ESG Lab Spotlight

Minimizing VDI Storage Costs with Citrix Provisioning Services

Date: November 2014 **Author:** Tony Palmer, Senior Lab Analyst

Abstract: *Citrix Provisioning Services (PVS) enables desktops to be provisioned and re-provisioned in real time, from a single shared disk image. The 7.1 release of PVS includes a new write cache option—cache on device RAM with overflow to disk—which allows the write cache to seamlessly overflow to a disk should the RAM cache become full. This spotlight summarizes ESG Lab's validation of the performance benefits of the new write cache option.*

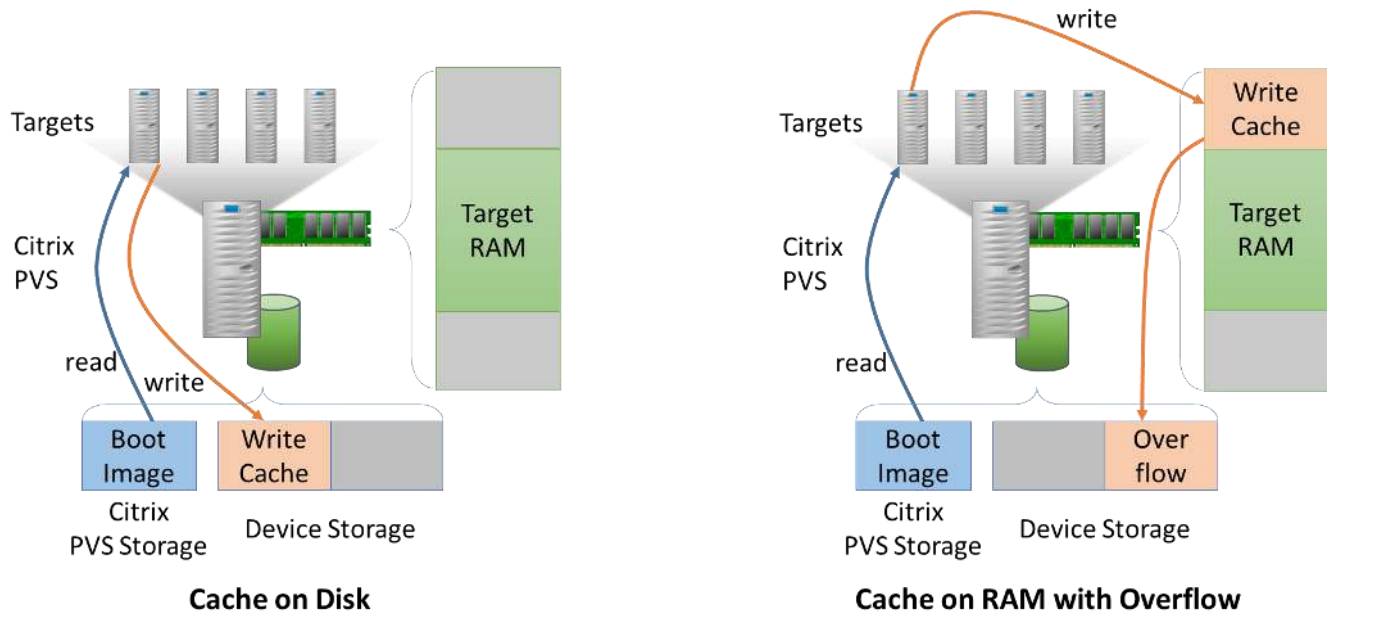
Citrix PVS and RAM Cache with Overflow to Disk

Citrix Provisioning Services (PVS) technology enables single instance virtual desktop image management. Administrators create a single master image of the boot disk for all virtual desktops. This disk image is streamed to each desktop as it boots. Updates to the operating systems or applications are applied only once to the master image. When the virtual desktop, or target, reboots, it automatically receives the latest version of the master image, and thus is automatically up to date with bug fixes and newly installed applications.

Targets are created on the virtual server without a boot drive, and are configured to boot from the network. When the target boots, the PVS server streams the boot disk image directly to the client while simultaneously caching the boot disk image in the PVS server RAM. Subsequent boots of additional targets using the same image stream the image from PVS server RAM rather than from the disk, accelerating boot performance and reducing boot times.

Since PVS is streaming a single disk image to multiple computers, the disk image must be read-only to avoid corruption. Because Microsoft Windows cannot boot from a read-only device, PVS implements a write cache, which is a unique location for the write data associated with each target.

Figure 1. Citrix PVS Write Cache Architectures



Source: Enterprise Strategy Group, 2014.

