IBM XIV® Storage System

Ideal Storage for the New Media Era

White Paper

August 2008

info@xivstorage.com
www.xivstorage.com

Contents

Introduction

The last decade has seen a revolution in the way in which digital information is generated and organized. There has been a clear shift from centralized content creation, such as by commercial creators of content, to a distributed model (also known as Web 2.0), in which every user creates content. This trend, combined with the phenomenal growth in digital imaging storage requirements, has created a huge demand for storage solutions that can handle all the data. The result is a call for a new type of storage architecture — one that is scalable, reliable, and easy to manage.

Unfortunately, the businesses that make their livelihood from new media-type applications, such as web-based mail service providers, digital video, music, and image repositories, and social-networking sites, operate on a completely different revenue model from more traditional industries, such as finance, telecom, and retail. Based mostly on advertising, this new model necessarily imposes strict limits on IT spending.

Today's breakthroughs in disk drive technology deliver solutions to the growing needs: namely, the combination of perpendicular recording and SATA-based disk drives produces the huge capacities and density required. SATA technology is already available on 1 terabyte (TB) drives, and even bigger disks are expected in the future.

What lies at the crux of today's problem is the architecture on which today's storage solutions are based. The layer that integrates disk drives into a single network storage system is inadequate in meeting today's new business demands and unable to leverage the outstanding advances in disk drive technology. The storage solutions available until now to new media and Web 2.0 companies have been either beyond their financial reach or severely limited in capabilities and scalability.

The IBM XIV® Storage System provides a revolutionary solution for these requirements, in that it combines unlimited scalability and unprecedented ease of management with groundbreaking Total Cost of Ownership (TCO).

We investigate here the requirements of the new applications so as to understand why they are different from traditional requirements and how current storage architectures fall short of their needs. We will also explore how the XIV system, with its entirely new storage paradigm, provides a solution for these needs.

New Industry, New Requirements

Web 2.0 and new media applications provide new challenges for the storage world, presenting demands that are very different from those in the more traditional IT, financial, telecom, and cellular domains. These requirements can be summarized as follows:

- **Scalability.** Today, end users create content, much of which is based on digital images (stills and video). Home devices, with their improved ability to generate high quality images and video, further contribute to the huge growth. The result is that the new media industry requires storage solutions to be much more scalable than current solutions.

- **Cost Sensitivity.** Unlike the financial, telecom, and cellular markets, for the Web 2.0 industry, services involving storage account for a substantial portion of business revenues. Hence, the total cost of storage ownership is not just something to be reduced, as for other sectors. Rather, a reasonably-priced storage solution can be the difference between a viable and non-viable business model.
- **Single Architecture.** Web 2.0 is mainly unpredictable. Services considered an obscure need of a niche market can become, overnight, mainstream applications relevant to all users. From a business opportunity point of view, this means that even the smallest application should be supported with a storage solution that can scale easily, with practically no limitations.

The Promise: Huge Disk Drives

As noted, disk drive technology has already answered the storage challenge presented by digital imaging and Web 2.0 applications by developing high-density SATA technology. This breakthrough is supported by two technological developments:

- **Perpendicular Recording.** Changes the direction of the magnetic signals, providing a huge gain in capacity and density.
- **Slower Disks.** Rotate at 7200 RPM (instead of 10K and 15K, as with the previous generation), allowing for higher density and consuming less power.

In just two years, these new technologies have led to a four-fold increase in standard disk drive capacity: from 250 GB to one TB, with higher capacities promised soon.

For the home user connecting a single disk to a single computer, this level of capacity is adequate. Today, each of us can store an almost infinite number of documents and digital images. For business and IT users, these technological advances pose hard questions: How can today's systems tap into this phenomenal leap in capacity? Is this even possible using disk arrays designed when disk drives were just a few GB?

Not That Simple to Get a Bigger System

A storage system uses a multitude of disk drives and presents them as a logical storage space. A naïve approach would suggest that doubling the disk drive size or doubling the number of disk drives doubles the storage system's capacity. These assumptions might work with regard to a specific range of capacities, but falter at a certain point. Let's look at the limitations:

Bigger Drives, Long Rebuild Times, and Reliability

Disk bandwidth, which has increased only modestly, has not kept pace with the dramatic increase in disk capacity. Consequently, the time required to write the contents of an entire disk has grown enormously. During disk rebuild, the hot spare disk is completely rewritten, causing the process to take several hours and, in some cases, even more than a day.

But the impact of the rebuild process should be measured in terms other than just time. The high stress that disk drives undergo during a rebuild makes double failure a much more likely event than the actual exposure time would suggest.

Another impact of the long rebuild process is its effect on performance. While the rebuild is in effect, much of the resources of the disk drive are dedicated to the rebuild process, significantly impairing the system's performance level. When performing a rebuild in a system using RAID-5 and/or RAID-6 type protections, the need to serve read requests by reading multiple blocks from multiple drives and compute an XOR slows down the system that much more.

The XIV system solves these problems by introducing a revolutionary redundancy architecture, which is based on several principles:

- The data on each disk drive is split into granular stripes, with each stripe mirrored on a different disk drive.
- Each stripe is mirrored on two disk drives, located on different modules.
- Every two disk drives in the system store together some small part of the data (except for any two disk drives on the same module).
- There is no disk drive dedicated for a hot spare. Instead, a certain part of the capacity of each disk drive is reserved for a hot spare.

As a result, rebuild time after a failure of a 1 TB disk drive is 30 minutes or less. Furthermore, since the rebuild process is performed by all remaining disk drives and modules concurrently, there is practically no performance overhead.

