

How the HP Integrity NonStop server leverages the HP StorageWorks XP disk array



The enterprise storage disk array	2
LDEVs, LUNs, and mirrors	2
Connecting to multiple systems	3
Distributing a database for disaster tolerance.....	4
Host-based versus disk array-based replication.....	5
Leveraging the StorageWorks XP disk array with NonStop RDF/ZLT software	7
Meeting your continuity requirements	9
For more information.....	10

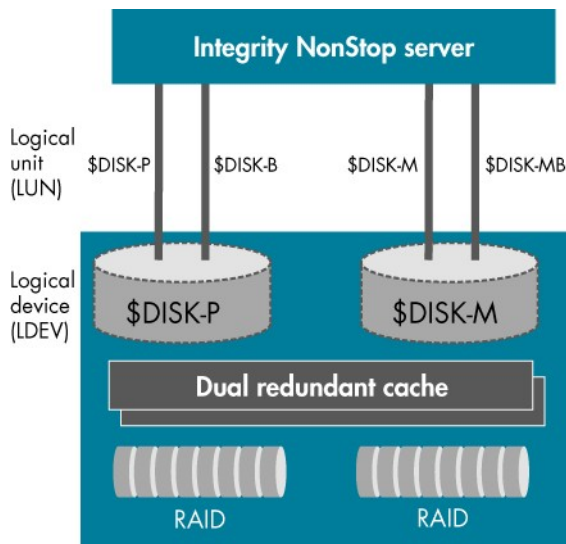
The enterprise storage disk array

The HP StorageWorks XP disk array consists of hundreds of physical disk drives supported by a highly available fault-tolerant infrastructure. The XP disk array's processors, power supplies, cooling, data paths, and most important, the disk cache are all separate and redundant. There is no single point of failure, and most upgrades can be done online. The XP disk array creates virtual images of the physical disks so that disk volumes of almost any size can be presented to the attached heterogeneous computer systems. This virtual volume functionality is similar to that offered by HP NonStop Storage Management Foundation (NonStop SMF) software, but with higher performance and no compatibility issues. Through the use of RAID technology, single and multiple disk failures are completely masked from the attached systems, and alternate disks can automatically replace failed ones.

LDEVs, LUNs, and mirrors

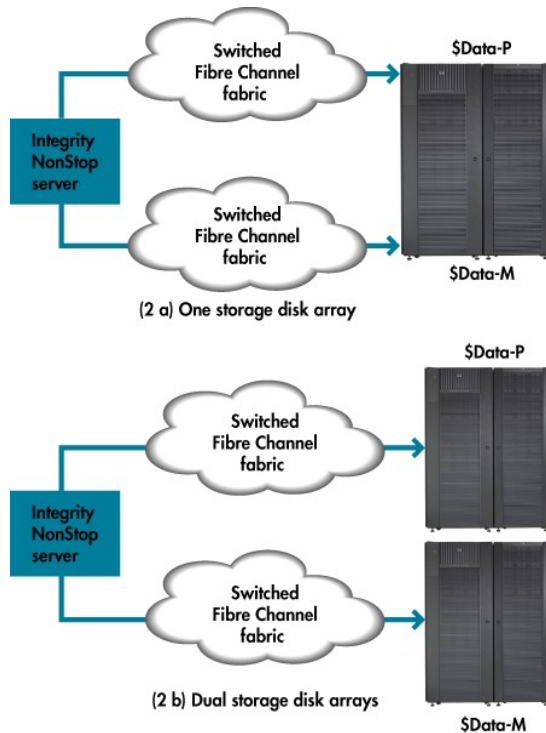
An XP virtual disk volume appears to the Integrity NonStop server as a logical device (LDEV). An LDEV can be nonmirrored, or two LDEVs can be combined into a mirrored pair, for example, \$DISK-P and \$DISK-M (see figure 1). The physical path from the HP Integrity NonStop server to an LDEV is called a logical unit number (LUN). Each LDEV has two LUNs, a primary and a backup for fault tolerance, for example, \$DISK-M and \$DISK-MB. The LUN paths should be kept completely separate—all the way from the Integrity NonStop server to the StorageWorks XP disk array—by routing each LUN via a different HP ServerNet fabric and a different Fibre Channel System Adaptor (FCSA) port, and traversing different storage area network (SAN) fabrics.

