



“ For over a decade Coolspirit have been supplying the UK’s top organisations with storage products and solutions so be assured we will meet your requirements head on.

It’s all about getting things right first time, quickly and simply! ”

**Damon Robertson**  
Coolspirit Ltd

**Our address**

24 The Bridge Business Centre  
Beresford Way  
Chesterfield  
S41 9FG

**Get in touch**

Call us on: 01246 454222  
Email us: [web@coolspirit.co.uk](mailto:web@coolspirit.co.uk)  
Find us: [View location map](#)  
Web: [www.coolspirit.co.uk](http://www.coolspirit.co.uk)

**Office hours**

mon - thurs 8:30am - 5:30pm  
fri 8:30am - 5pm  
sat - sun Closed

“ Boost your storage buying power...  
use ours! ”

The logo for Tintri, featuring the word "TINTRI" in a bold, blue, sans-serif font. A small, stylized green leaf-like shape is positioned above the letter "I".

# TINTRI

## Next-generation Tintri VMstore™

TECHNICAL WHITE PAPER

Vince Guan – Architect

Ed Lee – Architect

Pratap Singh – MTS

The bottom half of the page features a decorative background with flowing, wavy lines in shades of light green and blue. The lines curve upwards from the bottom left towards the right, creating a sense of movement and modernity.

### Why VM-aware storage?

Virtualization owes its success in transforming data centers to the power of the virtual machine abstraction. Unfortunately, storage for virtual machines (VMs) has increasingly become a bottleneck. A VM may be run on a generic pool of shared hardware resources, and its CPU and memory usage easily monitored and modified.

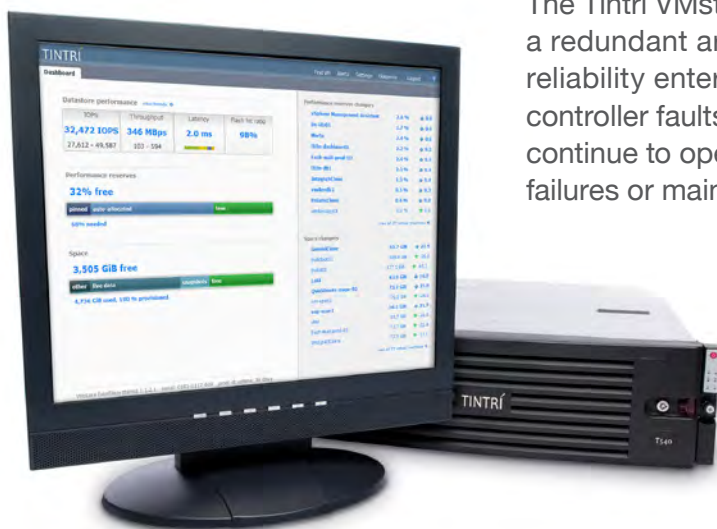
However, the problem arises with traditional storage: Designed before the advent of virtualization, traditional storage must be configured and tuned separately. This mismatch pulls virtual applications back into the physical realm.

Tintri™ overcomes these limits. Purpose-built for VMs and focused specifically on the problems of VM storage, Tintri VMstore™ provides management at the same level of abstraction as the rest of the virtual infrastructure. Specifically designed for virtualized environments, Tintri VMstore is the most innovative virtual machine storage solution on the market.

### Key enhancements

Key enhancements enable even broader adoption. Starting with a robust dual-controller implementation, Tintri also adds two industry-first technologies: instant bottleneck visualization, and VM auto-alignment. These new additions are a direct outgrowth of Tintri’s custom VM-aware file system, and address long-standing pain points in virtualized environments.

### Dual controllers for HA



The Tintri VMstore T540 is a dual-controller system that uses a redundant architecture to deliver the high availability and reliability enterprise IT environments demand. It eliminates controller faults as a single point of failure, and allows VMs to continue to operate from the VMstore without disruption from failures or maintenance events. A key design goal of Tintri’s dual-controller system is achieving higher levels of availability than a single controller system.