Adding Disk Drives Grows Capacity, Not Performance

Every storage system has a specification for a capacity limit, usually defined by the maximal number of disk drives and the capacity of the largest disk drive the system supports. The problem is that when there is a need for more capacity, there is most likely a need for more performance. For example, assume we have a system for an online image archive, supporting 1M users. Several growth scenarios can be examined:

- If people accumulate more images over time, more capacity to store them will be required.
- If people browse the images more frequently, more performance will be required.
- If the size of the average image grows, as in today's reality, more capacity and performance will be required.
- If more users register with our service then, again, more capacity and performance will be required.

This simple example shows how traditional demands for more storage capacity are almost always accompanied by a demand for more performance. Supporting these increased performance needs with traditional storage architectures does not work:

- **In Mid-range System Architectures.** The dual node approach dictates a strict limit on performance levels, and any disk shelf added reduces the performance--per-capacity ratio. This is a severe limitation of all mid-range architectures.
- **In High-end System Architectures.** Higher performance levels can be provided but, as noted earlier, often at prices that are unaffordable for Web 2.0-type companies.

Managing Storage

Another key aspect of storage cost and effectiveness is the management effort required. Systems based on traditional storage architectures are a challenge to manage and optimize, with significant effort spent manually controlling the layout of logical volumes on the physical disk drives. This manual control approach is one of the most acute problems of traditional data centers and a nearly insurmountable business obstacle for new media data centers. The huge capacities used in today's new applications, together with the dynamic changes in space consumption and usage patterns, makes manual tuning and layout management an impossible task.

Another aspect of storage management that contributes to the burden of managing new media applications is the constant need to resize volumes. The dynamic and unpredictable needs of these applications make space planning a budgetary nightmare in which the ideal amount of space is "as much as possible," with simple and non--disruptive storage scaling essential to the company's ability to thrive.

Total Availability

The world of Internet applications has created a strict requirement for availability 24 hours a day, 365 days a year. This world has no maintenance windows of inactivity, not even on weekends, nights, or holidays. An obvious implication for storage is the need for uptime through any hardware failure. A less obvious requirement is the need to maintain high performance levels, even through disk or controller failures. Many of today's architectures actually fail to support applicative needs upon a hardware failure, due to either the overhead of the rebuild process or the switch to write-through mode on a hardware failure.

Traditional Storage: the Bottom Line

The bottom line of the above-mentioned problems is simple: traditional storage options do not provide an adequate answer to the needs of the new Internet and digital media applications. The exposure to double failure, the need for high performance, and the requirement to limit storage, power, and management costs, require a completely new approach to storage.

In fact, two Internet giants, Google and Amazon, have created their own storage solutions, with in-house development teams and using standard components. This is sad testimony to the failure of the storage industry to meet their needs. Neither felt the need to develop its own operating system, database, or networking system, but was compelled to do so to address their storage needs. Needless to say, for just about anyone else, a homegrown storage solution is not a practical option.

IBM XIV Storage: New Paradigm for a New Era

Unlike other storage systems, most of which originated in the late 1980s and early 1990s, the XIV system was designed in the 21st century, using the latest technologies available and is optimized for handling huge capacities by means of large disk drives.

The XIV system brings exceptional benefits to new media and Web 2.0 applications:

- **Single Architecture for All Needs.** Even the smallest XIV system can be expanded to the largest configuration. This means that any uncertainty regarding future needs has no financial impact on the present. It gives Web 2.0 companies a safe, economically viable choice in the early phases of each application, with the ability to scale to any size as demand grows, even to mass usage levels.
- **Consistently High Performance, and Without Management Effort.** The XIV system provides unmatched levels of performance, even through any hardware failure. In addition, its performance levels are available automatically, without any management tuning or planning. This means that Web 2.0 companies are assured of getting full support for an application's needs, without the constant effort involved in tuning and monitoring, and without vulnerability to failures.
- **Thin Provisioning.** Setting logical volume size as larger than the system's physical size, and adding physical capacity only when really needed, reduce system costs while making management easier. This makes the scaling of applications from small capacities to a huge capacity even simpler — applications are simply not aware that the physical capacity installed is smaller than the logical capacity viewed by hosts.
- **Reliability.** The XIV system's unique redundancy scheme enables a rebuild of a 1 TB drive within 30 minutes, with virtually no performance hit. The scheme reduces the probability of double failure to much below that of any other architecture, while maintaining performance levels intact. It provides any new media, Web 2.0, or digital archive application with a high level of reliability and availability.
- **Ease of Management and Built-in Virtualization.** The XIV system completely frees the storage administrator from any planning of logical volumes over physical components, such as disk drives or modules. For Web 2.0 applications, freedom from planning means a major reduction in the management effort, together with better responsiveness of the storage service.

Summary

Web 2.0 and digital archive applications are finding today's storage offering inadequate for their needs. The combined requirement for scalability, performance, and ease of management, within reasonable costs, has been an unanswered challenge. Traditional storage system architectures were designed around low-capacity FC drives and do not scale well with the huge capacities offered by SATA drives.

In fact, two of the leading Internet companies, Google and Amazon, have chosen to develop their own storage architecture solutions in-house, basing them on standard components.

The XIV system employs a revolutionary architecture that is designed to deliver performance, scalability, and ease of management while harnessing the high capacity and cost benefits of SATA drives. As such, the XIV system provides everything a Web 2.0 application requires, while keeping the Total Cost of Ownership down.