Figure 1. LDEVs, LUNs, and mirrors.



Because the StorageWorks XP disk array uses RAID technology to mask disk failures, do Integrity NonStop server customers still need to implement traditional host-based mirroring? In a word, yes. Host-based mirroring provides better protection against SAN, fiber, switch, and connector problems, taking full advantage of the error recovery built into the disk process (DP2). DP2 implements end-to-end checksums to ensure that the data read from a disk is the same as the data that was written to the disk. If the checksum doesn't match, the mirror volume is read, and the data is corrected on the bad disk. This is true whether the mirrors are present on one XP disk array (see figure 2 a) or two XP disk arrays (see figure 2 b).

Figure 2. Host-based mirroring.



Connecting to multiple systems

Unlike a traditional internal disk enclosure, multiple computer systems can connect to and mount LDEVs on the same StorageWorks XP disk array. However, only one Integrity NonStop server can mount a specific LDEV at a time. This is not a loss of functionality; two Integrity NonStop servers or NonStop S-series servers could not mount the same internal disk at the same time either. Two systems can still open the same file via HP Expand networking software.

NonStop Transaction Management Facility (NonStop TMF) software supports the XP disk array just as it supports the internal disk, thus ensuring database integrity. Before the database is updated, NonStop TMF writes the before and after images of the affected database rows to its transaction log, which can be on an internal disk or the XP disk array.

As with an internal disk, a volume must be cleanly dismounted from one system before it can be mounted on another system. That is, all of the files on the volume are closed, the volume is disabled in NonStop TMF, the disk cache is flushed, and the volume is removed through the Subsystem Command Facility (SCF). However, there are several caveats for moving volumes from one system to another, which are discussed in "Host-based versus disk array-based replication."

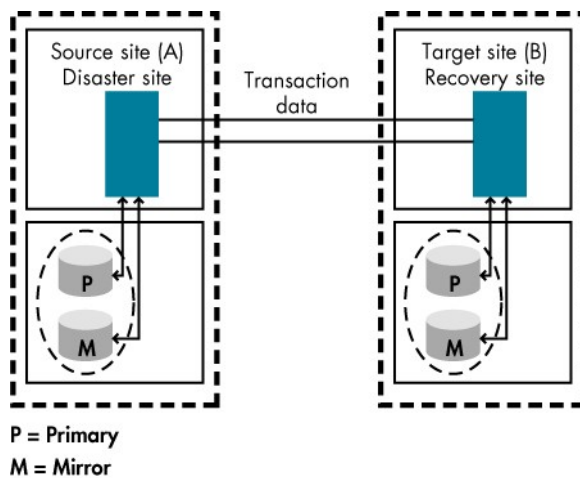
Distributing a database for disaster tolerance

If a current copy of the same database must be present on multiple systems for disaster recovery, NonStop Remote Database Facility (NonStop RDF) software should be used. NonStop RDF implements host-based replication that streams database changes from one Integrity NonStop server to another. Third-party data replication software also can be utilized for heterogeneous database replication or data transformation of databases on the disk array.

Before the primary database is updated, before and after images of the change are physically written to the NonStop TMF transaction log. If a database change is not present in the transaction log, it therefore never happened. By using this log for data replication, NonStop RDF avoids having to rely on data volume cache flushes. This means that system performance remains high, and the database is not rendered inconsistent if a failure prevents flushing of the data volume cache to the disk.

In addition, NonStop RDF replicates only those files designated as critical by the administrator and then sends one copy of only the changed fields (not the entire record) from the source system to the target system. Communications overhead is minimized so that slower and not distance-limited links can be employed between systems. NonStop RDF understands the state of every transaction it is replicating (see figure 3). This is true whether the transaction is wholly contained on one system or spans multiple systems.

Figure 3. NonStop RDF software knows the transaction state.



During takeover processing, NonStop RDF software backs out any transaction with an unknown final state while applications on the target continue processing. This ensures complete database consistency on a single target system or across an entire complex of target systems. The end result is that there is little to no impact on the target system during a takeover operation. There are no reboots and no need for a transaction manager or database recovery tool to scan logs for incomplete transactions.

