

Quantifying the Impact of Virtual GPUs

Benchmarking the User Experience in VMware Virtualized Environments with Dell EMC and NVIDIA nVector

August 2019

H17917

White Paper

Abstract

This white paper describes a new alternative approach to sizing VDI solutions with an emphasis on designing the VDI experience to meet the needs of the modern workforce. An introduction to initial findings by Dell EMC using NVIDIA nVector, this document also explains NVIDIA methodology, the benefits of measuring each element, and sizing considerations.

Dell EMC Solutions

Copyright

The information in this publication is provided as is. Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2019 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Intel, the Intel logo, the Intel Inside logo and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries. Other trademarks may be trademarks of their respective owners. Published in the USA August 2019 White Paper H17917

Dell Inc. believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

Contents

Executive summary.....4

Business challenge.....5

Going beyond the numbers6

Methodology: Factors for measuring UX.....7

Modeling knowledge worker behavior at scale7

Simulating many users and many behaviors9

Automating and running the workload at scale11

Test environment11

Results overview13

Test results: Measuring the end user experience.....13

Don't let the latest web standards like WebGL slow down your users19

When graphics intensive applications are used, GPUs increase
server density by up to 60%19

Conclusion.....22

Executive summary

Employees will not accept a compromised user experience (UX). Existing benchmarking tools provide limited insights on the quality of the user experience since they measure the response time on the virtual desktop and don't consider the responsiveness at the end user's access device.

Benchmarking the user experience at scale has become increasingly challenging considering changes to the way employees work today - including the increased use of graphics intensive applications, video and multimedia, and the rising use of multiple, higher resolution monitors. Application vendors are including more robust 3D features in their applications and using latest web standards like WebGL, all of which are very taxing to the CPU. Additional uncertainty lies in each release of new operating systems and updated versions of common office productivity software. Even newer versions of remoting clients and protocols may trade increased CPU utilization for improved features like image quality. IT is challenged to ensure they build infrastructure that can support the needs of their users today and in the future.

To address the need for better insights into the actual end user experience, NVIDIA has developed a benchmarking tool based on a methodology that measures key aspects of the user experience, including end user latency, frame rate, image quality, and server utilization. This enables IT to size the VDI infrastructure based on utilization thresholds that are relevant to the environment.

Dell EMC is now incorporating the NVIDIA UX profiling application into our test methodology to provide more insights into experience (see article [Creating Clarity through Multiple Lenses](#)). The results from testing with the updated workload type and client-side measurements show that a GPU-enabled VDI environment significantly improves the user experience as well as the server density when compared to a CPU-only VDI environment:

- 15% improved end user latency enabling faster response times
- 25% improved frame rate for better fluidity
- Uncompromised image quality

The testing also shows that the NVIDIA T4 GPU with NVIDIA GRID software fully supports the needs of today's demanding digital workplace. It provides enough GPU headroom, encoder bandwidth, and frame buffer to support today's modern productivity applications and alternative work environments.

The Dell EMC PowerEdge R740xd rack server with NVIDIA virtual GPU technology is optimized for VDI. The PowerEdge R740xd was designed to accelerate application performance, leveraging up to six NVIDIA T4 graphics accelerator cards and storage scalability. The 2-socket, 2U platform has the optimal balance of compute resources to power the most demanding environments.

This white paper details how Dell EMC used NVIDIA's benchmarking tool to measure the quality of the end-user experience and quantifies the benefits of NVIDIA GRID on VMware virtualized environments running on Dell PowerEdge R740xd servers.

Business challenge

In recent years everyday business applications like Microsoft Office, Google Chrome, Skype, and PDF readers have evolved to offer graphics-rich features, providing users with better interactivity and improved overall performance. For example, the wide adoption of standards such as WebGL are making it easier to render interactive 2D and 3D graphics within any compatible web browser. To further enhance their productivity, users often have multiple monitors and are upgrading to 4K resolution monitors, which have become more mainstream as they have become more affordable. Between 2019 to 2024 the market for 4K displays is expected to grow at a compound annual growth rate (CAGR) of 23%.¹ Due to these trends, users are now consuming more CPU resources and active memory than ever before.

In tandem with these upgrades, the graphics requirements of operating systems have also increased. Windows 10 is a good case in point. It has the highest graphics requirements of any operating system to date, with a 50% increase in CPU requirements compared to Windows 7.² Offloading graphics workloads currently performed by CPUs and moving them to virtual GPUs using NVIDIA GRID frees up CPU resources on server hardware and enables a local-desktop-like experience to remote users.

The business challenge is, therefore, to design a virtual desktop environment that users love.

NVIDIA's performance engineering teams have developed a methodology and set of benchmarking tools that simulates, at scale, the end user workflow. It measures the metrics described in Figure 1, which Dell EMC has leveraged in testing the PowerEdge R740xd VDI solution. This benchmarking tool is called NVIDIA nVector.

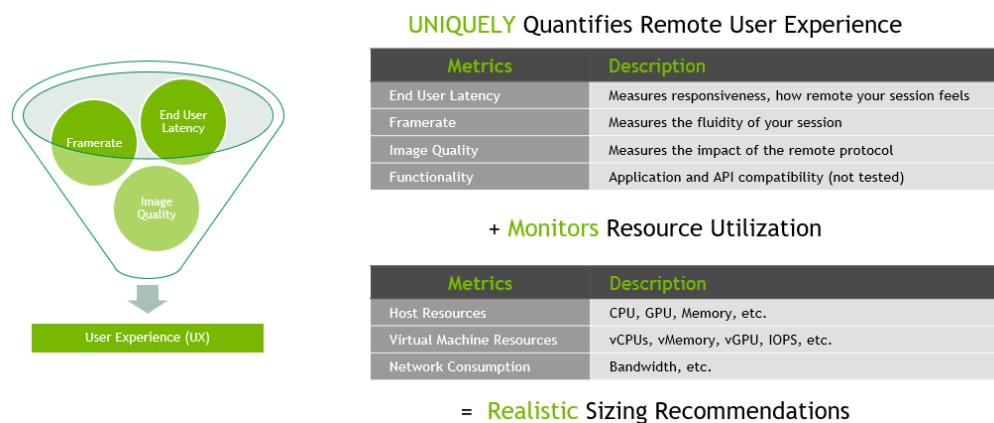


Figure 1. Quantifying user experience and scale

¹ Mordor Intelligence. March 2018. ["4K Display Resolution Market - Segmented by Product \(Monitor, Smart TV, Smart Phone\), Vertical \(Media & Entertainment, Retail, Consumer Electronics\), and Region - Growth, Trends and Forecasts \(2018 - 2023\)"](#)

² Percent of time consuming GPU (DirectX or OpenGL), comparing Windows 7 to Windows 10 builds in 2018 (builds 1803 and 1709)

Dell EMC measured the quality of the user experience across three specific metrics (Figure 1):

- **End-User Latency** — Measures the responsiveness of a remote desktop or application. It measures the lag that an end user experiences when the user interacts with a remote desktop or application.
- **Framerate** — Measures the perceived “smoothness” of the UX. It actually measures the rate at which frames are delivered on the screen of the endpoint device.
- **Image Quality** — Measures the impact of remoting on image quality. It quantifies the impact using the Structural Similarity Index (SSIM) between a frame that is rendered on the target virtual machine and the frame that is visible to the end user.

