VMware Horizon 6 RDSH Performance and Best Practices

Performance Study

TECHNICAL WHITE PAPER

Table of Contents

Executive Summary

VMware Horizon 6® Advanced and Enterprise versions deliver several types of remote access including remote desktops and remote applications, which are the focus of this paper. Remote applications run seamlessly on the user's client and are served using Microsoft Remote Desktop Services hosts (RDSH). These hosts can be consolidated on a single server and run as VMware vSphere® virtual machines. This paper shows the performance of remote sessions (desktops and seamless applications) and the performance of the RDSH virtual machines, paying particular attention to the sizing of the Horizon 6 environment. The paper also reveals the competitive results of leading remote display network protocols. Finally, the paper presents some of the best practices to tune the RDSH performance. All performance tests were run using the VMware View Planner 3.5 benchmark.

Introduction

VMware Horizon 6 delivers remote desktops and applications from a single platform. With Horizon 6, IT can deliver seamless applications and users can access their desktops or applications from anywhere on devices running Windows, Linux, MAC, iOS, and Android operating systems [1] [2]. This paper specifically covers RDS hosted applications and RDS desktops, but Horizon 6 also serves virtual desktops using VMware View technology and applications packaged with VMware ThinApp®.

There are several benefits to using RDS hosting in Horizon 6, including support for hundreds of end users from a single server [2]. RDSH has comparative lower storage demands because it caches various common data in main memory.

An advantage of using virtual machines as RDS hosts is the ability to consolidate multiple RDS hosts, run as virtual machines, onto a single server machine. Likewise, many end users can be consolidated onto a single virtual machine. This paper presents the performance results of sizing the number of virtual machines per server, the number of vCPUs in each virtual machine, and the number of users that can be served by each RDSH virtual machine.

Another advantage to using RDSH desktops and applications on Horizon 6 is the availability of PC over IP (PCoIP), a remote display network protocol that is optimized and adaptive under different network conditions. This paper later shows how PCoIP is competitive with remote display protocols used by Windows Server 2012 (RDP 8) and Citrix XenDesktop/XenApp (ICA).

Test Environment

The system under test included a vSphere server hosting several RDSH virtual machines, each running Windows 2012 R2 Server and each configured with 2 – 16 vCPUs and 16 – 96GB memory. The RDSH virtual machines communicated with the client virtual machines using the PCoIP remote display protocol. Each client virtual machine ran either a remote desktop or a set of remote applications. Each client virtual machine ran 32-bit Windows 7 with 1GB RAM.

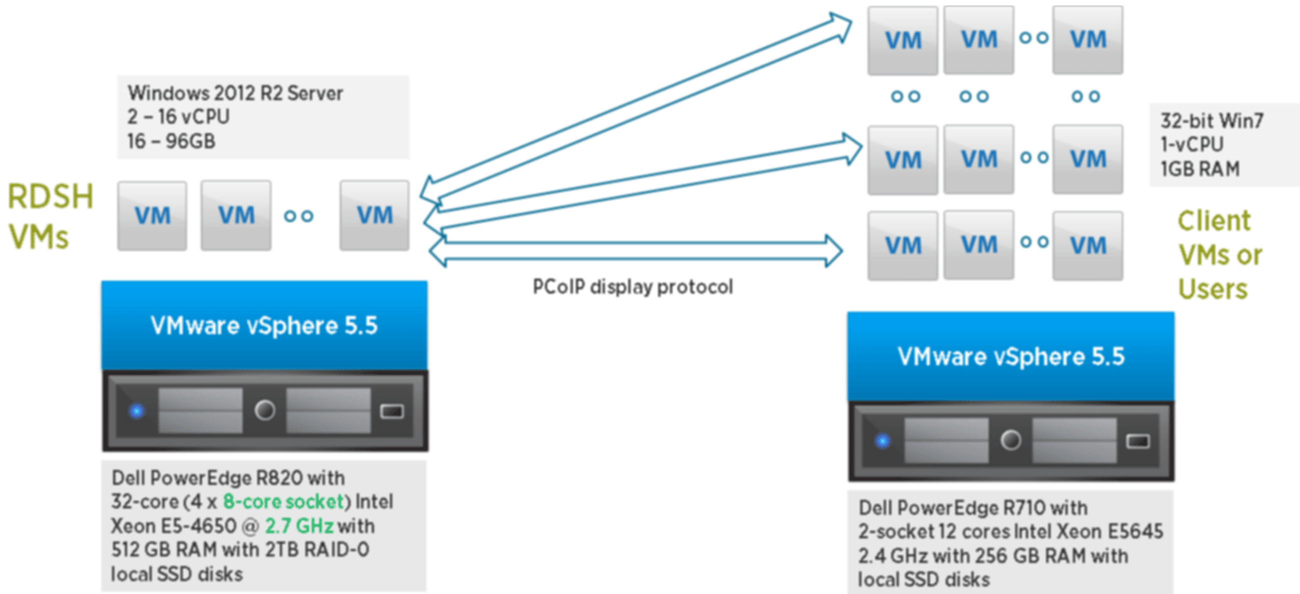

Figure 1. System under test (SUT)

The list of hardware and software used in this setup appears in Table 2 and Table 3 in the appendix.

Benchmark

The VMware View Planner 3.5 [3] benchmark was used to study the performance and sizing of Horizon 6 RDSH. View Planner is a workload generator and sizing tool built specifically for Horizon remote desktop services. By running its workload against a system under test, View Planner reports on platform characterization for CPU, memory, and storage; it evaluates the user experience (how many seconds it takes to perform certain remote desktop and remote application tasks); and it lends an understanding of scaling issues and identifies bottlenecks in performance. Figure 2 shows the architectural design of VMware View Planner 3.5.

