



# VMware<sup>®</sup> Horizon View<sup>™</sup> 5.2 and Hardware Accelerated 3D Graphics

Performance Study

TECHNICAL WHITE PAPER

## Table of Contents

Introduction.....	3
Hardware Accelerated 3D Graphics.....	3
3D Desktops in View.....	3
Supported Configurations.....	3
Best Practices.....	4
Maximizing VM Consolidation.....	4
Optimizing Resource Sharing.....	4
Tracking Compute Resources.....	5
Client Sizing.....	5
vSGA Performance Tests.....	5
Test Bed Architecture.....	7
Performance Metrics.....	8
Scalability Results.....	9
Light 3D Workload.....	9
Interactive User Experience.....	9
Light CAD Workload.....	11
CAD Workload 1 (SolidWorks Viewer).....	12
CAD Workload 2 (Solid Edge Viewer).....	13
Conclusion.....	15
References.....	16

## Introduction

The significant cost savings that can be realized by leveraging a virtual desktop infrastructure (VDI), coupled with the rapid growth in cheaply available bandwidth (both on LAN and WAN), has meant that VDI no longer need be constrained to low-resolution desktops associated with task workers, but can be leveraged to cost effectively bring high-fidelity, high resolution, multi-monitor 3D desktops to an ever wider audience.

This whitepaper discusses the support for hardware accelerated 3D graphics that debuted with vSphere 5.1 and View 5.2 and presents performance and consolidation results for a number of different workloads, ranging from knowledge workers using 3D desktops to performance-intensive CAD-based workloads. Further, given that the intensity of a 3D workload will vary greatly from user to user and application to application, rather than highlighting specific case studies, we demonstrate how the solution efficiently scales for both light- and heavy-weight 3D workloads, until GPU or CPU resources are fully utilized. This paper also presents key best practices to extract peak performance from a 3D VMware® Horizon View™ 5.2 deployment.

## Hardware Accelerated 3D Graphics

In response to user demand for an ever richer set of applications to be supported in the virtual environment, VMware has enhanced View to support hardware accelerated 3D graphics.

Support for non-hardware accelerated 3D graphics was introduced in VMware vSphere 5.0. This enabled VMware View 5.0 to support VMs running Windows Aero and enabled basic 3D application use (for example, Google Earth).

In the next phase of VMware's 3D vision, vSphere 5.1 introduced GPU virtualization, enabling multiple VMs to simultaneously share a single, physical GPU. This feature, termed vSGA (Virtual Shared Graphics Acceleration), is compatible with all key VMware technologies, including vMotion, and enables the recently released View 5.2 to support GPU-backed virtual desktops.

The support for hardware accelerated 3D graphics delivers significantly higher performance in a 3D environment. This expands the 3D application space that can be successfully run in View 5.2 to include key technologies such as CAD and medical imaging.

### 3D Desktops in View

Support for 3D desktops in View can be administered by using the VMware Horizon View Administrator console, and can be enabled on a per-pool basis or controlled on a per-VM basis using the VMware vSphere client. For complete details on managing and configuring 3D desktops, refer to "VMware Horizon View 5.2 Documents" [1].

### Supported Configurations

Currently, vSphere supports DirectX v9 and OpenGL 2.1 applications running in both Windows 7 and Windows 8 VMs. vSphere 5.1 provides virtualization support for a range of Nvidia GPUs. Currently, the supported GPUs are:

- Quadro 4000
- Quadro 5000
- Quadro 6000
- Tesla M2070-Q
- Grid K1
- Grid K2

## Best Practices

View 5.2 and PCoIP dynamically adapt to the available CPU and bandwidth resources to present the optimal user experience. Even when tens of VMs are sharing a single physical GPU, vSphere ensures that the resource is fairly shared between the different VMs. As a result, very little out-of-the-box configuration is required deliver peak performance:

- Configure VMs to use VMXNET3 NICs.
- Consider disabling PCoIP's build-to-lossless mode.
- Set the MaxAppFrameRate (see Optimizing Resource Sharing).

The next sections describe additional considerations for maximizing performance and efficiency of a 3D desktop workload.

## Maximizing VM Consolidation

The administrator console provides the ability to configure the amount of graphics or video RAM (termed VRAM) allocated to each VM. The default per-VM VRAM allocation is 128MB. While increasing the per-VM VRAM allocation might deliver higher performance, it will limit the number of VMs that can simultaneously share a GPU; the GPU's memory is sub-divided between the VMs. Accordingly, if the GPU has 4GB of memory, and the VMs are configured to use the default 128MB of VRAM, it is possible for up to 32 VMs to simultaneously share the GPU. Similarly, if the VMs are configured to use 512MB each, the per-GPU consolidation ratio may be reduced to 8. To further increase the number of VMs supported per server, vSphere 5.1 can support multiple GPUs per system.

For Aero and basic 3D applications, use of the default VRAM allocation is sufficient and is recommended to ensure maximum consolidation ratios. Higher allocations can be reserved for VMs that will run more intensive 3D operations, where additional GPU resources will deliver improved performance.

For pools containing more VMs than can be simultaneously supported by the GPU, additional VMs will not boot once the GPU resources have been exhausted. In this situation, rather than explicitly setting the pool to use hardware 3D in the administrator console, it is advantageous to use the "automatic" option; this enables any additional VMs that might be needed, over and above the GPU's capabilities, to be supported using vSphere's software renderer solution.

In contrast, for situations where two distinct groups of users share a server—one group requiring hardware 3D acceleration and one not—it is best to configure these two groups of VMs as separate pools. Use the administrator console to explicitly configure the hardware group to use hardware accelerated 3D and the other group to use either software 3D or even no 3D, as appropriate. Changes to a pool's 3D strategy are handling automatically by View and vSphere and do not require manual configuration of the desktop VMs.

## Optimizing Resource Sharing

In contrast to a physical workstation that has sole use of its GPU, in the virtualized environment GPUs become a shared resource. As a result, it is important to ensure that each VM does not use the GPU resource in a wasteful manner. For instance, in many situations it often does not make sense for a 3D application to render hundreds of frames per second if View is configured to remote at a lower frame-rate (30fps is the default setting). For these situations, View provides a registry setting to limit the maximum application frame rate. This can either be configured in the template VM or on a per-VM basis, and the value should typically be set to the maximum frame rate that is being used by PCoIP. This configuration is achieved by using the following registry setting (REG\_DWORD):

```
HKLM\SOFTWARE\VMware, Inc.\VMware SVGA DevTap\MaxAppFrameRate
```

Setting this registry entry for a 3D workload has been found to significantly improve the performance and consolidation ratios achievable.

