

The disaster tolerance continuum

Technical brief



Contents

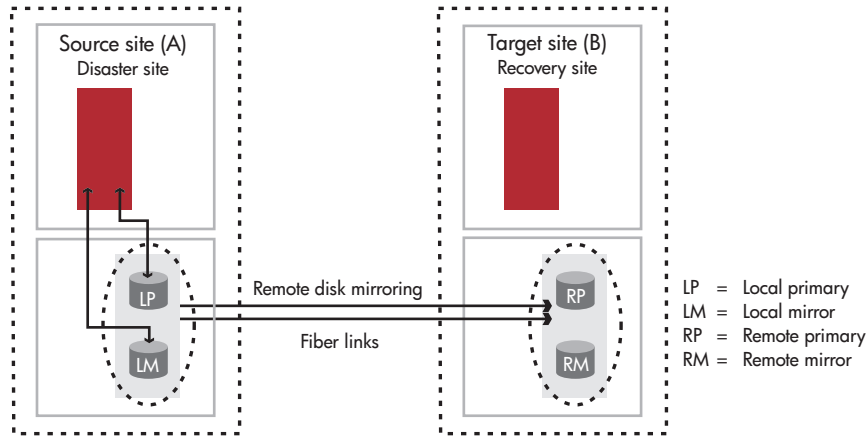
- Evaluating recovery requirements 3
- Challenges to RPO 4
- Challenges to RTO 4
- Typical systems, typical solutions 5
- A complex approach to recovery 5
- Disaster tolerance versus disaster recovery 5
- Lowering RTO and RPO 6
- Application uptime is no accident 7

HP NonStop RDF software helps to create a disaster-tolerant environment to enable continuous application availability.

Disaster tolerance is becoming a “must have” for an increasing number of applications. But disaster tolerance is not an absolute. There are levels of tolerance to system and site failures, just as there are levels of tolerance to any other unwanted stimuli—noise, heat, you name it. With many applications, it is enough to react quickly to a site disaster and recover from it. But for others, recovery is not nearly good enough. Recovery, no matter how fast and how complete, implies that some damage has already been done—services disrupted, money lost.

Customers with applications that are unable to tolerate any type of loss—in time, data, transaction integrity, and so on—require continuous application availability. This brief presents basic concepts about the disaster tolerance continuum. It also compares and contrasts two high-end approaches to disaster protection: the replicated enterprise storage system (ESS) and storage area network (SAN) approach, which is largely hardware-centric, and the environmental approach recommended by the HP NonStop Enterprise Division, which, through the use of HP NonStop Remote Database Facility (NonStop RDF) software and distributing processing across multiple sites, removes the need for recovery entirely.

Figure 1. ESS replicates bits.



Note: RP and RM are connected to the target site upon failure of the source site

Evaluating recovery requirements

There are two essential metrics to keep in mind when evaluating business process availability requirements:

- **Recovery time objective (RTO)**—the tolerable maximum length of time that a business process can be unavailable
- **Recovery point objective (RPO)**—how much work in progress can be lost and (by extension) how critical is the impact of that loss

RTO and RPO are interrelated, but should be evaluated independently. The interplay between the two goals needs to be carefully judged for each business process when defining the functional requirements for a disaster protection plan.