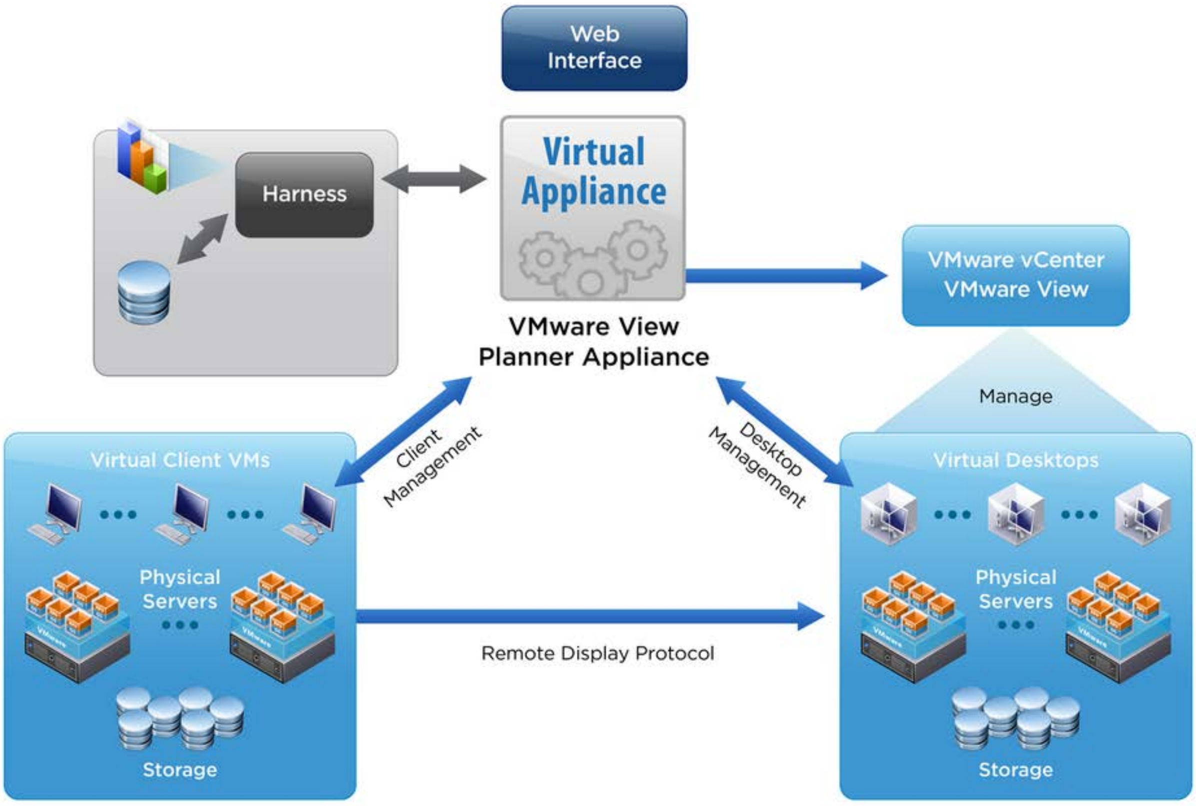

Figure 2. VMware View Planner

Workload

The workload in View Planner 3.5 leverages several different applications which are run as remote services (either in a remote desktop or as individual hosted applications). The user operations are mainly divided into three groups, as described in the *VMware View Planner Installation and User's Guide* [4] (see page 25): Group-A: Interactive operations, Group-B: I/O Operations, Group-C: Background load operations. Group-B operations were used for the test results in this paper because Group-B first reached the threshold criteria of 6 seconds [4]. Group-B operations include:

- Adobe Acrobat Reader 10 open PDF file
- Microsoft Internet Explorer 11 open a file served by Apache, open a file in a Web album
- Microsoft Office 2010:
  - Excel open and save file
  - PowerPoint open a file
  - Word open and save a file
- Microsoft Outlook 2010 open program and save an attachment
- Mozilla Firefox 3.6 open file

A think time of 5 seconds was used to simulate the time a user takes to pause during and between performing actions in these applications.

Performance Metrics

The View Planner benchmark was run for five iterations, with each iteration completing all the user operations in the chosen group. Iterations are divided into three phases: ramp-up (first iteration), steady-state (iterations between first and last), and ramp-down (last iteration). During each iteration, View Planner reports the latencies for each operation performed within each virtual machine. For scoring purposes, the first and last iterations were discarded and only the steady-state iterations were scored. Hence, the test was run for 3 iterations for all the experiments while scoring the middle iteration.

Quality of Service (QoS), determined for Group-A user operations as 1 second, and Group-B user operations as 6 seconds, is the 95^{th}-percentile latency of all the operations in a group.

Sizing for Best Performance

The goal for sizing is to consolidate as many sessions on a particular infrastructure with a focus on quality. In order to achieve this goal, three aspects of performance were examined:

1. How many users/sessions per physical core?
   – Desktop session
   – Seamless application session
2. How many vCPUs should be used for an RDSH virtual machine?
3. How many RDSH virtual machines are needed?

The answers to questions 2 and 3 can be found by testing various sizes of virtual CPU over-commitment when paired with different numbers of virtual machines, which you will see in the following results. Table 1 shows an example of the combinations that were tested. In the 1:1 ratio, no vCPUs are over-committed. For example, 8 VMs each with 4 vCPUs uses all 32 cores on the hardware. In the 1 ½ : 1 ratio of over-commitment, 8 VMs each with 6 vCPUs uses 48 "cores"—all 32 cores are used with 24 cores being hyper-threaded. In the 2:1 ratio, 8 VMs each with 8 vCPUs fully uses all 32 cores each split into 2 hyper-threads.

vSphere host has 32 cores (4 sockets with 8 cores each). Hyper-Threading yields 64 "cores."		
1:1 ratio	1 ½:1 ratio over-commitment	2:1 ratio over-commitment
2 VMs x 16 vCPUs 4 VMs x 8 vCPUs 8 VMs x 4 vCPUs } 32 cores	3 VMs x 16 vCPUs 4 VMs x 12 vCPUs 6 VMs x 8 vCPUs 8 VMs x 6 vCPUs 12 VMs x 4 vCPUs } 48 "cores"	4 VMs x 16 vCPUs 8 VMs x 8 vCPUs 16 VMs x 4 vCPUs } 64 "cores"

Table 1. Possible combinations of vCPUs per VM to fully utilize available cores

Right-Size Users/Sessions per Physical Core

The first consideration is how many users (sessions) can be consolidated on each core. Figure 3 shows the optimal number of users (sessions) per physical core that can be served on the RDSH virtual machines. This gives the megahertz per user, which appears in the "Best Practices for Horizon 6 Performance" section later in this document.

Desktop Session

The x-axis shows the number of users per core that were supported in the workload run and this is correlated against the y-axis, which shows the response time in seconds. Thorough testing in VMware's Performance Engineering team has shown that 6 seconds is the point at which users can tolerate delay in an action made in the remote desktop (or remote application). Nine users per core is the point that intersects with a 6 second response time.

It is later shown that the number of users per core for seamless applications is approximately 8. See Figure 5.

