



I D C T E C H N O L O G Y S P O T L I G H T

Addressing the Storage Challenges in Virtualized Environments

February 2011

Adapted from *Survey Results: Trends in Storage for x86 Virtualized Environments* by Noemi Greyzdorf and Benjamin Woo, IDC #225059

Sponsored by Xitech

The drive to achieve greater resource utilization has pushed many organizations to adopt server and desktop virtualization. Virtualization has the promise of delivering lower operational costs, higher asset utilization, agility of infrastructure, and greater responsiveness to the market. Though virtualization has seen significant adoption, storage is the variable that is holding companies back from benefiting in full from virtualization. The increasing adoption of virtualization and the higher density of virtual machines per physical host are creating new storage challenges. This Technology Spotlight examines the storage issues many organizations face and discusses the role that Xitech plays in this important market.

Market Trends

The need for greater efficiency in the datacenter isn't new. The globalization of markets and the increasing dependence on information systems as a way to communicate with customers, partners, and vendors; process transactions; and maintain inventories demand consistent application performance and availability. At the same time, organizations are buckling under the deluge of new and copied data, which is growing over 50% year over year, and are forced to seek ways to store all this data without drowning in cost and complexity. If IT managers had unlimited budgets and no environment constraints, the task of storing all this data and ensuring application performance wouldn't seem so challenging. Unfortunately, IT has seen much tighter budgets in recent times, driving managers to seek new approaches to delivering IT services that keep their enterprises competitive.

The road to greater efficiency didn't start with virtualization, but virtualization saw an increase in adoption as a result of efficiency initiatives. Server virtualization became the engine behind server consolidation, capex savings, and ease of deploying test and development systems in 2005. By 2008, virtualization entered the production environment, bringing with it greater virtual machine mobility, responsiveness to the needs of the organization, and streamlining of operations related to server provisioning. The adoption of virtualization continues to expand, with more than half of all workloads virtualized by the end of 2010. The expectation is that nearly 70% of all workloads will be virtualized by 2013. At the same time, virtual machine density — the number of virtual machines per physical host — continues to increase, averaging six virtual machines per host in 2009. Virtualization holds the promise of greater efficiency on a number of levels: higher resource utilization, operational efficiency through standardization, faster time to market, higher application uptime, and improved responsiveness to market demands. The future datacenter is fully virtualized with internal or external customers subscribing to IT services as they do already to telephony. This service-oriented datacenter is efficient and cost effective. Virtualization is seen as the heart of such a datacenter. As always, there is a big *but*; before the vision of the future can be realized, organizations must face the challenges created by virtualization in the datacenter, in particular around storage.



Top Storage Challenges

The adoption of virtualization has created new challenges in storage, in particular the trade-off often inadvertently imposed between capacity and performance. A recent IDC survey showed that the top two storage challenges in virtualized environments are application performance and capacity growth. What's driving these challenges?

- The aggregation of multiple virtual machines on a single physical host is expressed as highly randomized reads and writes that put greater demands on the spinning media. These workloads create I/O bottlenecks on the storage system, causing application performance degradation or inconsistency. Load balancing in this environment is complicated by multiple abstraction layers.
- Traditional architectures are not flexible enough and often can't respond quickly enough to the changing needs of the application.
- Provisioning storage is often time consuming and is not aligned with the ability to swiftly provision a virtual machine. The delays caused by storage provisioning may impact an organization's ability to deploy necessary resources to drive revenue and retain customers.
- Data growth and virtual machine sprawl are consuming increasing amounts of capacity. Managers don't have systems to dynamically change allocated storage capacity, reclaim underutilized capacity, or minimize waste.
- The adoption of capacity optimization technologies impacts the overall performance of the storage system. Increasing performance by adding more drives decreases the overall utilization, negatively affecting return on investment.
- Optimizing production environments negatively impacts a manager's ability to protect data using traditional approaches to data protection. Additionally, the abstraction layer that is virtualization makes granular backups and restores more complicated, requiring more performance and often more capacity than would otherwise be necessary.

