

Customer Focused Testing: Ten steps to improved SQL Server 2005 replication

Microsoft SQL 2005 Server provides a built in, application aware, host-based feature called Database Mirroring as an alternative to existing array-based remote replication solutions such as the HP Continuous Access solution. This document provides 10 recommended steps that should be used when planning to deploy remote replication solutions for SQL Server 2005 in an enterprise environment. This document is meant as a companion to the HP whitepaper, "High Availability for SQL Server 2005 Using Array-Based Replication and Host-Based Mirroring". By leveraging these recommendations and best practices, customers can select the optimal replication technology for their environment, can optimize performance, and can accelerate implementation, thereby reducing costs and personnel resources.

1 Understand your environment

Before implementing any replication solution, it is important to understand your recovery goals and the limitations of the available infrastructure. For SQL Replication, the key information includes:

- Recovery Point Objective – This is a measure of how much data loss can be tolerated. For an RPO of zero, array based replication solutions such as HP Continuous Access should be implemented in synchronous mode, whereas synchronous replication using SQL Server database mirroring, being a log-shipping solution, has some potential for data loss. Asynchronous replication can provide advantages but should not be considered if an RPO of zero is required.
- Recovery Time Objective – This is a measure of the amount of time it takes a user to get their backup site running after a complete failure at the primary site. The RTO includes the time to recover, bring the backup database online, and redirect any applications to the backup database server. For the Continuous Access solution, it includes the time to fail over the replicated LUNs to the backup EVA; for the SQL Server database mirroring solution, it includes the time for the Mirror to transition to the role of the Principal.
- Workload (IO activity) – Heavy workloads will put more of a strain on the available resources (servers, storage, and the intersite link), whereas lighter workloads are more forgiving. Keep in mind that for replication, it is only the amount of writes that matters. Databases with low percentage of changes (typically <5%) will provide more flexibility than those with a higher quantity of writes.
- Intersite Link – Both the bandwidth and latency of the intersite link have a direct influence on replication performance. The ISL has to support the ability to transport the writes to the remote site while still providing disk write response times of <20ms. Asynchronous results are relatively unaffected by the ISL latency and bandwidth; however, the link can have a direct impact on recoverability. In a heavy workload configuration, response times will typically increase sharply with greater ISL latencies. In a light workload configuration, more linear and predictable increases are typically observed.



© Copyright 2008 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. Inc. Microsoft and Windows are U.S. registered trademarks of Microsoft Corporation. Oracle is a registered trademark of Oracle Corporation and/or its affiliates. The information in this document is subject to change without notice.

5697-7434

1st edition: March 2008



2 Choose the correct replication technology

Based on your RPO and RTO, a comprehensive review of array-based and host-based replication technologies for SQL Server 2005, and the practical considerations for selecting which method to deploy, should be completed prior to implementation.