The metrics that quantify user experience, when combined with resource utilization information from the host, virtual machine and network, provide a holistic picture that enables IT to better size their VDI environments for scale, while continuing to ensure a great user experience.

This paper outlines how the test environment was set up to simulate the day-to-day tasks that knowledge workers perform on standard applications. Then, it explains how end user experience was measured, as well as the trade-offs between user experience and server density. This data will help you design a smarter VDI infrastructure with the best possible user experience while optimizing user density.

Going beyond the numbers

While NVIDIA nVector can quantify user experience across metrics of end user latency, frame rate, and image quality, one metric that is not quantified is application functionality. Some features within applications just won't work without GPUs, or require GPUs for acceleration. One example is the 3D tilt view in Google Maps, which is simply not available to a user with CPU-only VDI (see Figure 2). Another example is Adobe Photoshop. Features like Scrubby Zoom, which enables you to zoom into a specific part of a photo for more accurate editing, is also not available to a user with a CPU-only VDI.³ It's important to understand the applications your users work in and which features they require. Think about future-proofing your investments today to avoid potential issues in the future as application vendors design more robust 3D features into their applications.

³ Adobe. December 31, 2017. [Photoshop Graphics Processor \(GPU\) Card FAQ](#).

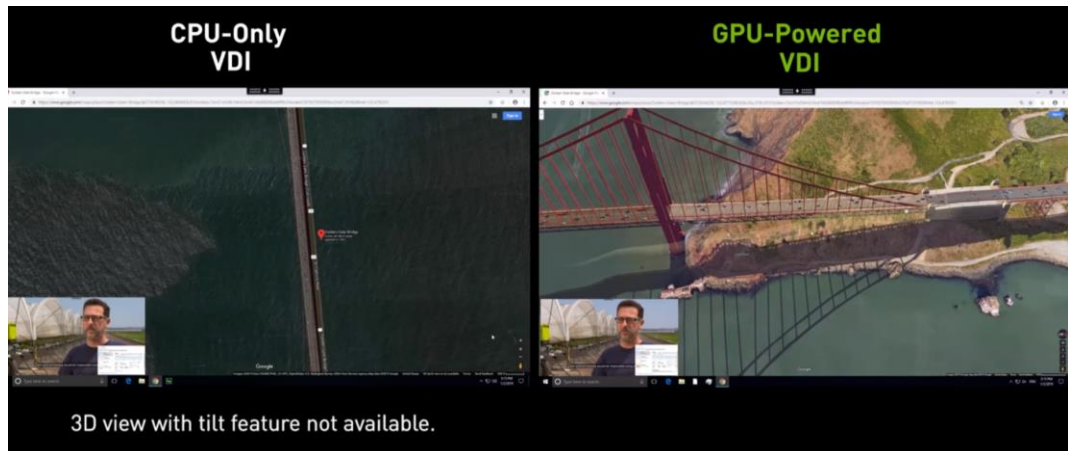


Figure 2. CPU only VDI vs GPU Accelerated VDI in a dual monitor setup running Google Maps.

Methodology: Factors for measuring UX

Typical VDI deployments have two conflicting goals: Achieving the best possible user experience and maximizing user density on server hardware. Problems arise as density is scaled up, however, because after a certain point it negatively impacts user experience.

NVIDIA nVector measures these trade-offs by modeling how knowledge workers use applications and what happens to performance when workloads are run at scale.

Modeling knowledge worker behavior at scale

To measure the quality of the knowledge worker experience on VDI, Dell EMC used a workload that simulates user behavior for a set of applications that are a good representation of knowledge workers' most widely used applications (Figure 3, Figure 4, and Figure 5). The sample set includes Microsoft Word 2016, Microsoft Excel 2016, Microsoft PowerPoint 2016, Google Chrome for web browsing, video streaming, and browsing PDF documents. The following images are screenshots taken of the workflows tested including web browsing, video viewing, and viewing/editing content.

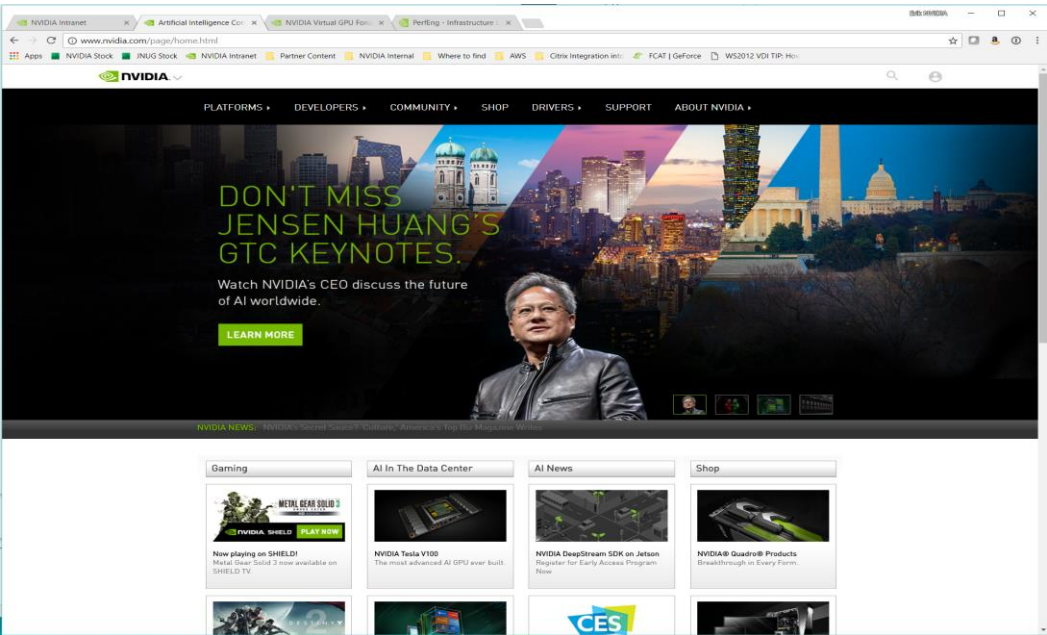


Figure 3. Web browsing with modern web browser Google Chrome

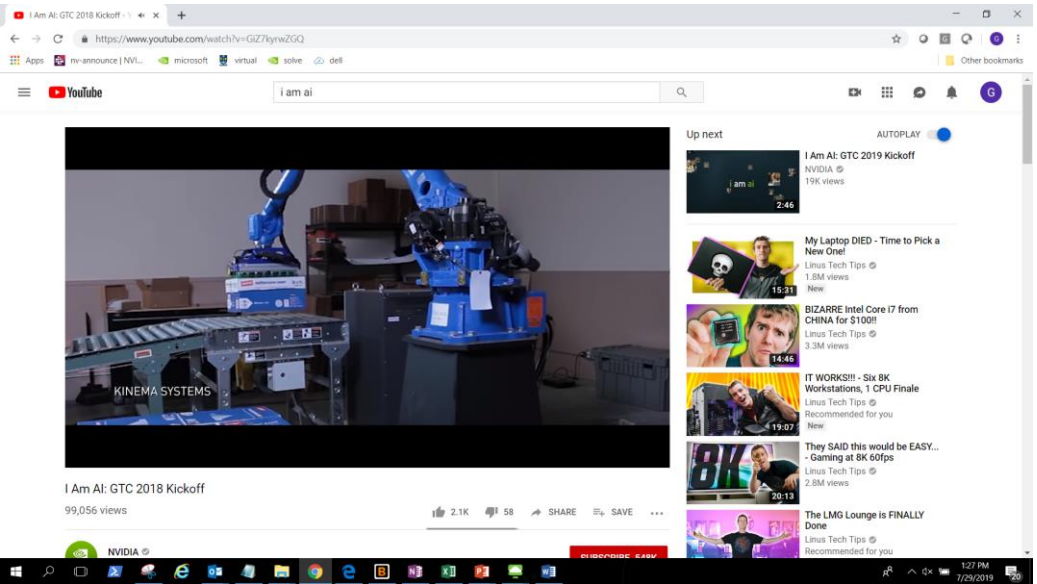


Figure 4. Viewing videos (web or local) using Google Chrome or Windows Media Player