The drive to greater efficiency has challenged storage managers to find new ways of delivering storage to applications. To achieve the best results, managers must first start evaluating storage solutions based on what they need. If the application demands performance, then they need to understand how much it would cost to achieve a desired performance parameter with each approach under consideration. If the application demands capacity, then they need to understand how much it would cost to acquire and maintain such resources. Organizations often sacrifice capacity utilization in order to achieve performance. Today, there no longer needs to be such an expensive and complex trade-off. Some technologies for organizations to consider are:

- Capacity optimization technologies such as deduplication and compression reduce the data footprint, which increases overall perceived usable capacity. Deduplication and compression may be especially effective with static data that exists in multiple copies. This is common in virtualized environments where the same operating system may be stored multiple times, representing each virtual machine.
- Thin technologies are another way to address the need for more storage capacity. Thin provisioning allows administrators to fool the application into thinking it received more capacity than it actually did. Capacity is allocated as data is being written; the remaining capacity stays in the general pool from which each application draws. Thin replication, thin migration, and thin snapshots are also examples of capacity optimization technologies where only bits that have meaning or blocks that hold actual data are replicated or copied, saving storage capacity as well as network bandwidth.

- Short stroking the drive or deploying more fast-spinning drives has been an acceptable approach to achieving greater performance. The adoption of these techniques often requires the sacrifice of capacity. This has often been the explanation behind low capacity utilization rates such as 30%.
- Dynamic storage tiering is a more recent approach to improving performance without sacrificing capacity. This is achieved by gathering heuristics on multiple data parameters and moving hot blocks to faster media. Dynamic tiering does this bidirectionally; cold data is moved to slower, denser storage media.
- Solid state drives (SSDs) and Flash have gained traction as a way to achieve higher performance of the overall system. A typical solid state drive offers fewer gigabytes but exponentially more IOPS than a hard disk drive (HDD) on a gigabyte per dollar basis; as a result, using SSDs and Flash to accelerate applications may reduce the amount of capacity traditionally wasted on an HDD to achieve the same levels of performance. SSDs and Flash can be deployed either as a tier of storage or as cache. The SSD and Flash technology is significantly more expensive than an HDD per gigabyte and can become less reliable if used improperly. The best results are achieved when SSDs and Flash target the right workload.

Building a storage solution that addresses the demands of a server virtualized environment starts with designing a system that is flexible, responsive, consistent, and reliable with standardized processes for provisioning and allocation of storage. Once the building blocks are in place, it becomes much easier to add features and functionalities such as capacity optimization, management, and serviceability.

Benefits

Applying technologies such as capacity optimization, thin technologies, dynamic tiering, and SSDs and Flash can significantly improve performance, reduce waste, and contain capital expenditures. The use of these technologies, when deployed in an intelligent, automated way, may also reduce operational complexity and enable closer alignment with the trends in server virtualization.

Considering Xitech Hybrid ISE

ISE, Intelligent Storage Element, has been designed by Xitech to consistently and reliably deliver capacity and performance without the trade-off. The system consists of hardware and software that work in concert to extract optimal performance and resilience from the commodity components.

The hardware design addresses rotational vibration, effective cooling, and redundant power. Xitech's tests show up to 98% reduction in vibration solely due to the design of the disk enclosure or datapac. Drives that are shipped to the customer are initially tested to weed out early mortality drives. Each ISE brick comes with a dual controller, two multidrive enclosures (datapacs), two fans, and two batteries that can sustain the system for up to 96 hours.

The software, RAGS, is deployed in an active-active controller. It is responsible for mapping data and ensuring redundancy and consistent performance. Within each datapac, data is laid across all drives for greater resiliency and shorter rebuild times. Additional software manages each drive, logging errors and responding to them before they cause performance problems, data corruption, or data loss. The ability to monitor and manage drives sector by sector allows ISE to apply fixes to problems that would otherwise cause the entire drive to fail. ISE can fail a sector or a whole head, keeping the rest of the platters, heads, and sectors operational. ISE also ensures data integrity through DIF, data integrity field, running end-to-end checksum on every block and every drive. These techniques extend the life of the drives and the overall system, allowing Xitech to offer a five-year warranty on ISE compared with the industry standard of three years.

Hybrid ISE

The second generation of ISE, announced in January 2011, adds a new feature that directly addresses the performance challenges created by virtualization. Hybrid ISE supports two types of media in the datapac: HDD and SSD. SSDs are not a spinning media and don't have the physical limitations of a hard disk drive. A typical SSD can support over 10,000 IOPS compared with a typical 15,000 RPM HDD that supports less than 200 IOPS. The best results per price point are achieved with SSDs when the drives are used to support read-intensive workloads. When an SSD must handle too many writes, the drive can experience a shorter life span and lower performance overhead due to "garbage collection." Understanding these nuances, Xiotech designed Hybrid ISE so that data is always first written to hard disk drives. Once data has been written, the system begins to monitor blocks for usage patterns. As some blocks are accessed more frequently than others, the hottest blocks with the best cost benefit are moved to the SSD. No data is moved off the SSDs until the drives are 100% full. This approach may seem counterintuitive, but it reduces the amount of writes and garbage collection that must occur on the drives. Once SSDs reach their full capacity, blocks that have cooled off are moved back to the hard disk drives while new blocks that have become hot are moved to the SSDs. This storage tiering approach, called Continuous Adaptive Data Placement (CADP), delivers high performance where it is needed most while using media in an optimal way to extend life and ensure resiliency.

