VDI Graphics Acceleration on a Dell EMC MX7000 Modular Chassis

May 2020

H18278

White Paper

Abstract

This white paper describes performance testing results and provides density guidance for a VMware Horizon 7 virtual workstation solution configured on a Dell EMC MX7000 modular chassis. The document provides a SPEC benchmark score for the virtual workstations on MX7000 tested with a SPECviewperf 13 workload (Maya and 3ds Max viewsets).

Dell Technologies Solutions

Copyright

Contents

Executive summary

Business challenge

The media, entertainment, and gaming (MEG) industry is going through a major transformation to meet consumer demand for high-quality content. These organizations are looking for ways to improve production quality cost-effectively and maximize profits. Investing in IT infrastructure that is agile, scalable, efficient, and cost-effective helps content producers outpace the competition and thrive in this era of digital disruption.

Creative professionals in the MEG industry typically work on physical workstations for creating, editing, animating, and rendering 3D modeling workflows. However, physical workstations pose challenges in terms of mobile and remote staff. Working with massive 3D sets across a WAN network is challenging and can put your organization's data security at risk.

Another challenge is working with a globalized team of creative professionals. The projects in MEG industries often follow an iterative design approach to refine the prototypes. Collaborating over different geographical areas causes a delay in the projects. For organizations with IT budget constraints, buying and maintaining many physical workstations is not cost-effective.

Virtual desktop infrastructure (VDI) presents an opportunity for the MEG industry to address its need for greater mobility and collaboration among creative professionals. NVIDIA Quadro vDWS enabled virtual workstations in a VDI environment offer performance that equals that of physical workstations. With VDI, IT teams can better align with budget targets while still focusing on innovation. In addition to the productivity gains for creative professionals, IT teams benefit from the following VDI capabilities:

- Easy onboarding of partners and contractors
- Centralized management of workstations
- Improved security controls
- Live migration capabilities of workstations, which eliminates business disruption during maintenance activities

The MEG industry uses many graphics-intensive applications, including Autodesk Maya and 3ds Max. These applications require both high-end CPU and GPU processing power. Many VDI platforms with GPUs used for graphics-intensive applications do not exhibit the expected performance gains due to CPU bottlenecks. For example, Autodesk Maya benefits from a high clock speed CPU core while creating, modifying, and animating 3D models.

However, achieving a smooth and responsive experience when rotating, zooming, or panning while editing and rendering 3D models with massive datasets requires a higher frame rate in the viewport that is only achievable by using a GPU. Applications in the MEG industry require a VDI graphics acceleration solution that offers the right balance of CPU and GPU compute.

Our virtual workstation solution based on VMware Horizon 7 running on a Dell EMC MX7000 modular platform meets these challenges. The PowerEdge MX7000 Modular Chassis, which is equipped with up to 16 NVIDIA T4 GPUs (part of the NVIDIA Tesla

product line) contained in an Amulet Hotkey CoreModule, brings the balance of high-speed processors and high-performance graphics that are required for compute acceleration for virtual workstations. The Amulet Hotkey CoreModule is a unique core expansion module that brings graphics virtualization to the PowerEdge MX7000 Modular Chassis platform. Organizations in the MEG industry can now benefit from the agility, flexibility, and scalability of a modular environment along with a graphics platform that provides better CPU and GPU performance.

Dell EMC Isilon storage, a winner of the 71st Annual Technology & Engineering Emmy Awards, is at the core of the production media workflows in the media and entertainment industry. Isilon is a key member of the Dell Technologies Media and Entertainment solution portfolio. Our customers in these industries can now use Isilon unstructured data storage with Horizon 7 virtual workstations running on a Dell EMC MX7000 modular environment to achieve the compute acceleration that is required for their graphics-intensive applications.

Solution overview

The Horizon 7 virtual workstation solution running on the MX7000 modular environment combines the performance of Intel Xeon Scalable processors from MX7000 blade servers and the powerful graphics acceleration enabled by Amulet Hotkey CoreModules equipped with NVIDIA T4 GPUs. The solution offers excellent performance for the graphics-intensive applications that are used in the MEG industry. 3D modeling applications benefit from the high-performing CPU and GPU cores in this solution. The latest version of these applications comes with advanced rendering engines capable of using both CPUs and GPUs for rendering.

