

IBM XIV Storage System



Technical Description

IBM XIV Storage System

Storage Reinvented



Performance

The IBM XIV® Storage System offers a new level of high-end disk system performance and reliability. It is a core component of the IBM Information Infrastructure which helps clients address their needs for availability, security, compliance and retention of information. The XIV system provides consistency under all conditions, immunity to hotspots, and consistent performance in the face of hardware failure. Several revolutionary practices contribute to this achievement:

Perfect Load Balancing

Each logical volume in the IBM XIV Storage System is divided into multiple stripes of one megabyte. These stripes are distributed evenly across all disks in the system using a sophisticated pseudo-random distribution mechanism.

This unprecedented approach to load balancing ensures:

- The equal use of all disks and modules, regardless of access patterns. Although applications may access some volumes or certain parts of a volume more frequently than others, the load on the disks and modules remains perfectly balanced.
- The maintaining of load balance regardless of changes in access patterns, such as the adding, deleting, or resizing of volumes, or the adding or removal of hardware.

Cache and Disk into Every Module

Unlike the design of traditional storage systems, the XIV system embeds the read/write cache in the same hardware module as the disks. This unconventional design aspect produces several advantages:

- **Distributed cache.** The cache is distributed evenly across all modules, enabling all the modules to serve host I/O and perform cache-to-disk I/O concurrently. This ensures that the cache never becomes a bottleneck.
- **High cache-to-disk bandwidth.** Aggressive pre-fetching is enabled by the fact that cache-to-disk bandwidth is the internal bandwidth of a module, providing dozens of gigabytes per second per rack.
- **Powerful cache management.** The XIV system reads a large cache slot per disk read, while managing least-recently used statistics in small cache slots. This unique combination is possible due to the system's huge CPU and high cache-to-disk bandwidth.

Huge CPU Power

Each data module is equipped with its own quad-core processor, giving the XIV system dozens of CPU cores. The system uses this vast processing power to execute advanced caching algorithms that support small cache slots, enable powerful snapshotting, and ensure high performance through higher cache hit rates.

High Performance Even During Disk Rebuild

A distributed rebuild mechanism engages all disks in the system during the rebuild process. The end result is a rebuild process that has minimal impact, keeping high performance levels intact.

High performance is enabled by:

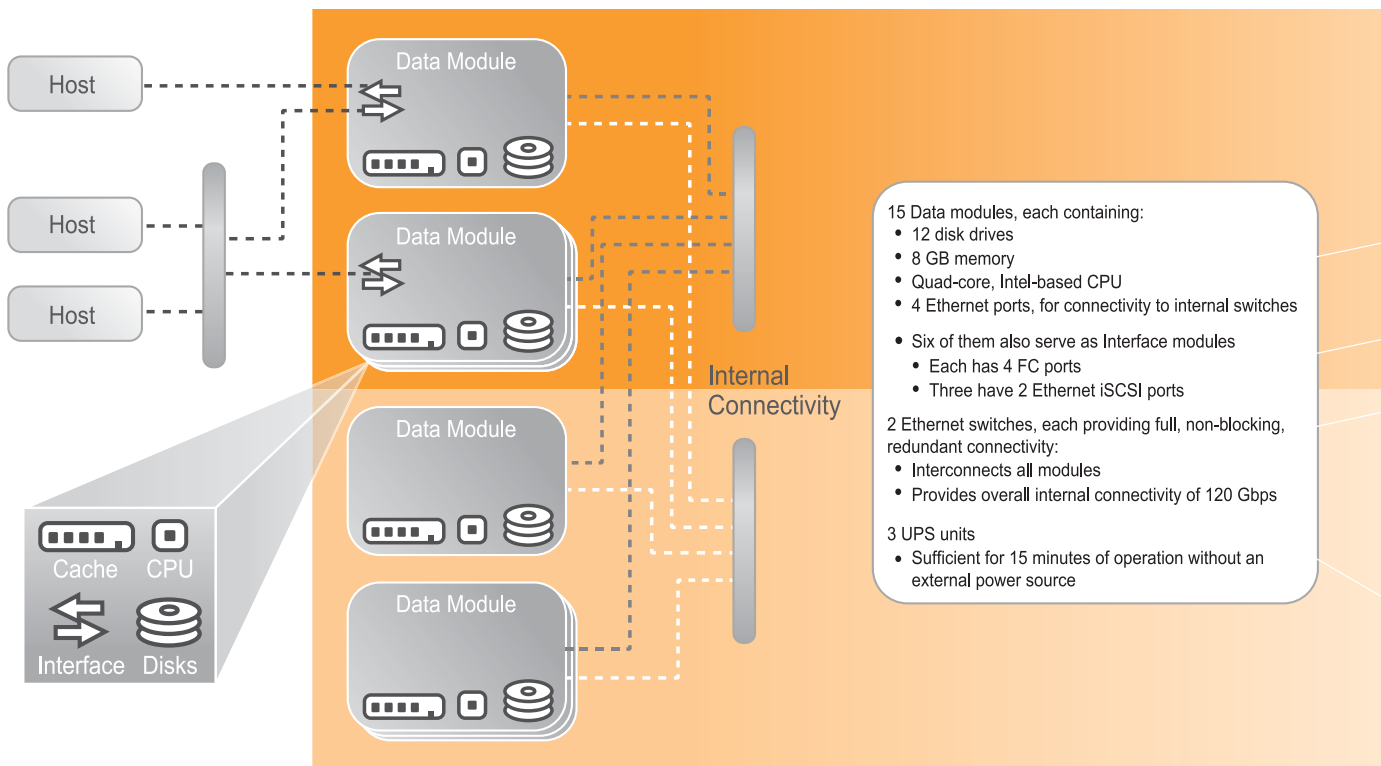
- *Perfect load balancing at all times*
- *Utilization of all system resources*
- *Innovative cache architecture*
- *Massive CPU power*