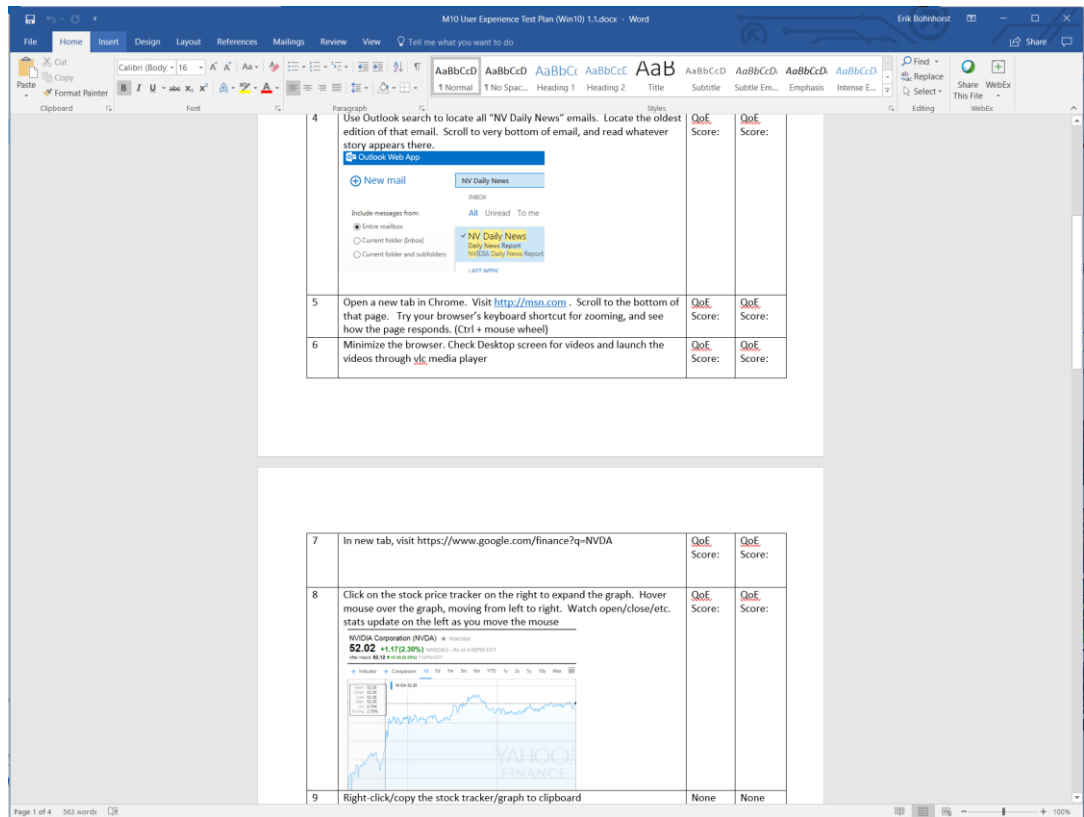


Figure 5. Viewing/Editing content with Microsoft Word 2016, Excel 2016, PowerPoint 2016, Google Chrome

The simulated workload also takes into account how these applications are used and is built with human speed input and scrolling. For example, all of these applications typically go through several stages when they're being used. The end-user will:

1. Open the application to either create new content or load pre-existing content.
2. Modify the content.
3. Review the changes by scrolling and/or zooming in and out of the content.
4. Save the content.
5. Copy content to the clipboard and use it in a different application (optional).
6. Close the application.

Simulating many users and many behaviors

The table below shows the workflow of each user. The graph in Figure 6 shows a cumulative increase in the number of users running workloads over time. Multiple users are tested at a time to simulate scale, with start and end times staggered to be more representative of real VDI environments.

Table 1. User workflow order of operations variants:

User #1	User #2	User #3	User #4	User #5
Microsoft Excel	Microsoft PowerPoint	Web Browsing	Web Video	...
Web Browsing	Microsoft Excel	Web Video	Microsoft Word	...
Web Video	PDF Viewer	Microsoft Word	Web Browsing	...
Microsoft Word	Web Browsing	Microsoft Excel	PDF Viewer	...
PDF Viewer	Microsoft Word	Microsoft PowerPoint	Microsoft PowerPoint	...
Microsoft PowerPoint	Web Video	PDF Viewer	Microsoft Excel	...

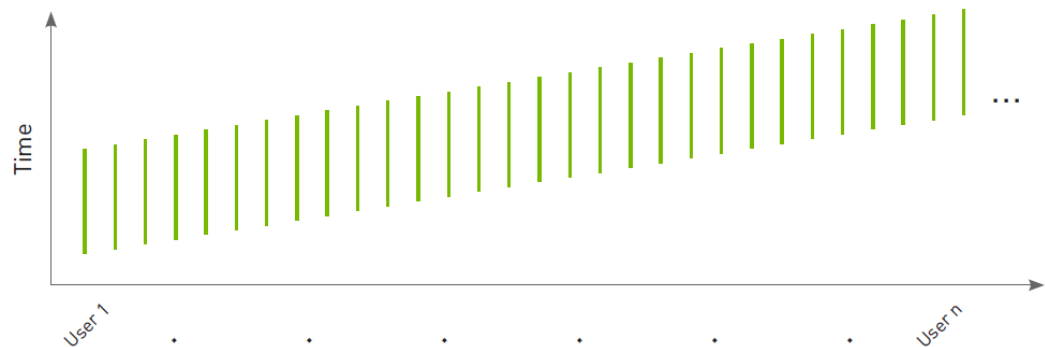


Figure 6. Characteristics of NVIDIA nVector.

The automated workload runs various applications in varying amounts of time (Figure 7). The variety of applications and length of time is modeled based on what a typical knowledge worker would do on a day-to-day basis.