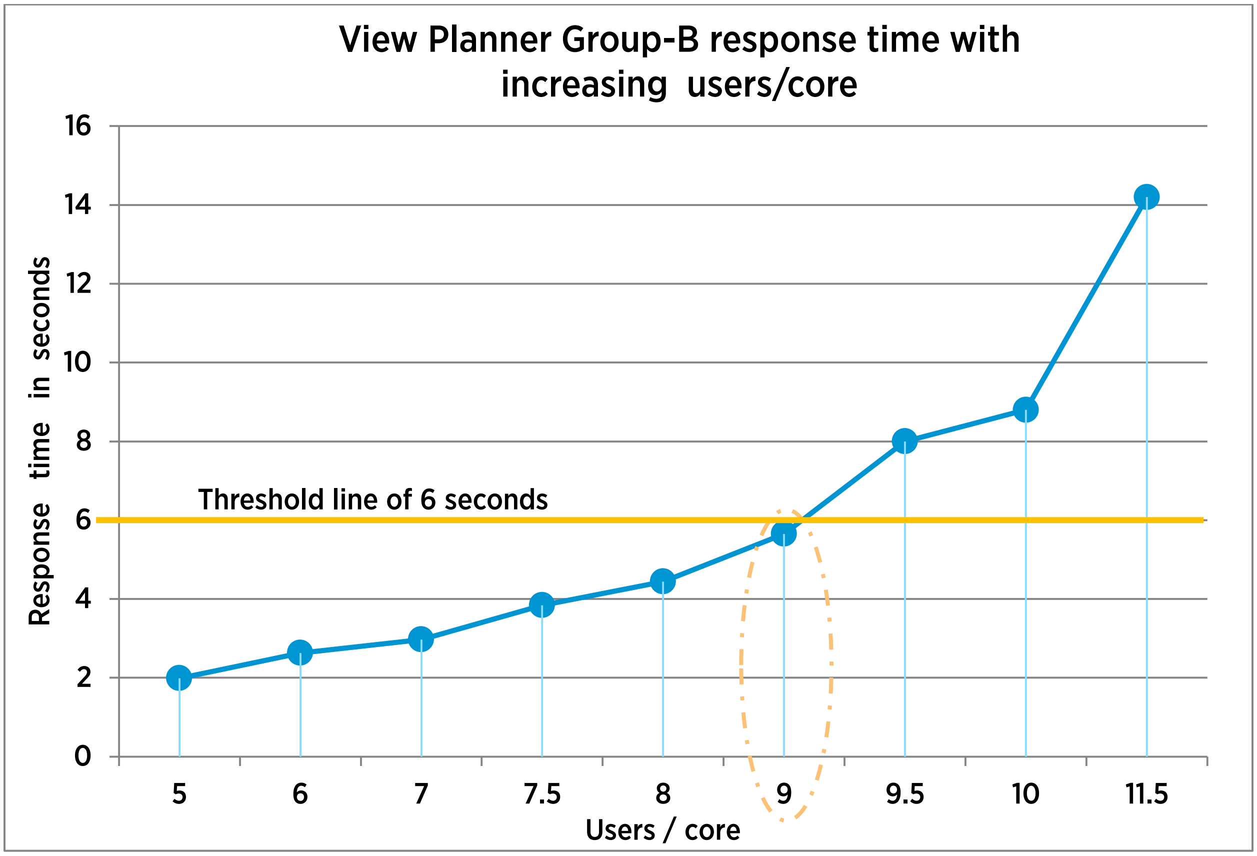

Figure 3. View Planner Group-B response time with increasing users per core

Ideally, this number of users per core is about 90% so that the users are almost fully utilizing, but not overusing each core to best take advantage of the compute power. Figure 4 shows this to be true—90% of each core is used when 9 users are served content from the RDSH virtual machine. Starting with 9.5 users, the CPU utilized is approximately 95%, which is too close to 100%.

To find the number of users that can be served by one RDSH virtual machine with acceptable performance, multiply the number of cores by the number of users per core to get 32 cores x 9 sessions/core = 288 sessions per host. Later, we present how to achieve this density using an optimal number of RDSH virtual machines.