High-end Open Storage Based on SATA Drives

The XIV system innovatively provides high-end open storage based on SATA drives. It leverages their huge capacity, density, power, and cost advantages, while ensuring high performance and reliability through several means:

- Attains high-end open performance through comprehensive parallelism, with all drives used concurrently.
- Optimizes performance through a sophisticated caching architecture and algorithms
- Ensures a minimal number of disk faults through uniform activity distribution
- Achieves high reliability through a rapid rebuild process that takes just minutes, ensuring virtually zero exposure to multiple disk failure





Storage Reinvented

Reliability

The XIV system provides outstanding reliability, enabled by its unique grid architecture and distributed rebuild mechanism. These qualities immunize the XIV system against failures.

Active-Active N+1 Redundancy

The XIV system can survive single failure without affecting host I/O. Every disk, module, switch or UPS unit is redundant and protected through an active-active N+1 redundancy scheme. Each of these components is hot-swappable—replaceable without system shutdown.



30-Minute Rebuild Time

The system is designed to minimize the risk of impact from disk failure. It does so by involving every disk in the system in the rebuild process, greatly shortening rebuild time. Using its unique distributed rebuild approach, the XIV system thin-slices all data into 1 MB stripes, and distributes and mirrors each stripe on different disks. On a fully utilized XIV system based on 1 TB disk drives, the double disk failure exposure is a mere 30 minutes of rebuild time. The result is a significant reduction in the potential risk of data loss—by orders of magnitude in comparison with other storage systems.

Rebuilding of Only Real Data

Standard storage systems perform disk rebuild on the block level, completely rebuilding the failed disk. The XIV system performs rebuild on only the data that is allocated to volumes and, within volumes, only the data actually written. In actuality, the rebuild time is even much less than 30 minutes for 1 TB disk drives since, in most cases, not all capacity is allocated and that which is allocated is not necessarily in use.

Self-Healing upon Module Failure

The XIV system employs self-healing even after module failure: the system automatically initiates a rebuild process and returns to full redundancy.

IBM XIV Storage System

Scalability

XIV technology scales in every aspect—capacity, cache size, and processing power—and is designed for seamless hardware transitions. It enables a near-infinite number of snapshots, and with virtually no performance overhead, opening the door to new storage paradigms based on snapshotting.