- For heavy workloads, Continuous Access produces better transactional performance and is less sensitive and more predictable than SQL Server database mirroring, regardless of the replication mode—synchronous or asynchronous.
- SQL Server database mirroring performance is not affected when small bandwidths (OC-3) are used; however, Continuous Access may not be a viable solution for low bandwidth implementations.
- Checkpointing is key for both Continuous Access replication and SQL Server database mirroring, especially with a heavy workload.
- Both Continuous Access and SQL Server database mirroring require the administrator to monitor the appropriate log mechanism if the ISL is out of service for a period of time.
- Continuous Access will copy all SQL Server components; SQL Server database mirroring will copy only the actual production database. When using the SQL Server database mirroring solution, the administrator must be sure to manually place the system databases, logins, and jobs on the Mirror server.
- Continuous Access is application neutral and is the preferred option when the administrator needs to replicate different applications simultaneously.
- SQL Server database mirroring can use an existing Ethernet infrastructure; Continuous Access requires the deployment of a dedicated connection (either a dark fibre link or an FCIP network).
- SQL Server database mirroring utilizes the server processor about half as much as Continuous Access.
- Better database performance is expected when replicating in asynchronous mode rather than in synchronous mode, unless there are low latencies on the ISL. At latencies below 10 ms, synchronous replication typically achieves a higher transaction rate (TPS) than asynchronous replication. If the network link has 6 ms or less, HP generally recommends using synchronous replication. (Tests indicate that at ISL latencies of 20 ms and beyond synchronous replication solutions may be non-viable.)
- Although there are benefits to using an asynchronous solution, there is the potential for data to build up in either the send queue or the write history log on the primary site, directly impacting the RPO. This data will be lost if there is a failure of the primary storage system and the array cannot be recovered.
- Asynchronous replication can use more of the available ISL bandwidth than synchronous replication, at high workloads and low latencies. This is an expected behavior, as asynchronous replication does not rely on an acknowledgement back from the destination array before sending the next I/O. As the ISL latency increases, the asynchronous bandwidth utilization decreases significantly due to the inability to drive additional I/O across the link. At in ISL latency of 50 ms and above, the bandwidth utilization results are typically the same for both light and heavy workloads due to significant I/O restrictions. For light workloads ISL latency plays little part in bandwidth utilization.
- Higher bandwidths (OC-12) deliver consistently better transactional performance in asynchronous solutions, regardless of latency.
- Asynchronous replication can use more of the available ISL bandwidth than synchronous replication, at high workloads and low latencies. This is an expected behavior, as asynchronous replication does not rely on an acknowledgement back from the destination array before sending the next I/O. As the ISL latency increases, the asynchronous bandwidth utilization decreases significantly due to the inability to drive additional I/O across the link. At in ISL latency of 50 ms and above, the bandwidth utilization results are typically the same for both light and heavy workloads due to significant I/O restrictions. For light workloads ISL latency plays little part in bandwidth utilization.
- Higher bandwidths (OC-12) deliver consistently better transactional performance in asynchronous solutions, regardless of latency.

3 Configure the EVA

The database data disks should be separated from the transaction log disk using two separate disk groups on the EVA. The separation of disk groups is beneficial for two main reasons:

- It guards against data loss due to problems with a specific disk group.
- The disk mechanics are such that disk performance suffers when serial I/O and random I/O patterns are mixed on the same disk. The transaction log data is always written to disk in a serial fashion, and OLTP data usually is written in a random fashion. Thus, the separation optimizes disk performance.

NOTE:

The recommendation for two disk groups applies only when sufficient numbers of spindles are available to the configuration. It is not generally recommended for configurations of 28 or fewer disks.

For workloads where tempdb is not heavily used, it is acceptable to leave in the default location on the shared storage and the associated SQL Server default databases. For application scenarios that require significant use of tempdb, such as queries and table copies, use a database layout where tempdb is placed on its own LUN.

The EVA 8000 array is capable of running typical SQL Server workloads with minimal tuning. Therefore, it should be sized for capacity requirements first, then for performance. Parity-based VRAID5 is entirely appropriate, although VRAID1 remains acceptable if increased availability is imperative. Use the one disk group configuration for simplicity and ease of management.

With Continuous Access, it is recommended that all LUNs associated with a single database in a Data Replication (DR) Group reside on one owning controller. As this may unbalance the EVA controller load, consider assigning the majority of non-replicated LUNs hosted by the EVA to the controller with the least load, or consider multiple database replication streams. Since Continuous Access uses a single controller for each DR group, it is important to verify, prior to implementation, that there is sufficient controller performance bandwidth to handle the workloads.

4 Configure the SQL Server cluster

Implementation of the database servers for the Continuous Access EVA and SQL Server database mirroring solutions each include Microsoft Cluster Server (MSCS) local site for failover capabilities on site A. However, the site B configurations differ slightly:

- Because replication requires an independent SQL Server instance on site B, the Continuous Access solution includes the server on site B in the cluster configuration, effectively creating a geographically dispersed cluster. The synchronous Continuous Access configuration provides the option of including solutions such as HP StorageWorks Cluster Extension Enterprise Virtual Array (CLX) in the configuration to assist in case of a major component failure on site A. This enables Microsoft Cluster Server (MSCS) to control the failover operation of the EVA LUNs associated with the SQL Server in tandem with the other related cluster group resources.
- The SQL Server database mirroring configuration consists of a two-node, single-quorum-based cluster on site A, with one clustered instance of SQL Server mirrored to a stand-alone, dedicated SQL instance on site B.

