

# IBM XIV Storage System Gen3 and the Microsoft SQL Server I/O Reliability Partner Program

*Eric B. Johnson*

*IBM Systems and Technology Group ISV Enablement  
June 2013*

**Stay connected with IBM System Storage**

 [Twitter](#) |  [LinkedIn](#) |  [Facebook](#) |  [YouTube](#) |  [RSS feeds](#) | → [Forum](#) | → [Community](#) | → [More communities](#)



## Table of contents

<b>Abstract</b> .....	<b>1</b>
Disclaimer .....	1
<b>Introduction</b> .....	<b>1</b>
<b>IBM XIV Storage System Gen3 compliance</b> .....	<b>2</b>
Core 1.00: Windows logo certification .....	2
Core 1.01: Windows application programming interface (API) support.....	2
Core 1.02: Stable media .....	2
Core 1.03: Forced unit access (FUA) and write-through .....	2
Core 1.03.01: Multi-level data protection .....	3
Core 1.04: Asynchronous capabilities .....	3
Core 1.05: Write ordering .....	3
Core 1.06: Torn I/O protection .....	4
Core 1.07: New Technology File System (NTFS) support .....	4
Core 1.08: Testing .....	4
Advanced 2.01: Write ordering (remote storage solutions) .....	4
Advanced 2.01.01: synchronous data replication .....	5
Advanced 2.01.02: Asynchronous data replication .....	5
Advanced 2.02: Transactional sector or block rewrites .....	6
Advanced 2.03: Virtual backup device interface (VDI) .....	6
Advanced 2.04: Clustering.....	7
Advanced 2.05: File streams .....	7
Advanced 2.06: Protection.....	7
Advanced 2.07: Hardware virtualization .....	7
<b>Conclusion</b> .....	<b>8</b>
<b>Resources</b> .....	<b>9</b>
<b>Trademarks and special notices</b> .....	<b>11</b>



## Abstract

---

*This white paper describes the IBM XIV Storage System Gen3 membership in the Microsoft SQL Server I/O Reliability Partner Program.*

*Further details about the Microsoft SQL Server I/O Reliability Partner Program are located at the following website:*

[www.microsoft.com/en-us/sqlserver/solutions-technologies/mission-critical-operations/io-reliability-program.aspx](http://www.microsoft.com/en-us/sqlserver/solutions-technologies/mission-critical-operations/io-reliability-program.aspx)

## Disclaimer

This document has been produced independently of Microsoft® Corporation. Microsoft Corporation expressly disclaims responsibility for, and makes no warranty, express or implied, with respect to the accuracy of the contents of this document.

Also, refer to the IBM disclaimer in the “Trademarks and special notices” section located at the end of the paper.

## Introduction

---

The Microsoft SQL Server I/O Reliability Partner Program, formerly known as the Microsoft SQL Server Always-On Storage Solution Review Program, defines multi-level storage subsystem I/O criteria that help to ensure Microsoft SQL Server optimal database compatibility, integrity, and high availability. The IBM® XIV® Storage System Gen3 complies with the program guidelines and this white paper reveals the Microsoft partner requirements for storage solution participation.



## IBM XIV Storage System Gen3 compliance

---

The following sections list the IBM XIV Storage System compliance details according to the Microsoft SQL Server I/O Reliability Partner Program. Further details about the Microsoft SQL Server I/O Reliability Partner Program specifications can be found in the “Resources” section of this paper.

The following IBM XIV Storage System meets the Microsoft SQL Server I/O Reliability Partner Program requirements:

- IBM XIV Storage System Gen3 11.x or later

### Core 1.00: Windows logo certification

As indicated by the Windows Server Catalog website (at [www.windowsservercatalog.com/results.aspx?text=ibm+xiv&bCatID=1282&avc=10&ava=0&OR=5&chtext=&csstext=&csstext=&chbtext=](http://www.windowsservercatalog.com/results.aspx?text=ibm+xiv&bCatID=1282&avc=10&ava=0&OR=5&chtext=&csstext=&csstext=&chbtext=)), the IBM XIV Storage Systems are Microsoft Windows® logo certified. The Windows logo certification testing matrix includes IBM XIV storage multiple path support using the native Microsoft Windows Multipath I/O (MPIO) feature which is enabled through the IBM XIV Host Attachment Kit for Windows. Additionally, there are numerous Microsoft certified cluster solutions with various IBM System x® servers and IBM XIV Storage Systems. Windows logo certification helps to ensure that partner solutions conform to Microsoft's highest standards for operating system fundamentals and platform compatibility.