### Tracking Compute Resources

When consolidating multiple GPU accelerated 3D VMs onto a server, it is important to track both CPU and GPU utilization. When considering a VM's CPU utilization, it is important to consult host-level information in order to ensure that the contributions of all virtual machine components are considered. This is readily achieved either by using `esxtop` or by consulting the appropriate plots in the vSphere client or vCenter. GPU utilization can be determined by leveraging `nvidia-smi` on ESXi. This command returns information, as illustrated in Figure 1, showing the GPU's memory usage and utilization.

```

~ # nvidia-smi -l
Wed Feb 27 23:25:59 2013

+-----+
| NVIDIA-SMI 4.304.46   Driver Version: 304.46       |
+-----+-----+
| GPU  Name           | Bus-Id      Disp. | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap | Memory-Usage | GPU-Util  Compute M. |
+-----+-----+-----+-----+-----+-----+
|  0  Quadro 4000     | 0000:0B:00.0  Off |           N/A       |
| 36%   55C   P12    N/A /  N/A | 28%  575MB / 2047MB |    40%    Default   |
+-----+-----+-----+-----+-----+-----+

+-----+
| Compute processes:                                   GPU Memory |
| GPU      PID  Process name                             Usage       |
+-----+-----+-----+-----+-----+
| No running compute processes found                   |
+-----+

```

Figure 1. Using `nvidia-smi` to check GPU utilization in ESXi

By examining both the GPU and CPU utilization, it is simple to determine what, if any, resource is first exhausted, and begins to restrict further scaling.

### Client Sizing

3D workloads are frequently more graphically intensive than traditional office worker applications. As a result, they have the capacity to place higher computational demands on the user's client device. Accordingly, some thin clients may lack the computational resources to deliver a high-quality 3D experience to the end user. While the exact client device requirements are driven by the specifics of the workload running on the remote desktop, for the workloads undertaken in this white paper, a single-core Intel Atom-based thin client was found to be adequate with significant idle CPU time still available.

## vSGA Performance Tests

The aim of this vSGA performance whitepaper is to demonstrate the scalability of a VDI solution that uses vSGA to support 3D graphics. Accordingly, the paper focuses on four different workloads that stress the vSGA solution in different ways. VMware View Planner [2] was used to measure the scalability—defined in terms of the consolidation ratio and the corresponding response time or frame rate during the runs. The chosen workloads represent typical customer use scenarios. The four workloads are:

- **Light 3D Workload:** This workload is composed of common desktop applications, including Office 2010, Adobe Acrobat, 720p video, IE9x static content, IE9x displaying a Web album, and Google Earth running in the Chrome browser. All these applications are launched at the beginning of the run and remain open for the duration of the run. Throughout the duration of the test, the workload performs a variety of different operations using these applications. The exact ordering of the operations differs from desktop to desktop to mimic real-world workloads. The desktop VMs run Windows 7 at a resolution of 1600x1200 pixels and have

Aero enabled. This workload represents a use-case scenario typical of a knowledge worker.

- **Light CAD Workload:** This workload adds the SolidWorks CAD viewer to the “Light 3D Workload.” The SolidWorks CAD viewer is typical of applications used by a CAD content consumer, and this workload represents the use case where CAD viewers are used occasionally in conjunction with typical office applications. In this test, the SolidWorks CAD viewer is used to run two models: a sea-scooter (as illustrated [Figure 2](#), upper left) and a cross-section of a shaft (as illustrated in [Figure 2](#), upper right). The models are run sequentially using a single SolidWorks viewer.
- **CAD Workload 1:** The SolidWorks CAD viewer discussed in the prior use-case is run in isolation; that is, without any other applications for the duration of the test. This workload uses the same sea scooter and cross-section models to demonstrate the capability of the system to generate and remote frames when using these models in a manner that is typical of the use-case scenario for a CAD content consumer.
- **CAD Workload 2:** A Solid Edge CAD viewer is run in isolation for the duration of the test. During the test a 3-1 reducer model was used, as illustrated in [Figure 2](#), bottom. This workload is run in a manner similar to that of “CAD Workload 1”.