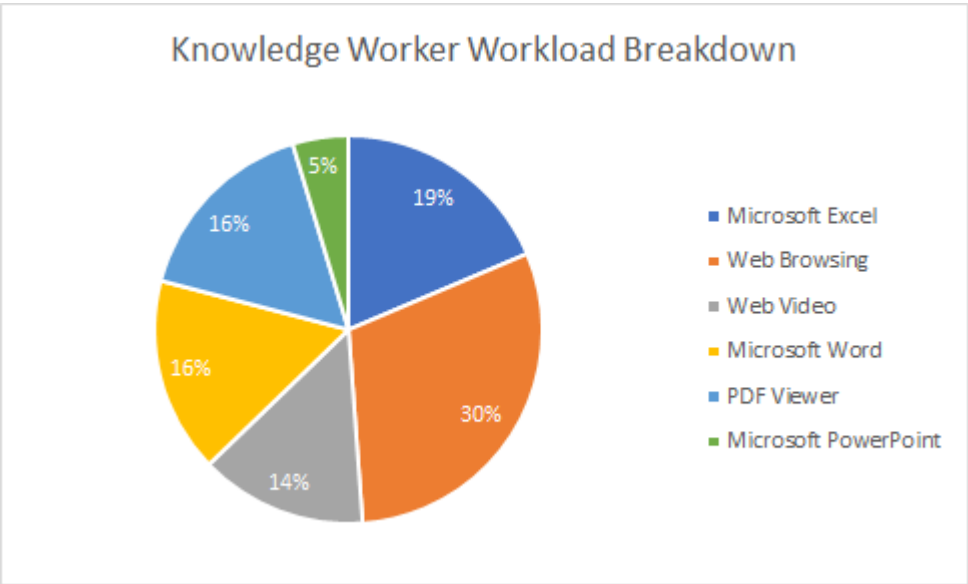


Figure 7. NVIDIA Profiler workload breakdown

Automating and running the workload at scale

NVIDIA nVector runs the simulated workflow of a knowledge worker, at scale. This part of the test requires performance monitoring to measure resource utilization. Acting as an execution engine, NVIDIA nVector orchestrates the necessary stages that are involved in measuring the end user experience for a pre-defined number of VDI instances (Figure 8):

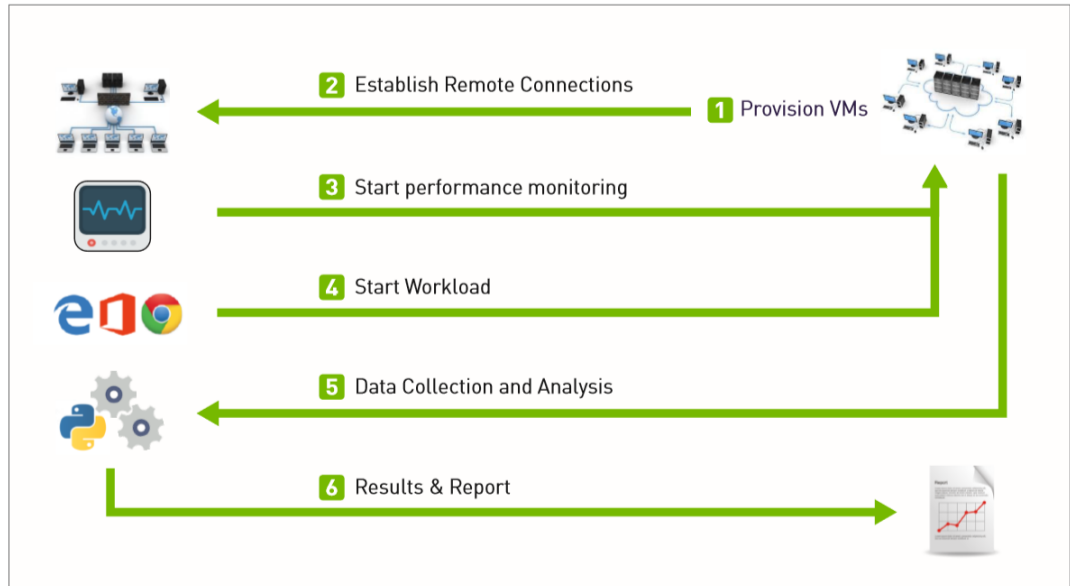


Figure 8. Stages involved in NVIDIA nVector in measuring UX

Here are the nVector profiling sequence steps that we performed in setting up our test environment:

1. Provision a number of VDI instances with predefined settings like vCPU, vRAM, and framebuffer, and provision an equal number of virtual machines that act as virtual thin clients.
2. Establish remote connections using the virtual clients to VDI desktops.
3. Measure resource utilization statistics on the server, as well as the guest OS of the VDI desktop.
4. Run a workload that emulates the knowledge worker on all of the VDI instances.
5. Collect and analyze performance data and end user experience measurements.
6. Generate a report that reflects the trade-off between end user experience and user density (scale).

Test environment

We quantify the benefits of NVIDIA GRID software 9.0 running on a Dell EMC PowerEdge R740xd rack server with NVIDIA T4 GPUs. The virtualized environment includes VMware Horizon 7.9 and VMware vSphere 6.7 U3. The test modeled 32 VDI users each using a 4K UHD monitor (Table 1). Our results demonstrated significant improvements with the

introduction of NVIDIA GRID Virtual PC (GRID vPC) and T4 GPUs on end user experience, as well as dramatic resource savings.

Table 2. ESX Server configuration:

Host configuration
Dell EMC PowerEdge R740xd
Intel(R) Xeon(R) Gold 6248 CPU @ 2.50GHz
VMware vSphere ESXi 6.7 U3
vGPU Manager: GRID 9.0 GA
Number of CPUs: 40 (2 x 20)
Memory: 384 GB
Storage: Local SSD
Hyperthreading: Turbo boost
Power setting: High Performance
GPU: 6 x T4

Table 3. Virtual machine configuration for 4K VDI Desktops

VM configuration (4K GRID vPC)	VM configuration (4K CPU only)
vCPU: 4	vCPU: 4
vRAM: 4096 MB	vRAM: 4096 MB
NIC: 1 (vmxnet3)	NIC: 1 (vmxnet3)
Hard disk: 32 GB	Hard disk: 32 GB
vGPU: t4-2b	vGPU: NA (CPU Only)
Virtual Hardware: vmx-13	Virtual Hardware: vmx-13
FRL enabled: no	FRL enabled: no
VMware Horizon 7.9	VMware Horizon 7.9
Blast Extreme YUV 4:2:0	Blast Extreme YUV 4:2:0
vGPU Driver: GRID 9.0 GA	vGPU Driver: NA (CPU Only)
Screen Resolution: 3840x2160	Screen Resolution: 3840x2160
Guest OS: Windows 10 Enterprise 1803	Guest OS: Windows 10 Enterprise 1803
Number of VMs: 32	Number of VMs: 32
Workload: MS Office 2016, Video, Web, PDF	Workload: MS Office 2016, Video, Web, PDF

Table 4. Table 4: Virtual machine configuration to simulate Endpoints with 4K resolution

Virtual client configuration (4K)
vCPU: 4
vRAM: 6144 MB

Virtual client configuration (4K)
NIC: 1 (vmxnet3)
Hard disk: 40 GB
vGPU: m10-2b
Virtual Hardware: vmx-13
VMware Horizon Client 5.1
vGPU Driver: GRID 9.0 GA
Screen Resolution: 1920x1080
Guest OS: Windows 10 Enterprise 1803

Results overview

Let's look at the overall comparison between NVIDIA vGPU accelerated and CPU-only. The comparison is made considering VDI desktops running common-place business applications for knowledge workers at high resolution of 3840x2160 (4K).

The benefits of GPU-accelerated VDI vs CPU-only VDI include:

- Up to 15% improved end user latency enabling faster response times with GRID vPC when compared to CPU-only
- Up to 25% improved frame rate for better fluidity with GRID vPC when compared to CPU-only
- Uncompromised image quality
- Up to 60% more users supported on the server when running graphics intensive applications

Test results: Measuring the end user experience

This paper has demonstrated how Dell EMC leveraged NVIDIA nVector to model knowledge worker behavior, as well as the orchestration stages that were involved in executing this workload at scale. This next section will further examine how the end user experience is measured and the results that were obtained.