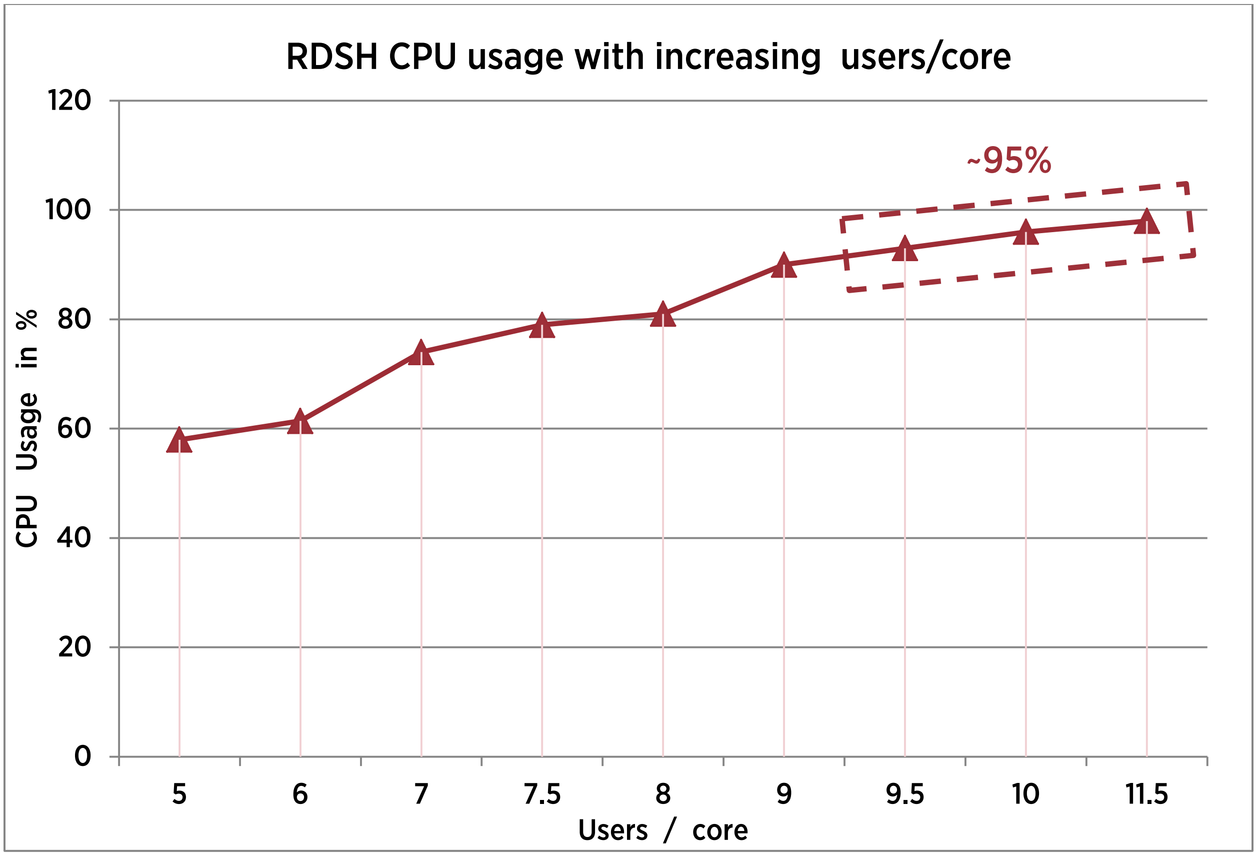

Figure 4. RDSH CPU usage with increasing users per core

Seamless Application Session

Similar to the remote desktop session, all the applications were run as remote seamless applications. The application sessions were increased from 40 to 68 on an 8-vCPU RDSH virtual machine. Figure 5 shows that density is seen between 8 and 8.5 users per core: (~8.25). This is slightly lower than in the desktop sessions, which was 9 users per core. While video was selected as one of the applications to compare the results with desktop session scaling, it is very likely that in typical use-case scenarios, video is less likely to be used as a remote application.

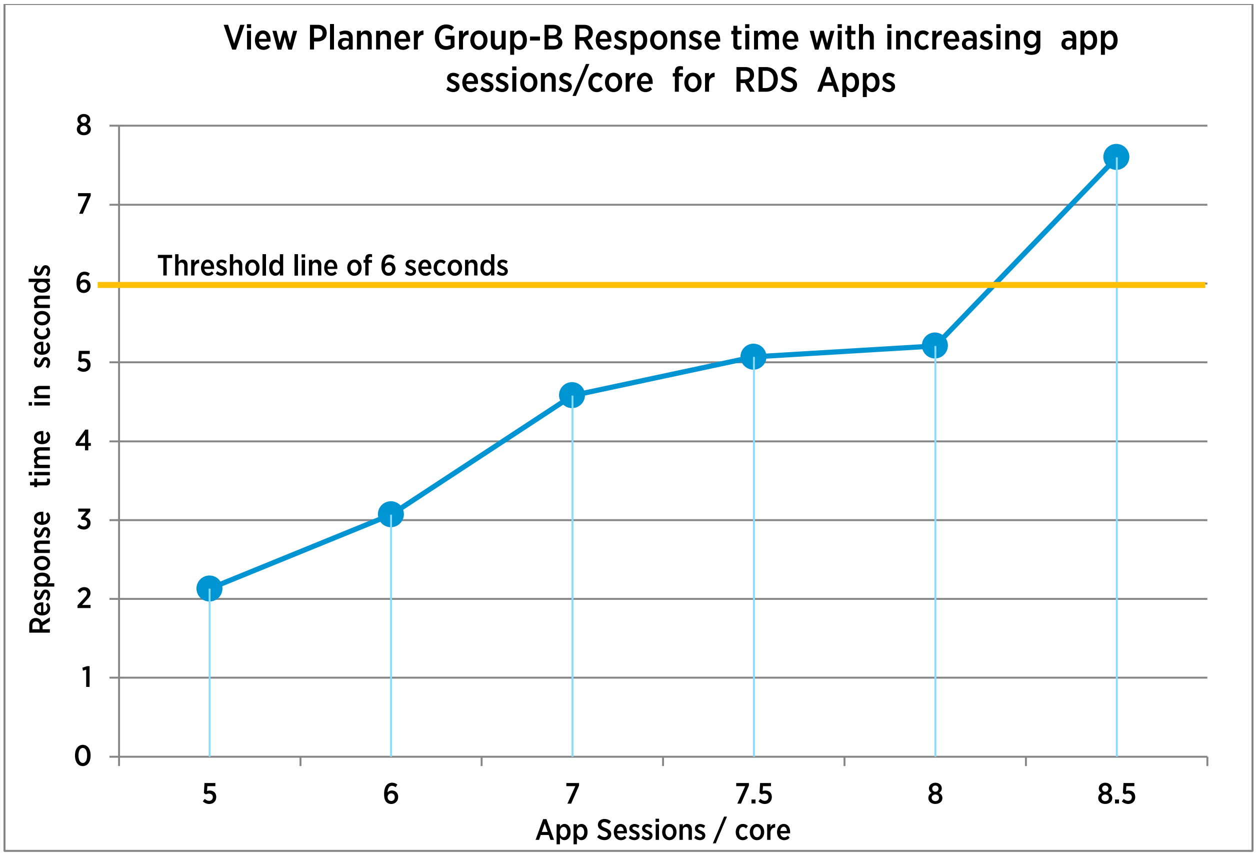

Figure 5. View Planner Group-B response time with increasing application sessions per core for RDS applications

Right-Size vCPU for an RDSH VM

The Group-B response times are used again to determine the number of virtual CPUs needed for best performance for an RDSH virtual machine. The response times were compared against the 6-second metric of acceptable user experience—anything lower than 6 seconds is good, and lower is better. Anything above 6 seconds is too high and therefore unusable. Figure 6 shows that 6- and 8-vCPU virtual machines provide better performance than other numbers of vCPU.

