

Tintri FlashFirst™ Design: Storage Intelligence for Virtualized Environments

Built on the industry's first and leading application-aware storage architecture, Tintri VMstore delivers performance for thousands of VMs with sub-millisecond latencies using a patented hybrid flash and disk storage architecture. This technical paper discusses how Tintri FlashFirst™ Design including VM-level QoS works and how it makes virtualization simple, predictable and efficient.

Executive Summary

Performance and management complexity are the two biggest storage bottlenecks for increased adoption of virtualization in modern datacenters. Storage designed for virtualization addresses the performance and efficiency mismatch that plagues traditional storage and virtualization, and eliminates many mundane storage management tasks, allowing IT to focus on innovation.

Tintri VMstore™ is a hybrid storage solution with a combination of flash-based SSDs and high-capacity disk drives. Based on patented FlashFirst™ design, VMstore can service over 99 percent of all IO from flash, providing very high levels of throughput and consistent sub-millisecond latencies. VMstore also delivers VM-level QoS to guarantee every VM gets the performance it needs without any manual tuning or configuration — simply put, smart. VMstore also delivers non-disruptive and automatic VM alignment, which eliminates the performance-robbing overhead that can come from misaligned VMs on storage.

This white paper reviews the performance needs for storage in a virtual environment and details the capabilities of Tintri VMstore that make it the industry's leading storage solution for virtualization and cloud.

Drivers for Change in Virtualized Data Centers

Enterprise IT is under unrelenting pressure to reduce capital and operating expense, driving it to virtualize infrastructure to improve hardware utilization and scalability, and advance toward enhanced operational efficiency and flexibility. While virtualization has dramatically improved utilization of resources in the data center, the complexity of virtualized infrastructure gives rise to concerns around, among other things, application performance.

When IO requirements and VM behavior are poorly understood, a painful trial-and-error process ensues. Storage and VM administrators must coordinate to ensure each application not only has the space it needs, but also enough IO performance for the expected load.

Optimum application performance requires solutions for virtualization's unique performance challenges. The latest generation of servers can easily support upwards of tens of virtual servers, each of which can generate its own IO stream. The resulting IO patterns virtual environments generate are far more random than those generated by applications running on bare-metal servers. This "IO blender" effect translates into strong degradation of performance on traditional storage, which often means IT defers virtualizing IO-intensive tier-1 applications, and manually isolates some workloads even if they are virtualized.

Application performance guarantees are a concern if an enterprise wants to run tier-1 applications using a virtualized infrastructure, since many applications built using shared, virtualized infrastructure tend to present a noisy neighbor problem, where application performance can suffer if another application has a spike in demand and draws on shared resources.

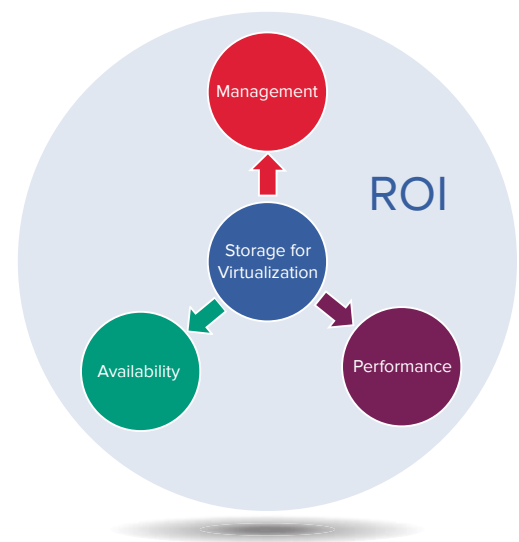


Figure 1: Storage that is engineered for virtualization must be optimized for high availability, minimal management overhead, and VM-granular performance. Each of these attributes have a dramatic effect on ROI.

Flash for Virtualized Environments

Flash is often considered the second-most disruptive technology after virtualization in the datacenter. Flash offers IT architects a way around the storage bottleneck caused by virtualization's extremely random IO needs, which could choke legacy disk-centric storage.

Using single-level cell (SLC) technology, the first wave of flash deployments in the datacenter was targeted at customers who needed performance at any cost. In this wave, flash was as much as 20 times the cost of disks and was mostly used as the end point for data in SANs or direct attached storage. Even with flash used as a read cache, the high cost and configuration restrictions severely limited adoption for virtualization and mainstream application.