NVIDIA nVector packages a utility that has the capability to emulate additional display devices within a virtual machine. This enables us to measure user experience on the virtual endpoints at different number of screens and/or at different resolutions. This utility uses NVIDIA graphics driver's APIs which requires a vGPU in client for our vGPU as well as CPU-only tests.

Metric 1: End user latency

The “end user latency” metric defines the level of response of a remote desktop or application. It measures the lag that an end user experiences when interacting with a remote desktop or application.

NVIDIA has developed a unique way to measure end user latency across any application using a unique technique. A lightweight application creates a small screen overlay which consists of a predetermined sequence of colors. It stays on top of the application that the user is interacting with, and receives a signal every five seconds. As the colors on the overlay update, NVIDIA nVector measures the time it took for the user to see the shift in colors on the remoting client (see Figure 9). The more activity on the screen, the longer the potential shift in color is seen as many other pixels need to be rendered back to the client. This method factors in the intensity and graphical nature of the set of applications evaluated while deriving responsiveness of the overall remote VDI session. In this testing, the workload run was representative of typical knowledge worker actions such as typing in Microsoft Word, transitioning between slides in PowerPoint, scrolling on a multimedia rich website, or viewing a PDF with images.

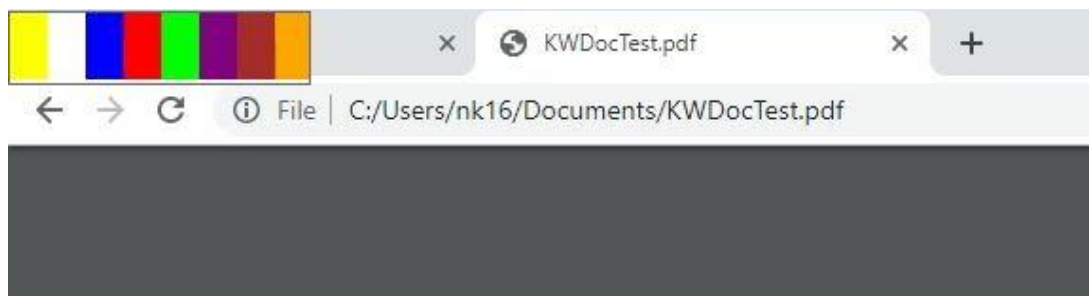


Figure 9. Screenshot snippet of the screen overlay of a predetermined sequence of colors that helps measure end user latency

How GPUs improve end user latency by 15%

Each VDI session collects approximately more than five hundred samples of end user latency, which will vary depending on the application and workflow. This data is then aggregated to get a full picture of latency for all users. The average of the entire sample set, which consists of thousands of samples, was used as a good measure of end user latency. Results demonstrated that GPU-accelerated VDI sessions saw a 15% improvement in latency (Figure 10). CPU-only VDI users experienced latency of 175 milliseconds (ms), while GPU-accelerated VDI users experienced latency of 148 ms, or a difference in 27 ms

While 27 ms might not seem significant, it does add up. A study on latency impact to human performance has found that “increasing latency above 13 ms, in a relative sense, has an increasingly negative impact on human performance for a given task. While imperceptible at first, added additional difference in latency continues to degrade a human’s processing ability until approaching 75 to 100 ms. Here we become very conscious that input has become too slow and we must rely on adapting to conditions by anticipating input rather than simply reacting to input.”⁴ We have seen real world implications of the impact of just a few milliseconds of latency. For example, a broker

⁴ PubNub Staff. [How Fast is Realtime? Human Perception and Technology](#). Feb 9, 2015

could lose \$4 million in revenues per millisecond if their electronic trading platform is 5 milliseconds behind the competition.⁵

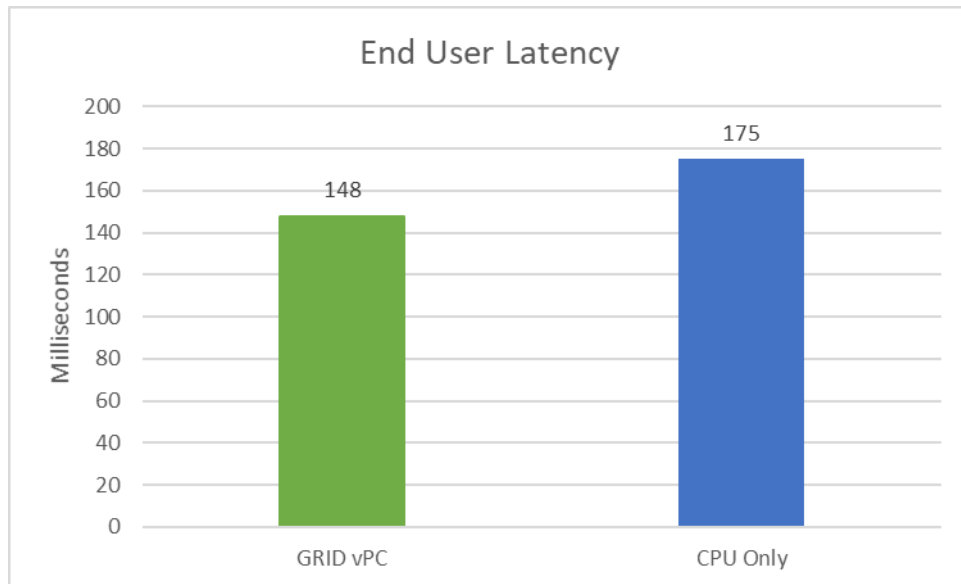


Figure 10. End-user latency of a GPU-accelerated vs CPU-only VDI environment (lower is better)

Metric 2: Frame rate

Frame rate is a common measure of user experience and defines how smooth the experience is. It measures the rate at which frames are delivered on the screen of the endpoint device. For the entire duration of the workload, NVIDIA nVector collects data on the 'frames per second' sent to the display device on the end client. This data is collected from thousands of samples and the value of the 90th percentile is taken. The greater this number, the more fluid the user experience.

How GPUs increase remoted frames by 25%

GPU-accelerated VDI sessions experienced 25% more frames/second (Figure 11), resulting in a more fluid and native-like experience. Fewer rendered frames delivers sluggish response times and sub-optimal interactivity. The average is calculated across the time period of workload execution which involves several idle periods. The frame rate during idle periods yields smaller numbers. These idle periods emulate the user behavior of pausing to take a break or thinking and understanding the information displayed on the screen, and hence are a necessary element in modeling user behavior. However, during active periods when a user is watching a webcast or scrolling through an Excel spreadsheet, the GPU acceleration empowers the VDI desktop to render and encode more frames per second, thereby ensuring a smooth and near-local-desktop-like user experience.

⁵ Yoav Einav. [Amazon Found Every 100ms of Latency Cost them 1% in Sales. Gigaspaces](#). January 20, 2019.