### Core 1.01: Windows application programming interface (API) support

The IBM XIV Storage System Gen3 fully supports the core Windows APIs. This enables secure data caching and storage for Microsoft SQL Server throughout the various solution stack layers including: server, fabric, and storage.

### Core 1.02: Stable media

The IBM XIV Storage Systems contain stable media that honor the Write-Ahead Logging (WAL) protocol which Microsoft SQL Server relies upon in order to provide atomicity, consistency, isolation, and durability (ACID) database properties as described in chapter 2 of the SQL Server I/O Basics documentation at <http://technet.microsoft.com/en-us/library/cc917726.aspx>. The latter helps to maintain data integrity in the event of system failures or unintentional restarts, including power outages. In addition to stable media, these storage systems provide volume consistency group functionality to further strengthen ACID database properties.

### Core 1.03: Forced unit access (FUA) and write-through

The IBM XIV Storage Systems fully support the FUA and write-through requirements using a virtual storage design built upon a unique grid or modular architecture and data protection foundation.



### Core 1.03.01: Multi-level data protection

The IBM XIV virtual storage design includes multi-level data protection starting with simple host writes. To expound, a host write is temporarily mirrored between the cache of different modules before the XIV Storage System acknowledges, the write is received and the data is destaged to the member disks of each individual module. Consequently, the data is always protected against possible module failures even before the data is destaged using 1 MB XIV partitions.

IBM XIV partitions are the fundamental building blocks used to create volumes and incorporate primary and secondary mirrored data copies mapped to separate disks. Using sophisticated data distribution algorithms that preserve data redundancy and performance equilibrium, XIV partitions are automatically and dynamically distributed across all disks (up to 180). As the distribution algorithms pseudo randomly distribute the data across all of the spindles, there is no logical unit number (LUN) locality of reference to a particular disk or subset of disks, essentially eliminating hot-spots regardless of workload changes over time. As a result, the IBM XIV Storage System self tunes in response to application I/O patterns and configuration and capacity changes, all while maintaining maximum data protection.

Data protection is also reinforced using IBM XIV predetermined data distribution algorithms, which help ensure fast recovery from major and minor faults using prefailure detection and proactive corrective healing. In the event of module or disk failures, global spares striped across all disks quickly redistribute data back to a fully redundant state. During such events, the performance impact is notably minimized and further enhanced by the XIV core, enterprise-class data protection mechanisms.

IBM XIV multi-level data protection employs active/active N+1 redundancy of all data modules, disks, interconnect switches, and battery backup uninterruptible power supply (UPS) units. The XIV also contains an automatic transfer switch (ATS) for external power supply redundancy. A built-in UPS complex consisting of three UPS units protects all disks, cache, and electronics with redundant power supplies and fans, which further promotes Microsoft SQL database availability and reliability.

### Core 1.04: Asynchronous capabilities

The IBM XIV Storage Systems support the necessary SQL Server asynchronous I/O operations. Also, to meet this program requirement, the storage systems do not attempt to convert asynchronous into synchronous I/O.

### Core 1.05: Write ordering

To support the write-ahead logging protocol and prevent database corruption with local or remote mirror solutions, the IBM XIV Storage Systems implement write ordering. With regards to remote mirror solutions, this ensures that the primary site master mirror volume write order is identical to the secondary site slave mirror volume for both synchronous and asynchronous storage environments.



## Core 1.06: Torn I/O protection

In order to help prevent torn I/O and improve host performance with Windows Server 2003 and earlier, a partition offset of 1024 KB is recommended for IBM XIV Storage Systems, which equals the size of an XIV partition as previously referenced. With the alignment changes in Windows 2008 or later versions with a default offset of 1024 KB, partition offsets are no longer required for XIV LUNs. The IBM XIV Storage Systems ensure SQL Server all-or-none transaction atomicity requirements are enforced (which include torn I/O protection). Thus, only full and intact I/O operations are written to disk.

## Core 1.07: New Technology File System (NTFS) support

The IBM XIV Storage Systems fully support all Windows NTFS capabilities including (but not limited to) sparse files, file streams, encryption, compression, and all security properties.

## Core 1.08: Testing

The SQLIOSim utility was used to perform 24-hour data durability and reliability stress tests. The following storage configuration was used:

IBM XIV Storage System Gen 3 with 11.2.0.a code

- Fifteen 512 GB solid-state drives (SSDs),– global read-only cache
- One hundred and eighty 2TB serial-attached SCSI (SAS) drives

All required testing successfully passed.