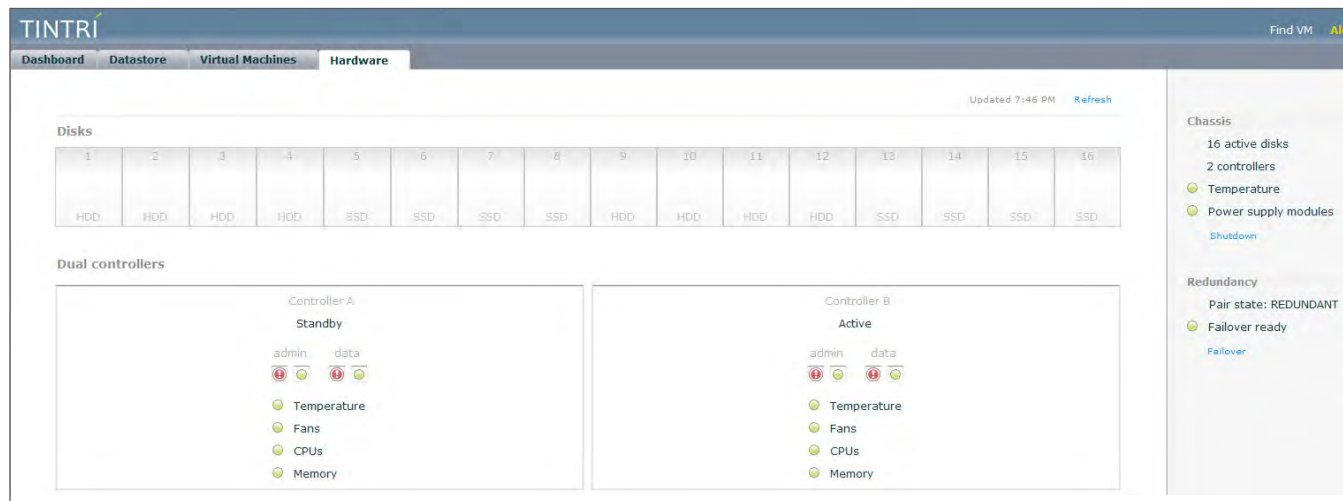
The T540’s underlying dual-controller architecture provides two independent and fully redundant sets of computing resources for the Tintri file system. Any hardware failure, either external or internal, that may make one set of resources unavailable causes the

other to take over. For example, if the network connection to the first controller has problems, or the first controller encounters hardware failures (e.g., memory ECC errors, disk controller errors), the second

controller takes over the operation and continues providing access to the VMs without disruption. This hardware redundancy enables the system to tolerate any component failure and significantly increases availability.

### Robust active-standby design

The T540 uses an active-standby model where VMs are actively serviced from only one of the controllers, while data and state information are synchronized between the controllers so that either can become active at any time. This design ensures that both controllers are not equally stressed at the same time and software is not executed in the same manner on both controllers. This eliminates the possibility of coordinated failures, and increases overall system reliability. In contrast, active-active configurations generally have more complicated and failure-prone failover and give-back behavior. Active-active also requires more administrative management to avoid potentially catastrophic performance drops, while typically extracting only 50 percent to 70 percent performance from the secondary controller.



### Simplified maintenance

Tintri's redundant active-standby architecture makes upgrade and maintenance easy. System software upgrades are non-disruptive. The upgrade process first updates the standby controller; once updated, the standby takes over the active role; at this point the originally active controller updates. This guarantees operations transition seamlessly without any disruption to the VMs. For maintenance, users can easily choose which controller is active from the GUI and perform a failover to switch a controller from active to standby (or vice versa) at any time in seconds. This makes it simple for IT administrators to plan and schedule regular maintenance in enterprise environments without disruption.

## Instant bottleneck visualization

Administrators dread troubleshooting storage performance problems. Users complain that their VM is slow. You suspect the problem may be with the storage, but how do you verify this when the VM is sharing a LUN with a dozen other VMs and the LUN is a slice of a RAID array that contains many other LUNs? Unfortunately, the legacy array provides no statistics on a per-VM basis. Perhaps the problem is not really the storage, and has something to do with the ESX host or the storage network, or even the user’s application.

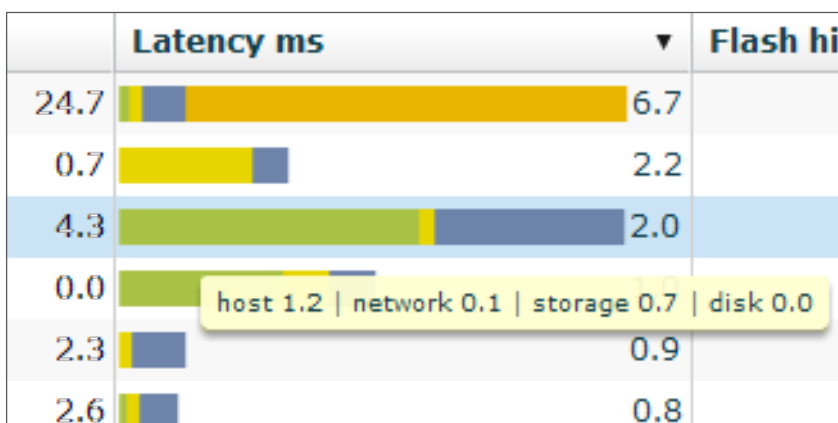
### Background

Today, identifying the cause of performance bottlenecks is a time consuming, frustrating and sometimes inconclusive process that requires iteratively gathering data, analyzing the data to form a hypothesis and then testing the hypothesis. In large enterprises, this process often involves coordination between several people and departments, typically spanning many days or even weeks. Many IT professionals refer to these interactions as “pin-the-blame” meetings!