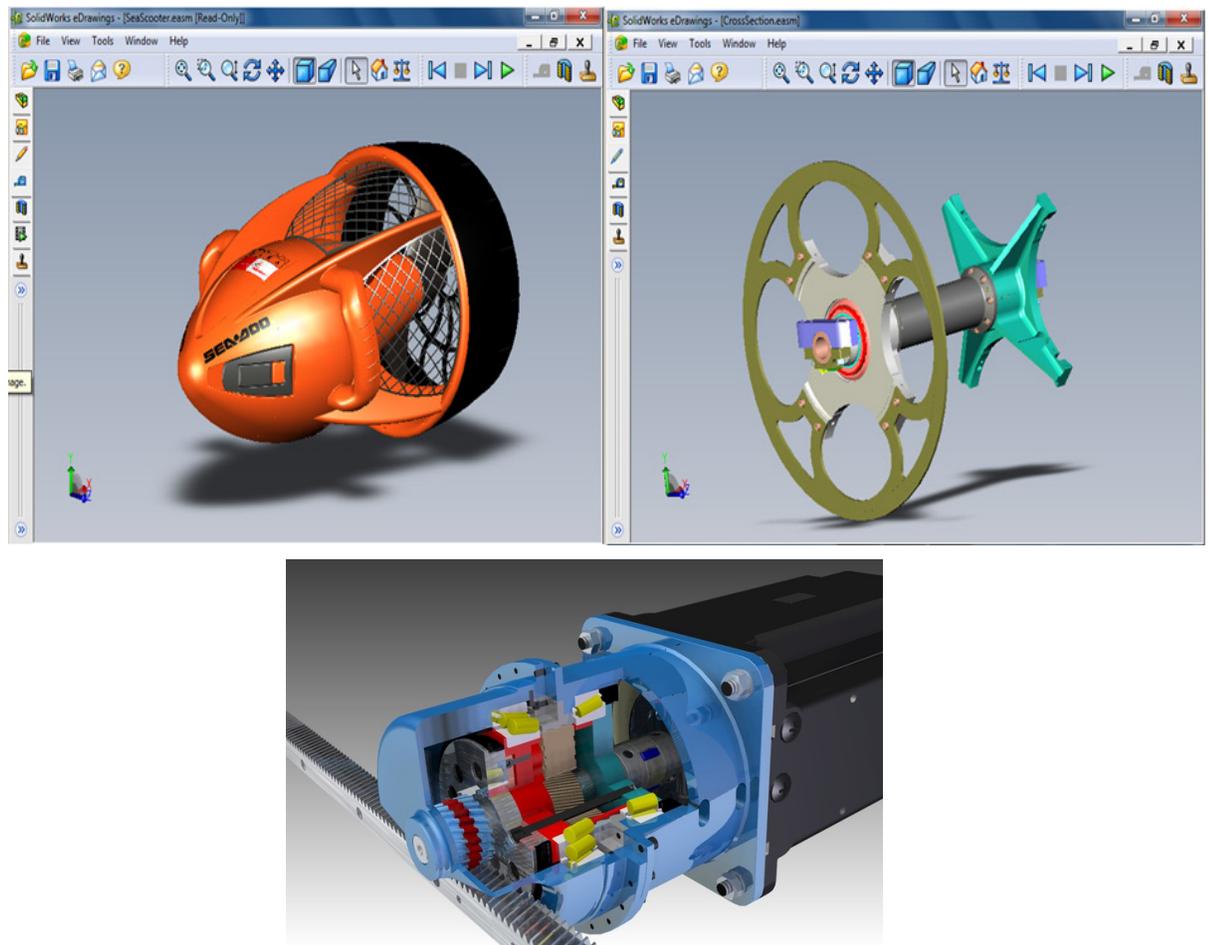


Figure 2. Shows the “SeaScooter” (top left) and “CrossSection” (top right) models that were used in performance tests with the SolidWorks CAD viewer and the “3-to-1 reducer” (bottom) model that was used in performance tests with the Solid Edge Viewer.

In initial performance testing, it was quickly discovered that the sophisticated image caching techniques in View 5.2 ensured that any repetitive interaction with the CAD applications was rapidly cached such that, in some cases for the remainder of the test, View was able to source up to 90% of the total remoted pixels from the image cache. Accordingly, simple model rotations or model animations are not suitable operations for examining the real-world performance of the system. Time was spent devising a more real-world interaction with the 3D models. The goal was not to completely defeat the View image caching, but to manipulate the model in a way that more closely mimics the potential usage by an actual CAD user. After studying how users tend to interact with 3D models, an automated interaction with the model was devised that approximates this process and this method is used in the CAD workloads that are presented in this whitepaper.

### Test Bed Architecture

The test bed architecture, shown in Figure 3, uses the View Planner architecture [2], and is composed of three major logical components:

- The workload is executed on Windows 7 VMs. These VMs are referred to as the desktop VMs and each are configured with 2 vCPUs, 1.75GB of memory and 128MB of VRAM. The desktop VMs are all located on a single physical server as shown in Figure 3. The load on the server and the GPUs is varied by changing the number of desktop VMs.
- The simulated users use VMware View software clients running on Windows XP VMs to connect to their Windows 7 desktop VMs. These Windows XP VMs are referred to as client VMs, and, as illustrated in Figure 3, are located on a second physical server. The simulated users use 2-vCPU VMs as their client machines. Guidance on client sizing is discussed in the Best Practices section.
- The DNS server, the View Planner appliance, the VMware View-broker, and VMware vCenter are run using VMs located on a third physical server, not shown in Figure 3.

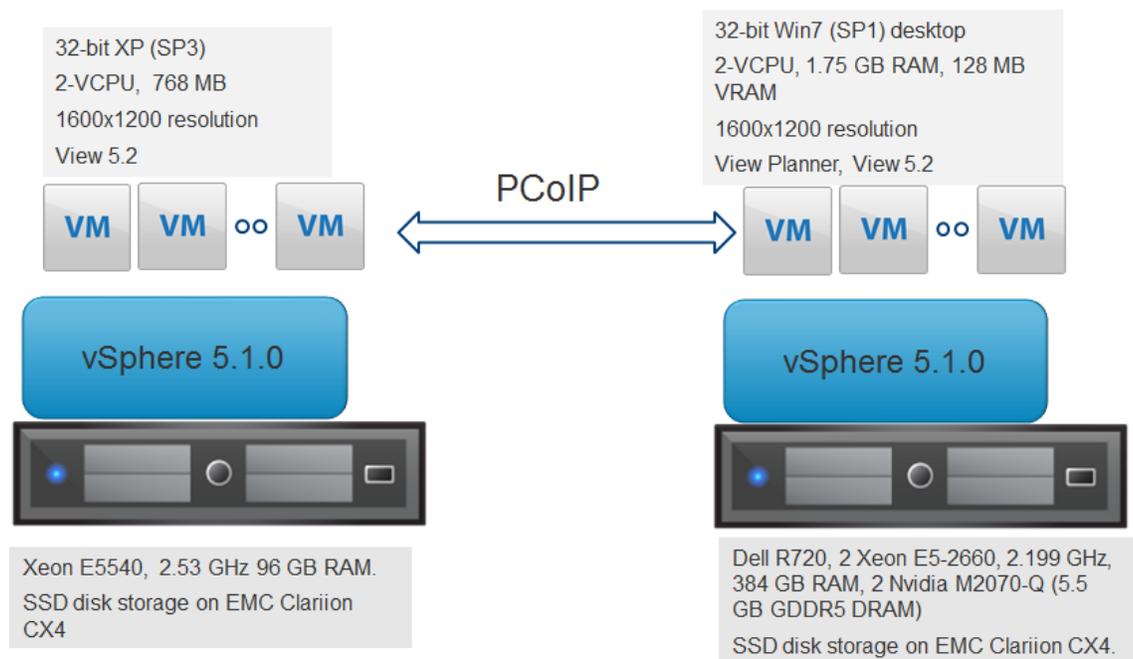


Figure 3. Experimental setup for measuring the performance of the vSGA stack and View 5.2.

The tests use a modified version of View Planner 2.1 that supports higher 1600x1200 screen resolutions and incorporates 720p video streams. A “think time” of ten seconds was chosen for the View Planner runs, and the CAD workloads were implemented using View Planner’s support for custom apps.