To bring down costs, the second wave of flash deployment leveraged multilevel cell (MLC) flash technology. MLC flash is cheaper, but has a fraction of the endurance of SLC flash. To make MLC and enterprise MLC (eMLC) flash reliable enough as a data end-point, vendors coupled it with clever algorithms such as wear leveling, and in some cases dual-parity RAID. But the random nature of IO from virtualization, wear-leveling and RAID algorithms, along with the unique way data is written on flash, caused issues around write amplification. Storage designed with either SLC or MLC flash can suffer from write amplification, where the amount of data written to flash is a multiple of the actual data that needs to be written, but the impact is more pronounced in MLC flash.

Many flash-only and hybrid flash-disk vendors tried to solve write amplification problems using a flash file system incorporating garbage collection, which involves reading and rewriting data to flash memory. Poorly implemented garbage collection algorithms can cause latency spikes and limit the effective utilization of flash capacity, even when flash is being used as a cache.

On the cost front, in virtual environments even with data reduction techniques such as compression and deduplication, eMLC and MLC based flash-only storage is substantially more expensive than equivalent disks and hybrid flash-disk storage. Due to advances in density and capacity, hard disk drives have been able to maintain a substantial cost lead over flash.

So is hybrid flash-disk storage the way to cost-effective performance for virtual environments? Unfortunately the answer is not clear, as all hybrid storage solutions are not created equal. A major consideration for hybrid storage solutions, especially in virtual environments supporting thousands of VMs, is the ability to serve IO from flash. Even in hybrid architectures that use flash only as a read cache, this is an important metric as flash misses require data to be pulled from disk. In extremely random IO environments created by virtualizing hundreds or thousands of VMs, hybrid storage systems must be able to cost effectively serve 99 percent of IO from flash. Storage must also ensure every VM gets the performance it needs. Otherwise VM performance will suffer, counteracting the benefits of virtualization.

Accomplishing this without added complexity or cost requires intelligent storage that understands and operates at the VM level, tracking and maintaining active data for each individual VM in flash in real time. This is what Tintri Zero Management Storage delivers with its patented FlashFirst Design.

Tintri VMstore for Virtualization and Cloud

Tintri VMstore is smart storage that sees, learns and adapts. Tintri VMstore was built from the ground up based on the industry's first and leading VM-Aware Storage architecture to deliver always-optimized storage for dynamic virtual environments. It has storage intelligence for sub-millisecond latencies, VM-level QoS optimization, and deep visibility into the operational characteristics of each VM. This means highly effective monitoring and troubleshooting and VM-granularity data management, such as instantaneous snapshots, zerospace cloning and efficient replication. The result is higher predictability, unparalleled performance and efficiency and higher levels of productivity.

This intelligent Tintri VMstore functionality delivers unparalleled performance and density without the complexity.

Storage Intelligence

Tintri VMstore's approach automatically ensures every VM gets the performance it needs using unique FlashFirst Design, VM-granular QoS and automatic nondisruptive vDisk alignment. Importantly, VMstore appears in the virtualization layer as an independent datastore. Operating as individual storage datastores makes it easy to scale and control each node as part of a VMware Storage DRS cluster.

Tintri FlashFirst Design: VMstore is a hybrid storage solution. It uses a combination of flash-based SSDs and high-capacity disk drives for storage. The difference is Tintri VMstore integrates flash as a first-class storage medium rather than as a bolt-on cache or tier to fully leverage continued improvement in flash price and performance. Tintri's patented FlashFirst Design incorporates algorithms for inline deduplication, compression and working set analysis to service more than 99 percent of all IO from flash (Figure 5), for very high levels of throughput and consistent sub-millisecond latencies for both read and write operations.

Flash-first design minimizes swap between SSD and HDD by leveraging data reduction in the form of deduplication and compression to increase the amount of data that can be stored on flash. This is complemented by detailed profiling of all active VM IO, to ensure metadata and active data are kept in highperformance flash. Only cold data is evicted to disk, which does not impact application performance. It takes advantage of the fact that each VM has an active working set, which is a fraction of the overall VM. Using a flash-only approach means all data must be stored on high performance (and expensive) flash, whether it needs to be there or not.

