hp storage

november 2002

invent

technical blueprint

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

table of contents

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.

The solution was designed and tested for two approaches to replication. Depending on your needs, you can replicate the entire SAP database or just the SAP database redo log information. Either implementation of the solution provides a failover/recovery time at a remote computing site measured in minutes instead of hours or days.

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.

solutions 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 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.

business needs

eCommerce is a critical component of business regardless of market segment and presents an increasing pressure on businesses as well as software providers to run these applications on a 24 × 7 × 52 basis. As a result, scheduled downtime has rapidly begun having a negative business impact, and any unforeseen disruption for any length of time at a computing site could be potentially disastrous. Thus the need for data replication techniques evolved. In the SAP environment, data replication applies primarily to data distribution in applications such as disaster recovery, data warehousing, and decision support.

As customer applications and 24 × 7 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.

solution design and design rules

Datasafe for mySAP.com (Windows 2000/Oracle) consists of 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 Microsoft Windows 2000 Cluster Service (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 nondisaster 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.

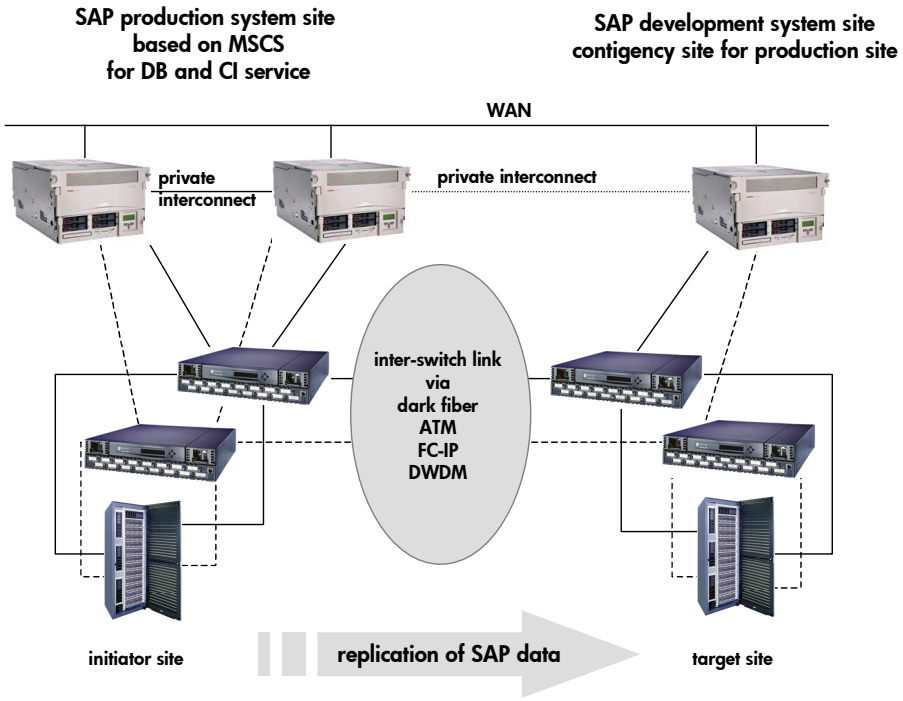

figure 1

The StorageWorks 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.

component review

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 Fibre Channel (FC) switches between the initiator and target sites. Dark fiber and Fibre Channel-Internet Protocol (FC-IP) gateways are certified as well as ATM (Asynchronous Transfer Mode) and DWDM (Dense Wavelength Division Multiplexing). 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 fiber 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 on-premises 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 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 self-test suite for 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. www.compaq.com/products/storageworks/ma8kema12k/index.html

Microsoft Windows 2000 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 for 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 it 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.