Scalability in Every Dimension

- Thanks to perfect load balancing, the XIV system fully uses all disks on all racks, without administrative intervention
- Cache is embedded in each data module enabling cache size to grow linearly with disk capacity. Cache bandwidth and, as such, performance levels, grow linearly with cache size.
- The number of host interfaces and their bandwidth grow proportionally to system capacity, increasing linearly the number of applications supported and total throughput
- Internal switching capacity stays unchanged regardless of system size, avoiding bottlenecks and ensuring that throughput grows proportionally to capacity
- Processing power is embedded in each module, ensuring that system size has no impact on performance during snapshots, caching, and self-healing

Easy Migration and Upgradeability

- Any XIV system can scale from the minimal to maximal configuration
- The system automates the transfer of data to new hardware, automatically load-balancing all data across new and existing hardware
- The use of standard components enables the system to capitalize on the latest industry technologies
- The XIV grid architecture allows upgrades to leverage—and mix—future hardware technologies, ensuring investment protection
- All hardware upgrade activities can take place without interruptions to production

Thin Provisioning

The XIV system was built to provide thin provisioning. It powerfully allows organizations to use thin provisioning to trim physical capacity buying to an as-needed-only basis:

- Provides the flexibility to define logical volume sizes as larger than physical capacity, while physical capacity need be just larger than the data actually written
- Enables the reclaiming of disk space no longer in use
- Automatically shrinks volumes upon migrating data from a non-XIV system, offering great power and space savings

Total Cost of Ownership

The XIV system offers numerous direct and indirect cost advantages:

- Power, floor space, and cooling expenses are significantly reduced due to the use of SATA drives—without compromising performance or reliability
- Ease of management greatly reduces administrator activities and related costs
- Differential snapshots, elimination of orphaned space, and thin provisioning capabilities enable data centers to use less system capacity to achieve the same goals
- Investment protection is ensured thanks to XIV's grid architecture, which scales from a minimal to maximal configuration and supports the mixing of hardware technologies as they emerge





Ease of Management

Workload-Reducing Virtualization

The XIV system is fully virtualized, freeing the user from the need to plan how logical volumes are spread over physical disks. The system assigns physical resources automatically, ensuring optimal resource utilization. The user sees newly added hardware components as additional free space.

Powerful Management Tools

GUI—A simple and intuitive GUI allows all administrative operations to be carried out quickly and easily, with minimal training and knowledge

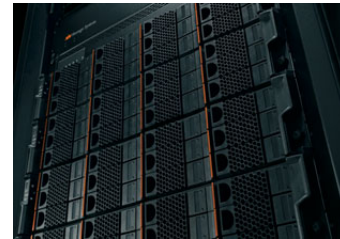
CLI—A powerful command line interface tool enables the writing of complex scripts for high-level system administration and integration with hosts and applications

The XIV system's virtualization and simplicity, together with its powerful GUI management, significantly reduce the time and effort required for managing storage resources.

Single Tier Solution

- High-end open performance, reliability and features
- Mass storage affordability
- Avoids the cost of software and administration overhead required for some tiered storage solutions

IBM XIV Storage System



© Copyright IBM Corporation 2008
IBM Systems and Technology Group
Route 100
Somers, New York 10589
Produced in the United States
August 2008
All Rights Reserved

IBM, the IBM logo, ibm.com, System Storage, XIV, and the XIV logo 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 ibm.com/legal/copytrade.shtml

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

This document could include technical inaccuracies or typographical errors. IBM may not offer the products, services or features discussed in this document in other countries, and the product information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. The information contained in this document is current as of the initial date of publication only and is subject to change without notice. All performance information was determined in a controlled environment. Actual results may vary. Performance information is provided "AS IS" and no warranties or guarantees are expressed or implied by IBM. Information concerning non-IBM products was obtained from the suppliers of their products their published announcements or other publicly available sources. Questions on the capabilities of the non-IBM products should be addressed with the suppliers. IBM does not warrant that the information offered herein will meet your requirements or those of your distributors or customers. IBM provides this information "AS IS" without warranty. IBM disclaims all warranties, express or implied, including the implied warranties of noninfringement, merchantability and fitness for a particular purpose or noninfringement. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

<http://www.xivstorage.com>

TSF03012-USEN-02