The Dell EMC PowerEdge MX7000 Modular Infrastructure is built for use in a software-defined data center. Dell EMC PowerEdge MX7000, with its new kinetic infrastructure technology, provides better resource granularity, agile management, and future-proof capabilities than traditional high-density blade servers. It is a high performance, modular infrastructure that enables customers to flexibly configure and optimize their IT infrastructure for a wide variety of traditional, new, and emerging workloads. Customers can run a VDI workload in an MX7000 modular environment along with a wide variety of other workloads to benefit from the agility, flexibility, and scalability offered by this platform.

The PowerEdge MX7000 modular chassis is optimized for VMware vSAN software-defined storage environments and combines easy management with better capacity and performance for VDI workloads. It provides highly scalable, direct-attached storage capabilities housing up to 112 HDD/SDD drives per chassis in addition to six internal compute sled HDD/SDD drives. The Dell EMC OpenManage Integration for VMware vCenter (OMIVV) streamlines the management processes in your data center by allowing you to use VMware vCenter for managing your PowerEdge server infrastructure.

Key benefits

Benefits of the Horizon 7 virtual workstation solution on MX7000 infrastructure include:

- **High performance**–The solution combines the high-performing CPU compute capabilities of the Intel Xeon Scalable processors in the MX7000 chassis and the graphics acceleration offered by Amulet Hotkey-enabled NVIDIA T4 processors. The seamless integration of Amulet Hotkey CoreModules into the MX7000 provides

greater efficiency and performance for graphics workloads running in a modular environment.

- **Agile management**–The Dell EMC OpenManage Enterprise Modular software running on an MX7000 chassis delivers automated systems management and offers a secure and unified interface to manage the compute, storage, and fabric layers. Simplified end-to-end life cycle management accelerates operations and increases team effectiveness.
- **Flexible architecture**–The disaggregated architecture of the MX7000 chassis allows customers to allocate the on-demand ratios of shared pools of compute, storage, and fabric as needed.
- **Optimized vSAN performance**–The MX7000 chassis enhances vSAN performance with scalable, local-attached storage and a low-latency network. The high-speed, east-west network improves VMware vSphere vMotion responsiveness for VDI workloads running on vSAN clusters.

Document purpose

This white paper highlights the performance of Horizon 7 virtual workstations running in an MX7000 modular environment. The document provides the SPEC benchmark scores for 3D graphics applications that are used in the MEG industry such as Autodesk Maya and 3ds Max and evaluates the performance and efficiency of virtual workstations running these applications. It includes the test environment configuration and explains the test results from the performance testing carried out on the environment using the NVIDIA nVector tool running SPECperfview13 workloads.

Audience

This white paper is intended for decision-makers, managers, architects, consultants, developers, and technical administrators of IT environments in the MEG and other industries who want an understanding of the Horizon 7 virtual workstation solution running on the PowerEdge MX7000 modular environment.

We value your feedback

Dell Technologies and the authors of this document welcome your feedback on the solution and the solution documentation. Contact the Dell Technologies Solutions team with your comments.

Authors: Anand Johnson, Nicholas Busick

The following website provides additional VDI documentation: VDI Info Hub for Ready Solutions.

Technology overview

Introduction

This solution uses the following hardware and software components.

Dell EMC PowerEdge MX7000

The Dell EMC PowerEdge MX portfolio delivers a fully managed, high-performance system that frees up valuable IT resources and personnel so you can focus on innovation. The PowerEdge MX, with its kinetic infrastructure, offers real flexibility by disaggregating compute, storage, and networking fabric resources into shared pools that can be available for on-demand allocation and re-allocation. This kinetic infrastructure, which is designed for the modern software-defined data center, delivers optimal utilization, productivity, and efficiency. With an industry-leading no midplane layout and scalable fabric architecture, the PowerEdge MX supports new processor technologies, storage types, and connectivity innovations well into the future.