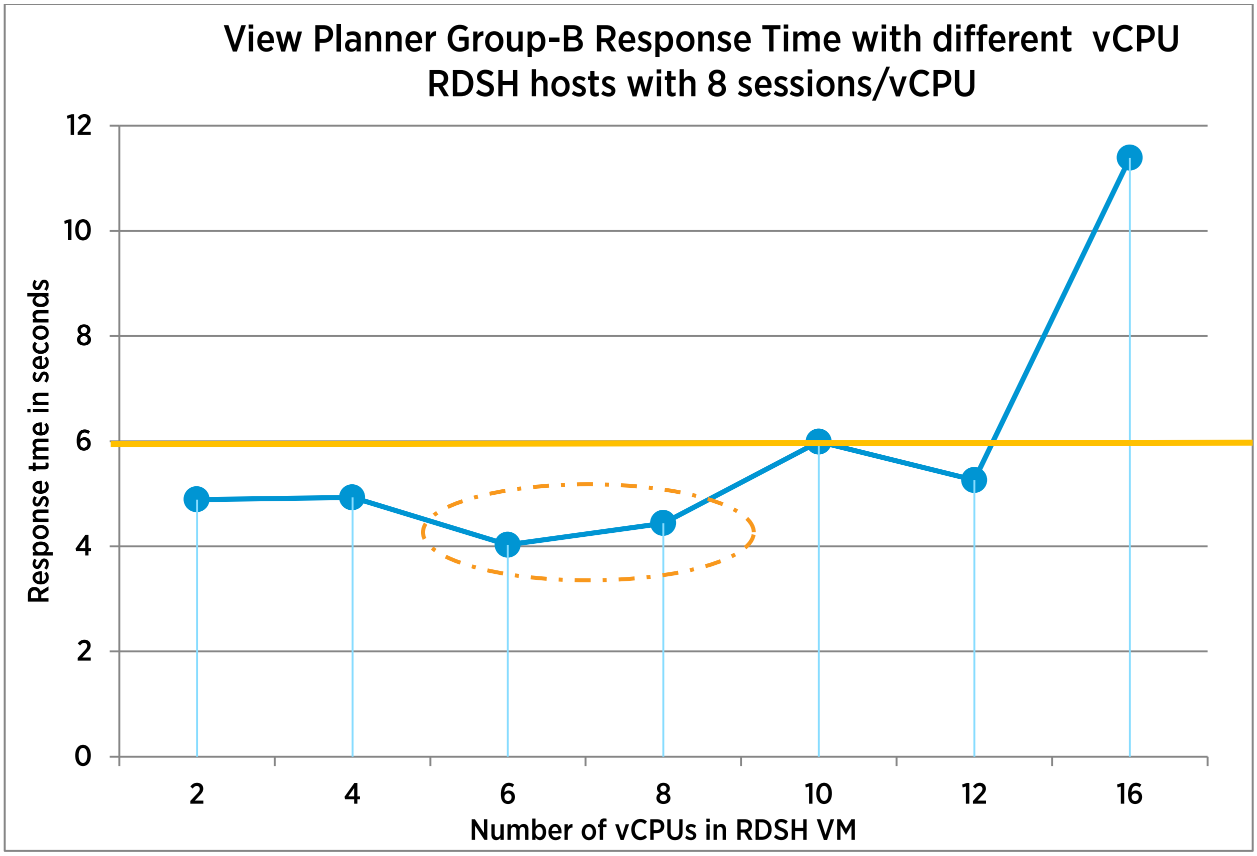

Figure 6. View Planner Group-B response time with different vCPU RDSH VMs with 8 sessions / vCPU

Right-Size Number of RDSH Virtual Machines

The final metric to determine is how many instances of each RDSH virtual machine should be run for optimal performance. This is achieved by taking the response time for 288 desktop sessions run for 14 different vCPU configurations, which is shown in Figure 7.

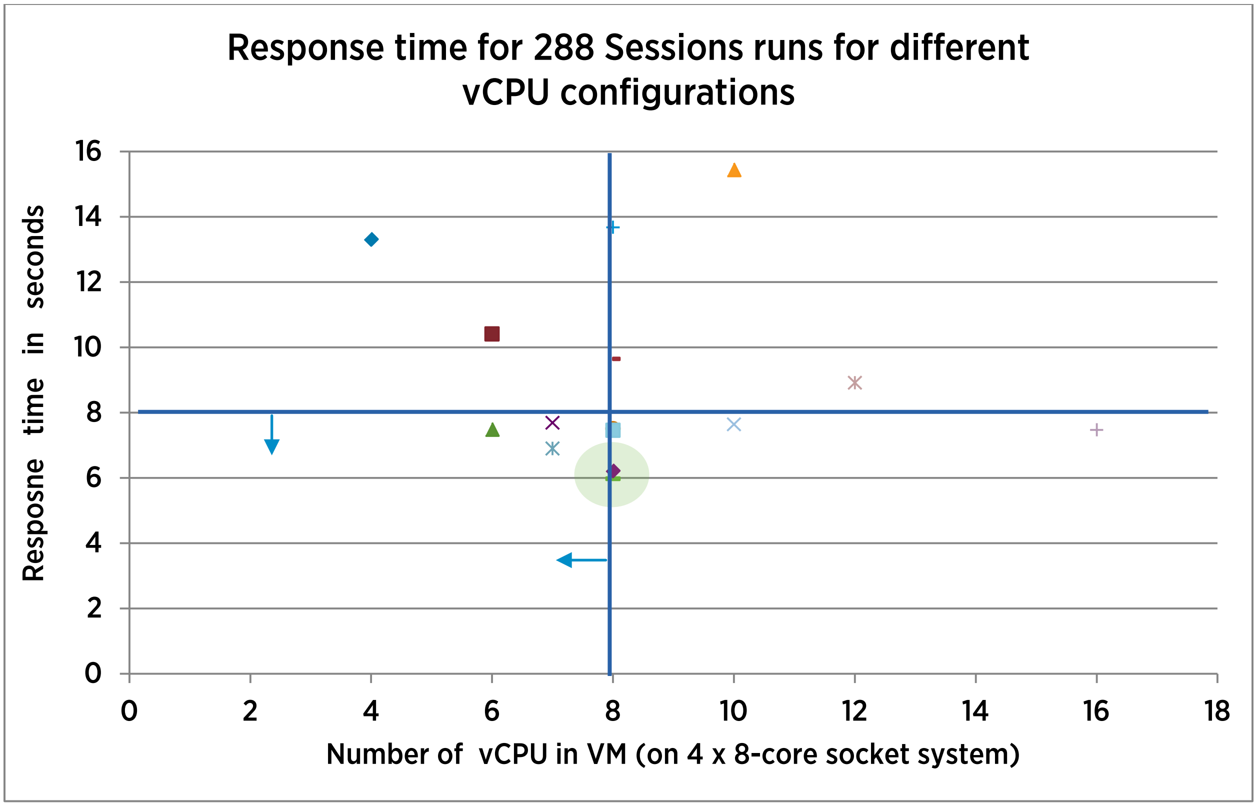

Figure 7, Response time for 288 sessions run for different vCPU configurations

One of the points inside the green oval represents 8 VMs x 8 vCPUs and is shown at the response time of 6 seconds (the highest amount of time before user experience noticeably degrades). So, 8 virtual machines, each with 8 vCPUs give the best performance. Some other configurations of 7-vCPU and 8-vCPU performance were also close to the best case.

According to Table 1, the configuration of 8 VMs x 8 vCPUs is a 2:1 over-commitment ratio. Figure 8 shows the over-commitment ratio test for the different VM x vCPU configurations, and the result does yield a 2:1 over-commitment ratio, as seen in the green oval.