## Advanced 2.01: Write ordering (remote storage solutions)

The IBM XIV Storage Systems fully adhere to the Microsoft SQL Server I/O Reliability Partner Program write ordering requirement for remote storage using either synchronous or asynchronous data replication for Fibre Channel (FC) or iSCSI protocols. To meet program write ordering requirements, consider using IBM XIV consistency groups that are supported for both types of data replication.

IBM XIV consistency groups allow administrators to take simultaneous or synchronized snapshots of multiple volumes concurrently used by SQL. However, each IBM XIV volume can only belong to a single consistency group and there are a few different SQL consistency group snapshot strategies to consider based on backup and recovery preferences. For full database recoveries, place data and log volumes in the same consistency group. For point-in-time recoveries, create two separate consistency groups, one for logs and one for data. This ensures that the logs do not get overwritten, thus permitting point-in-time transaction log restores. Also, IBM XIV consistency group snapshots can be used in combination with preferred or alternate Microsoft SQL Server data protection practices.

### Advanced 2.01.01: synchronous data replication

IBM XIV synchronous mirroring, as depicted in Figure 1, is designed for metropolitan area networks that span shorter distances (up to 200 km) to minimize the I/O delays inherent to geographically dispersed sites. Typically, a host writes volume data to the cache of a local IBM XIV interface module and it redundantly duplicates the data to the cache of another local IBM XIV module. After the local cache operation completes, synchronous data replication copies the local write to the cache of a remote IBM XIV interface module. Once again, the remote IBM XIV interface module redundantly writes to the cache of another remote XIV module. The entire data replication (serial) caching process involves four IBM XIV modules: two local and two remote. The synchronization is considered complete after redundantly writing to the IBM XIV cache at both sites and a write acknowledgement is sent to the host independent of destaging data to the disks. This significantly expedites the synchronization process as writing to cache is much faster and more efficient than writing to disk before returning the write acknowledgement.

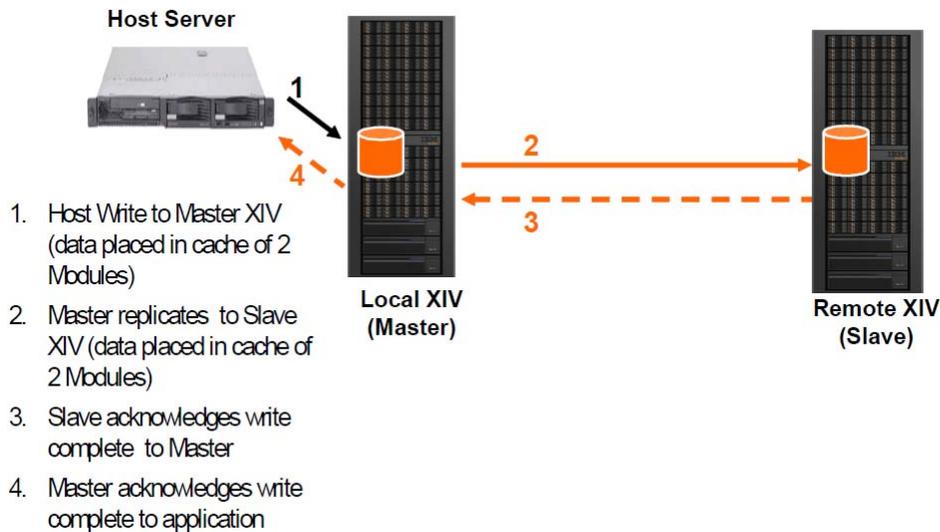


Figure 1: IBM XIV synchronous mirroring

### Advanced 2.01.02: Asynchronous data replication

IBM XIV asynchronous mirroring, as depicted in Figure 2, is designed for wide area networks (WANs) or similar solutions that span long distances (greater than 200 km) and are prone to greater I/O delays. The synchronization is considered complete after writing to the local IBM XIV modules' cache. Rather than continuous data synchronization between local and remote sites, IBM XIV asynchronous data replication is periodically synchronized automatically using user-defined intervals or manually using synchronization (sync) job triggers. Asynchronous data replication requires sync jobs that use IBM XIV snapshots to ensure consistent mirrored volume data between sites. The sync job only updates changes to the master mirror volumes that occur after the creation of the most recent and last replicated snapshots that were successfully synchronized.