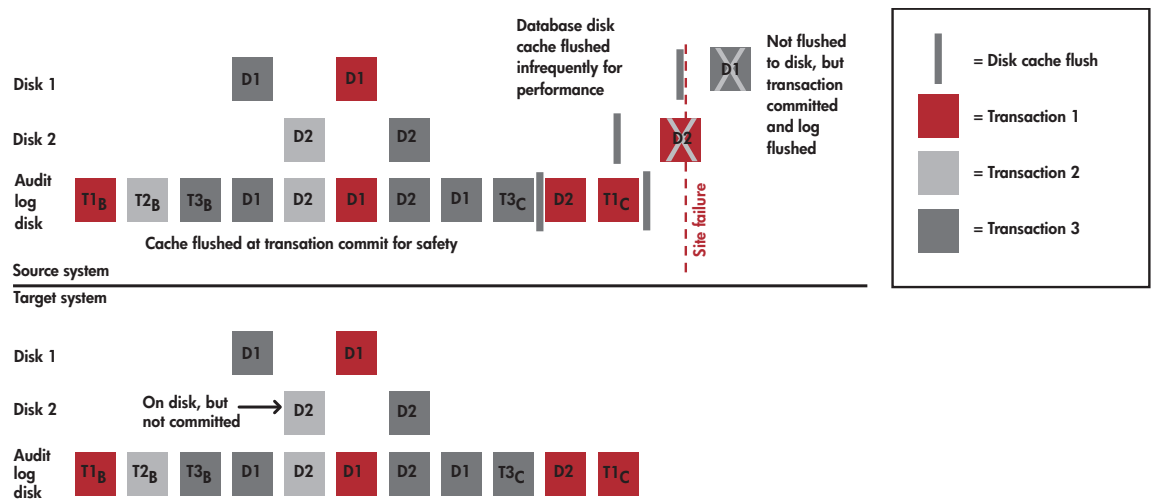
With RTO and RPO in mind, let's look at automated disaster protection alternatives. Given the minimal RTO and RPO characteristics of most applications running on NonStop servers, it is likely that tape backup and restoration is not an adequate option, nor is shipping transaction logs to another system and then applying them to the backup database in a batch manner. Assuming that they will be supported on NonStop servers in the future, replicated enterprise storage systems can provide a reasonable alternative in some cases. With an ESS or storage area network solution, the backup database must be on physical disk, and to obtain the lowest RTO, those disks should be able to

be connected to a target system that is ready to take over from the source system and pick up the application load. As bits are changed on the disks that hold the primary database, the primary ESS streams the changes to the backup disks in either near real time (asynchronous) or real time (synchronous or lockstep).

ESS and SAN solutions replicate data by sending track-by-track changes from the source ESS at the primary site to a target ESS at a remote site. When this replication occurs in synchronous mode, ESS vendors guarantee that the bits on the disks of the target ESS are identical to those on the disks of the source ESS at any point in time (see figure 1).

The RPO is subject to two factors: how often data is flushed from the host computer system to the ESS module and how fast the data can be sent over the communications link to the target ESS module. Lockstepping requires a fast connection, typically using dedicated (dark) optical fiber, between the local and remote ESS units. However, this is an expensive and not always available option. Moreover, there is a distance limitation to maintain the real-time nature of the replication. If these caveats are not an issue, customers will find that, with asynchronous hardware replication, attached computer systems are not affected by replication overhead or latency. This is because once data is written to the source ESS by the source computer system, it is copied to the target ESS typically from the source ESS system's disk cache buffers. Let's look at how this can affect application uptime during a failure situation.

Figure 2. The physical disk does not equal the logical database.



Challenges to RPO

Computer systems buffer substantial amounts of disk information in an internal disk cache held in memory because frequent writes to any kind of external storage substantially slows down an application. In a disaster recovery scenario, if the data has not yet been flushed (written) from the computer system's disk cache to the local or source ESS, it is lost during a local site failure. The bigger the cache buffers (for higher performance), the more data can be lost and the further the RPO is from the time of actual failure.

Something else to consider is when the attached computer systems do flush their internal disk caches, all of the buffers are not flushed synchronously. This scenario is shown in figure 2. The application begins three transactions (T1, T2, and T3). As records are written to disk 1 and disk 2, they are recorded in the audit log. Until the records are flushed from the computer system cache to the disk module, they cannot be replicated. Even though the T1 and T3 transactions are committed, one record on disk 1 from the T3 transaction and one record on disk 2 from the T1 transaction are not flushed before the system fails and are not replicated. On the other hand, the T2 transaction is not committed, yet one record is flushed to disk and replicated. This illustrates how parts of uncompleted logical transactions may be flushed to disk and replicated, whereas parts of completed transactions may not be flushed and therefore not replicated. Hence, not only can the ESS be several minutes behind the logical state of the database as the application sees it, but the contents of the ESS are "fuzzy" and may not reflect a consistent state of the database either. After a source system failure, the target system may be required to make the database consistent before it can take over processing.