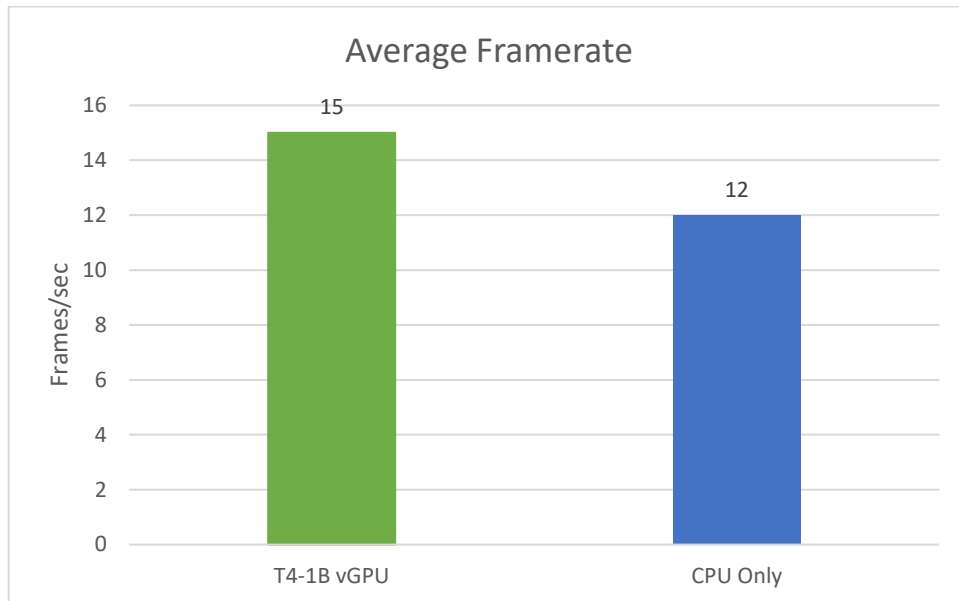


Figure 11. Average Frame rate of a GPU accelerated vs. CPU-only VDI environment (higher is better)

Metric 3: Image quality

NVIDIA nVector uses a lightweight agent on the VDI desktop and the client to measure image quality. These agents take multiple screen captures on the VDI desktop and on the thin client to be compared later. The structural similarity (SSIM) of the screen capture taken on the client is computed by comparing it to the one taken on the VDI desktop. When the two images are similar, the heatmap will reflect more colors above the spectrum shown on the right with an SSIM value closer to 1.0 (Figure 12). As the images become less similar, the heatmap will reflect more colors down the spectrum with a value less than 1.0. More than a hundred pairs of images across an entire set of user sessions are obtained. The average SSIM index of all pairs of images is computed to provide the overall remote session quality for the entire population of all users.

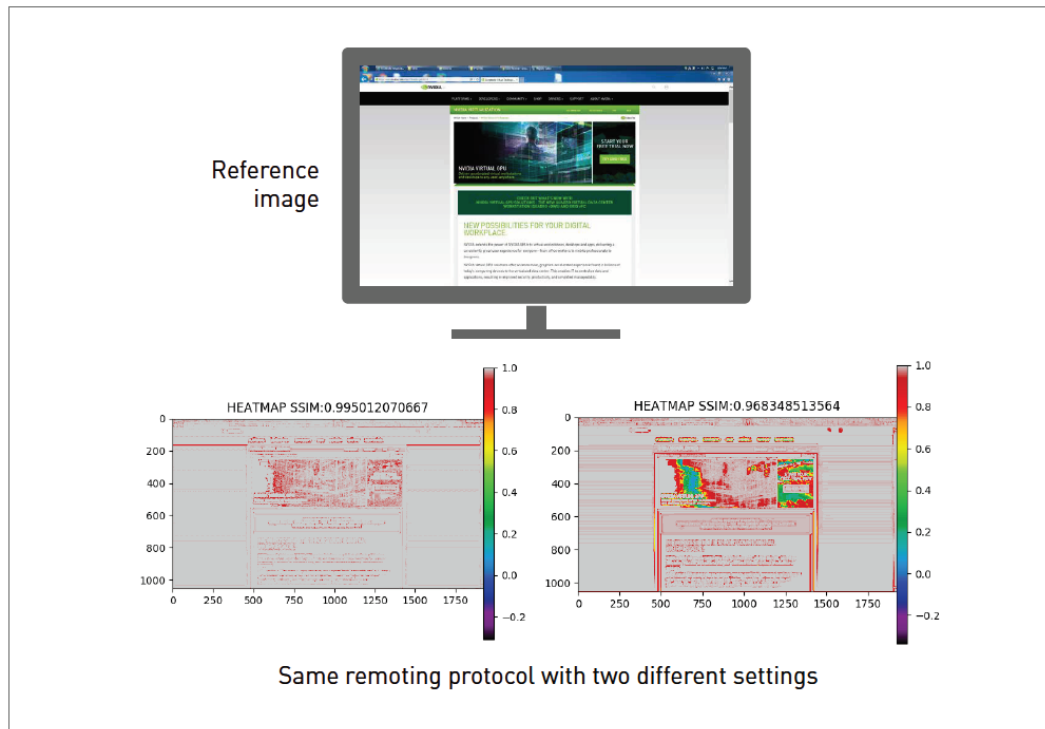


Figure 12. SSIM as a Measure of Image Quality

How GPUs ensure uncompromised image quality

Image quality is dictated by the remoting protocol, as well as by the configuration and policies set in the VDI environment. Our test demonstrates that GPU-accelerated VDI users will experience uncompromised image quality, as structural similarity (SSIM) of the screen capture are both above 0.99 for both CPU and GPU-accelerated VDI environments (Figure 13).

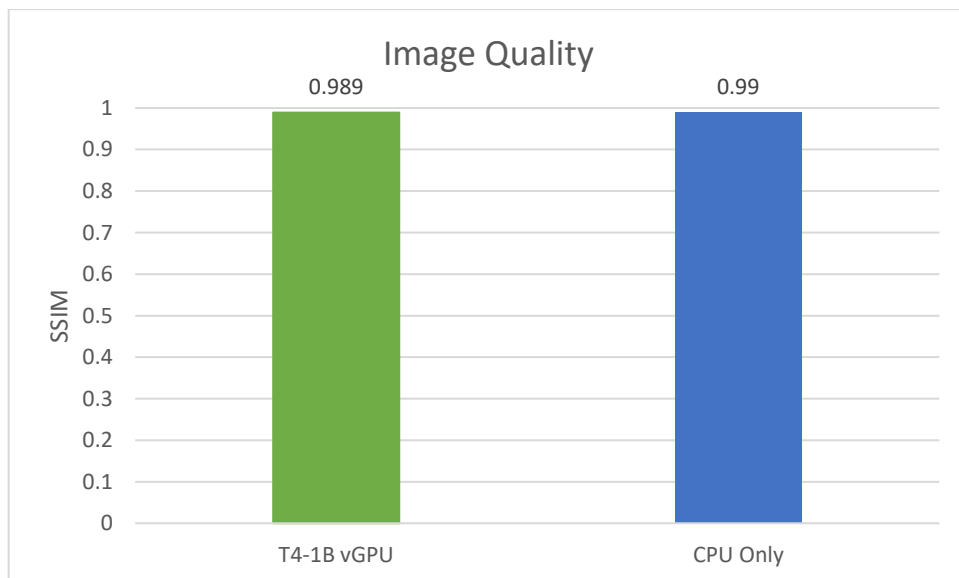


Figure 13. Image quality of a GPU accelerated vs. CPU-only VDI environment (higher is better)

Metric 4: Server utilization

Observing overall server utilization allows us to understand user experience in the context of resource utilization. NVIDIA nVector periodically samples ESX Server's CPU core utilization while the workload is actively running for all user sessions. To determine the 'steady state' portion of the workload, samples are filtered, leaving out the times when users have all logged on and the workload start ramps up and down. This yields us the middle third of samples measuring CPU Core utilization on ESX server. This period covers the time during which all user sessions were actively running workload. Once steady state has been established, all samples are aggregated to get the total CPU core utilization on the server.

