

Microsoft® SQL Server “Always On” Technologies

Hitachi Data Systems Contributes
“Always On” Storage Solutions

A Partner Solutions White Paper

By Rick Andersen and Simon Pengelly

December 2006

Microsoft®
GOLD CERTIFIED
Partner

Executive Summary

Application availability is paramount to today's businesses. To support mission-critical applications and ensure their availability, Microsoft is offering the SQL Server Always On Technologies program. The program involves partners to provide organizations with a full range of options, such as online operations, failover clustering, database snapshots, database mirroring, log shipping, and peer-to-peer replication technologies—to help them achieve and maintain appropriate levels of application availability for the SQL Server 2005 Enterprise Edition.

Hitachi Data Systems provides highly available and reliable end-to-end storage solutions that meet SQL Server Always On Technologies partner requirements. This partner solution white paper examines the storage product, features, and configuration settings that meet these technical requirements and ensure Hitachi Data Systems storage solution compliance with the SQL Server SQL Server Always On Storage Solution Review Program.

Contents

Overview	1
Microsoft SQL Server “Always On” Technologies	1
Program Requirements	1
Microsoft Windows Logo Certification	1
Core Windows Application Programming Interface (API) Support.....	2
Stable Media	2
Forced Unit Access (FUA) and Write-through	2
Asynchronous Capabilities.....	2
Write Ordering.....	2
Torn I/O Protection	2
NTFS support.....	3
Hitachi Data Systems Solutions	3
Operating Microsoft Windows Logo Certification	3
Core Windows API Support	4
Stable Media	4
FUA and Write-through	4
Asynchronous Capabilities.....	5
Write Ordering.....	5
Replicated Solutions and Write Ordering	5
Torn I/O Protection	6
NTFS Support	6
Summary	6
References	7

Microsoft® SQL Server “Always On” Technologies

Hitachi Data Systems Contributes “Always On” Storage Solutions

A Partner Solution White Paper

By Rick Andersen and Simon Pengelly

Overview

The purpose of this partner solution white paper is to address the requirements outlined in the SQL Server Storage Solution Review Program, which is specific to the SQL Server Always On program, using Hitachi storage systems. This paper describes how Hitachi storage complies with this Microsoft program’s expectations. For more information about Hitachi storage systems for SQL Server, please refer to the reference section at the end of this document.

Microsoft SQL Server “Always On” Technologies

The Microsoft SQL Server Storage Solution Review Program is a specific SQL Server program that enables storage solution providers to highlight those storage solutions and configurations, via the SQL Server “Always On” labeling, that they have successfully reviewed against core functional Microsoft SQL Server storage requirements. The core requirements defined herein must be met for reliable, highly available SQL Server storage systems.

The program does not define or review the performance characteristics of storage solutions. Please note that the “Always On” Storage Solution Review Program is not a Microsoft certification, qualification, or logo program. Microsoft makes no warranties or representations with regard to third-party storage solutions, including without limitation regarding the supportability of such third-party storage solutions.

Program Requirements

The “Always On” Program details the core requirement for the SQL Server Storage Solution Review Program. Storage system capabilities are divided into two categories: required or recommended. The required capabilities are detailed below with a description of the requirement.

Microsoft Windows Logo Certification

Microsoft Windows logo certification helps ensure the safety of Microsoft SQL Server data by testing various aspects of the products. To be compliant with the SQL Server Storage Solution Review Program for “Always On,” all storage solutions must pass the Certified for Windows logo testing.

Core Windows Application Programming Interface (API) Support

SQL Server utilizes several APIs to enable secure data storage. A storage solution must ensure that a system supports specific API properties throughout the various layers and implementations of the I/O solution.

Stable Media

SQL Server relies on the Write-Ahead Logging (WAL) protocol to maintain the Atomicity, Consistency, Isolation, and Durability (ACID) properties of the database and to ensure data integrity. WAL relies on stable media capabilities. A solution must comply with this stable media intention.