Host-based versus disk array-based replication

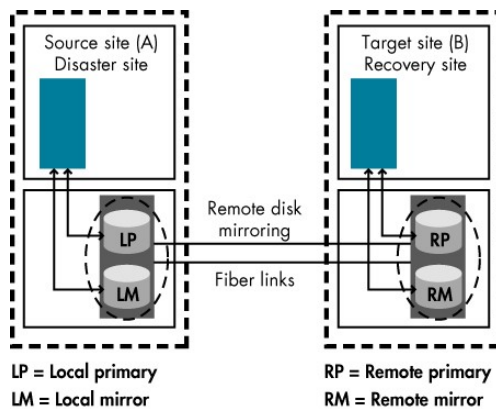
The StorageWorks XP disk array offers its own data replication facilities through StorageWorks Business Copy XP software, which creates a local copy of a disk volume, and through StorageWorks Continuous Access XP software, which streams disk volume changes from a local disk array to a remote disk array, usually via a high-speed interconnect (see figure 4). This begs the question, “Why not use these native StorageWorks XP replication facilities instead of NonStop RDF software?”

First, some background. The Integrity NonStop server was designed as an integrated hardware/software stack, including the disk subsystem, transaction monitor, database, and database replication, to maximize data integrity. NonStop TMF software, in addition to ensuring the integrity of critical data and maintaining a log of every transaction performed on the Integrity NonStop server, also monitors the status of every disk volume on the Integrity NonStop server including mounting and unmounting. This ensures that access to inconsistent data is prohibited.

Using Business Copy XP on an active data volume is similar to taking an ordinary tape backup. Files are being created and deleted, and transactions and database block splits are in flux. In other words, the disk volume and the database on it are in a corrupt state. Because of its transaction log, NonStop TMF supports online dumps which, unlike a backup of an active disk volume, can be made consistent when they are restored to the system. And when a system is restarted after a crash, NonStop TMF will automatically use its log to ensure that committed transactions are completed and that indeterminate transactions are backed out before allowing access to the database. If a business copy is made from an active disk volume and mounted as a new volume, there is no transaction log for it that can be used to make the volume and database consistent. For this reason, Business Copy XP only is supported on an Integrity NonStop server for a disk volume that is in a quiescent state (active files on the volume are closed, the volume is disabled in NonStop TMF, and the disk cache is flushed).

The use of Continuous Access XP has similar issues. Continuous Access XP replicates the bits from a source disk volume to a target disk volume, and as with Business Copy XP, an active volume is corrupt at any point in time. Additionally, the design of the disk subsystem on an Integrity NonStop server forces disk writes and cache flushes in a very specific order to ensure that information will be recoverable in almost any circumstance. But like Business Copy XP, Continuous Access XP can be used to copy a volume that is in a quiescent state, except that there is an additional limitation. NonStop SQL software and some applications write system-specific information into their databases in such a manner that even a cleanly dismantled disk volume may not be usable on a different system.

Figure 4. Continuous access replication.



For offline disaster recovery, it is possible to implement remote host-based mirroring for an entire disk configuration—including the NonStop TMF audit volumes. This is done by placing all the primary disk volumes on one StorageWorks XP disk array collocated with the source system and placing all the mirror disk volumes on a second disk array at a remote location, possibly collocated with a target or backup system. After a failure of the protected system, all of the mirror disks are switched to the backup system (with the same node name/number as the source system), and the source system's NonStop TMF configuration is brought up to start volume recovery. Unlike the seconds to minutes that NonStop RDF takes to perform a failover operation (see figure 5), this can take hours and any workload on the target system first must be jettisoned. If there was a NonStop TMF configuration on the target system, it is no longer in use, and therefore all of the dumps are unusable. Volumes on the target system need to be integrated into the switched NonStop TMF configuration and new dumps taken before the application is brought up.

Unlike the Integrity NonStop server databases, not all databases have an integrated replication facility, and in fact, many were designed to rely on the native replication capabilities of a disk array. This is why the use of Business Copy XP and Continuous Access XP are so popular on systems other than the Integrity NonStop system. After a source system outage, the replicated disks are switched to the target system, the operating system scans and repairs the disk structures, and then the database manager on the target system uses its logs on the replicated volumes to make the database consistent. Assuming that there is no system-specific information in the files, the copy can be used by an application on another system. For many computer systems, the Continuous Access XP feature of the StorageWorks XP disk array is the ideal solution for disaster recovery.