PVS supports four different write caching options:

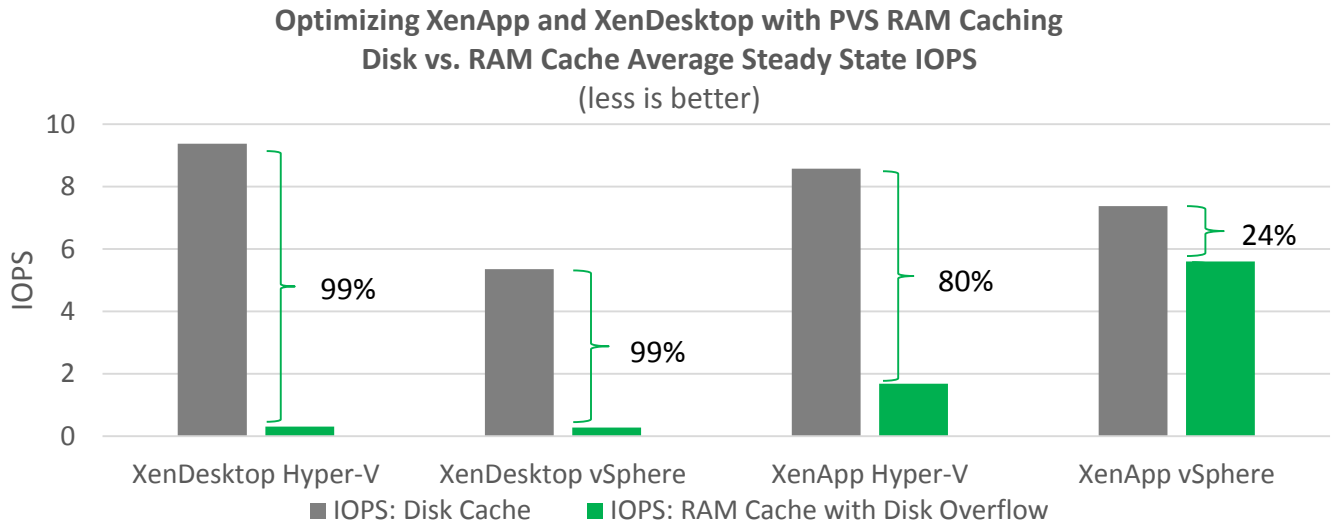
- Cache on Server – The write cache is placed on the PVS server storage. This option provides the lowest performance.
- Cache on Device Disk – The write cache is placed on the target's virtual hard disk.
- Cache in Device RAM – A portion of the target's RAM is reserved for the write cache, and is therefore unavailable for use by the operating system. Since the write cache is ephemeral, this option results in stateless desktops. In addition, should the RAM cache overflow, the target will crash.
- Cache in Device RAM with Overflow to Disk – New for the 7.1 release of Citrix PVS, this configuration option combines the best features of both Cache on Device Disk and Cache in Device RAM. Unlike Cache in Device RAM, target memory is not reserved. Instead, the write cache is mapped to non-paged pool memory and used when needed. Any unused cache memory is returned to the O/S. Once the write cache in RAM is full, stale data is written to the disk cache, freeing up memory for additional writes. Additional performance is gained by sequencing writes to disk and using larger disk blocks than used for the other write caching options.

Performance Advantages of PVS RAM Cache with Disk Overflow

ESG Lab audited Citrix's performance validation of PVS RAM cache with disk overflow. Performance tests were run on two dual processor HP servers with 384GB RAM and an HP SmartArray storage controller with 16 300GB 15,000 RPM SAS drives. Testing was accomplished with the LoginVSI (Login Virtual Session Indexer) benchmarking tool.[1] LoginVSI was designed to compare the performance of desktop virtualization solutions using different software and hardware configurations by mimicking typical user behavior using well-known desktop applications like Microsoft Office, Internet Explorer, Adobe Acrobat Reader, and Java/Freemind. Testing used the "medium" workload, which simulates the behavior of a typical "knowledge worker."

LoginVSI was used to determine the performance of the RAM cache configurations for both Citrix XenApp application virtualization infrastructure and Citrix XenDesktop virtual desktop infrastructure. The Citrix infrastructure platforms were tested running on both Microsoft Hyper-V and VMware vSphere hypervisors. The IOPS per target during the steady state portion of the LoginVSI test sequence was averaged for two runs for each Citrix platform on each hypervisor, and the results are shown in Figure 2.

Figure 2. Optimizing XenApp and XenDesktop with PVS RAM Caching



[1] www.loginvsi.com

Using PVS RAM cache with overflow, Citrix XenApp on VMware vSphere showed a 24% reduction in IOPS, and Citrix XenApp on Microsoft Hyper-V showed an 80% reduction in IOPS, which correlates to a reduced load on the storage system. On both Microsoft Hyper-V and VMware vSphere, the PVS RAM cache with overflow configuration for Citrix XenDesktop resulted in a 99% reduction in steady state IOPS, to less than 0.1 IOPS per target. As a result of using this option and reducing the storage workload, administrators can decrease the number of disks required to support the infrastructure. Alternatively, administrators can increase the number of users and applications per server and per disk, reducing the cost of implementing a Citrix XenApp infrastructure.