Forced Unit Access (FUA) and Write-through

To support WAL, SQL Server uses `FILE_FLAG_WRITE_THROUGH` and `FlushFileBuffers` to open files. Both of these options must be supported by storage solutions. All components in a solution must honor the write-to-stable media intent. This includes, but is not limited to, caching components. It is not enough to only honor WAL for SQL Server log files. Data files and backup streams also depend on WAL behavior. Some storage products include battery-backed caching mechanisms as part of the write-through stable media support. If these caching mechanisms are present in the solution, the SQL Server “Always On” solution white paper should document the practical limits of the write-through stable media protection for a production environment.

Asynchronous Capabilities

SQL Server performs most of its I/O using asynchronous capabilities. If a request specifies asynchronous operation, no API call should cause a synchronous condition. Synchronous I/O can cause unexpected scheduler and concurrency issues. Therefore, a SQL Server “Always On” solution must provide asynchronous I/O capabilities.

Write Ordering

A tenet of the WAL protocol is write ordering or order preservation. An SQL Server “Always On” solution must provide write-ordering capabilities. The write-ordering requirement applies to both local and remote I/O destinations. If a database is split among physical paths, all the paths must honor the ordering across all files related to database. To satisfy this ordering requirement, products sometimes use a user-defined “consistency group” that encapsulates all the database files.

Torn I/O Protection

An SQL Server “Always On” solution must provide sector alignment and sizing in a way that prevents torn I/O, including splitting I/Os across various I/O entities in the I/O path. Additionally, a solution must accurately report sector size to Windows I/O APIs. Accurately reporting sector size helps prevent sector size mismatches and avoid torn writes. For example, a drive that does 4KB writes reports 512 bytes while the drive performs a read/write of the 4KB sectors. This inaccuracy in reporting sector size can create a condition in which data is lost and exposed as a torn write. An SQL Server “Always On” solution

must document configurations in such a way that uses sector sizes from the sector size list that is supported by Microsoft SQL Server: 512 bytes, 1024 bytes, 2048 bytes, and 4096 bytes.

To indicate when a torn-write situation occurs, it is recommended that the solution log appropriate warning events.

NTFS support

NTFS capabilities must be supported. This includes but is not limited to the following:

- :: Sparse Files
- :: File Streams
- :: Encryption
- :: Compression
- :: All Security Properties

Sparse files must be supported on NTFS-based file systems. Microsoft SQL Server 2005 uses sparse files in support of DBCC CHECK* commands and snapshot databases.

Common copy and compression utilities may not honor sparse file metadata but instead may copy all bytes, ignoring the sparse allocations and requiring full storage space. Storage solution providers may choose to provide utilities to copy or move sparse files without destroying the sparse file intent.

Hitachi Data Systems Solutions

Operating Microsoft Windows Logo Certification

The following Hitachi storage systems are qualified/certified under the Windows Logo Certification Program for both iSCSI and Fibre Channel protocols:

- :: Hitachi TagmaStore® Universal Storage Platform models USP100, USP600, USP1100
- :: Hitachi TagmaStore Network Storage Controller model NSC55
- :: Hitachi TagmaStore Adaptable Modular Storage models AMS200, AMS500, AMS1000
- :: Hitachi TagmaStore Workgroup Modular Storage model WMS100
- :: Hitachi Thunder 9500™ V Series modular storage systems
- :: Hitachi Lightning 9900™ V Series enterprise storage systems
- :: Hitachi Lightning 9960™ and Hitachi Lightning 9910™ storage systems
- :: Hitachi Thunder 9200™ modular storage systems

For more information, please visit the Window Server catalog at <http://www.windowservercatalog.com/results.aspx?text=hitachi&bCatID=1282&OR=5>

Core Windows API Support

Hitachi storage supports all native Windows file system calls when accessed via block-based iSCSI and Fibre Channel protocol storage. This includes full support for overlapped I/O, sparse files, and direct and buffered or delayed I/O calls.

Stable Media

The Hitachi lines of storage comply and adhere fully to the SQL Server WAL protocols and meet the ACID requirements as defined by Microsoft.