Finally, all of the performance tests discussed in this white paper were run with default settings, except for ensuring that the registry setting “MaxAppFrameRate”, described in the section titled “Optimizing Resource Sharing”, was set to 30 frames per second to match the default PCoIP frame rate.

## Performance Metrics

For VDI deployments, performance is typically measured by the number of users that can be supported with a certain level of remote desktop *responsiveness*. Responsiveness is defined by a variety of metrics, including application response time and remotely delivered frame-rates. The following metrics are used to quantify View 5.2 performance:

1. **Consolidation ratio:** The number of users that can be supported concurrently on a server. The consolidation ratio is also frequently presented as the number of VMs per processor core.
2. **Response time:** View Planner measures the response time for non-I/O-bound operations while the workload is running and reports the 95<sup>th</sup> percentile of these response times. In this white paper, the results are normalized to the maximum allowable View Planner threshold [2] response time.
3. **Remotely delivered frame rates:** VMware has patent-pending techniques that accurately measure the number of remotely delivered frames that correspond to frame updates generated by 3D applications and videos.

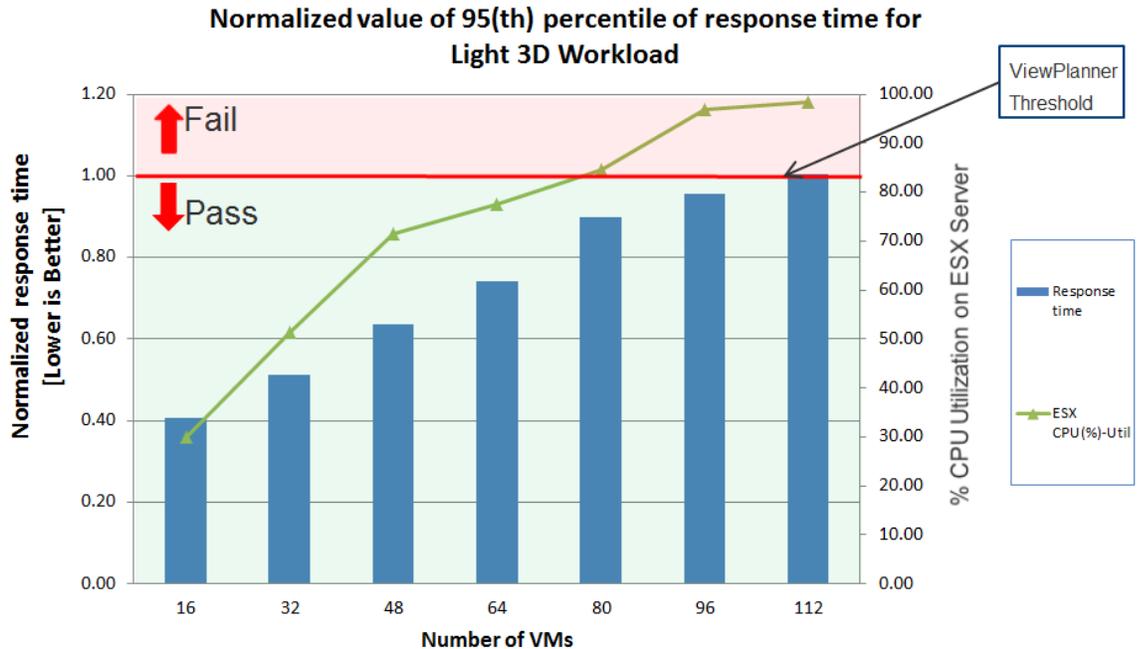
We report the consolidation ratios achieved in conjunction with either response time or remotely delivered frame-rate metrics. As a result, there is no single consolidation ratio, rather a range of consolidation ratios that can be reconciled with a user’s perception of acceptable performance. For the typical office workloads run by View Planner, based on extensive testing, a View Planner threshold [2] response time was found that represents the upper limit on an acceptable response time. A detailed description of this selection process is available in the View Planner discussion forum [3]. In this paper, the maximum consolidation ratio that can be achieved is represented by the maximum number of VMs that can be run without violating the View Planner threshold.

The scalability of VMware’s VDI solution was investigated by running these workloads on a single Dell R720 server with different VM consolidation ratios. The number of VMs that can be supported per GPU can be dictated by either the GPU’s compute resources being exhausted or the GPU’s available memory being exhausted. For the light 3D workload and the light CAD workload, it was found that GPU memory is the limiting factor on the GPU. Accordingly, the test system leveraged 2 GPUs, each with almost 6GB of RAM, to maximize the number of VMs that could be supported. When running with 2 GPUs, the major bottleneck to further scaling was found to be the CPU. CAD workloads were experimented upon with both 1 GPU and 2 GPU setups. This testing found that, in both scenarios for the dual-socket server under test, the CPU represented the bottleneck to further scaling.

The CAD workloads and light CAD workloads stress the CPU and show how well the vSGA solution scales with CAD workloads running at peak load either in isolation or together with other applications.

## Scalability Results

### Light 3D Workload



**Figure 4.** The bar chart presents the normalized values of the 95th percentile of the response time as the number of VMs is increased. Scaling is continued until performance falls below the View Planner response time requirements. The corresponding CPU utilization, as measured using `esxtop`, is shown by the line graph. The peak GPU utilization was 20%.

In this initial test the light 3D workload was run using View Planner and the number of VMs gradually increased until the View Planner response threshold was exceeded. These results are presented in [Figure 4](#). Based on this data, it is clear that the vSGA stack can support 7 users per core on this system while each user is executing the light 3D workload; running on higher performance processors will typically deliver even higher consolidation ratios.

As discussed previously, the results shown in [Figure 4](#) were obtained using desktop VMs configured with 128MB VRAM. Since the test bed had two Nvidia M2070-Q GPUs with about 5.5GB DRAM each, only around 80 desktop VMs can be supported by the GPUs. As the number of VMs is increased further, the additional VMs are supported using the software-renderer. For this light 3D workload, the maximum consolidation ratio achieved on the dual-socket server under test was 112: the test was stopped when, at 112 VMs, the VM responsiveness reached the upper limit allowed by the View Planner responsiveness threshold. At this point, both the GPU memory and CPU resources of the system are exhausted and further consolidation, while maintaining acceptable user experience, is not possible.