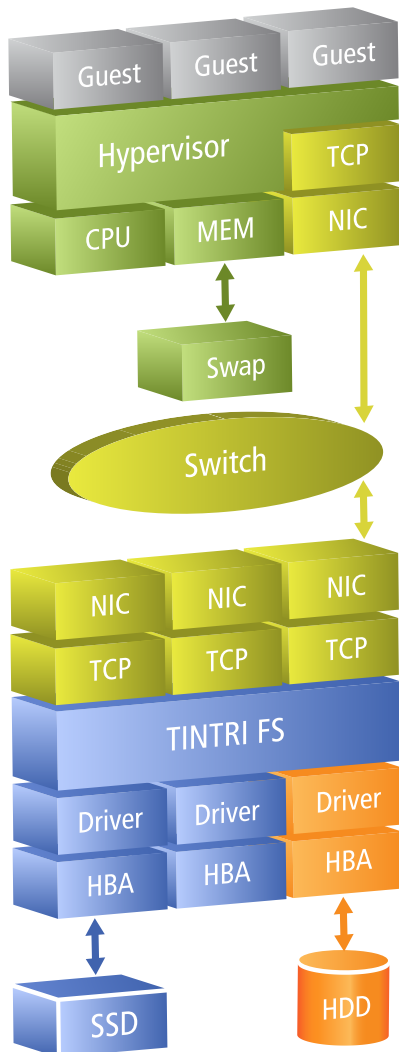
### Tintri’s bottleneck visualization

Fortunately, Tintri’s new instant bottleneck visualization automates this troubleshooting process. For each VM and vDisk stored on the system, Tintri displays a breakdown of the end-to-end latency all the way from the guest OS down to the disks within the Tintri appliance. For any VM or vDisk, you can see at a glance how much of the latency was spent in the ESX host, the network, the Tintri file system, or accessing the disk. Moreover, a history of this information is automatically stored and can be displayed as a graph over time, so you can see the bottleneck for each VM at any given point over the last seven days.

This visualization is generated by automatically collecting per-VM hypervisor latency stats and correlating them with per-VM storage stats that the Tintri VMstore collects for each VM (see diagram on next page). The hypervisor latencies are obtained using standard vCenter APIs, while the network, file system and disk latencies are provided by Tintri VMstore, which knows, for each IO request, the identity of the corresponding VM.



### Per-VM Statistics



■ **HOST** = Hypervisor + CPU Ready + Swap Wait

■ **NETWORK** = TCP/IP + NIC + Switch

■ **STORAGE** = Tintri FS + Driver + SSD

■ **DISK** = Driver + HDD

**HOST** includes delays due to hypervisor overhead, VMs unable to run due to low CPU reserves, or swapping due to low memory.