synergy of components

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

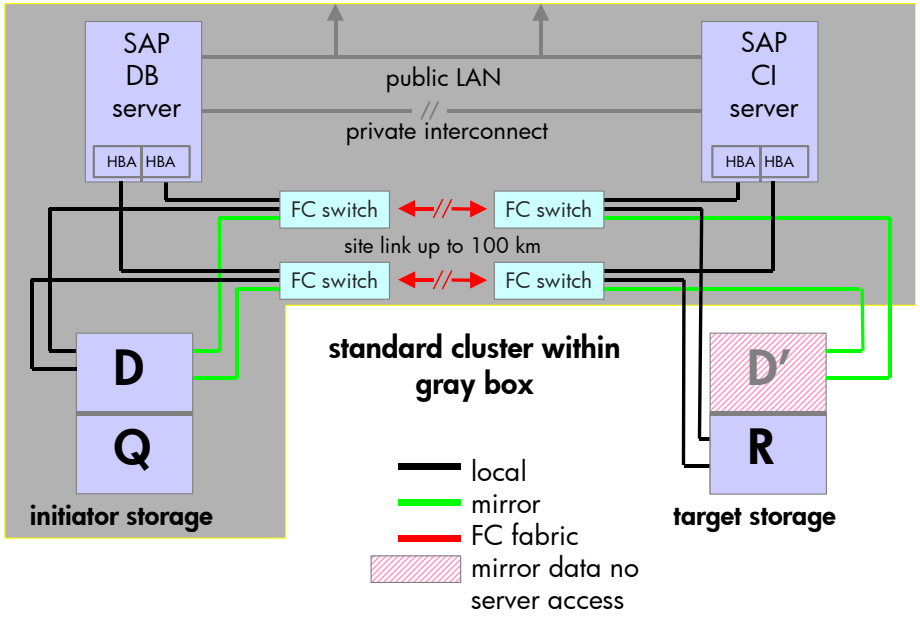

figure 2

replicating the entire Oracle database

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 command-line interface (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.

replicating Oracle redo log information only

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.

managing DRM failover and failback

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.

adaptable, extensible, controllable

adaptable: for SAP and Microsoft cluster environments

The StorageWorks Datasafe solution for mySAP.com enhances the Microsoft high-availability features with the disaster-tolerant capabilities of HP StorageWorksData Replication Manager, maintaining application performance dependent 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%.

extensible: to enable you to do business they way you choose to

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 100-Mbit network for a write-intensive workload showed an overhead in the range of 3% for a full database replication scenario when compared to direct Fibre connection. Once the quality of service for the IP network is reduced through bandwidth limitations and route delays, the DRM overhead will increase to compensate for this.

Two approaches to replication were tested in the solution. 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 StorageWorks 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 disaster-tolerant (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.

controllable

An important consideration in an SAP customer's disaster-tolerance 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 30 MB/s per HSG80 with direct Fibre ISLs (inter-switch links) 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 nonclustered, 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.

solution-specific configuration hardware

hardware sample configuration

(The solution is not limited to PL8500 servers or a specific amount of memory or the number of switches. Every HA/F500 enhanced DT configuration meets the requirements of Datasafe.)

	initiator site	#	target site	#
server	PL8500	2	PL8500	1
	CPU 4 GB memory	8	CPU 4 GB memory	8
	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
		Supporting BOTH SITES		
FC IP gateways		CNT StorEdge Router 1000		
SAN management		HP OpenView Storage Management Appliance		
client (SAP)		ML370		
network infrastructure	The servers and the client are connected via 10/100 NICs			

software

software	version
Windows 2000 Advanced Server and Data Center	SP3
HP StorageWorks Secure Path	4.0
HP StorageWorks Array Controller Software (ACS)	8.7P.0
Fabric OS	2.6c
SAP R/3	4.6D
Oracle	8.1.7
HP StorageWorks Command Scripter	1.0A
HP StorageWorks Data Replication Manager failover scripts	1.6

why hp

- **ensures continuous uptime** by systematically eliminating single points of failure across the board—from the hardware right up to the SAP application level; HP solutions for SAP provide instant, automated switchover for hardware, OS, database, and SAP components, ensuring service continuity in the event of a failure
- **delivers rock-solid security** with solutions that perform both encryption for secure transactions and instant authorization/authentication checks
- **prepares you for sudden load peaks** by keeping extra CPU and storage capacity on standby for instant activation whenever needed
- **enables rapid deployment of mySAP.com solutions** because HP consultants, cooperating closely with SAP, use a structured approach based on SAP best practices to help customers design and implement the IT infrastructure that their enterprise needs to ensure a smooth, speedy rollout of mySAP.com solutions
- **provides end-to-end control** of the entire mySAP.com environment with management tools and support services that manage every component, from hardware to application—even in distributed, Internet-based system environments
- **enables faster recovery time** to ensure that the customer's SAP environment is restored with minimal impact to their business
- **provides the highest level of storage performance in the industry**, which contributes to higher productivity of the customer's SAP resources

glossary

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

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.

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

CLI (command line interface): 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.

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.

dark fiber: Optical fiber that is dedicated to a single customer, who is responsible for attaching the voice or data equipment and lasers needed to "light" the fiber.

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 (either full copy or fast-failback) is determined by the enabling of logging or Failsafe.

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.

initiator: 1. A SCSI device that requests an I/O process be performed by another SCSI device, namely the SCSI target. The controller is the initiator on the device bus.
2. (For subsystems using the Data Replication Manager disaster tolerance 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.

LUN: Logical unit number. A value that identifies a specific logical unit belonging to a SCSI target ID number.

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

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

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 (DT) applications such as the Data Replication Manager. CLI commands available are ADD REMOTE_COPY_SETS, SET *remote-copy-set-name*, SET *controller* REMOTE_COPY.

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.

synchronous mode: A mode of operation of the remote copy set whereby 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.

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.

for more information

To learn more about HP storage solutions for SAP, contact your local HP sales representative or visit our Web site at: www.hp.com

11/02