The PowerEdge MX7000 is a 7U form factor modular chassis that accommodates various compute and storage sled combinations. These combinations are connected by high-speed fabrics, share power and cooling, and are managed by embedded Open Manage Enterprise – Modular Edition systems management. The key capabilities of the MX7000 chassis are:

- Includes a 7U modular enclosure that has eight front-accessible, single-width bays that accommodate various compute and storage sleds
- Supports three I/O fabrics, each with redundant modules
- Hosts flexible blocks of server and storage resources while providing outstanding efficiencies through shared power, cooling, networking, I/O, and management within the chassis
- Provides nondisruptive upgrades by using the unique no midplane design that ensures easier future technology upgrades

For a VDI environment running on the MX7000 modular environment, Amulet Hotkey CoreModules enable graphics acceleration. The Dell EMC PowerEdge MX7000 modular chassis is equipped with up to 16 NVIDIA T4 GPUs contained in Amulet Hotkey CoreModules. This configuration brings high-density graphics acceleration to VDI environments. MX7000 compute sleds combine the performance of Intel Xeon Scalable processors and NVIDIA T4 GPUs with the flexibility, scalability, and efficiency of the MX7000 platform.

For more information, see the PowerEdge MX7000 documentation. For the technical specification of the MX7000, see Appendix A: Technical specifications of MX7000 modular platform.

The following figure shows the MX7000 platform:

Figure 1. Dell EMC PowerEdge MX7000

The following figure shows the MX740c compute sled:

Figure 2. MX740c compute sled

The MX740c compute sleds that we used for this testing are fully configurable, single-width compute sleds powered by up to two 28-core 2nd Generation Intel Xeon Scalable processors. Each PowerEdge MX740c server supports 24 DDR4 DIMM slots and up to six 2.5-inch SAS/SATA (HDD/SDD) drives. A maximum of eight MX740c servers can be installed in the PowerEdge MX7000 chassis. Flexible memory configurations of up to 1.5 TB (RDIMM) or 3 TB (LRDIMM) are available.

NVIDIA T4 GPU

The NVIDIA T4 is a universal GPU that serves various workloads. This GPU is based on Turing architecture and comes with 2,560 CUDA cores and a 16 GB DDR6 memory. The T4 operates at 70 W, providing higher energy efficiency and lower operating costs than its predecessors. It has a single-slot PCIe form factor. The T4 comes with an enhanced NVIDIA NVENC encoder that can provide higher compression and better image quality using the H.264 and H.265 (HEVC) video codecs. The NVIDIA T4 NVENC encoder provides up to 25 percent bit rate savings for H.265 and up to 15 percent bit rate savings for H.264.

NVIDIA Quadro vDWS

NVIDIA vGPU software creates virtual GPUs that are mapped to underlying physical GPUs. These virtual GPUs are assigned to each virtual workstation. NVIDIA Quadro vDWS licenses enable the powerful Quadro graphics driver for professional 3D applications. The driver, which is installed on each virtual machine (VM), enables graphics commands from the VMs to pass to the GPU hardware. Quadro vDWS comes with multiple profile size options that address different use cases and varying graphics requirements. Select the correct GPU and vGPU profile (frame buffer) size for your professional graphics use cases. Quadro vDWS supports applications with the highest resolution graphics and enables professionals to work with multiple 4K, 5K and 8K displays. Quadro vDWS software supports up to four 4K displays or two 8K displays, enabling creative professionals in the MEG industry to work on graphics easily.

Amulet Hotkey CoreModule

Amulet Hotkey CoreModules enable GPU acceleration in a PowerEdge MX modular environment. A PowerEdge MX740c compute sled within the Dell EMC PowerEdge MX7000 communicates over an Amulet Hotkey Mezzanine card to an Amulet Hotkey CoreCartridge in the CoreModule. The CoreCartridge houses up to two NVIDIA T4 GPUs. The Amulet Hotkey Mezzanine card enables a direct GPU PCIe pass-through connection to the compute sleds. A single CoreModule can provide GPU acceleration for up to eight MX740c compute sleds. The CoreModule is installed in fabric B1 or B2 of the MX7000 chassis. With two fully populated CoreModules in a 7U form-factor, the MX7000 chassis can host up to sixteen NVIDIA T4 GPUs. The Amulet Hotkey CoreModule product line is supported by the mission-critical expertise of the Amulet Hotkey Global Support organization.