### Interactive User Experience

To both accurately and automatically quantify and measure interactive user experience at scale, VMware has developed a test suite composed of six tests: three simple 3D rendering tests, a fourth test that drags a window across the screen, a fifth that scrolls the contents of a window, and a final test that maximizes and minimizes a window, all of which are controlled by mouse events sent by the View client. These tests are constructed using patent-pending techniques that enable performance agents installed on the View clients to accurately identify

each of the frames that are rendered to the View client's display [5]. The performance agent installed on the View client leverages the frame identity information, in conjunction with its knowledge of when each frame was actually rendered to the client's display to compute a user experience metric. This metric can be used to quantitatively characterize the remote VMs responsiveness in the response to mouse events as the consolidation ratio is increased.

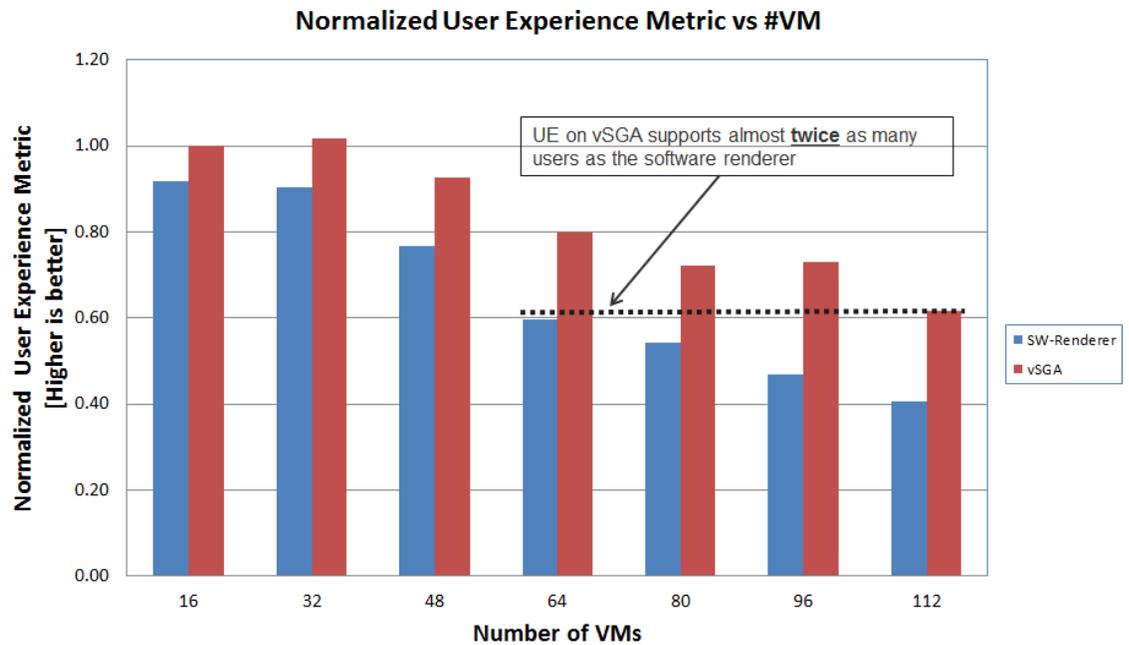


Figure 5. Presents normalized user experience metrics for both the vSGA and software renderer solutions as the number of desktop VMs is increased. The data clearly illustrates that the user experience observed with 112 users on a vSGA based system is equal to that observed with only 64 users when using a pure software renderer solution.

User experience was measured while running the light 3D workload on both vSGA and the software renderer. The normalized values of user experience are shown in Figure 5, and clearly illustrate:

- Using hardware accelerated 3D improves responsiveness in comparison with a software solution, even at low consolidation ratios, where CPU is not exhausted.
- Adding GPUs to an existing software-renderer solution enables the VM consolidation ratio to be almost doubled while maintaining user experience.

Light CAD Workload

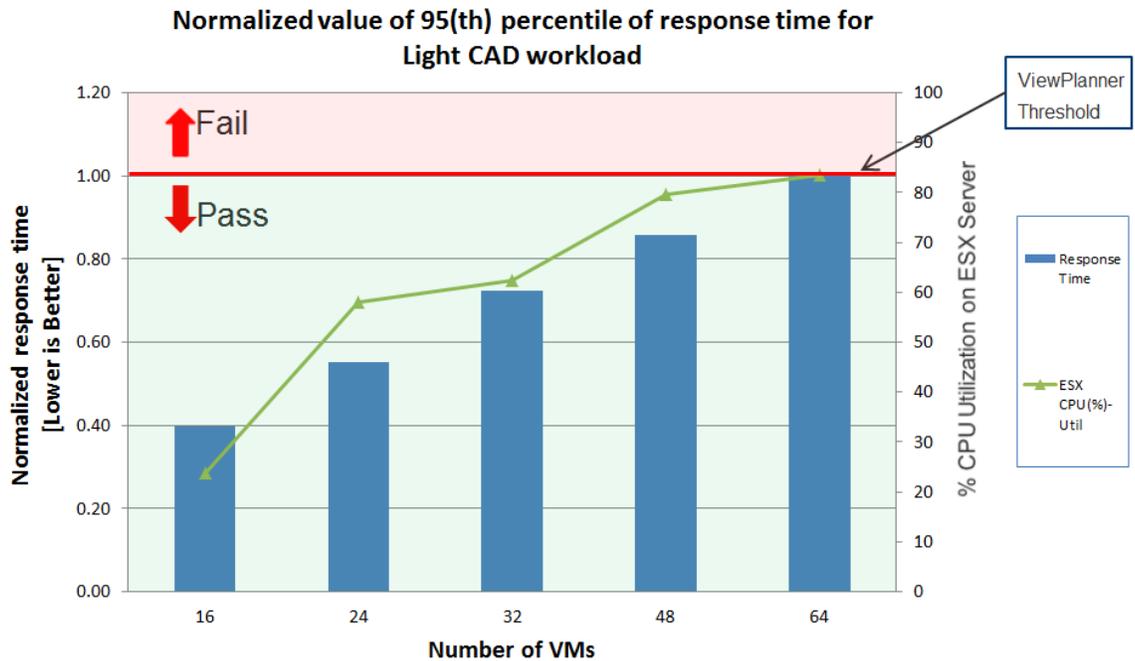


Figure 6. The bar chart presents the normalized values of the 95<sup>th</sup> percentile of the response time for the light CAD workload as the number of VMs is increased. The corresponding CPU utilization, as measured using `esxtop`, is shown by the line graph. Peak GPU utilization was 35%.