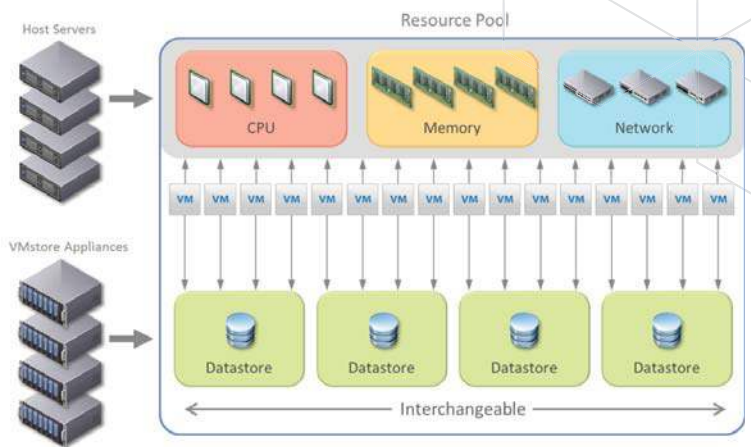


Figure 2: Tintri VMstore maps directly to the VM and vdisk.

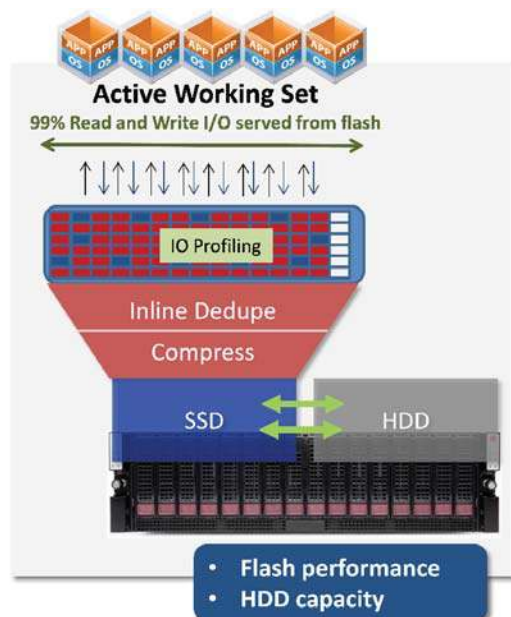


Figure 3: VMstore FlashFirst design delivers 99 percent+ IO from flash while enabling disk storage economics.

Unlike flash-only products, 100 percent of the operational flash capacity on a Tintri VMstore can be used without concern about running out of space and having applications come to a screeching halt. In addition, the Tintri VMstore is operationally far simpler and more cost-effective than flash-only products.

Legacy storage systems often incorporate flash to an existing disk-based architecture, using it as a cache or bolt-on tier, while continuing to use disk IO as part of the basic data path. In comparison, VMstore services 99 percent of IO requests directly from flash, thereby achieving dramatically lower flashlevel latencies, while delivering the cost advantages of disk storage.

Tintri's innovative FlashFirst design addresses MLC flash endurance and other issues that previously made it unsuitable for enterprise environments: Flash suffers from high levels of write amplification due to asymmetry between the size of blocks being written and the size of erasure blocks for flash. Unchecked, this reduces random write throughput by more than 100 times, introduces latency spikes and dramatically reduces flash lifetime.

FlashFirst design uses a variety of techniques to handle write amplification, ensure longevity and safeguard against failures, such as:

- Data reduction using deduplication and compression: Reduces data before it is stored flash resulting in fewer writes.
- Intelligent wear-leveling and garbage collection algorithms: Leverages not only information on flash devices, but also real-time active data from individual VMs for longer life and consistently low latency.
- SMART (Self-Monitoring, Analysis and Reporting Technology): Monitors flash devices for any potential problem, issuing alerts before they escalate.
- High performance dual parity RAID-6: Delivers higher availability than RAID-10 without inefficiency of mirroring, or the performance hit of traditional RAID-6 implementations.

VM QoS: Tintri VMstore is designed to support a mixed workload of hundreds of VMs, each with a unique IO configuration profile. VMstore can analyze and track each IO request to individual VMs and vDisks. This enables VMstore to isolate the VMs, queue and allocate critical system resources such as networking, flash/SSDs and system processing to individual VMs. Tintri VMstore's QoS capability is complementary to VMware's performance management capability. The result is consistent performance where it is needed. And, all of VM QoS functionality is transparent, so there is no need to manually tune the array or perform any administrative touch.