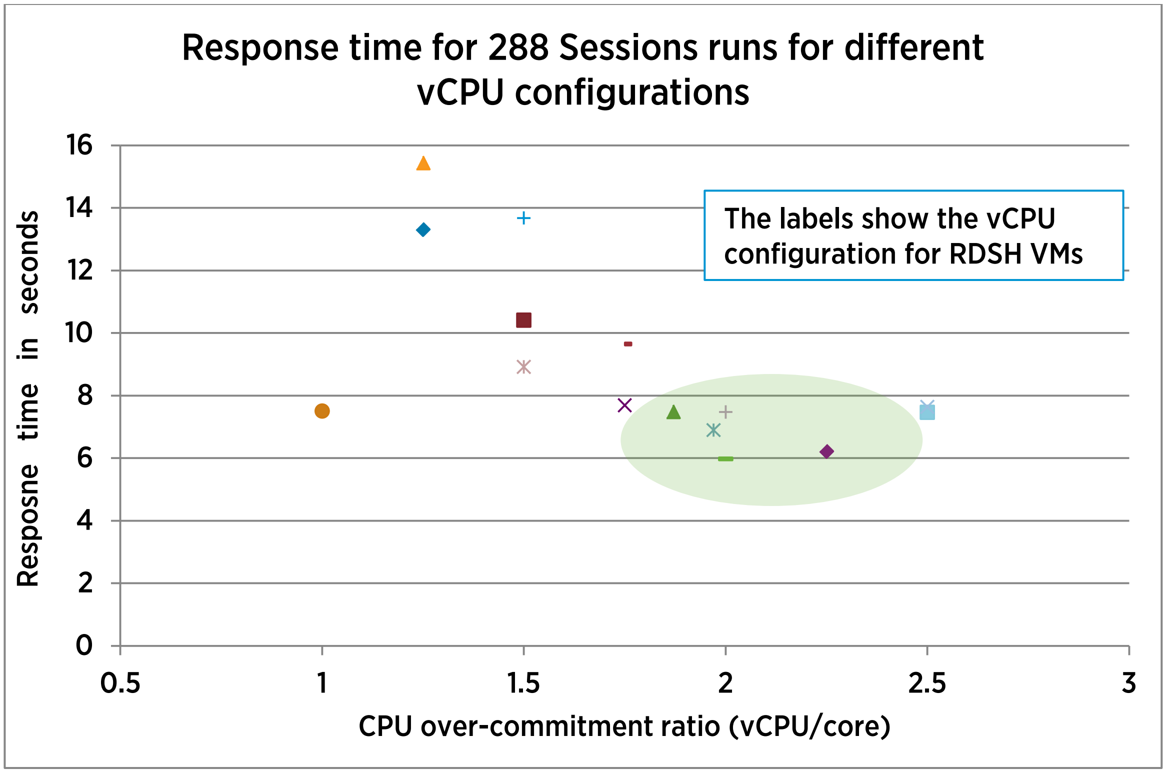

Figure 8. Response times for 288 sessions run for different vCPU configurations (vCPU/core)

So, the 2:1 ratio of over-commitment provided the best performance.

Custom Application Performance with Visio

Instead of the Group-A, Group-B, and Group-C operations usually run as the View Planner workload, custom applications can be created to run within the View Planner environment. (See "Using Custom Applications in View Planner" on page 121 of the View Planner installation and user guide [4].)

For this experiment, Visio operations were configured to run as the workload (as shown in Figure 9). This application represents a popular software tool used in an enterprise environment.

This test used the same setup as the previous tests, except for the RDSH virtual machine, which was a 4-vCPU Windows 2008 R2 server with 10GB virtual RAM.

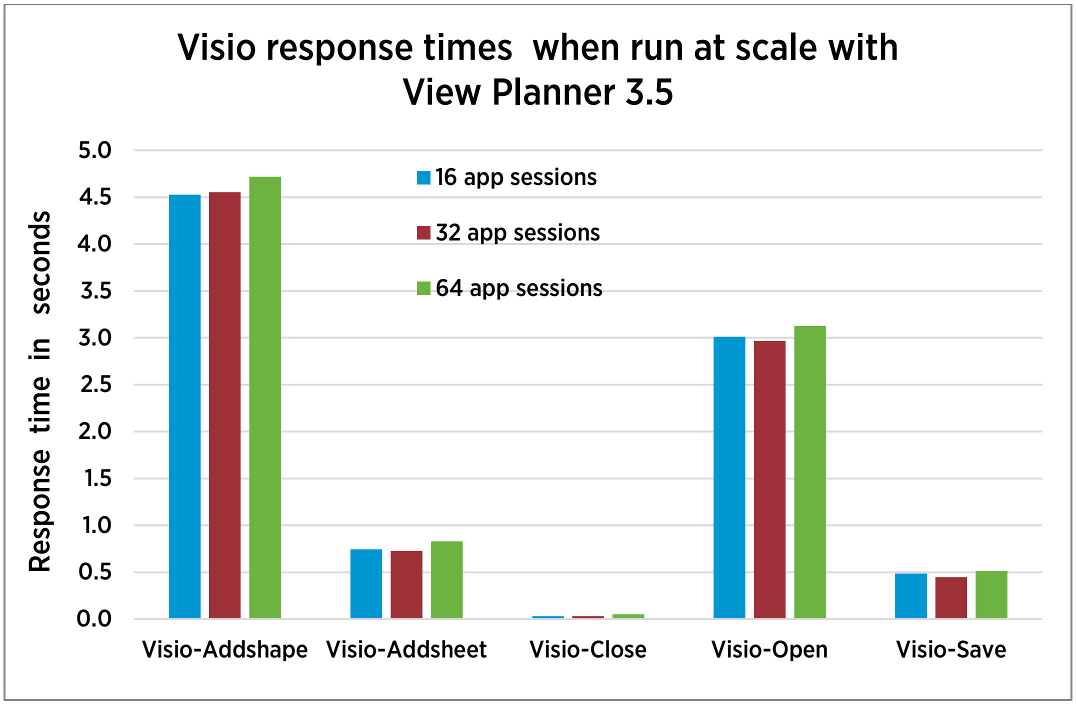

Figure 9. Visio response times with View Planner 3.5 when scaling out application sessions

This test shows the performance of some common activities taken place in Visio remote application sessions. Figure 9 shows how scaling the application sessions from 16, to 32, to 64, does not affect performance very much. There is only about a 5% increase in response time when moving up from 32 sessions to 64 sessions. This shows that a typical Horizon 6 remote application like Visio scales well.

Remote Protocol Performance

One of the important aspects of using remote applications is network performance. Horizon 6 uses the remote viewing network protocol PCoIP. This section compares PCoIP performance (VMware Horizon 6) to RDP 8 (Microsoft Windows 2012 R2 server) and ICA (Citrix XenApp/XenDesktop 7.1).