All writes to Hitachi storage systems are committed to battery backed nonvolatile mirrored cache and then committed or hardened to RAID protected volumes within the storage system.

Hitachi storage systems are protected from data loss through the utilization of redundant power supplies and battery backup of both cache, and, in the case of the Universal Storage Platform, battery backup for the disk drives, to allow for the de-staging of data in cache to the drives.

When a Hitachi storage system is restarted after a power failure, all writes held within the battery-backed cache will be de-staged to disk prior to the disk volumes coming online.

FUA and Write-through

All Hitachi storage systems comply with the FUA and Write-through requirements.

Hitachi storage systems ensure that when data is written directly from SQL Server 2005 in a non-buffered fashion (when files are opened with the FILE_FLAG_WRITETHROUGH flag), it is written to stable media immediately. In contrast, when data is buffered on the host and then subsequently flushed to disk in a batch operation, it will always be written to stable media when the data flush (FlushFileBuffers) call takes place. In the event of a failure in the storage system during the flush process, the SQL Server 2005 host will be notified of a failure to write backup data to disk.

The architecture of all Hitachi storage ensures that all write data within the battery-backed cache subsystem is preserved. In order to accomplish this, the following functions are implemented:

Mirrored Cache for Writes

With the Hitachi storage system the cache is mirrored or duplexed to provide availability of write data in the unlikely event of a cache board failure. This implementation protects against a failure since all updates are duplicated within the cache subsystem.

Battery Backup

All Hitachi storage systems have battery backup designed to ensure that all cache write data is preserved and that all I/O is properly de-staged to the disk storage system. Hitachi storage systems provide a full eight-hour backup of the storage system resources.

The batteries are monitored by the automated Hi-Track® “call home” service/remote maintenance tool to provide 24/7 notifications to IT service personnel in the event of a battery discharge.

Asynchronous Capabilities

All asynchronous I/O from either a host or application will be honored in the order in which they were issued. Hitachi storage systems will never modify an asynchronous I/O to a synchronous or serial I/O.

Write Ordering

In all cases writes issued to the Hitachi storage systems will be written in the order that they are issued from the host.

Replicated Solutions and Write Ordering

For those SQL solutions requiring replication of data between geographically separated sites and storage systems, Hitachi Data Systems provides Hitachi Remote Replication software, including the Hitachi TrueCopy™ software modules that serve the various Hitachi storage models, as a business continuity solution. TrueCopy software is fully compliant with the SQL Server 2005 “Always On” requirements.

Hitachi TrueCopy software enables a storage-based hardware solution for disaster recovery, which enables fast and accurate system recovery. Once TrueCopy software operations are established, duplicate copies are automatically maintained for backup and recovery purposes. During normal TrueCopy software operations, the primary volumes remain online to all hosts and continue to process both read and write I/O operations. In the event of a disaster or system failure, the secondary copy of data can be rapidly invoked to allow recovery with a very high level of data integrity. Duplicate copies of data are automatically maintained for backup and disaster recovery.

TrueCopy software offers two modes of operations for storage-based replication:

TrueCopy Synchronous software (which serves all Hitachi enterprise and modular systems) provides volume-based real-time data backup and is ideal for high-priority data backup, duplication, and migration tasks. In the event of a disaster or system failure at the primary site, the secondary TrueCopy Synchronous data can be rapidly invoked to allow recovery at the volume level with an extremely high level of data integrity. TrueCopy Synchronous software writes data to both cache in the primary storage system and the secondary storage system prior to signaling the application that the I/O is complete.

TrueCopy Asynchronous software (which serves all Universal Storage Platform and Network Storage Controller models) and **TrueCopy Extended Distance software** (which serves the AMS1000 and AMS500 models) represent unique and outstanding disaster recovery solutions for large amounts of data that span multiple volumes and storage systems across long distance. These TrueCopy modules offer a group-based update sequence consistency solution that enables fast and accurate database recovery, even after a “rolling” disaster, without the need for time-consuming data recovery procedures. TrueCopy volume groups at the remote site can be recovered with full update sequence consistency. Further, asynchronous capabilities of these TrueCopy software modules provide update sequence consistency for user-defined groups of volumes (for example, databases, and associated logs) as well as protection for write-dependent applications in the event of a failure.