The following figure shows the Amulet Hotkey CoreModule for MX7000:

Figure 3. Amulet Hotkey CoreModule for MX7000

VMware Horizon

VMware Horizon 7 provides the centralized management, agility, and simplicity that is required for your virtual desktop infrastructure. With Horizon 7, your workstations reside inside the data center premises, which makes the provisioning, maintenance, and recovery of virtual workstations easier. Horizon 7 with VMware Just-in-Time Management Platform (JMP) can provision and deliver virtual desktops and applications in a fast, flexible, and personalized manner. JMP uses Instant Clones for ultrafast provisioning of desktops, App Volumes for real-time application delivery, and Dynamic Environment Manager for contextual policy management to deliver an experience with the simplicity of non-persistent management.

When you implement VDI using Horizon 7, the IP of your organization is secure inside your data center premises. A lightweight Horizon 7 client is installed on the endpoint devices that communicate with virtual workstations in the data center. The Horizon 7 client is compatible with most devices on the market. It provides flexibility for employees working from home, as well as contractors and partners working from any device anywhere, thus increasing collaboration and productivity.

For graphics-accelerated VDI workloads, Dell Technologies recommends using the Blast Extreme Display protocol, which provides an enhanced remote session experience for professional graphics applications, even in a low latency network. The Blast Extreme protocol supports the H.264 and H.265 codecs, which can encode the graphics content from a virtual workstation display. NVIDIA T4 GPUs come with an advanced NVENC encoder. This encoder can offload H.264 or H.265 encoding from server processors, providing lower latency and better performance for professional graphics applications.

Test environment and configuration

Test environment

This section describes the hardware and software components that we used to validate the solution.

We built a three-node VMware vSAN cluster using Dell EMC MX740c compute sleds for this testing, as shown in the following table:

Table 1. MX7000 test environment

Node number	Use	Model
Node1	Management	MX740c
Node 2	Compute	MX740c
Node 3	Workstation testing performed by our partner, Amulet Hotkey. See Appendix B: Amulet Hotkey Modular Workstation–Data Center Optimized.	MX740c

Hardware component configuration

The following table shows the hardware components that we used in the solution configuration:

Table 2. MX740c hardware component configuration for the compute node

Enterprise platform	CPU	Memory	RAID controller	Drive config	Network	GPU
MX740c	Intel Xeon Gold 6254 (18 Core @ 3.10 GHz)	384 GB @2933 MT/s	Dell HBA 330 Adapter	2 x 120 GB M.2 1x 800 GB SSD (WI) 2x 1.92 TB SSD (RI)	Qlogic 25 GbE 2P rNDC	2 x NVIDIA T4 GPU cards per compute sled

Software component configuration

The following table lists the software component version details:

Table 3. Software components

Component	Description/version
Hypervisor	VMware vSphere ESXi 6.7 U3
Broker technology	VMware Horizon 7.10.1
Broker database	Microsoft SQL 2014
Management VM operating system	Microsoft Windows Server 2016
Virtual desktop operating system	Microsoft Windows 10 Enterprise 64-bit 1803
Office application suite	Microsoft Office 2016 Professional Plus
Test Software	NVIDIA nVector (SPECviewperf 13) + nVector Lite 1.0

Component	Description/version
Thin Client (End Point)	Dell Wyse 5070 WIE10 1809 + Pentium Silver J5005 processor (4C @ 1.5 GHz, 10W TDP)

Virtual workstation configuration

The following table shows the configuration of the VDI virtual workstation VM:

Table 4. Workstation VM configuration

vCPUs	ESXi memory configured	NVIDIA vGPU profile	Screen resolution	Operating system
6	32 GB	T4-8Q	1920 X 1080	Windows 10 Enterprise 64-bit

Performance testing process

Overview

To ensure the optimal combination of end-user experience (EUE) and cost-per-user, we carried out this performance analysis and characterization (PAAC) test using an industry best-practices methodology that monitors both hardware resource utilization parameters and EUE during load-testing.