As discussed in the initial workload descriptions, the light CAD workload is run using View Planner and involves adding SolidWorks as a custom application. Each user continues to utilize a variety of typical office applications, but also periodically interacts with SolidWorks. The results are shown in Figure 6, and illustrate that up to 64 VMs can be supported.

It is important to note that scaling was stopped at 64 VMs, even though the CPU utilization was less than 100%. This was necessary because higher consolidation ratios exceeded the View Planner response time threshold and were deemed to have failed the test’s performance criteria. This illustrates a key aspect of this paper’s approach to scaling: scale-up does not continue until a 100% CPU utilization is reached. Rather, scale-up is only continued as long as the response time meets the View Planner threshold response time, ensuring that users at the peak consolidation ratio see acceptable performance.

CAD Workload 1 (SolidWorks Viewer)

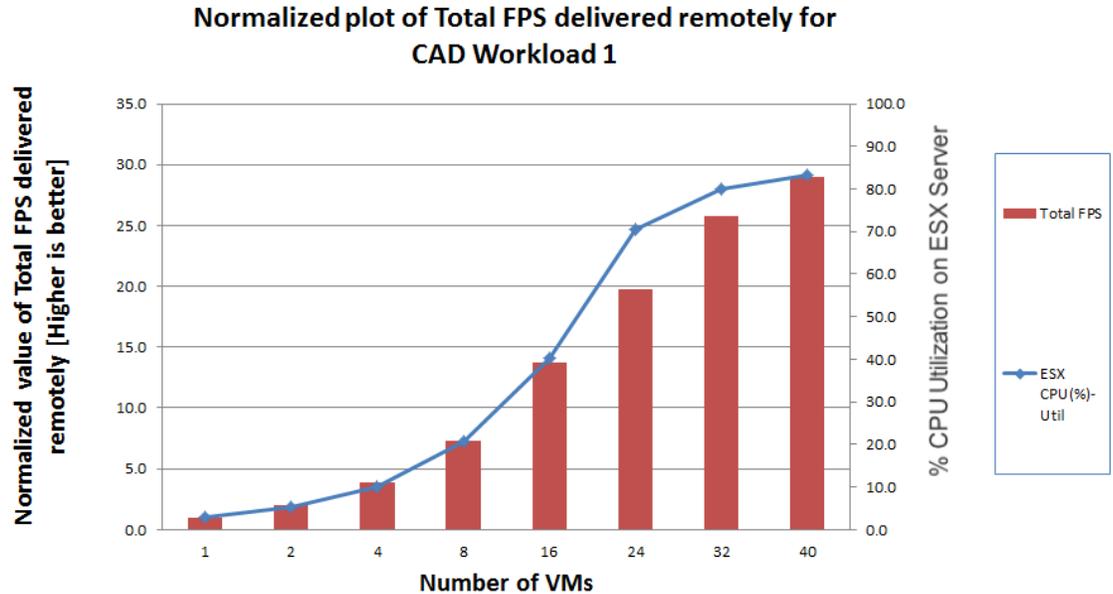


Figure 7. The bar chart presents the aggregate frame rate delivered by the system to the remote clients as the load on the server is increased. The results are normalized and the frame rate observed with just one VM running on the server is defined as the basis for comparison. The corresponding CPU utilization as measured using `esxtop` is shown by the line graph. GPU utilization was observed to be around 37%.

In this workload the office applications used in the prior tests are dropped and attention is focused on the 3D CAD application. The workload consists of the SolidWorks CAD viewer interacting with the two previously discussed models. The normalized system performance (where performance is defined as the normalized aggregate frame rate delivered by the system to the remote clients) as the number of VMs is increased is presented in Figure 7 and illustrates the scalability of the vSGA stack and View 5.2: the aggregate performance with 32 desktop VMs is almost 26X the single VM performance.

Rather than presenting this performance data in terms of the server aggregate, Figure 8 presents the same data on a per-VM basis. From the data it is apparent that the per-VM frame-rates are only modestly impacted as the consolidation ratio is increased from 1 to 40 VMs. The performance criteria for this test are very stringent and while the per-VM performance with 40 VMs currently running is only just less than 80% of the single VM performance, it is nevertheless considered to be insufficient and a consolidation ratio of 32 VMs was considered to be the maximum for this workload and the system under test.

The results in Figure 7 and Figure 8 clearly illustrate the strength of the vSGA solution for 3D workloads: a dual-socket server that might traditionally be provided on a per CAD-user basis is now capable of supporting up to 32 CAD users.

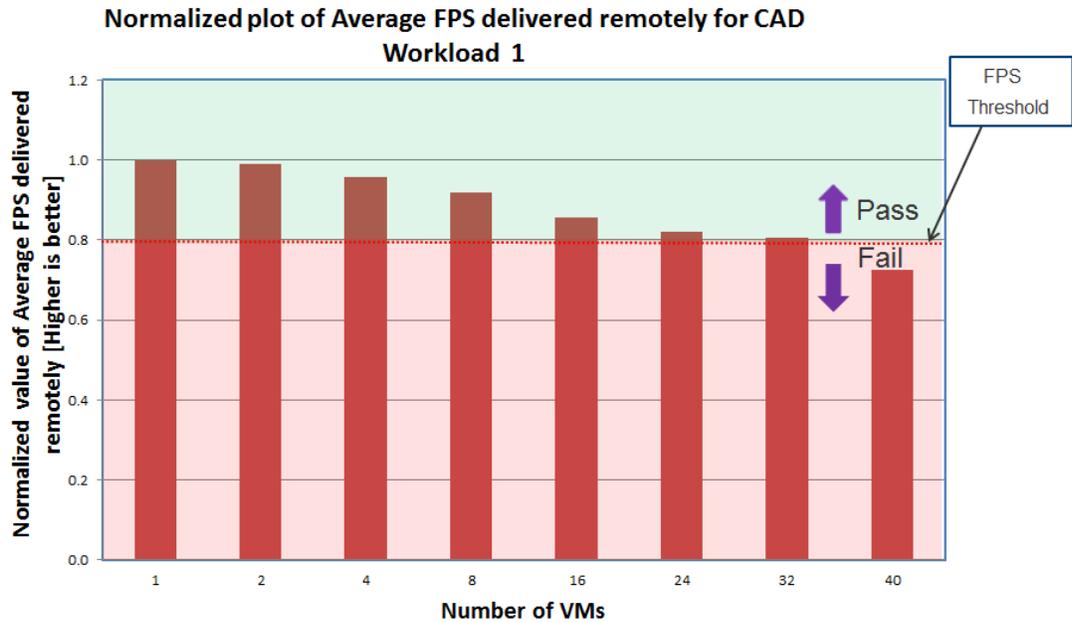


Figure 8. The bar chart presents the average frame rates delivered to the View clients as the load on the server is increased. The results are normalized and the frame-rate observed with just one VM running on the server is defined as the basis for comparison.