Torn I/O Protection

Hitachi storage systems fully support standard NTFS sector sizing and reporting requirements (512 bytes, 1024 bytes, 2048 bytes, and 4096 bytes). Sector sizing is automatically aligned for Windows systems during the logical (LUN) creation process. Use of diskpart.exe to align sectors is not required unless a more optimum size is recommended for the specific application.

NTFS Support

To ensure a stable “Always On” SQL Server environment, storage vendors are required to support the following NTFS features:

All features of sparse files

- :: File streams
- :: Encryption
- :: Compression
- :: All security properties

The Hitachi lines of storage provide full support for all areas described above when the storage systems are accessed via Fibre Channel and iSCSI protocols.

Summary

Hitachi storage systems offer scalable, enterprise-class storage environments for SQL Server databases, meeting or exceeding all the requirements for an “Always On” solution. For further information about the comprehensive hardware and software suite of offerings for SQL Server environments, please visit the Hitachi Data Systems Web site at www.hds.com

References

Microsoft

- :: SQL Server I/O Basics, Chapter 2
- :: SQL Server 2000 I/O Basics

For the latest information about Microsoft SQL Server 2005, see the SQL Server site:
www.microsoft.com/sql/default.mspx

Hitachi Data Systems

- :: Intelligent Virtual Storage Controllers: Hitachi TagmaStore® Universal Storage Platform and Network Storage Controller Architecture Guide
- :: Hitachi Tagmastore® Adaptable Modular Storage and Workgroup Modular Storage Architecture Guide

Knowledgebase Articles

- :: KB230785—SQL Server 7.0, SQL Server 2000, and SQL Server 2005 logging and data storage algorithms extend data reliability
- :: KB917047—Microsoft SQL Server I/O subsystem requirements for tempdb database

 **Hitachi Data Systems Corporation****Corporate Headquarters**

750 Central Expressway
Santa Clara, California 95050-2627
U.S.A.
Phone: 1 408 970 1000
www.hds.com
info@hds.com

Asia Pacific and Americas

750 Central Expressway
Santa Clara, California 95050-2627
U.S.A.
Phone: 1 408 970 1000
info@hds.com

Europe Headquarters

Sefton Park
Stoke Poges
Buckinghamshire SL2 4HD
United Kingdom
Phone: + 44 (0)1753 618000
info.eu@hds.com

Hitachi Data Systems is registered with the U.S. Patent and Trademark Office as a trademark and service mark of Hitachi, Ltd. The Hitachi Data Systems logotype is a trademark and service mark of Hitachi, Ltd.

Tagmstore and Hi-Track are registered trademarks and Application Optimized Storage, TrueCopy, Thunder 9200, Thunder 9500, Lightning 9900, Lightning 9960, and Lightning 9910 are trademarks of Hitachi Data Systems Corporation.

All other product and company names are, or may be, trademarks or service marks of their respective owners.

Hitachi Data Systems has achieved a Microsoft Competency in Advanced Infrastructure Solutions.

Notice: This document is for informational purposes only, and does not set forth any warranty, express or implied, concerning any equipment or service offered or to be offered by Hitachi Data Systems. This document describes some capabilities that are conditioned on a maintenance contract with Hitachi Data Systems being in effect, and that may be configuration-dependent, and features that may not be currently available. Contact your local Hitachi Data Systems sales office for information on feature and product availability.

Hitachi Data Systems sells and licenses its products to certain terms and conditions, including limited warranties. To see a copy of these terms and conditions prior to purchase or license, please go to http://www.hds.com/products_services/support/warranty.html or call your local sales representative to obtain a printed copy. If you purchase or license the product, you are deemed to have accepted these terms and conditions.

©2006, Hitachi Data Systems Corporation. All Rights Reserved.

WHP-235-01 KL December 2006