Challenges to RTO

Getting back to RTO, although replication latency is minimized (once the buffers are flushed), there are still issues concerning the remote or backup database end. Before the target system can begin using data on the remote ESS database, recovery must be done using the database manager's inherent power-on "restart" functionality to change the fuzzy database to a consistent database. This is akin to what happens after a system power failure when all in-memory information is gone. Consequently, all applications on the remote system, replicated or not, may need to be brought down for a time for the cutover to take place.

But what if the transaction logs needed for recovery have been off-loaded to tape and are no longer on the system (or are located at the primary site)? Obviously, there is an effect on RTO if you have to bring the logs back online to complete the database recovery and an effect on RPO if the logs are unavailable.

Typical systems, typical solutions

For applications that do not require extremely low RTO or RPO or both, replicated enterprise storage can be a viable disaster protection alternative. It is clearly superior in terms of RTO and RPO to tape backup, and unlike software replication solutions, ESS does not require any processing power on the source or target systems. Replicated storage can also serve as the basis for a hybrid solution that incorporates some software replication capabilities to lower the RTO. But for applications that call for minimal RTO or RPO, or particularly where *both* RTO and RPO must be minimized, a different approach is required.

A complex approach to recovery

As mentioned previously, track-by-track changes, including disk directory updates and changes to noncritical files, are sent across the ESS link between sites. Because the database transaction log is required for recovery purposes, the ESS is replicating more than twice as much data—the changes in the database plus the log of those changes—as is needed. The result is excessive overhead.

If transactions span multiple source computer systems, database recovery needs to be coordinated across all backup systems so that the entire multisystem environment is kept consistent. Do you know if the database 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 system(s) will be affected during recovery operations?

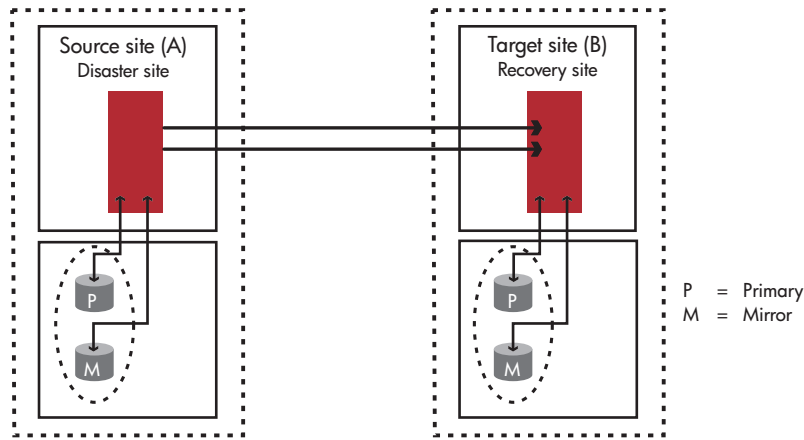
Disaster tolerance versus disaster recovery

Organizations that require continuous application availability need a computing environment that has business continuity built in from the ground up, not plugged in later on.

The NonStop server's transaction management software and remote database replication software cooperate at the lowest levels of the platform. Thus, organizations can literally process straight through catastrophic disasters without interruption of service.

HP NonStop Transaction Management Facility (NonStop TMF) software delivers uncompromising transaction protection and database integrity by tracking before and after images of every transaction. No transaction is ever lost during an application fault or hardware failure in a NonStop server environment. Even the most complex and highly distributed transactions are fully completed or fully backed out, so databases are always consistent. NonStop RDF software extends this protection across multiple systems for business continuity, while ensuring complete database consistency across any number of nodes.

Figure 3. NonStop RDF software understands transactions.



Most important, sharing processing in real time provides a much greater measure of protection against disaster-induced downtime than merely replicating bits. With continuous processing, any node can take over the work of any other node in a matter of seconds. There is no need for system reboots, application restarts, or database reconfigurations. Nor do local workloads on the target node have to be jettisoned.

