandwidth between switches is handled by adding additional connections between the switches, to a maximum of two connections.

mission critical: A term applied to information systems upon which the success of an organization depends and the loss of which results in unacceptable operational or financial harm.

normal member: A mirrorset member that, block-for-block, contains exactly the same data as that on the other members within the mirrorset. Read requests from the host are always satisfied by normal members.

normalizing: A state in which, block-for-block, data written by the host to a mirrorset member is consistent with the data on other normal and normalizing members. The normalizing state exists only after a mirrorset is initialized. Therefore, no customer data is on the mirrorset.

normalizing member: A mirrorset member whose contents are the same as all other normal and normalizing members for data that has been written since the mirrorset was created or since lost cache data was cleared. A normalizing member is created by a normal member when either all of the normal members fail or all of the normal members are removed from the mirrorset. *See also* copying member

other controller: The controller in a dual-redundant pair that is not connected to the controller serving your current CLI session with a local terminal. *See also* this controller *and* local terminal.

planned failover: As applied to the Data Replication Manager, an orderly shutdown of the controllers for installation of new hardware, updating the software, and so on. The host applications are quiesced and all write operations permitted to complete before the shutdown. The controllers must be in synchronous operation mode before starting a planned failover. *See also* synchronous mode *and* unplanned failover.

reliability: A measure of how dependable a component or system is once it is in use. Availability might be considered the sum of reliability and data integrity.

remote copy sets: A feature that allows data to be copied (mirrored) from the originating site (initiator) to a remote site (target). The result is a mirror copy of the data (remote copy set) at two disparate sites. Used in disaster tolerant applications such as the Data Replication Manager. CLI commands available are ADD REMOTE_COPY_SETS, SET *remote-copy-set-name*, SET *controller* REMOTE_COPY.

remote copy set metadata: Data that describes the remote copy set membership and state. To assist with site failover, this metadata is located in the mirrored write-back cache on the controller where each member resides. Backup copies of the metadata reside in the controller NVRAM at each site. Only the initiator modifies the metadata and ensures all copies are subsequently updated.

site failover: The process that takes place when storage processing is moved from one pair of controllers to another. All processing is shifted to the target (remote) site. This is possible because all data generated at the initiator site has been replicated at the target site, in readiness for such a situation.

storage array: An integrated set of storage devices. Storage arrays can be manipulated as one unit with a single command.

storage unit: The generic term for storagesets, single-disk units, and all other storage devices that are installed in a subsystem and accessed by the host. A storage unit can be any entity that is capable of storing data, whether it is a physical device or a group of physical devices.

storageset: 1. A group of devices configured with RAID techniques to operate as a single container. 2. Any collection of containers, such as stripesets, mirrorsets, striped mirrorsets, JBODs, and RAIDsets.

surviving controller: The controller in a dual-redundant configuration pair that serves its companion's devices when the companion controller fails.

synchronous mode: A mode of operation of the remote copy set whereby the data is written simultaneously to the cache of the initiator subsystem and the cache of the target subsystem. The I/O completion status is not sent until all members of the remote copy set are updated. *See also* asynchronous mode.

target: A SCSI device that performs an operation requested by another SCSI device, namely the SCSI initiator. The target number is determined by the device's address on its SCSI bus. For subsystems using the disaster-tolerant Data Replication Manager solution, data processing occurs at the initiator site and the data is replicated or mirrored to the target site. In the event of a system outage, the database is recovered from the target system. *See also* initiator.

this controller: The controller that is serving the current CLI session through a local or remote terminal. *See also* other controller.

unit: A container made accessible to a host. A unit may be created from a single disk drive or tape drive. A unit may also be created from a more complex container, such as a RAID set. The controller supports a maximum of eight units on each target.

unplanned failover: As applied to the Data Replication Manager, recovery from an unplanned outage of the controllers. This may occur when the site communication is lost or it may be due to some other failure whereby remote copy sets cannot be implemented. The controllers do not perform an orderly shutdown. *See also* planned failover.

write history logging: As applied to the Data Replication Manager, the use of a log unit to log a history of write commands and data from the host. Write history logging is used for mini-merge and fast-failback. *See* mini-merge *and* fast-failback.

zoning: As applied to the Data Replication Manager, an optional, licensed feature of the SilkWorm switch that allows a finer segmentation of Storage Area Networks (SANs) by allowing ports or WWN addresses to confine access to devices that are in a common zone.

for more information

HP Links

For further information on mySAP.com, please refer to the following HP publications:

Network Storage Solutions:

http://www.compaq.com/products/sanworks/drm/index.html

http://www.compaq.com/storage/whitepapers.html#soft

Data Replication Manager Design Guide - Application Notes

Oracle Databases Replication and Solutions

ftp://ftp.compaq.com/pub/solutions/customsystems/en-orasolrep-wp-02.pdf

Oracle Storage Compatibility Testing - Remote Mirroring White Paper

http://www.compaq.com/products/storageworks/library/whitepapers/RF-1208-03-010301.html

Disaster Tolerance -The Technology of Business Continuity

ftp://ftp.compaq.com/pub/products/sanworks/techdoc/drm/12D3-0500A-WWEN.pdf

Application Notes – Booting Windows from a Storage Area Network

CNT Links

http://www.cnt.com/products/ultranet_edge_storage_router/

Oracle Links

ORACLE 8i Standby Database

http://technet.oracle.com/deploy/availability/pdf/stby8i_twp.pdf

Microsoft Links

http://www.microsoft.com

http://support.microsoft.com/default.aspx?scid=kb;EN-US;q305547

SAP Links

SAP Documentation Library

http://help.sap.com/

SAP OSS Note 559730 *hp StorageWorks datasafe for mySAP.com (Windows 2000/Oracle)*

All brand names are trademarks of their respective owners.
Technical information in this document is subject to change without notice.

11/2002