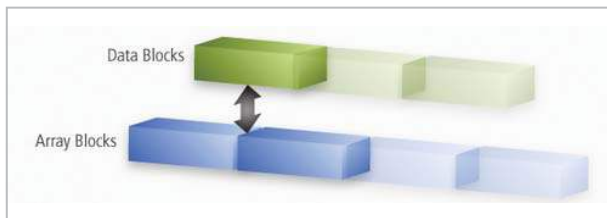


Figure 4: IO overhead due to misaligned blocks

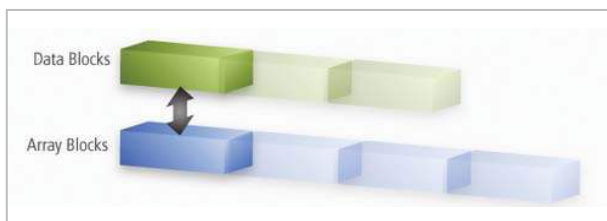


Figure 5: No IO overhead due to aligned blocks.

QoS is critical when storage must support high performance databases generating plenty of IO alongside latency sensitive virtual desktops. This is commonly referred to as the noisy neighbor problem in legacy storage architectures that are flash-only and lack VM-granular QoS. Tintri VMstore ensures database IO does not starve the virtual desktops, making it possible to have thousands of VMs served from the same storage system.

VM auto-alignment: VM alignment poses real challenges as virtualization deployment expands across enterprise data centers. Misaligned VMs magnify IO requests, consuming extra IOPS on the storage array. The impact snowballs as the environment grows with a single array supporting hundreds of VMs. At this size, performance impact estimates range from 10 percent to more than 30 percent.

Every VM writes data to disk in logical chunks. Storage arrays also represent data in logical blocks. When a VM is created, the block boundaries of the VM and storage do not always align automatically. If the blocks are not aligned, VM requests span two storage blocks, requiring an additional IO operation (Figure 4).

Storage administrators in virtualized data centers attempt to address this issue by aligning VMs to reduce the impact of misalignment on performance. Unfortunately, realigning a VM is a manual, iterative process that generally requires downtime.

VMstore offers VM auto-alignment. Rather than the disruptive approach of realigning each VM, Tintri VMstore dynamically adapts to the VM layout (Figure 5). Tintri VMstore automatically aligns all VMs as they are created, migrated or cloned — with zero downtime. An IT administrator can now eliminate this arcane task and enjoy performance gains with no VM downtime, and zero administrator intervention.

Scaling Beyond a single Tintri VMstore

Each VMstore appears as a high-capacity datastore in the virtualization layer. This eliminates performance interdependencies between VMstore systems. Tintri Global Center™ can be used as a centralized control point for monitoring and reporting on multiple VMstore systems and resident VMs. In the future Tintri Global Center will enable IT to set performance and capacity-based policies between different VMstore systems.

Conclusion: Tintri VMstore Smart Storage

Tintri smart storage helps IT organizations eliminate storage complexity and minimize costs for virtualized environments by addressing the mismatch between storage and virtualization. Tintri VMstore's storage intelligence delivers unparalleled performance and efficiency, and helps make virtualization predictable with zero management. Here is a summary of the performance benefit from Tintri features:

Performance and Efficiency

- Tintri FlashFirst design delivers as much as ten times better performance, and VM densities with sub-millisecond read and write latencies for virtualized workloads without any tuning or over-provisioning vs. traditional storage.
- VM-level QoS and performance allocation enables a single VMstore to support hundreds of different virtual workloads such as VDI and tier-1 applications, eliminating the noisy neighbor problem.

Making Virtualization Predictable

- FlashFirst design and VM-level QoS delivers predictable performance for critical VMs, even in dynamic environments. This also enables thousands of VMs to be consolidated on to a single Tintri VMstore system.

Enable Extra Productivity

- VMstore's simplified design and setup allows IT to get up and running in as few as eight minutes.
- Understanding and operating at the VM level eliminates the need to set up disk/RAID groups, LUNs and volumes and the need to maintain mapping between storage and VMs.
- Auto-alignment eliminates the need for disruptively realigning individual VMs to achieve optimal performance.