This is essentially a software-centric approach to disaster tolerance. There is no idle hardware, unless that's your preference. Each source system controls only its portion of the database, and NonStop RDF replicates it to another system that does not modify it until a takeover is declared. Two systems can replicate to each other, multiple systems can protect each other in a ring, or whatever is required by the continuity plan. And because the database is split across the sites, you never have two databases of record at the same time.

Lowering RTO and RPO

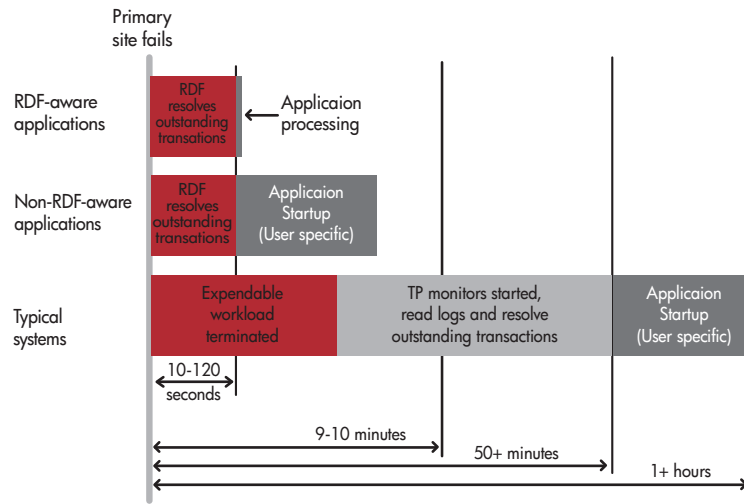
The shared-processing approach that is enabled by NonStop RDF software circumvents the limitations of replicated storage primarily through the use of audit logs generated by NonStop TMF. As the primary database is updated, before and after images of the change are physically written, in a serial manner, to the audit log disk. Because the audit trail is a guaranteed log of all database changes, only the audit trail needs to be replicated between systems. By leveraging the NonStop TMF audit trail for data replication, NonStop RDF software avoids having to rely on data disk cache flushes, so data is not lost; nor is the database rendered inconsistent if a failure prevents flushing of the disk cache.

In addition, NonStop RDF software replicates only those files designated as critical by the administrator. And then it 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 non-distance-limited links can be employed between systems—you're not necessarily limited to dark fiber and distances under 30 miles. Because it is part of the system software, NonStop RDF software 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, and even if it has not yet been flushed from system cache to physical disk.

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. In other words, minimal RTO (see figure 4). Rather, 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; thus, minimal RPO on top of minimal RTO. To achieve a similar RPO, an ESS implementation can be offline for hours as database recovery tools and transaction managers scan logs to make the database transactionally consistent.

In some cases, even minimal RPO is unacceptable to an organization. Instead of using only NonStop RDF software, a hybrid hardware/software solution can be implemented using NonStop RDF software in conjunction with synchronous hardware replication of the NonStop TMF audit trails to bring the RPO to zero while slightly increasing the RTO.

Figure 4. Recovery time objective time line.



Application uptime is no accident

Not all businesses, nor even all NonStop customers, face the exigencies of today's automated stock exchanges where RTO and RPO need to be calibrated to 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 as 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 weighed, both in and of itself, and versus the costs of protecting against it. In many cases, ESS may be a satisfactory solution when applications can be offline for minutes up to hours 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.

Technology should never be selected before a thorough risk analysis and business impact analysis are done on your key business processes. And then, whatever technology or combination of technologies are employed for disaster tolerance, they should fit within a larger continuity planning and process framework for ensuring business survival.

But disaster-tolerant computing is not the future—businesses can't wait that long. It exists today, using out-of-the-box components, and is already providing the ultimate in application continuity for numerous companies that are not willing to "bet the business" on mere data replication.

To learn more, visit www.hp.com/go/nonstopcontinuity.

© 2001, 2002, 2004 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.

5982-3502EN, 01/2004