But have you thought about what happens if transactions span multiple source computer systems? The database software must know how to coordinate recovery across all of the backup systems so that the entire multisystem environment is kept consistent. Do you know if the replication solution being proposed to support your application is capable of parallel recovery across multiple systems or nodes, how long the recovery will take, and to what extent the applications running on the backup systems will be affected during recovery operations? With NonStop RDF software you are assured that most any application environment protected by NonStop TMF software can be made disaster tolerant without any changes in coding.

Steps to switch an entire disk configuration: NonStop RDF versus Continuous Access XP

NonStop RDF

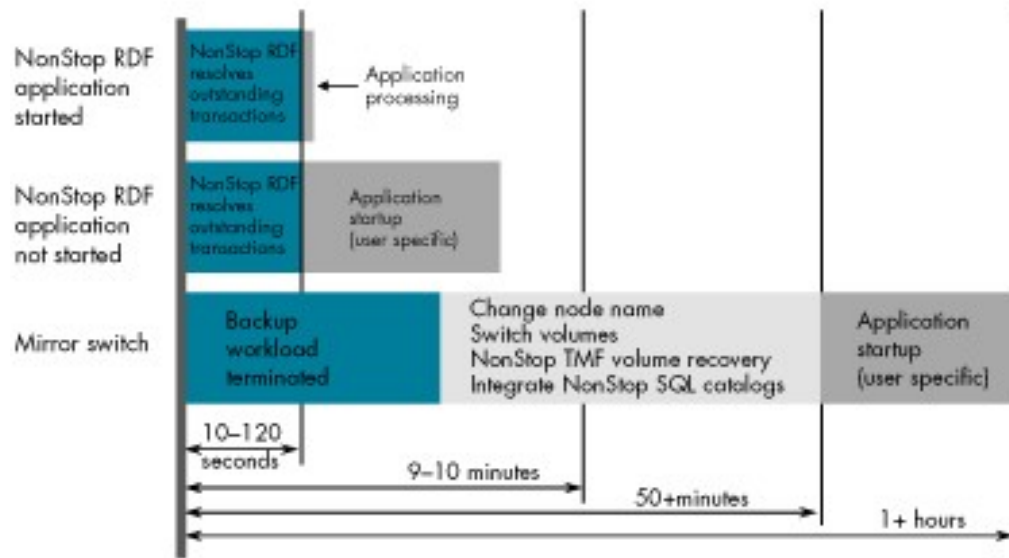
1. Issue NonStop RDF takeover command.
2. Start applications.

Continuous Access XP

1. Stop applications on the backup system.
 2. Stop NonStop TMF.
 3. Stop the operating system.
 4. Rename the system.
 5. Start the operating system.
 6. Switch disks to the backup system.
 7. Start NonStop TMF with the configuration from the failed system.
 8. Volume recovery by NonStop TMF.
 9. Integrate existing disks into the moved NonStop TMF configuration.
 10. Integrate both systems' tables into the same NonStop SQL catalog.
 11. Reconfigure the NonStop TMF tape catalog if necessary.
 12. Take new online dumps.
 13. Bring up applications.
-

Figure 5. Backup system takeover time line.

