Impact of NVIDIA Virtual GPU on Video Conferencing Tools

Elevating User Experience with NVIDIA Virtual PC (vPC)

Technical Brief

TB-10169-001_v02 | July 2021

Table of Contents

Intent of this Technical Brief

NVIDIA virtual GPU (vGPU) technology is enabling organizations to seamlessly shift to remote work. With this shift, video conferencing tools are becoming increasingly prevalent.

In this technical brief, you will find the impact of NVIDIA virtual PC (vPC) on video conferencing tools. As video and audio are offloaded to the graphics processing unit (GPU), you will see the difference in the user experience when executing common video conferencing features like web camera activation and screen sharing.

Executive Summary

As more organizations shift to remote work, video conferencing applications have become critical tool for employees to work seamlessly with remote teams. In order to be effective, users must have a great user experience when using video conferencing tools. Video and audio quality are key to productive real-time collaboration. Any delays in video or audio can lead to miscommunication and negatively impact overall productivity.

This technical brief provides an overview of the impact of NVIDIA virtual PC (vPC) on Microsoft Teams compared to CPU only virtual desktop infrastructure (VDI). Tests were executed to determine if vPC offers CPU offload when users are accessing video conferencing tools, similar to the results seen while testing the NVIDIA nVector knowledge worker workflow. As published within the existing *NVIDIA vPC Sizing Guide*, NVIDIA vPC offers improved server density as well as user experience satisfaction when comparing to a CPU-only virtualized environment. When introducing NVIDIA vGPU technology into the environment, virtual systems will no longer rely solely on the CPU for graphics processing. Therefore, systems without a GPU have higher overall CPU usage due to software application's inability to execute certain functions on the GPU, offloading CPU.

Overall, our video conferencing test results showed that by having vGPU present within the virtual machine (VM), there was a significant amount of vCPU offload which frees vCPU resources within the VM to run other tasks. The amount of CPU offload is dependent on how well the web conferencing application can take advantage of the GPU as well as network and web camera conditions. The CPU offload within VM's ultimately results in decreased CPU resource usage on host server due to the nature of virtualization using shared server resources. Therefore, NVIDIA vPC offers improved server density while boosting user experience.

Introduction

The NVIDIA virtual GPU (vGPU) solution provides a flexible way to accelerate virtualized workloads – from AI to VDI. The solution includes NVIDIA virtual GPU software and NVIDIA data center GPUs. There are two unique NVIDIA virtual GPU software licenses for virtual desktops and workstations, each priced and designed to address a specific use case:

- NVIDIA Virtual PC (vPC) and Virtual Applications (vApps) – accelerates office productivity applications, streaming video, Windows 10, RDSH, multiple and high-resolution monitors, and 2D electric design automation (EDA).
- NVIDIA RTX Virtual Workstation (vWS) – accelerates professional design and visualization applications, including Autodesk Revit, Maya, Dassault Systèmes CATIA, Solidworks, Esri ArcGIS Pro, Petrel, and more.

Decoupling the GPU hardware and virtual GPU software options enables customers to benefit from innovative features delivered in the software at a regular cadence, without a dependency on purchasing new GPU hardware. It also provides the flexibility for IT to architect the optimal solution to meet the specific needs of users in their environment.

NVIDIA Virtual PC

This technical brief will focus on NVIDIA Virtual PC (vPC). NVIDIA vPC software enables the delivery of graphics-rich virtual desktops accelerated by NVIDIA GPUs. NVIDIA vPC enables GPU sharing across multiple virtual machines, delivering a native-PC experience to knowledge workers while improving user density. Because tasks typically done on the CPU are offloaded to the GPU, the user has a distinguishable better experience than a CPU-only VDI VM. With an NVIDIA vPC enabled VDI environment, the server density is significantly improved too. Overall, NVIDIA vPC delivers great user experience that is nearly equivalent to its native PC counterpart. With lower latency and a higher frame rate than CPU-only VDI, applications are more responsive, providing users with best-in-class user experience.

Testing Methodology

Testing was executed in two phases to determine the impact of NVIDIA vPC. These phases represent different workflows that a typical knowledge worker performs when using video conferencing tools. Table 1 describes the two phases of testing as well as the VM configuration tested.