**NETWORK** includes delays in the hypervisor & Tintri TCP/IP stacks and NICs, plus network switch delays.

**STORAGE** includes all overhead for the Tintri FS to process request, plus driver and HBA overhead for SSD access.

**DISK** includes driver and HBA latencies to access HDD, plus time by the HDD to service requests.

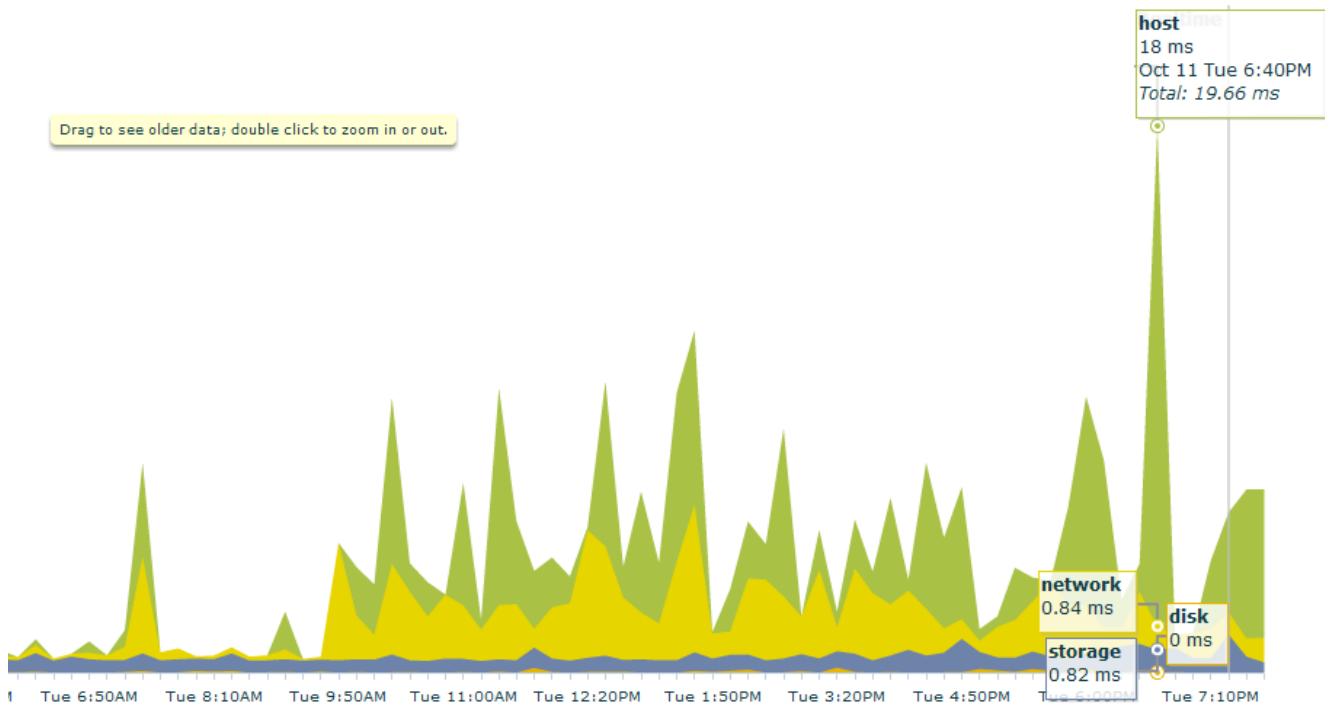
Tintri has been monitoring and using these latency stats ever since the first-generation system. The stats were so useful for triaging performance problems at production sites, we decided to expose them directly in our GUI. Using these stats since our earliest deployments, we discovered that a large fraction of the performance problems encountered in the field are due to hypervisor or networking issues, rather than storage problems

Based on customer interest, Tintri now exposes these latency statistics in an intuitive format. In an instant, you can see the bottleneck rather than trying to deduce where it is based on indirect measurements and time-consuming detective work.

Latency %	Latency ms	Flash hit %	Provisioned GiB	Used GiB	Host
0.3		100.0	100.0	97.1	esx13.tintri.com
0.0		100.0	20.0	5.9	esx13.tintri.com
0.0		100.0	50.0	10.9	esx9.tintri.com
0.0		100.0	81.5	79.9	esx-it02.tintri.com
0.0		100.0	50.0	10.1	esx-it01.tintri.com
0.0		100.0	500.0	8.1	esx13.tintri.com
0.0		100.0	64.0	35.8	esx13.tintri.com
4.2		100.0	100.0	95.9	esx13.tintri.com
0.0		100.0	1,024.0	495.0	esx-it01.tintri.com
0.0		100.0	32.0	29.3	esx13.tintri.com
0.0		100.0	22.0	22.4	esx13.tintri.com

Selected: 1 virtual disk, 5 IOPS, 0 MBps, 0.0 % reserves, 36 GiB [Hide graphs](#)

host network storage disk francis-win SCSI 0:0



## VM alignment

VM alignment is the to-do item almost every VM administrator has buried on a sticky note somewhere, but seems too daunting to tackle. And it's a problem that poses real challenges as virtualization spreads into more mainstream workloads. Misaligned VMs magnify IO requests, consuming extra IOPS on the storage array. At a small scale with a few VMs, the performance impact is small. However the impact snowballs as the environment grows, with a single array supporting hundreds of VMs. Performance impact estimates range from 10 percent to more than 30 percent.

### Background

Every guest OS writes data to disk in logical chunks. Storage arrays also represent data in logical chunks or blocks. When a VM is created, the guest OS writes logical data blocks to disk, but the block boundaries on the guest OS and storage don't always align automatically.