5 Tune the HBA

In Performance Monitor, there is an average disk queue length performance counter. Use this counter to study the disk queue length of the operating system. On the Continuous Access solution use the counter to determine if there is a significant buildup of requests held at the server. In general, the queue length should be limited to less than two times the number of physical spindles. It is important to test various (32, 64, 128, and 256) HBA I/O buffer settings (also known as execution throttle settings) to determine the effects of these settings on disk write operation performance and on transactional performance.

When examining the impact of HBA queue length in the mirroring solution, it is important to gauge the performance of both the Principal and the Mirror servers. In this configuration, the Mirror server is involved in most of the strenuous disk operations; therefore it was more important to monitor the disk write response time on the Mirror than on the Principal server. For example, definite benefits can be seen when increasing the execution throttle on the Mirror server, potentially resulting in a noticeable difference in the transaction per second processed by the solution.

 **NOTE:**

In the mirroring solution, there are more disk I/Os queued at the HBA on the Mirror server than the Principal server. This is due to the different write patterns between the two server roles. The redo queue on the Mirror server is processed consistently from the transaction log to the data files and this is done sequentially. For this reason, queues are expected to be larger than that of typical random disk I/Os from client updates which normally are received by the Principal server.

 **NOTE:**

You may be tempted to increase the I/O buffer settings on multiple servers on a SAN to aid server performance in general. If too many servers have a large I/O buffer setting on their HBAs accessing a common SAN, this will be at the detriment of the unmodified servers and the general SAN performance as the modified server HBA will take precedence on the SAN and could dominate the resources.

6 Tune the SQL Server Checkpointing

Customize the SQL Server checkpointing process to improve data-writing response time and performance when using either Continuous Access replication or SQL Server database mirroring for both synchronous and asynchronous modes. In synchronous replication, customizing the checkpointing process improves response times for Continuous Access and delivers improved transactional performance for both Continuous Access and SQL Server database mirroring. In asynchronous replication, it improves transactional performance for both Continuous Access and SQL Server database mirroring. This improvement becomes more significant as the workload increases.

Microsoft SQL Server uses a recovery interval to determine when to issue a checkpoint for a database in order to speed the recovery process after a database has had an outage. Effectively, the recovery interval setting is the default method of controlling the checkpointing, but the checkpoint command also can be executed manually or via a script. By configuring the checkpoint duration parameter, it is possible to control the elapsed time of the checkpoint operation. By designating a specific length of time, it is possible to manage the pattern of write operation traffic on the ISL.

The checkpointing operation causes significant write operations over the ISL at the beginning of its cycle, but these operations decrease to near zero at the end of each cycle. The ability to distribute the data more evenly over the complete period of the checkpoint will provide a more even disk response time, resulting in more consistent SQL Server transactional performance.

When you are replicating or mirroring data between two sites over an ISL, it is worthwhile to experiment with the checkpointing, and specifically with the duration times. However, extending the checkpoint intervals and durations even further in order to smooth out the ISL traffic and achieve better performance will increase the database recovery time. These parameters provide a compromise for balancing acceptable disk write response time and acceptable recovery time.

The reason for the performance improvement is different for the Continuous Access solution. Both solutions rely on data being written to disk on both the primary and secondary site storage systems. Since the Continuous Access solution replicates all disk writes (data and transaction log), it generates a lot of write activity; therefore, the ISL becomes the bottleneck. Checkpointing customization helps relieve this ISL bottleneck.

As the SQL Server database mirroring solution simply copies the transaction log updates, the actual amount of data being written to disk and subsequently sent over the ISL is relatively low. However,

the surge nature of checkpointing traffic tends to impact processes within the SQL Server mechanism, especially when the SQL Server database mirroring process is included in the configuration.

7 Allow the Lazy Writer to do some of your work

Using the default checkpointing settings causes dirty pages to be written to disk predominantly by the checkpoint operation. This does not take advantage of the Lazy Writer's ability to do more of this work as a background task. The Lazy Writer can assist in smoothing out the I/O patterns for the traffic over the ISL.