Table 1. Test Workflow

Phase	Workflow	vCPU	vGPU Profile	Memory
1	Web Camera	2	A10-2B	4 GB
1	Web Camera	2	N/A (CPU-Only)	4 GB
2	Screenshare	2	A10-2B	4 GB
2	Screenshare	2	N/A (CPU-Only)	4 GB

- Phase 1 - Web Camera
  - Incremental activation of three cameras
  - Incremental deactivation of three cameras
- Phase 2 - Screenshare
  - Screen sharing a PowerPoint presentation
  - No web camera activation

There was a total of three users executing tests during all phases of testing. The first test user acted as the video conferencing host, whereas the second and third test users were attendees and accessed video conferencing tools within two virtual machines. One configured with the NVIDIA vPC and the other was a vCPU-only VDI.

Microsoft Teams was the video conferencing tool used during each phase of the test.

Testing was not automated. Instead, test users followed a very well defined script to ensure performance logging was synchronized for both the vCPU-only VM and vPC VM. Testing focused on single VMs and was not executed to scale (i.e. 64 users) since enterprise users typically do not work exclusively with these tools but have multiple software applications active while interacting with video conferencing tools. The testing methodology primarily focused on the value of CPU offload which drives the value of NVIDIA vPC - improved server density and user experience satisfaction.

Test Setup

GPU Profiler is a commonly used tool that can quickly capture resource utilization while a workload is being executed on a virtual machine. This tool is typically used during a POC to help size the virtual environment in order to ensure acceptable user performance. GPU Profiler was run on a single VM with various vGPU profiles while the NVIDIA nVector knowledge worker workload was running. The following metrics were captured:

- Framebuffer %
- vCPU %
- RAM %
- Video Encode
- Video Decode

Figure 1 shows a screen capture of the GPU Profiler.

Figure 1. GPU Profiler Screen Capture

Table 2 summarizes the VM configuration for the NVIDIA vPC VM and the CPU-only VDI.

Table 2. VM Configuration

VM Configuration
vGPU Profile: A10-2B
vCPU: 2
vRAM: 4096 MB
NIC: 1 (vmxnet3)
Harddisk: 40 GB
Virtual Hardware: VMX-13
VMware Horizon: 8.2
vGPU Driver: 12.2 (Windows Driver 462.31)
Guest OS: Windows 10 Enterprise 20H2
Server: Intel Xeon Gold 6154 (2x @ 2.99GHz)

Variables Impacting Performance

Several variables can impact performance as well as the amount of resource usage within the VM's. These include:

- Network latency
- Web camera model, driver, and quality
- CPU speed
- Monitor setup – quantity and resolution
- Operating System

Additionally, how users configure their software such as the layout of video conferencing tools and the amount of user movement and activity can impact performance.

The most significant factor to consider when optimizing for performance is network latency and web camera quality. If either of these has poor quality, the advantages of increased software performance by the NVIDIA vPC will be less. Therefore, we used two standards to circumvent the impact of these variables:

- We used Logitech 1080p web camera
- We set the screen resolutions to 2560 × 1440 for both test cases