Takeover decision



Leveraging the StorageWorks XP disk array with NonStop RDF/ZLT software

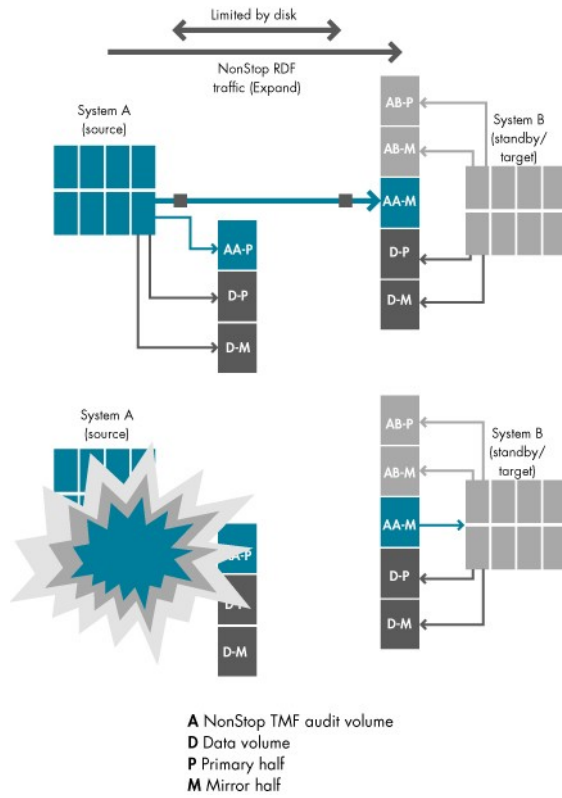
Transactions that have been committed on the primary system but not yet transmitted to the backup system can be lost after a catastrophic failure. For Integrity NonStop server customers that need ultimate protection, with no loss of committed transactions after a catastrophic failure, HP offers NonStop RDF/Zero Lost Transactions (NonStop RDF/ZLT) software, an add-on to the NonStop RDF/IMPX product. NonStop RDF/ZLT software is supported by dual disk arrays, one of which is used to locate half of the mirrored NonStop TMF transaction log volumes remotely from the source system (see figure 6).

The hardware and infrastructure consist of one or more remote disk mirrors and the communications infrastructure to support them. The disk mirrors containing the NonStop TMF audit trail are located remotely from the source system and are connected to both the source and *standby* systems but controlled by the source system. The NonStop RDF/ZLT standby system refers to the system that the remote disk mirrors will be connected to after a failure of the source system. The standby system can be the target system, a third system, or even the source system for testing purposes. The distance from the disks to the source and standby systems is dictated by the disk technology used in the configuration, but the source and target systems can be any distance from each other. The distance between the source and standby systems can be twice as long as the distance between the disk mirrors and each system because the remote disk mirrors can be located in between the systems. With a command to NonStop RDF/ZLT on the target system, the records not already transmitted to the target system are read from the audit mirrors and applied to the database.

In any configuration, the tradeoff is slightly longer recovery time versus no loss of committed transactions. Other than the physical connections to the additional disks, no changes are required to the application, source system, or target system. Transaction processing already running on any of the

systems can continue with little to no impact before a system failure, and only applications on the failed system are affected.

Figure 6. NonStop RDF/ZLT leverages StorageWorks XP.



Meeting your continuity requirements

Not all businesses, or even all Integrity NonStop server customers, face the exigencies of today's automated stock exchanges where the recovery time objective (RTO) and recovery point objective (RPO) must be zero. But it's a fact that many applications are becoming increasingly time-critical and that lost time is equating to ever larger amounts of lost dollars.

The RTO of a back-end system such as shipping or billing can be relatively relaxed because orders can always be printed out and mailed or faxed to the warehouse and credit cards can be cleared manually. But if the Web-based, customer-facing interface to the business is down, customers will go elsewhere. Companies need to look at each business process and each system at a time to determine what is acceptable in terms of disaster protection. The costs of downtime must be finely weighted, both in and of itself, and versus the costs of protecting against it. In many cases, switching the entire disk configuration may be a satisfactory solution when applications can be offline for hours to days and when extremely low RPO is not a decisive factor. In other cases, nothing short of unbroken business continuity will do, which argues in favor of the NonStop RDF software approach.

The real path to continuity is to create a disaster-tolerant environment that distributes the processing across multiple sites, removing the need for recovery. When a disaster strikes, surviving portions of the environment can immediately take over processing for the failed portions, maintain database consistency, and keep business-critical services online without being hampered by a lengthy recovery process.

Disaster-tolerant computing is not the future—businesses can't wait that long. It exists today and, in fact, has existed on the NonStop server for almost 20 years. Using out-of-the-box components such as the Integrity NonStop server and the StorageWorks XP disk array, HP is already providing the ultimate in application continuity for numerous companies that are not willing to take a chance on "almost works" data replication.

And one final note. Technology should never be selected before a thorough risk analysis and business impact analysis of key business processes. Then, whatever technology or combination of technologies is employed for disaster tolerance, it should fit within a larger continuity planning and process framework for ensuring business survival.

For more information

www.hp.com/go/nonstopcontinuity

Continuous application availability for HP Integrity NonStop servers data sheet

Continuous application availability for HP Integrity NonStop servers brochure

HP Integrity NonStop server security overview

HP NonStop Remote Database Facility Configurations technical brief

© 2005 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