For this performance study, 60 remote seamless application sessions were run on an 8-vCPU Windows 2012 R2 RDS virtual machine. All of the View Planner 3.5 applications were run, except video, with a 5-second think time in 3 iterations. The resolution was 1152x864, and the color depth was 32-bit.

The study looks at two primary factors that contribute to the speed of an application viewed remotely: CPU usage and bandwidth usage.

CPU Usage

Figure 10 shows the percentage of the guest (client) CPU used when performing the various workload tasks.

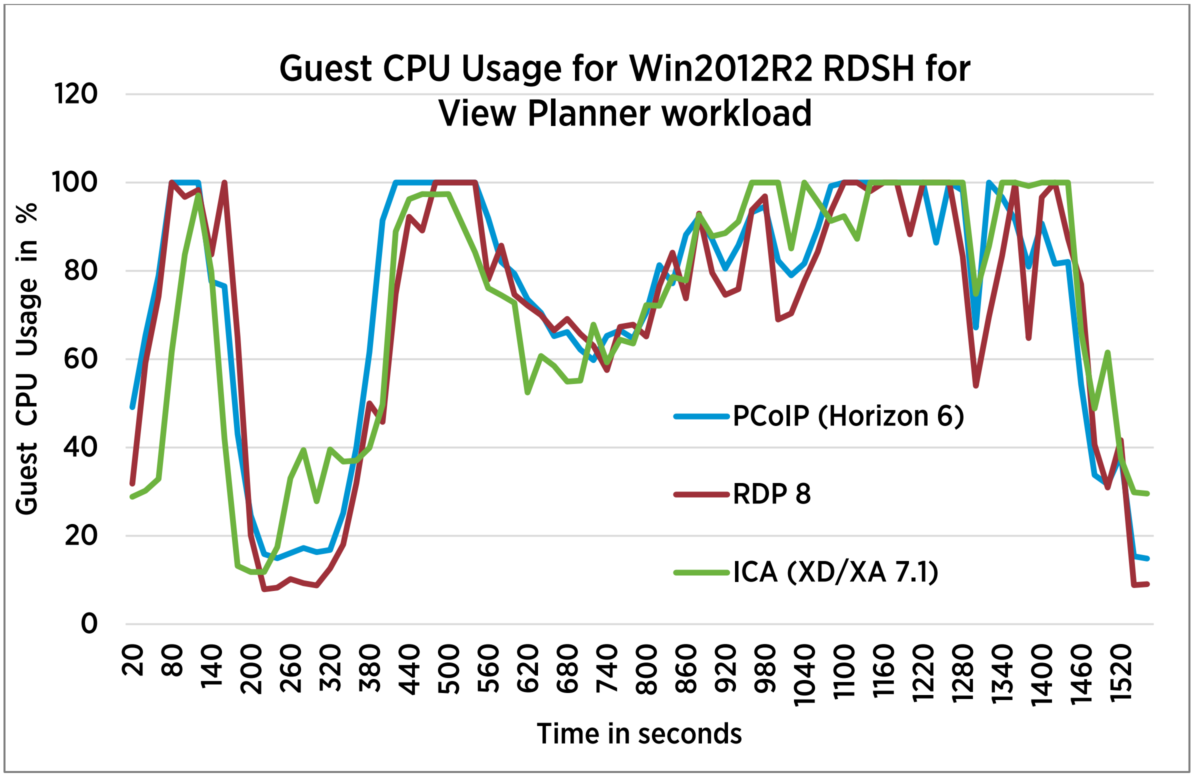

Figure 10. CPU usage in the RDH virtual machine for three competing remote display protocols

The results for the average CPU usage, when calculated, come out to PCoIP at 71.6%, RDP at 68%, and ICA at 71.2%.

This shows that RDP came in with the lowest CPU usage, but only 3% better than PCoIP or ICA. All the protocols use about the same amount of compute power on the guest operating system.

Bandwidth Usage

The average bandwidth usage in kilobits per second (Kbps) are calculated for the workload runs. The results show that PCoIP averaged 44.7 Kbps per session, RDP: 50.7 Kbps per session, and ICA: 48.4 Kbps per session. This reveals that PCoIP bandwidth performance is about 10% better than RDP and ICA bandwidth performance.

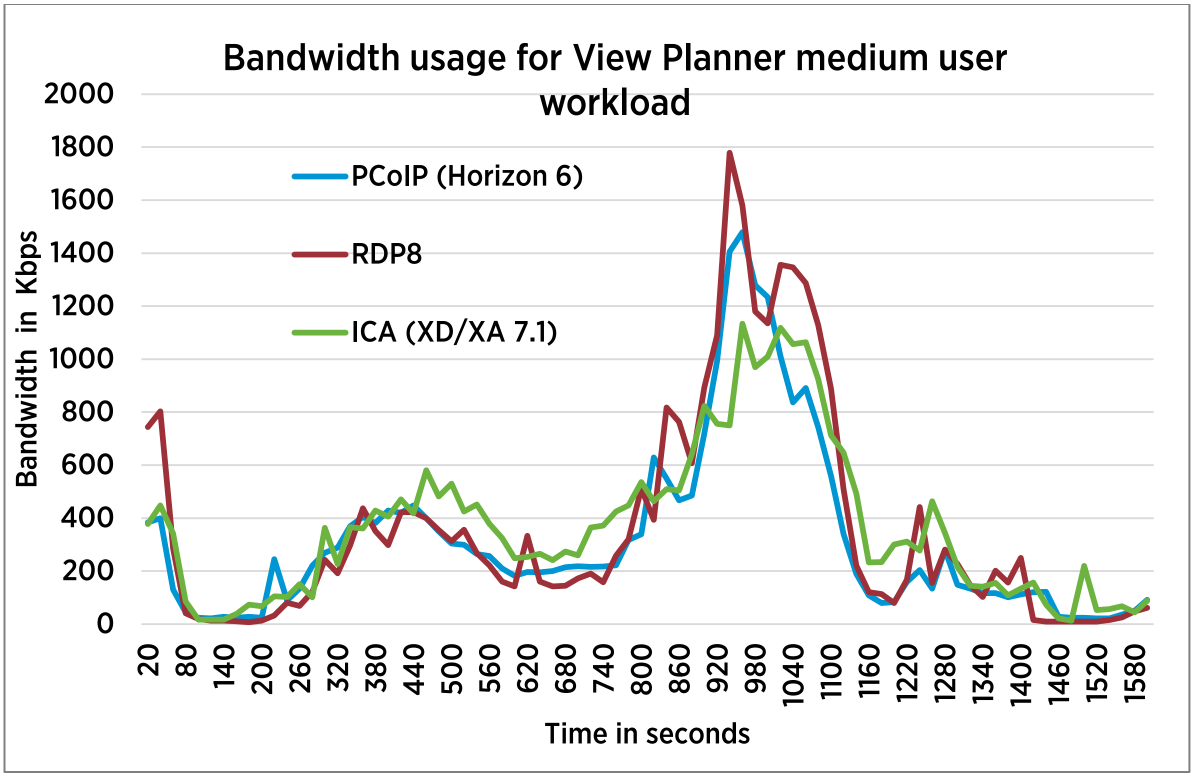

Figure 11. Bandwidth usage for View Planner medium user workload for three competing remote display protocols