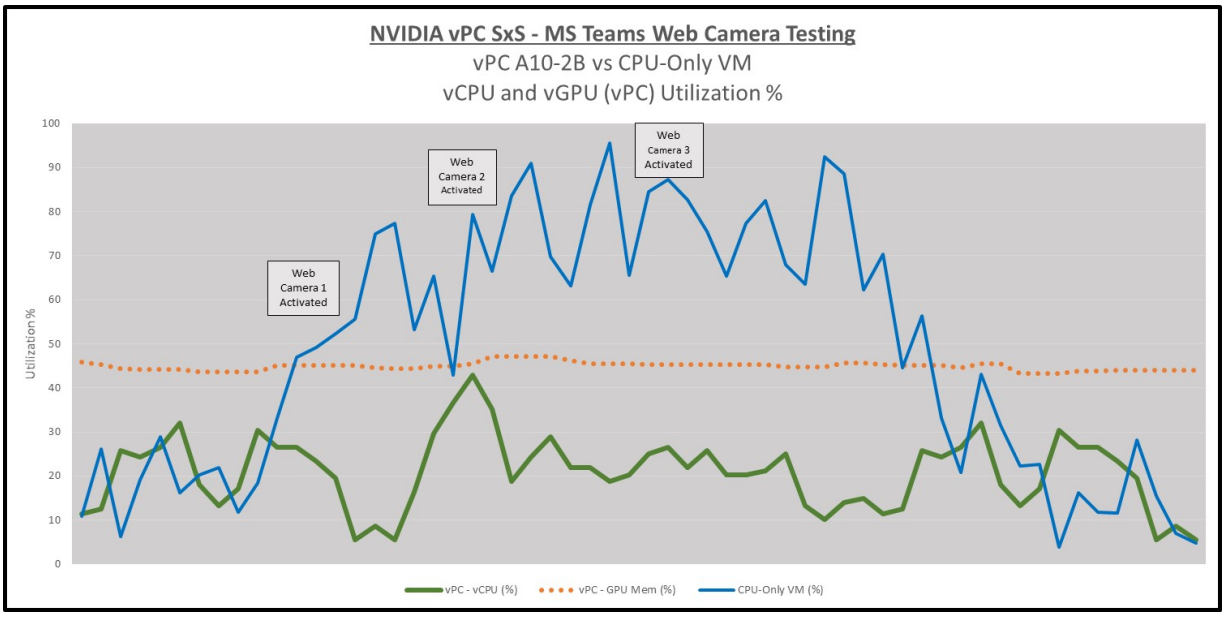
Video conferencing optimization management packs offered by different hypervisors and operating system platforms can also impact the amount of resource offload within the VDI. For instance, in NVIDIA vPC VM the Microsoft Hardware-Accelerated GPU Scheduling feature was enabled for the testing.

Impact of NVIDIA vPC

Overall, Microsoft Teams exhibited a certain level of CPU offload. The amount of CPU offload depends on how well the MS Teams software can take advantage of the GPU.

The user experience suffered on the CPU-only VM due to vCPU bottlenecks when accessing Microsoft Teams conferencing tools. The CPU-only VM (2vCPU) neared 100% utilization, resulting in poor video quality at times, due to pixel loss and stutter. Audio quality was also poor with audio delay and breakage. In order to improve the user experience within the CPU-only VM, the vCPU configuration should be increased (i.e. from 2 to 4vCPU) to more closely match the quality of user experience offered by the NVIDIA vPC VM. With this in mind, the increased amount of server density value prop really shines with the NVIDIA vPC since the vCPU count of two was adequate for the NVIDIA vPC VM. Figure 2 illustrates the CPU resource usage within the CPU only VM vs the NVIDIA vPC VM as well the utilization of the 2 GB GPU memory frame buffer.