CAD Workload 2 (Solid Edge Viewer)

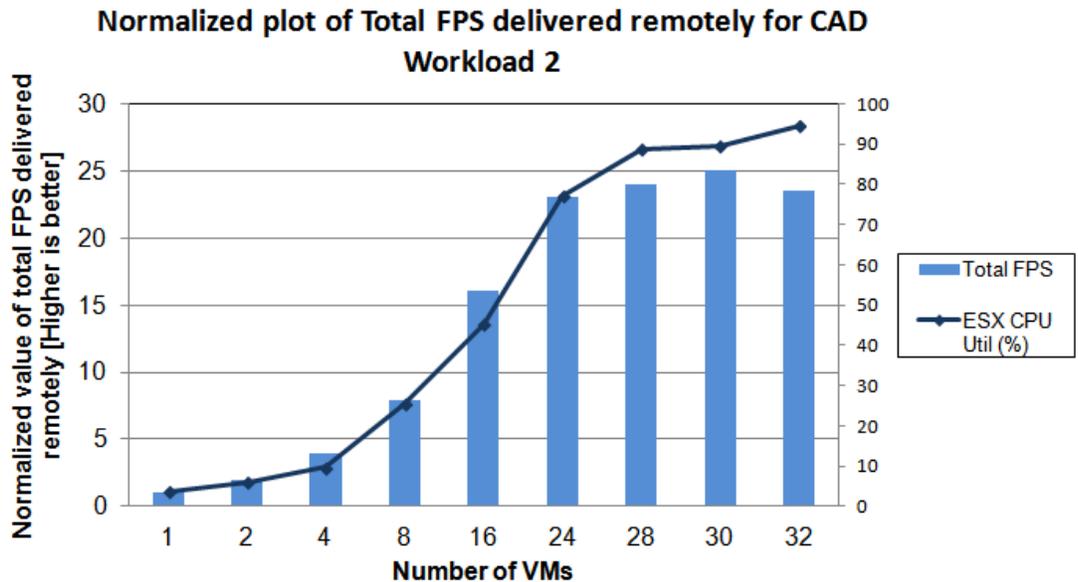


Figure 9. This bar chart presents the scalability of the vSGA solution as the load on the server is increased. The results are normalized and the frame-rate observed with just one VM running on the server is defined as the basis for comparison. The corresponding CPU utilization as measured using `esxtop` is shown by the line graph. GPU utilization was observed to be 85%.

In this workload, the Solid Edge viewer replaces the SolidWorks viewer and runs a single model: a 3-to-1 reducer (as shown at the bottom of Figure 2). As with the previous CAD workload, the simulated user’s interaction with the model is designed to mimic a real user’s potential usage pattern. Figure 9 illustrates the scalability of the vSGA and View solution; showing the aggregate remoted frame-rates (FPS) as the number of desktop VMs on the server is steadily increased. The results are fairly similar to that observed with CAD Workload 1; as the number of VMs is increased from 1 to 32 the aggregate remoted frame-rate increases by 25X. One difference to note between the results for CAD Workload 1 and CAD Workload 2 is the peak GPU load observed: with workload 1, it peaked around 37%, whereas with Workload 2, it reached as high as 85%. This difference results from the differing complexity of the 3D models, but highlights that the vSGA solution scales equally well when the GPU is close to fully utilized.

Figure 10 presents the same results as Figure 9, but presents the remoted frame-rate data on a per-VM basis. This view of the data highlights that the performance of the individual VMs sees little decrease as the number of VMs is scaled. This is further emphasized by the standard deviation results shown on the bar chart in Figure 10: not only does the average VM performance remain pretty constant as the number of VMs is increased, but performance is shown to be very consistent across all the VMs. This clearly highlights the ability of ESXi to fairly share the resources between the VMs, even under load, and ensure that the user experience is consistent across VMs.