When the default recovery interval is changed to any value above 0, the Lazy Writer instantly writes more of the dirty pages to disk. What is assumed is that the SQL Server understands the impact of altering the recovery interval and will increase the potential for the disk buffer pool to become full; it therefore instructs the Lazy Writer to write more dirty pages to disk to compensate for this. All of this has the effect of removing some of the write I/O load from the checkpointing operation and distributing it more evenly over time. This reduces the I/O contention at checkpointing times.

When using the 0 default recovery interval, the "lazy writes" are fewer than half of those observed when using any other value. It is assumed that SQL Server already has calculated recovery time and anticipated the checkpointing time and duration needed to write all the dirty pages to disk; therefore it does not employ the unpredictable Lazy Writer process.

8 Monitor your performance

The key to SQL Server engine performance management is to monitor the disk write response times for the database data and log disks. Acceptable SQL Server performance can be attained with disk write response times of up to 20 ms. As disk write response times increase, a near linear reduction in transactions per seconds is typically observed. Keep in mind that write performance will be directly affected by the replication technology utilized, the ISL Bandwidth, and the ISL latency.

Processor and network bandwidth utilization should also be examined. Because the network stack is managed by the operating system, processes such as interrupts, memory allocation, and encapsulation/de-encapsulation consume processor cycles. Therefore, as network utilization increases, processor utilization also increases. For SQL Server database mirroring, processor utilization of the Mirror server will be significantly less than that on the Principal server. This is because the Mirror server's only task in this configuration is to accept the mirrored transactions and to process them from the redo queue to the data files.

9 Validate Failover and Recovery

There are a number of failure modes that should be simulated prior to implementation in order to validate that the committed RPO and RTO objectives can be met. These include: server failure, storage failure, ISL failure, and full site failure. Each replication solution will behave differently in each situation and will also be impacted differently in the case of a planned failover versus an unplanned failover.

Continuous Access replication

Recovery time is directly impacted by the length of the checkpoint period.

If you have a lengthy storage or ISL failure, the size of the Write History Log could prevent a full site copy which could significantly impact performance during this period. When planning for this, make sure you understand the reliability of your ISL and be aware when you set the size of the Write History Log.

Although transactional performance of SQL Server may be impacted quite heavily after certain failure conditions, it is possible to manage recovery performance by managing user loads after the failure.

SQL Server database mirroring

After an outage, the solution has to be synchronized in order for it to be operational and able to accept new transactions. It will be in this state until all transactions have been played into the database.

To understand the condition of the SQL Server database mirroring configuration, use the Database Mirroring Monitor. This can be accessed via the SQL Management Studio interface and estimates the time needed to recover the database when synchronizing.

In an unplanned failover, there will be a period of time during which the database is not available. The duration of this unavailability depends on the actual workload entering the database prior to failure. Surprises can be minimized by using the mirror monitor or the documented formula for estimating the synchronization period.

After an ISL outage, monitor the transaction log size; it could fill to capacity as the log send queue grows, and could bring the database to a halt if left unaddressed.

10 Follow the documented HP Best Practices

SQL Server administrators

When replicating SQL Server databases to a contingency site, Continuous Access can process more TPS than SQL Server database mirroring with the same workload. This is the expected outcome because the dedicated hardware technology can off-load the resource from the SQL Server to accomplish this task.

Consider customizing the checkpointing operation when implementing either Continuous Access replication or database mirroring.

Consider using a duration option when implementing the checkpointing to regulate the surge-type nature of the traffic over the underlying architecture.

Failover of a clustered SQL Server instance between nodes will result in downtime for the database due to the failover mechanism and the database recovery process.

The SQL Server recovery time after an outage is affected by the checkpoint process and settings.

In order to satisfy the SQL Server engine when using Continuous Access to replicate databases between two sites, make sure disk write response time is less than 20 ms.

If the transaction delay is too high for acceptable application performance when mirroring a database between two sites in synchronous mode, consider adding more server memory and making it available to the SQL Server.

The time it takes to fail over a database between the Principal and Mirror depends on the database throughput prior to the failover.

When implementing SQL Server mirroring, to ensure that capacity will be available should the Mirror take over the Principal role, do not to install any resource-intensive application on the Mirror server.

Although there are clearly transactional performance benefits from using the asynchronous mode, RPO is compromised if there is a nonrecoverable failure of the Principal database.