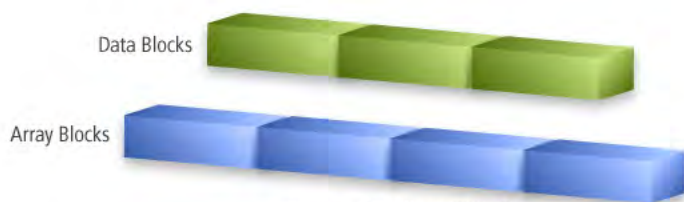


Figure 1: Misaligned blocks

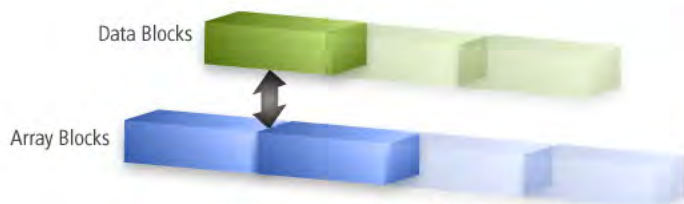


Figure 2: Overhead due to misaligned IO

If the blocks are not aligned, IO requests from the guest OS span two storage blocks, requiring additional IO. Figure 1 and Figure 2 below show an example of a misaligned layout and the IO impact.

This is not a new problem in the industry, and has been widely discussed by both vendors (VMware, EMC, NetApp, Microsoft) and bloggers like Duncan Epping, Josh Townsend and Chad Sakac. A quick Internet search finds message boards filled with discussions about alignment issues. The breadth and depth of the discussion makes it clear there's a very real and difficult problem with VM alignment.

### Alignment and Virtual Machines

In virtual environments, each VM's state is represented as a group of virtual disks that reside on a datastore. The datastore can be a storage array or local disks attached to each virtual server. The storage layout on a datastore is comprised of a set of blocks.

A VM runs a guest OS that creates one or more virtual disks to store state. The guest OS typically defines the layout of each virtual disk using some commonly used partition layout, such as a master boot record (MBR). The MBR stores information about how each virtual disk is partitioned into smaller regions, with its size and location. Except for Windows Server 2008 and Windows 7, blocks defined by the guest OS file system (NTFS, EXT3, etc.) do not typically align with the underlying datastore block layout.

Any IO misaligned by the guest OS amplifies the amount of IO required by the datastore. Since each datastore is used by many VMs, even a small amplification will likely exhaust a datastore’s resources with large numbers of misaligned IOs. Although this is a well-documented issue, many administrators avoid addressing the problem. For example, when we audited our own environment at Tintri, close to half of our VMs were misaligned. After proper alignment (more on that shortly), we experienced about a 30 percent performance improvement.

### Addressing VM alignment

So why are VMs misaligned? Certainly not due to a lack of awareness or attempts to address the issue. A variety of utilities can help align VMs and reduce unnecessary performance demands. Numerous blogs, whitepapers and knowledgebase articles describe why VMs should be aligned and provide step-by-step instructions.

But as administrators know, re-aligning a VM is a manual process. Worse, it generally requires scheduling substantial downtime.

### Tintri’s VM auto-alignment

Tintri’s VM-aware file system intrinsically “understands” each virtual disk. Building on this foundation, our second-generation Tintri VMstore offers VM auto-alignment. Rather than the conventional disruptive approach of re-aligning each guest, Tintri VMstore dynamically adapts to the guest layout. Nothing changes from the guest OS point of view.

Tintri VMstore automatically aligns all VMs as they are migrated, deployed, cloned or created — with zero downtime. A VM administrator can now completely eliminate this arcane storage-related task, and enjoy performance gains from 10 percent to more than 30 percent non-disruptively — with zero user interaction.



Figure 3: Aligned blocks

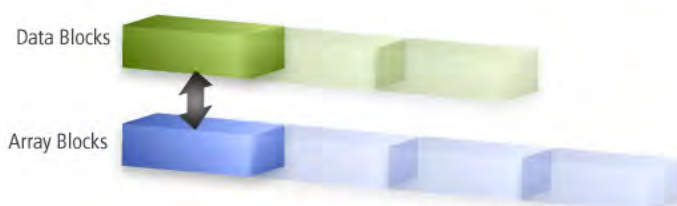


Figure 4: No overhead due to aligned IO

## Summary

Storage remains the primary obstacle to accelerating virtualization growth. In addition to a robust dual-controller implementation, Tintri’s second-generation system adds key new enhancements that eliminate major storage issues in virtualized environments. Instant bottleneck visualization and VM auto-alignment are a direct outgrowth of Tintri’s custom VM-aware file system. Tintri VMstore allows you to overcome the complexity, performance and cost obstacles preventing you from virtualizing more of your computing infrastructure.