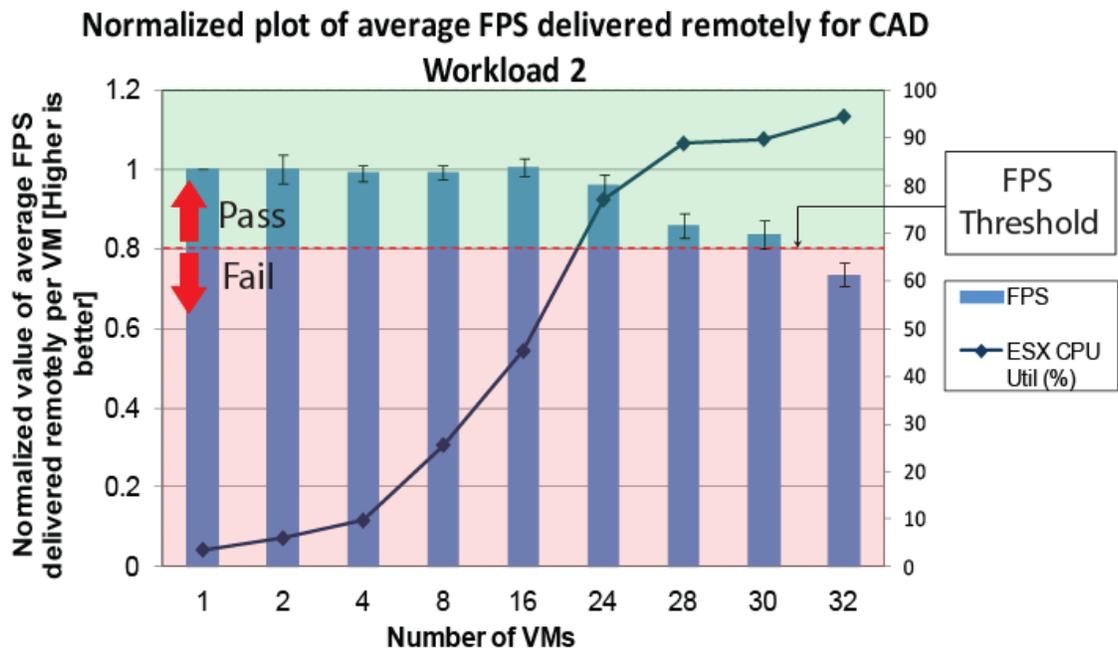
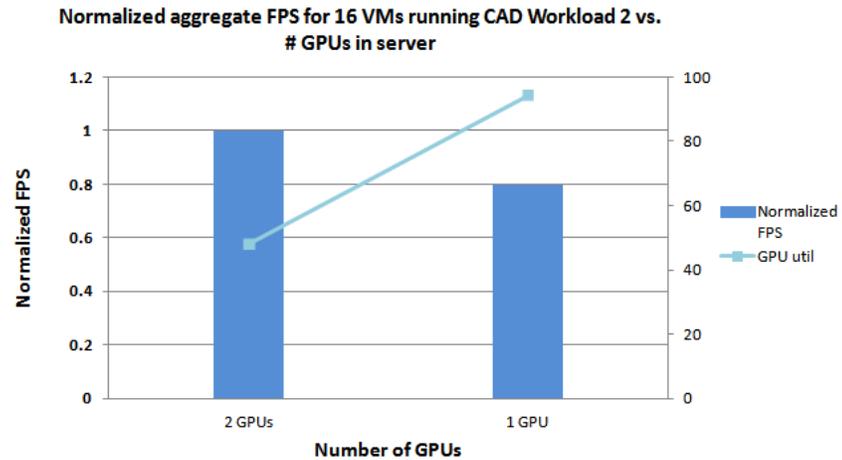


Figure 10. The bar chart presents the average per-VM remoted frame-rates (FPS) observed with the Solid Edge viewer as the number of VMs is increased. The results are normalized, with the frame-rate observed with a single VM used as the basis for comparison. In addition, the bar chart is also marked with the associated standard deviation. The corresponding CPU utilization as measured using `esxtop` is shown by the line graph.



**Figure 11.** The bar chart presents the remoted frame-rates observed with 16 desktop VMs when running with a single GPU and then with 2 GPUs. The results are normalized with the frame-rate observed with a single VM used as the basis for comparison. The corresponding GPU utilization is shown using the line chart.

Finally, as discussed previously, the server in the test-bed used to host the desktop VMs was equipped with two GPUs in order to support the GPU memory requirements of the light 3D workloads. However, for the CAD workloads, the available GPU memory is not a bottleneck. Accordingly, to demonstrate the performance of vSGA with just one GPU, the CAD workload 2 was also run with 16 VMs and a single GPU. The results are shown in [Figure 11](#) and clearly illustrate that, while the GPU utilization almost doubles with one GPU (as would be expected), the remoted frame-rate falls by only 20%, even though the single GPU is almost 100% utilized.

## Conclusion

This paper presents best practices and performance data for View 5.2's support for hardware accelerated 3D. The results clearly illustrate the ability of VMware's hardware-backed 3D support to scale efficiently to over 40 3D VMs per physical GPU and, even for performance-intensive 3D workloads, scale efficiently until GPU or CPU resources are exhausted. Specifically, it was demonstrated that using just a mid-range 2-socket x86 server configured with 2 GPUs that significant consolidation of 3D users can be achieved:

- Over 100 users running typical office applications on 3D desktops
- Up to 32 users running CAD applications such as Solid Works or Solid Edge

This clearly shows the benefits of GPU virtualization and the strength of VMware's 3D strategy.

Finally, the paper highlights that very little configuration is required to achieve optimal performance and scalability; vSphere 5.2 and View 5.2 work well out of the box to efficiently share the CPU and GPU resources between the VMs.

## References

- [1] VMware Horizon View Documentation. VMware Inc., 2013.  
[http://www.vmware.com/support/pubs/view\\_pubs.html](http://www.vmware.com/support/pubs/view_pubs.html)
- [2] B. Agrawal, R. Bidarkar, S. Satnur, T. M. Ismail, L. Spracklen, U. Kurkure, V. Makhija, "VMware View Planner: Measuring True Virtual Desktops Experience at Scale." *VMware Technical Journal*, Winter 2012.  
<http://labs.vmware.com/academic/publications/view-vmtj-winter2012>
- [3] VMware View Planner Community  
[http://communities.vmware.com/community/vmtn/servicessoftware/view\\_planner?view=discussions](http://communities.vmware.com/community/vmtn/servicessoftware/view_planner?view=discussions)
- [4] *VMware View Planner 2.1 Appliance and User Guide*. VMware, Inc.  
<http://communities.vmware.com/docs/DOC-15578>
- [5] L. Spracklen, B. Agrawal, R. Bidarkar, H. Sivaraman, "Comprehensive User Experience Monitoring," *VMware Technical Journal*, Spring 2012.  
<http://labs.vmware.com/academic/publications/spracklen-vmtj-spring2012>

## About the Authors

**Dr. Lawrence Spracklen** is a Senior Staff engineer at VMware. His current research focus is VDI deployments, specifically protocol optimization and multimedia quality analysis.

**Dr. Banit Agrawal** is a Sr. Performance Engineer at VMware. He has expertise and has filed several patents in the areas of VMware View, remote display protocols, VMware View Planner, and performance troubleshooting.

**Hari Sivaraman** is a Sr. Member of the Technical Staff at VMware. He works on 3D rendering performance and on CUDA support on ESX.

**Dr. Xing Fu** is a Member of Technical Staff in the Performance Engineering group at VMware. His work focuses on the performance of virtualization solutions.

**Rishi Bidarkar** is Sr. Manager in the performance team at VMware. He leads the VDI Performance and View Planner teams. He has filed several patents in the area of VDI performance and display benchmarking.

## Acknowledgements

The authors would like to extend their sincere thanks to Julie Brodeur for her comments, feedback, and improvements in the quality of this whitepaper. We also thank Warren Ponder, Pat Lee, and Rob Baesman for their comments on the early draft.