Best Practices for Horizon 6 Performance

RDSH Tuning and Optimization Strategies

This section collects all the RDSH sizing best practices obtained through extensive experiments and presents a brief summary below. Please tune the RDSH virtual machines using the available tool (RDSH optimizer, below), best practices for RDSH server VM, and the guide for guest VM best practices as described in *Horizon 6 Performance & Best Practices* whitepaper.

Best Practices for RDSH VM Sizing

As discussed in the RDSH sizing section earlier, several experiments were done with different configurations (y instances of x-vCPU RDSH server VMs on one host) to find the optimal values of x and y for a particular hardware. Based on the detailed analysis, it was found that scaling out the RDSH server provided the best performance; that is, using multiple RDSH server VMs instead of one big RDSH server virtual machine. Furthermore, it was discovered that the vCPU for the RDSH virtual machine should be less than or equal to the number of cores in the socket so that the virtual machine fits within the same NUMA node. For finding the number of instances, it was found that a 2:1 CPU over-commitment worked better (64 vCPU on 32 cores or 64 hyper-threaded cores in the experiments).

If the number of cores in the socket is less and the requirement is to use large vCPU RDSH VM such that it doesn't fit in the same NUMA node, you can try the preferHT=TRUE option in the VMX file of the RDSH server VM. This option forces the vCPUs to the fewest number of physical sockets by giving preference to hyper-threaded cores for scheduling purposes and increases the probability of the virtual machine staying in the same socket.

You can also set the CPU affinity where each vCPU is mapped to one hyper-threaded core. Since the recommendation is to do a

1:1 overcommit with hyper-threaded cores, setting the affinity in the VMX file can further improve the performance. Below is an example of the VMX configuration option, where an 8-vCPU virtual machine is mapped to cores 0-7. Similarly, the other 7 instances of 8-vCPU virtual machines can be mapped to 8-63 cores.

```
sched.cpu.affinity = "0,1,2,3,4,5,6,7"
```

Best Practices for RDSH Session Sizing

Based on detailed scaling results presented earlier, 9 desktop sessions per core were seen for a medium user (that of a typical office worker) on a 2.7 GHz processor, hence it comes down to about 300 MHz per user. For seamless applications, this will be about 350 MHz. So, setting the CPU MHz anywhere between 300 to 500 MHz per session will provide good user experience. The following summarizes the requirement for a typical session for a medium user:

- CPU: 300 to 500 MHz per session
- Memory: 400-500MB per session for 9 applications; it might be different for your applications, so please select appropriately based on the working set of your applications
- Disk space: 200-300MB per user in OS disk for profiles, temporary files, etc.
- Network: 50Kbps per session on average and please plan for peak bandwidth as well

RDSH server VM Optimization

To optimize the RDSH server virtual machine image, several registry and group-policy changes can be done. To do these optimizations, an RDSH optimizer tool is written and made available on the VMware Community at "Horizon 6 RDSH Optimization Tool." [5] If you want to see the registry changes, you run process monitor (procmon) and set the filter on WriteRegistry (regsetvalue, etc.) to see what optimizations are being applied.

Conclusion

This paper has provided sizing information for the Horizon 6 environment to achieve the best consolidation numbers for a virtual deployment of Horizon 6. Additional testing showed that a remote application, Visio, scaled out with good performance. Also, a test of three popular remote viewing protocols showed PCoIP, the protocol used in Horizon 6, to be competitive with other leading remote viewing protocols from Windows and Citrix.

Appendix A

Hardware and software for the system under test are shown in Table 2 and Table 3.

Hardware

Server	**Dell PowerEdge R820**
CPU	32-core (4 x 8-core socket) Intel Xeon E5-4650 @ 2.7GHz
RAM	512GB
Disk	2TB RAID-0 local SSD
Client	**Dell PowerEdge R710**
CPU	24-core (2 socket x 12 cores) Intel Xeon E5645 2.4 GHz
RAM	256GB
Disk	900GB RAID-0 local SSD

Table 2. Hardware for systems under test

Software

Hypervisor	VMware vSphere 5.5
VMs on server	Windows 2012 R2 2 - 16 virtual CPU 16 – 96GB RAM
VMs on client	92 VMs each connecting to 3 sessions (288 total sessions) Windows 7, 32-bit 1 vCPU 1GB RAM
Remote applications	Acrobat, Firefox, IE, Office (Excel, Word, PowerPoint), Outlook

Table 3. Software for systems under test

References

[1] VMware, Inc. (2014) VMware Horizon 6 Datasheet.
http://www.vmware.com/files/pdf/products/horizon-view/VMware-Horizon-View-Datasheet.pdf

[2] VMware, Inc. Cindy Heyer. (2014) Application-Delivery Options in VMware Horizon 6.0.
http://www.vmware.com/files/pdf/techpaper/vmware-horizon-view-workspace-application-delivery-options.pdf

[3] VMware, Inc. (2014, October) Download VMware View Planner 3.5.
https://my.vmware.com/web/vmware/details?productId=320&downloadGroup=VIEW-PLAN-300

[4] VMware, Inc. (2014, October) VMware View Planner Installation and User's Guide Version 3.5.
https://my.vmware.com/web/vmware/details?downloadGroup=VIEW-PLAN-300&productId=320

[5] VMware, Inc. Banit Agrawal. (2014, December) Horizon 6 RDSH Optimization Tool.
https://communities.vmware.com/docs/DOC-28524

[6] VMware, Inc. (2013, May) VMware Horizon View 5.2 Performance and Best Practices.
http://www.vmware.com/files/pdf/view/vmware-horizon-view-best-practices-performance-study.pdf

Authors

Dr. Banit Agrawal is a staff engineer at VMware. He has expertise and filed several patents in the area of VMware View, remote display protocols, VMware View Planner, and performance troubleshooting.

Julie Brodeur is a senior technical writer in the Performance Engineering group at VMware.

Acknowledgements

The authors would like to thank the View Planner team and Warren Ponder for their contributions to this paper.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com