In an asynchronous SQL Server database mirroring configuration, if there is a loss of the Principal database and the Mirror database has to be transitioned to become the Principal, the likelihood for some data loss is high if there is any buildup of transactions in the Principal's send queue.

Server administrators

For all solutions, consider adjusting the default HBA I/O buffer length (execution throttle settings) to relieve any potential disk I/O queues that can develop on the SQL Server, especially at checkpointing times.

When implementing a combination of Microsoft Cluster Server and SQL Server database mirroring, HP recommends that tests be conducted on the mirroring timeout value setting, including the application of a real-life workload to the database, to determine the correct number.

Even though it has been shown that the Mirror server processor utilization is significantly less than that of the Principal, be careful if considering placing other applications or workloads on the Mirror server because it will become the Principal if a failover occurs.

In testing, the SQL Server behaved in a predictable fashion and used all but 500 MB of available system memory (23.5 GB). If other programs or applications are installed on the server, consider placing an upper limit on the SQL Server's maximum server memory setting.

In an asynchronous SQL Server database mirroring configuration, after a local cluster node failover, limit client access to the Principal database in order to enable the databases to synchronize once the Principal is back online.

Storage administrators

When implementing Continuous Access, place all LUNs associated with a database in one DR group residing on one owner-controller. Because this might unbalance your EVA controller load, consider assigning the majority of the remaining LUNs hosted by the EVA to the controller with the least load.

Consider defining the write history log size and destination DR group when configuring Continuous Access to guarantee that space will be available if the ISL is dropped for a long period of time.

In a synchronous Continuous Access configuration, after there has been an outage on the ISL and the ISL link has been reinitiated, limit client activity when conducting replication. It is likely that transactional performance will be affected once the replication has been reestablished and is in a state of merging.

In an asynchronous Continuous Access configuration, loss of the source storage system when replicating in asynchronous mode is likely to result in data loss. Manual intervention will be required to bring the database back online on the destination site because automated failover is not supported.

In an asynchronous Continuous Access configuration, a majority node set based cluster can be configured to continue to run in the event of a storage system failure by not setting the quorum resource to be reliant upon shared storage.

Network administrator

SQL Server database mirroring seems to be more sensitive to latency on the ISL than the Continuous Access solution, especially under a heavy workload. If an ISL has latencies above 4 ms, consider using array-based replication instead of database mirroring. Alternatively, experiment with adding more server memory to the Principal.

If using array-based, block-level replication for a heavy SQL Server workload, an OC-3 bandwidth may not be sufficient when in synchronous mode.

From a performance standpoint, there seems to be no difference whether the SQL Server database mirroring solution uses an OC-3 or an OC-12 bandwidth.

If an ISL outage occurs, closely monitor the resynchronization mechanisms of all the solutions because this can have a significant impact on both the performance and the resilience of your database environment after the link is reestablished.

If the ISL latency is below 6 ms, HP recommends synchronous replication for Continuous Access because asynchronous mode may show some performance degradation due to the double caching mechanism being exposed at low latencies.

Continuous access may use more ISL bandwidth in asynchronous mode than in synchronous mode, especially in low-latency scenarios.

In a Continuous Access solution environment, after an ISL outage, the replication relationship enters a merging state. The impact on SQL Server transactional performance is less in asynchronous mode than in synchronous mode during the merging period.

If there is an ISL outage and latency on the link is high (>15 ms), monitor the write history log growth rate both during and after the outage.

NOTE:

A full site copy will occur if the write history log fills, so management of the incoming requests may be necessary until the write history log is depleted.

After an ISL outage, the SQL Server database mirroring solution may need to have client access restricted for a period of time to enable the solution to catch up and become synchronized; under some circumstances, with low ISL latency and a light workload, the solution will be capable of catching up while still allowing client access.

We value your feedback

In order to develop technical materials that address your information needs, we need your feedback. We appreciate your time and value your opinion. The following link will take you to a short 8-question survey regarding the quality of this paper: <http://hpwebgen.com/Questions.aspx?id=12046&pass=41514>

For more information, see Customer Focused Testing: <http://www.hp.com/go/hpcf>