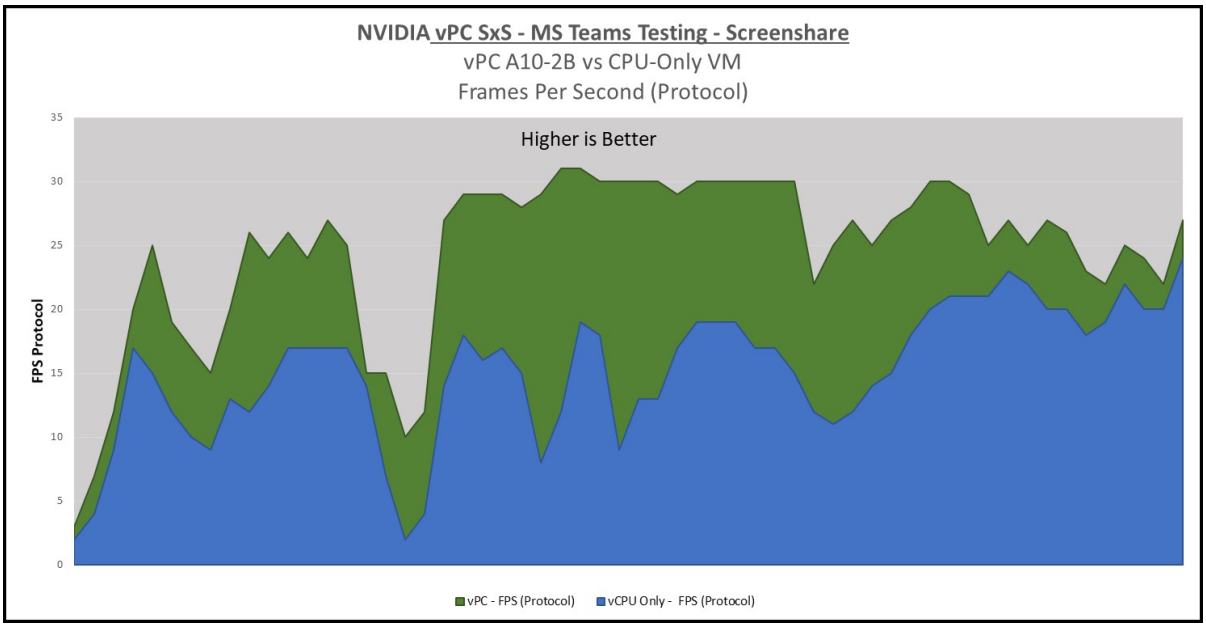
Figure 2. MS Teams Testing Web Camera



Note: MS Teams results tested on a server with Intel Xeon Gold 6154 CPU @ 3.0 GHz, NVIDIA vPC software, VMware ESXi 7.0 Update 2, VMware Horizon 8.2, host and guest driver 462.31, VM config, WIN10 (20H2), 2vCPU, 4GB Memory, A10-2B, VMware Blast Protocol. Test workflow incrementally turned on three web cameras, ending in a total of three web camera feeds (one or attendee), prior to concluding tests, web cameras were turned off in MS Teams.

The remote frames per second (FPS) were also much higher for the NVIDIA vPC VM, providing a better user experience while using fewer vCPU resources. Figure 3 highlights the difference in FPS between GPU-accelerated VDI and CPU-only VDI.

Figure 3. MS Teams Testing – Screenshare



Conclusion

Remote work is continuing to become more prevalent throughout businesses across many industries. For teams to be productive, video conferencing tools, like Microsoft Teams have become necessary for real-time collaboration and communication. It is not enough for organizations to provide video conferencing tools, they also must ensure a suitable user experience with optimal video and audio quality. Video breakage and audio delays can lead to miscommunication, loss of information, lower productivity, and overall frustration. NVIDIA GPU-accelerated VDI powered with NVIDIA vPC delivers a more consistent user experience when using video collaboration tools.

With CPU offload in mind, NVIDIA vPC offers increased server density since virtualized server resources are shared across multiple virtual desktops. This offload of CPU also frees up vCPU resources within the VM to run other tasks. With a CPU-only VM, more vCPU resources would need to be allocated to match the quality of the user experience offered by an NVIDIA vPC VM, increasing the amount of hardware to purchase, deploy, and manage.

Although optimization packs can improve the user experience with video collaboration tools in VDI, there is added complexity in administering and managing endpoints when using optimization packs. By using GPU-accelerated VDI for video conferencing applications, NVIDIA vPC, not only is the user experience improved, but the added benefit of centralized data and applications simplifies IT administration and management, saving time, costs, and IT resources.

Resources Links

NVIDIA Resources:

NVIDIA vPC Sizing Guide

Quantifying the Impact of NVIDIA Virtual GPUs

NVIDIA vPC Solution Overview

NVIDIA vPC webpage

Other Resources:

Try NVIDIA vGPU for free

NVIDIA vGPU 90 day evaluation

NVIDIA Virtual GPU Software Documentation

NVIDIA vGPU Certified Servers

Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation ("NVIDIA") makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer ("Terms of Sale"). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

NVIDIA products are not designed, authorized, or warranted to be suitable for use in medical, military, aircraft, space, or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death, or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer's own risk.

NVIDIA makes no representation or warranty that products based on this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer's product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this document. NVIDIA accepts no liability related to any default, damage, costs, or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this document or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.

Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

Trademarks

NVIDIA, the NVIDIA logo, and NVIDIA RTX are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

OpenCL

OpenCL is a trademark of Apple Inc. used under license to the Khronos Group Inc.

Copyright

NVIDIA Corporation | 2788 San Tomas Expressway, Santa Clara, CA 95051
http://www.nvidia.com