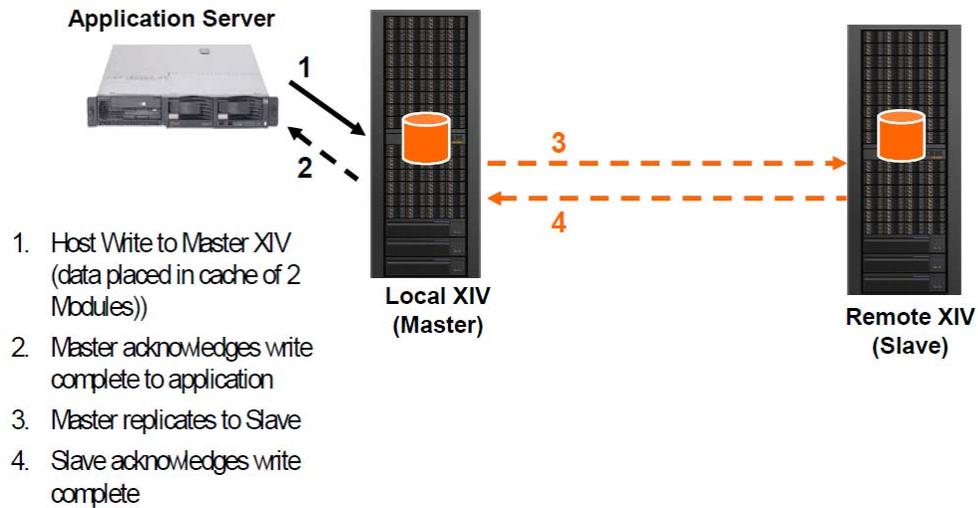


Figure 2: IBM XIV asynchronous mirroring

## Advanced 2.02: Transactional sector or block rewrites

The IBM XIV Storage Systems comply with the transactional sector or block rewrite I/O reliability requirement of the Microsoft SQL Server I/O Reliability Program.

## Advanced 2.03: Virtual backup device interface (VDI)

The IBM XIV Storage Systems fully support the SQL Server virtual backup device interface specification requirements of the Microsoft SQL Server I/O Reliability Partner Program. IBM also provides backup software that adheres to the VDI specification and integrates seamlessly with the IBM XIV and Microsoft SQL Server. Not only do IBM Tivoli® Storage Manager for Databases and IBM Tivoli Storage FlashCopy® Manager for Windows support the Microsoft VDI specification, but they also support the Microsoft Volume Shadow Copy Service (VSS) framework. Both IBM backup products contain Microsoft SQL plug-ins that enable rapid, Microsoft SQL-aware VSS backups that use the IBM XIV snapshot technology.

The IBM XIV Storage System snapshot technology uses the Redirect-on-Write (ROW) approach. With IBM XIV ROW snapshots, new host writes are written to different source volume locations while the original source data remains intact. Thus, the original source volume data is not copied during the ROW snapshot process. Instead, ROW snapshots simply alter volume metadata pointers which allow the volume and its snapshot to use the same data source for all portions that have not been modified. Consequently, near-limitless snapshots can be generated almost instantaneously and no further disk space is necessary unless the source volume data is modified.

Closely related to its snapshot technology and used for data protection, the IBM XIV family also provides controller-based volume copy functionality. Even though they are source volume point-in-time copies similar to snapshots, volume copies are completely independent of the source volumes. To clarify, a host can start using a volume copy immediately after the XIV Storage System modifies the source data pointer or metadata. After this, using the source volume data, a background copy duplicates the data to a new area on the disks and the target volume metadata eventually reflects the new data location upon



completion of the copy operation. There is minimal impact to the storage system and hosts maintain volume read/write access throughout the entire background copy process using ROW operations. Hence, they are similar to conventional volume copies, yet take advantage of the IBM XIV sophisticated metadata functionality.

#### Advanced 2.04: Clustering

The IBM XIV Storage Systems are qualified under the Windows Logo Certification program and meet the hardware compatibility requirements for clustering as required by the Microsoft SQL Server I/O Reliability Partner Program.

#### Advanced 2.05: File streams

The IBM XIV Storage Systems fully adhere to the Microsoft SQL Server I/O Reliability Partner Program file streams requirement.

#### Advanced 2.06: Protection

The IBM XIV Storage Systems fully adhere to the Microsoft SQL Server I/O Reliability Partner Program advanced protection recommendations.

#### Advanced 2.07: Hardware virtualization

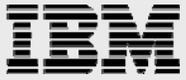
The IBM XIV virtual storage design complies with Windows Server Virtualization Validation Program or SVVP (refer to <http://windowsservercatalog.com/svvp.aspx?svppage=svvp.htm>) as outlined in the SQL Server virtualization support policy at <http://support.microsoft.com/KB/956893>.