How GPUs reduce server CPU load by 38%

Steady state was determined when all VDI sessions were active, and all 32 users were interacting with applications and data. When comparing CPU-only and GPU-accelerated VDI sessions during steady state, a consistent gap in server CPU utilization was observed, which averaged 38% and could go up to 57% (Figure 14). Even larger gaps were observed during some periods, which could be attributed to variations in the workload (video, multimedia, PowerPoint, etc.).

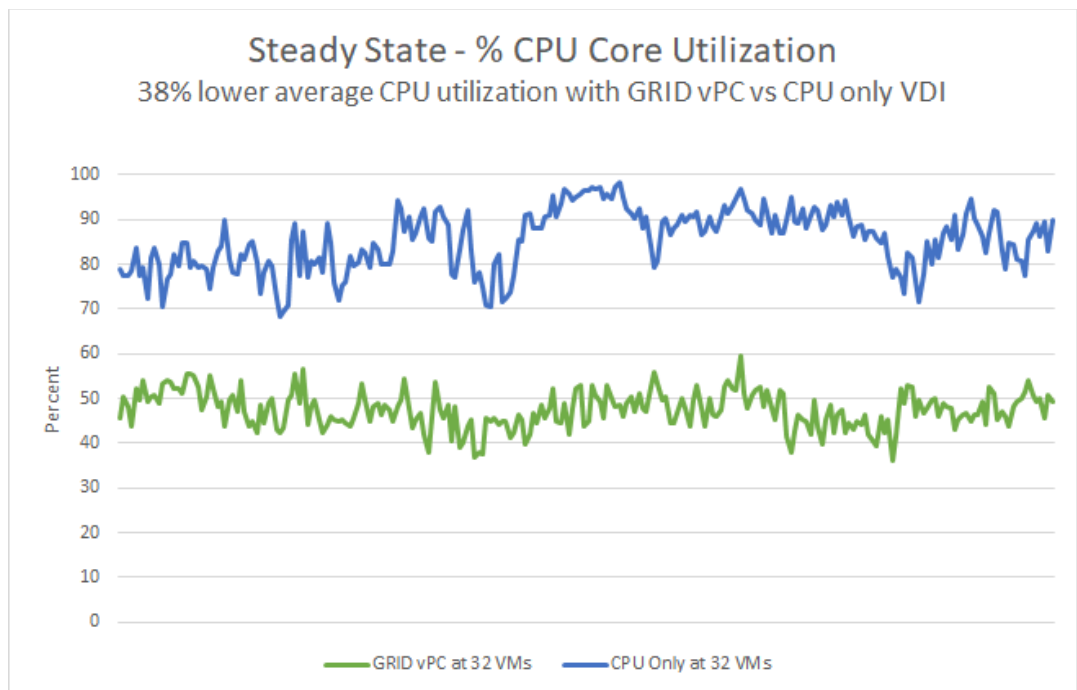


Figure 14. CPU utilization in a GPU accelerated vs. CPU-only VDI environment, with a single 4K monitor (lower is better)

Don't let the latest web standards like WebGL slow down your users

Flash, HTML5, and WebGL are all very taxing to the CPU. In fact, the latest standards like WebGL, currently used in 53% of the top 100 websites, can cause the CPU to hit 100% when just animating a simple scene.⁶ The reason is that WebGL is displayed on an HTML canvas. The HTML elements must overlay the rendered WebGL image, and the WebGL image must have transparent parts that allow the HTML elements beneath it to show through. The blending of HTML and WebGL is known as compositing. A lot of copying and compositing of images is being done by the CPU, thereby increasing its utilization significantly. The good news is modern browsers like Chrome and Edge are hardware accelerated by default, which means web developers can now leverage the GPU to render graphics.⁷

WebGL is being used widely by many business sectors, for example:

- The education sector creates interactive learning tools that have complex graphics
- The architecture sector uses WebGL to showcase properties on the web
- Ecommerce sites use WebGL to render products with intricate detail and enable product visualization in real-time

The current workload tested does not use WebGL, but CPU utilization of the CPU-only VDI has already reached close to 100%. In contrast, the GPU-accelerated VDI averages 38%, leaving plenty of headroom to support users' demanding applications when they need it.

When graphics intensive applications are used, GPUs increase server density by up to 60%

A single HD monitor setup is also very commonly used. For a single HD monitor, just like 4K use-case discussed in previous sections, our baseline is CPU-only evaluation. Because lower screen resolution reduces resource contention, we could scale up to 64 users. In order to do a valid comparison, we matched the scale of 64 VDI desktops that are accelerated with a T4-1B vGPU which gives the VDI instance one GB of graphics memory. Table 5 and Table 6 outline the configuration details of setup used for a single HD setup.

Table 5. Virtual machine configuration for 1080p VDI desktops

VM configuration (1080p GRID vPC)	VM configuration (1080p CPU Only)
vCPU: 4	vCPU: 4
vRAM: 4096 MB	vRAM: 4096 MB
NIC: 1 (vmxnet3)	NIC: 1 (vmxnet3)

⁶ Zhihao Yao, Zongheng Ma, Yingting Liu, Ardalan Amiri Sani, Aparna Chandramowlishwaran. 2018. [Sugar: Secure GPU Acceleration in Web Browsers](#). In *ASPLOS '18: 2018 Architectural Support for Programming Languages and Operating Systems*, March 24–28, 2018, Williamsburg, VA, USA. ACM, New York, NY, USA, 16 pages.

⁷ Tom Smith. [Everything you need to know about WebGL](#). Creative Bloq. June 11, 2018.

When graphics intensive applications are used, GPUs increase server density by up to 60%

VM configuration (1080p GRID vPC)	VM configuration (1080p CPU Only)
Hard disk: 32 GB	Hard disk: 32 GB
vGPU: t4-1b	vGPU: NA (CPU Only)
Virtual Hardware: vmx-13	Virtual Hardware: vmx-13
FRL enabled: no	FRL enabled: no
VMware Horizon 7.9	VMware Horizon 7.9
Blast Extreme YUV 4:2:0	Blast Extreme YUV 4:2:0
vGPU Driver: GRID 9.0 GA	vGPU Driver: NA (CPU Only)
Screen Resolution: 1920x1080	Screen Resolution: 1920x1080
Guest OS: Windows 10 Enterprise 1803	Guest OS: Windows 10 Enterprise 1803
Number of VMs: 64	Number of VMs: 64
Workload: MS Office 2016, Video, Web, PDF, WebGL*	Workload: MS Office 2016, Video, Web, PDF, WebGL*

Table 6. Virtual machine configuration to simulate endpoints with 1080p resolution

Virtual client configuration (1080p)
vCPU: 4
vRAM: 6144 MB
NIC: 1 (vmxnet3)
Hard disk: 40 GB
vGPU: m10-1b
Virtual Hardware: vmx-13
Number of screens: 1
VMware Horizon Client 5.1
vGPU Driver: GRID 9.0 GA
Screen Resolution: 1920x1080
Guest OS: Windows 10 Enterprise 1803

When the workload was run without WebGL, CPU utilization of the GPU-accelerated VDI was on average 16% lower, and up to 36% in some scenarios (Figure 15).