The heuristics that are gathered by the system as it monitors the characteristics of various workloads are used not only to manage the placement of data in real time but also to create templates for workloads with common characteristics. This allows the system to gain intelligence that can serve it to prefetch blocks into SSDs before they become hot. The prefetch capability further improves the overall performance of the system.

Hybrid ISE and Virtualization

The main challenges for storage in server virtualized environments are application performance and capacity growth. Hybrid ISE has been designed to address the changing demands of the virtualized workload by leveraging advantages inherent to ISE data placement technology and hardware datapac design, now further enhanced by the use of SSD to increase IOPS and reduce the cost of storage performance. The SSDs do not account for the overall capacity of the system and are used solely for the acceleration of the application. This creates an environment where capacity is added to store data and SSDs are added to enhance performance.

ISE offers performance, reliability, and consistency: ISE and Hybrid ISE can be deployed in three ways:

- Direct attached to a server via Fibre Channel (FC). The server would have access to all the capacity and performance of a single ISE system.
- Multiple ISE systems can be aggregated using a volume manager and attached via FC SAN.
- Xiotech's Block Storage Controller or File Storage Controller can be deployed in front of ISE as a gateway, exporting capacity and performance over standard network protocols along with the data management capabilities that enterprises want.

Leveraging storage virtualization, NAS, or a volume manager to give ISE a sort of personality allows end users to take advantage of efficiencies inherent in ISE and the data management capabilities offered through gateways.

Challenges

Xiotech is facing challenges that can also be seen as opportunities. The issues associated with designing systems with storage media in mind are not well known, understood, or mentioned. Rarely does the market talk about the performance-capacity trade-off made every day by storage administrators. Fewer know the impact of rotational vibration on disk drives over time.

- To gain mind share, Xiotech needs to educate the market so that when the solution is described, the listener understands why the capabilities are relevant.
- The market understands that achieving desired performance can be complicated and what needs to be done varies from use case to use case. Unfortunately, organizations still insist on purchasing capacity. It is critical for the conversation to be about cost of performance as the expense of capacity. This conversation will be made easier by the interest in and adoption of SSD and Flash technologies.
- Enterprises buy storage solutions for a particular use case; the requirements include data management capabilities, which Xiotech offers through its Block Storage Controller and File Storage Controller. These gateways must be competitive in features and functionality with other gateways and integrated solutions.
- Organizations are used to buying systems on a three-year refresh cycle. Xiotech offers a five-year warranty that can reduce capital expenditures. Because it is not the norm, it might be challenging for organizations to compare acquisition costs of ISE and competitive systems. A little education and changing the conversation's focus can help.

Conclusion

In an effort to achieve greater efficiencies in the datacenter, managers sought to consolidate resource and reduce waste. The first steps included the adoption of virtualization. Higher utilization and lower capital expenditures came with greater complexity on the storage side. The trade-off between capacity and performance resulted in wasted resources, higher costs, and less flexibility. Xiotech has designed a system that can serve as a building block of the virtualized datacenter while addressing the main challenges that virtualization creates. The ISE system incorporates a hardware design that reduces rotational vibration — one of the leading causes of disk failure over time — improves resiliency, and ensures consistency in performance and availability. ISE software responds to drive errors and attempts to fix them, thus almost completely eliminating drive failures. It also ensures data integrity and collects heuristics that help the system optimize performance for a given application. The Hybrid ISE datapacs add the ability to even more efficiently deliver performance and capacity without making unnecessary trade-offs or sacrifices.

When considering a solution for a virtualized environment, organizations should consider ISE as a building block for efficient and effective storage.

ABOUT THIS PUBLICATION

This publication was produced by IDC Go-to-Market Services. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Go-to-Market Services makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

COPYRIGHT AND RESTRICTIONS

Any IDC information or reference to IDC that is to be used in advertising, press releases, or promotional materials requires prior written approval from IDC. For permission requests, contact the GMS information line at 508-988-7610 or gms@idc.com. Translation and/or localization of this document requires an additional license from IDC.

For more information on IDC, visit www.idc.com. For more information on GMS, visit www.idc.com/gms.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com