## Conclusion

---

This white paper details the Microsoft SQL Server I/O Reliability Partner Program participation of the IBM XIV Storage System Gen3. The program defines numerous storage subsystem standards that consider normal business conditions and unexpected failures and help promote Microsoft SQL Server optimal database compatibility, integrity, and high availability for partner solutions. The program also offers core and advanced storage subsystem compliance guidelines that include the required and recommended categories. As part of the core program requirements, Microsoft partner storage solutions must undergo 24-hour, multifaceted stress tests. In summary, IBM XIV Storage System Gen 3 satisfies and even exceeds many of the Microsoft SQL Server I/O Reliability Partner Program criteria to help customers properly plan for and meet their business critical database objectives.

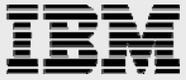


## Resources

---

The following websites provide useful references to supplement the information contained in this paper:

- Microsoft SQL Server IO Reliability Program Overview  
[download.microsoft.com/download/7/8/0/78004501-DEFC-467E-B6F5-560905AE9908/SQL\\_Server\\_IO\\_Reliability\\_Program\\_Overview.pdf](http://download.microsoft.com/download/7/8/0/78004501-DEFC-467E-B6F5-560905AE9908/SQL_Server_IO_Reliability_Program_Overview.pdf)
- Microsoft SQL Server IO Reliability Partner Program Requirements  
[download.microsoft.com/download/F/1/E/F1ECC20C-85EE-4D73-BABA-F87200E8DBC2/SQL\\_Server\\_IO\\_Reliability\\_Program\\_Review\\_Requirements.pdf](http://download.microsoft.com/download/F/1/E/F1ECC20C-85EE-4D73-BABA-F87200E8DBC2/SQL_Server_IO_Reliability_Program_Review_Requirements.pdf)
- SQL Server I/O Basics, Chapter 2  
[technet.microsoft.com/en-us/library/cc917726.aspx](http://technet.microsoft.com/en-us/library/cc917726.aspx)
- Windows Server Catalog  
[windowsservercatalog.com](http://windowsservercatalog.com)
- IBM XIV Storage System  
[ibm.com/systems/storage/disk/xiv/index.html](http://ibm.com/systems/storage/disk/xiv/index.html)
- IBM disk storage systems  
[ibm.com/systems/storage/disk/?lnk=mprST-dsys-usen](http://ibm.com/systems/storage/disk/?lnk=mprST-dsys-usen)
- IBM XIV Storage Tips for Microsoft SQL Server 2008 R2  
[ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP101758](http://ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP101758)
- Microsoft SQL Server 2012 OLTP Workload Benefits Using IBM XIV Storage System Gen3 SSD Cache  
[ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP102103](http://ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP102103)
- IBM XIV Storage System: Copy Services and Migration  
[ibm.com/redbooks/abstracts/sg247759.html](http://ibm.com/redbooks/abstracts/sg247759.html)
- Windows Server 2008 Hyper-V Geographically Dispersed Cluster with IBM XIV Storage  
[ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP101836](http://ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP101836)
- IBM Microsoft Windows Server 2012 Failover Cluster Configuration Program  
[ibm.com/support/entry/portal/docdisplay?brand=5000008&Indocid=MIGR-5092409](http://ibm.com/support/entry/portal/docdisplay?brand=5000008&Indocid=MIGR-5092409)
- IBM Tivoli Storage FlashCopy Manager  
[ibm.com/software/tivoli/products/storage-flashcopy-mgr](http://ibm.com/software/tivoli/products/storage-flashcopy-mgr)
- IBM Tivoli Storage Manager  
[ibm.com/software/products/us/en/tivostormana](http://ibm.com/software/products/us/en/tivostormana)



## Stay connected with IBM System Storage

 [Twitter](#) |  [LinkedIn](#) |  [Facebook](#) |  [YouTube](#) |  [RSS feeds](#) | → [Forum](#) | → [Community](#) | → [More communities](#)



## Trademarks and special notices

---

© Copyright IBM Corporation 2013.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

IBM, the IBM logo, and [ibm.com](http://ibm.com) are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Information is provided "AS IS" without warranty of any kind.

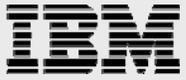
All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local IBM office or IBM authorized reseller for the full text of the specific Statement of Direction.

Some information addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in IBM product announcements. The information is presented here to communicate IBM's current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.



Photographs shown are of engineering prototypes. Changes may be incorporated in production models.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.