When graphics intensive applications are used, GPUs increase server density by up to 60%

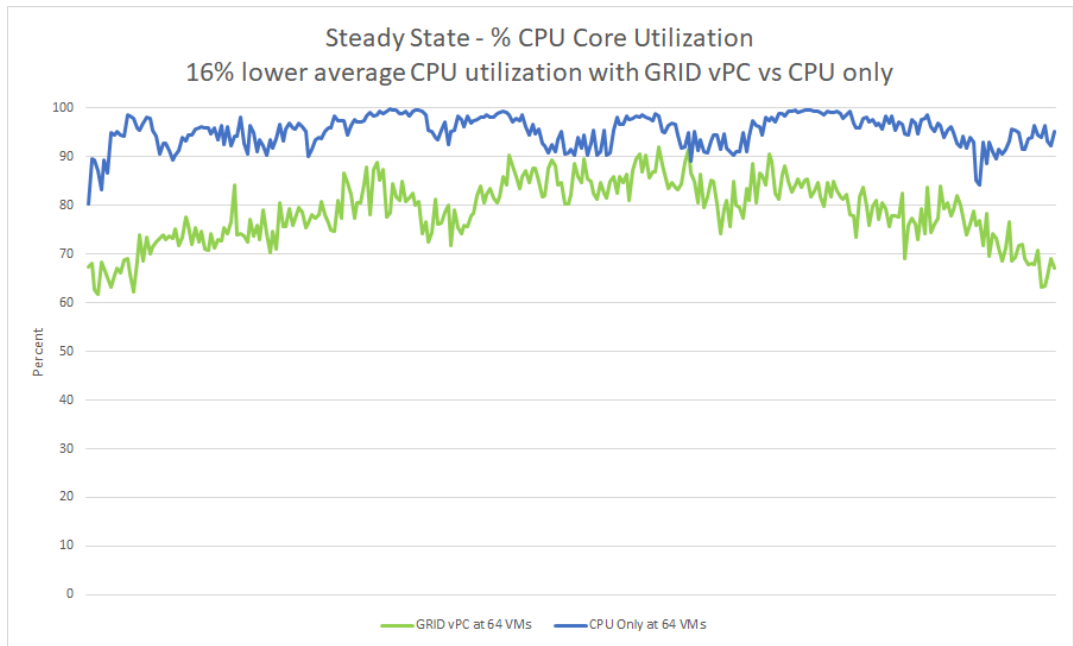


Figure 15. CPU utilization for 64 users in a CPU-only vs. GPU-accelerated VDI environment, with a single HD monitor without WebGL (lower is better)

When a minimal amount WebGL was added to the existing knowledge worker workload (8 minutes out of 51 minutes, or 16% of the total tasks executed), CPU utilization was significantly impacted (Figure 16).

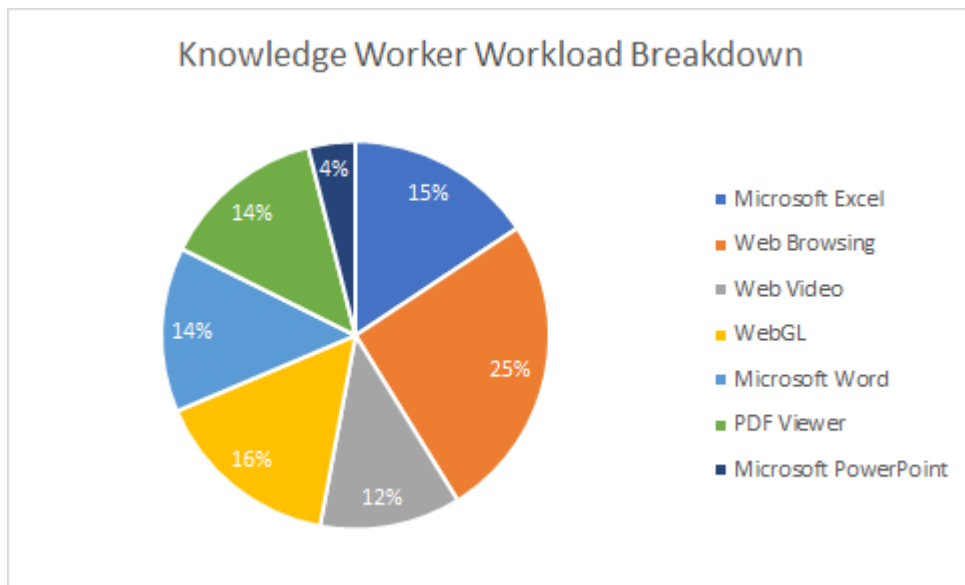


Figure 16. Knowledge worker workload breakdown with WebGL

When the number of VMs on the CPU-only VDI environment was lowered to 40 VMs, the CPU utilization dropped to the level of the GPU-enabled VDI environment with 64 VMs (Figure 17). Adding GPU increases the user density by allowing 60% more users than on non-GPU-VDI environments and creates a better user experience as mentioned earlier.

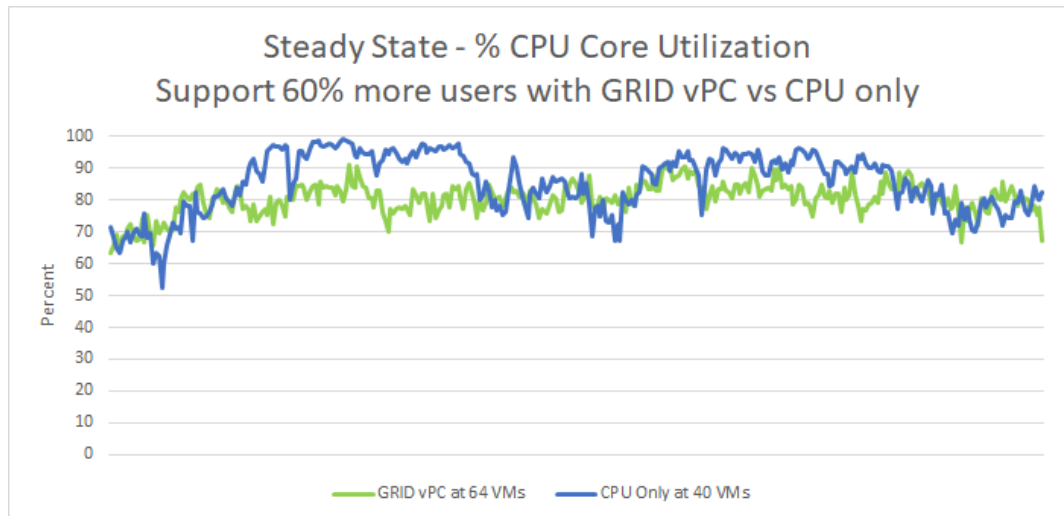


Figure 17. CPU utilization for 40 users in a CPU-only vs. 64 users with GPU-accelerated VDI environment, with a single HD monitor with WebGL (lower is better)

Conclusion

Our test results demonstrate that Dell EMC PowerEdge R740xd VDI environments enhanced with NVIDIA GRID and T4 GPUs deliver significant performance improvements. Ultimately, they provide an experience so streamlined and consistent that it allows users to be more productive than they would be on a traditional CPU-only VDI environment. These test results are particularly relevant to today's digital workplaces. With the prevalence of graphics-intensive applications and deployment of Windows 10 across the enterprise, adding graphics acceleration to VDI powered by NVIDIA virtual GPU technology is critical to preserving the user experience. Moreover, adding NVIDIA GRID with NVIDIA GPUs to VDI deployments increases user density on each server which means more users can be supported with a better experience.

To learn more about measuring user experience in your own environments, contact your Dell Account Executive.