The LoginVSI benchmark tool simulates the complete virtual desktop lifecycle, starting from the target virtual machine boot. Boots of multiple targets are staged to overlap in time, simulating the real-world experience of multiple users booting targets at the start of the work day. After booting a target, a virtual user is logged in. Next, the user workload is started. After the workload has run for a predetermined number of loops, the workload is stopped, and the user is logged out. For the entirety of the test, from boot through logout, the peak storage load and the 95th percentile of the peak storage load was captured. This data was used to calculate the number of disks needed to support a typical 500 user configuration for each Citrix platform on each hypervisor.

Table 1. 95th Percentile Load and Disks required for a 500 User Workload

Citrix Platform	Hypervisor	95th %ile IOPS Disk Cache	Disks Required	95th %ile IOPS RAM Cache	Disks Required	Potential Storage Cost Savings
XenDesktop	Hyper-V	7,455	41	133	3*	93%
XenDesktop	vSphere	4,968	28	73	3*	89%
XenApp	Hyper-V	6,788	38	1,538	9	77%
XenApp	vSphere	5,850	33	4,268	24	27%

* While only one disk is required to support the number of IOPS, a minimum of three disks in a RAID 5 configuration is recommended for data protection.

For these comparisons between disk and RAM caching, the configurations were kept identical. The physical server had more RAM than was allocated to all targets, and each target was allocated the same amount of RAM regardless of the caching option.

The results of the LoginVSI testing show that Citrix XenApp on Hyper-V requires a minimum of 38 disks to support the 95th percentile of IOPS load with disk-based write cache. This is reduced to just nine disks with RAM caching. XenDesktop demonstrated a similar reduction, from 41 disks down to just one disk (although a minimum of three disks in a RAID 5 configuration is recommended for data protection).

Thus, by choosing PVS RAM caching with overflow to disk, administrators can leverage available free memory to gain a true reduction in both steady state and peak load IOPS. This translates to a corresponding reduction in the number of disks required. Citrix XenApp demonstrated a reduction in the number of disks of 27-77%, while XenDesktop demonstrated an 89-93% reduction in the number of disks required to support peak workloads.

The new 7.1 release of Citrix Provisioning Server provides a new write caching option, utilizing target RAM for the cache, overflowing to disk when the RAM cache becomes full. Selecting this option reduces both peak and steady state workload on the storage infrastructure. This leads directly to 80% or more cost savings since administrators implementing Citrix desktop and application virtualization infrastructures can use smaller, potentially less powerful, less costly storage systems.

The Bigger Truth

Delivering a good user experience is essential to the success of VDI deployments because users will not tolerate diminished performance as compared to physical machines. As a result, fast, predictable performance and scalability are critical concerns. Virtual desktop environments can hammer an infrastructure with random, shifting I/O; bottlenecks can occur in the storage domain that will have an impact on performance. To resolve these issues, organizations have become accustomed to overestimating shared storage to assure end-user performance, resulting in not only higher equipment costs, but also higher costs for storage management, data center floor space, and energy for power and cooling. Ironically, the cost savings that VDI offers can be significantly reduced or even eliminated in the effort to maintain expected performance. Citrix is addressing this challenge by reducing the amount of storage required to support desktop virtualization.

Citrix Provisioning Services utilizes streaming technology to provide full image lifecycle management for XenApp and XenDesktop, while drastically reducing overall storage requirements through RAM-based caching techniques. This approach to storage enhances the user experience and makes VDI more cost effective by reducing the amount of storage organizations must purchase and maintain.

The new version 7.1 release of PVS bridges the gap with a new write cache option. RAM cache with overflow to disk implements a new algorithm for the RAM-based write cache, using non-paged pool memory. When available memory is exhausted, stale data is written to disk, freeing memory for use by the OS or for additional write cache. Additional performance is gained by sequencing writes to disk and using larger disk blocks.

ESG Lab has validated that RAM cache with overflow to disk provides significant performance improvements over the previous implementation of disk caching. Citrix XenApp demonstrated a 24% reduction in steady state IOPS on VMware vSphere and an 80% reduction on Microsoft Hyper-V. Citrix XenDesktop on both Hyper-V and VMware vSphere demonstrated a 99% reduction in steady state IOPS, to under 0.1 IOPS per target.

Similar results were demonstrated for storage IOPS during steady state operations for a larger pool of desktops. The reduction in storage workload allows for a decrease in the total number of disks required to support a given number of users. RAM caching with overflow to disk exhibited a 27-77% reduction in the number of disks required to support the workload for application virtualization, and 89-93% reduction in the number of disks required for desktop virtualization.

RAM cache with overflow to disk is designed to enable organizations to confidently support server-based computing and virtual desktop implementations while reducing both steady state and peak storage workload. This reduces storage sprawl and administration and maintenance costs while enabling administrators to leverage smaller storage systems with fewer disks, further lowering costs.

Organizations are turning to server-based cache solutions for virtualized desktop workloads that demand high performance. With the 7.1 release of PVS, Citrix has given administrators a new cache configuration that cost-effectively improves desktop density and increases overall storage infrastructure efficiency.

Learn more about PVS write cache options

www.citrix.com