Our testing used NVIDIA nVector, a performance testing tool from NVIDIA for benchmarking VDI workloads. The nVector tool creates a load on the system by simulating a workload that matches a typical VDI environment. It was designed to assess the experience at the endpoint device rather than the response time of the virtual desktop. The tool also captures the corresponding performance metrics that quantify user experience, including image quality, frame rate, and user latency from the endpoints. These metrics, when combined with resource utilization information from the servers under test, enable IT teams to assess their VDI graphics-accelerated environment needs.

We tested multiple runs for each user load scenario to eliminate single test bias. We used a pilot run to validate that the solution was functioning as expected and we validated that testing data was being captured. We then tested subsequent runs to provide data that confirmed that consistent results were observed.

To confirm true EUE, we logged into a VDI session and completed several tasks typical of a normal user workload. We feel confident that this small incremental load on the system did not significantly impact our ability to provide reproducible results. While the assessment undoubtedly is subjective, it helps to provide a better understanding of the EUE under high load. It also helps to assess the reliability of the overall testing data.

Load generation

The default workload for the NVIDIA nVector suite is the Knowledge Worker profile. However, we used the more demanding SPECviewperf 13 professional graphics workload with the nVector tool for this performance testing. The SPECviewperf 13 benchmark measures the 3D graphics performance of systems running under the OpenGL and Direct X application programming interfaces. Because SPECviewperf 13 is a standard

benchmark for measuring graphics performance for professional applications, including 3ds Max and Maya, it was the most appropriate choice for this testing effort.

The SPECviewperf 13 workload uses a series of viewsets taken from independent software vendor (ISV) applications to characterize the graphics performance of a physical or virtual workstation. It uses GPU-enabled launchers or endpoint VMs to connect a specified number of users to the available virtual desktops within the environment. After the simulated user is connected, a login script starts a test workload script on the user environment that is configured by the login script. Each launcher or endpoint system can launch connections to a single "target" machine (such as a VDI desktop). Mapping takes place from the endpoint to the desktop used by a specific user. The launchers and desktops are deployed and monitored using the nVector management VM where the framework runs. The data collected during the test is analyzed afterwards.

The SPECviewperf 13 benchmark's workloads, called viewsets, represent graphics content and behavior typical of actual applications. We used the Autodesk 3ds Max 2016 (3dsmax-06) and Autodesk Maya 2017 (Maya-05) viewsets that are available within the benchmark. For more information about the SPECviewperf 13 viewsets, see the SPEC documentation.

We used the following login and boot scheduling paradigm:

- Simulated users are logged in every 5 seconds
- All user desktops are prebooted before logins begin
- All Windows 10 virtual desktops were configured with Windows Defender anti-virus

Resource monitoring

The NVIDIA nVector SPECviewperf 13 test with 3ds Max and Maya viewsets generated the reported SPEC benchmark scores. SPEC scores indicate the speed and performance of the simulated graphics applications. A good SPEC score is usually associated with a system with good GPU acceleration capabilities.

We used VMware vCenter to gather key host utilization metrics, including CPU, memory, disk, and network usage from the compute host during each test run. We have not included results from the management host server in our reporting. The management server was monitored during testing to ensure it was not responsible for any testing bottlenecks. We used nVector to run a script on the ESXi host to gather GPU-related resource usage. We ran the script before starting the test run and stopped it when the test was completed. The script contained NVIDIA System Management Interface commands to query each GPU and log GPU utilization and GPU memory utilization to an output file for analysis.

Measuring user experience

We ran the NVIDIA nVector Lite tool while testing with the NVIDIA nVector SPECviewperf 13 workload for test case 3 to assess the user experience from the endpoints. We performed the test this way because endpoint monitoring is normally only available with the NVIDIA nVector full suite that contains the default Knowledge Worker workload. For the SPECviewperf 13 workload, we needed the nVector Lite tool and the nVector tool to capture endpoint metrics. The NVIDIA nVector Lite tool measured the following three key metrics from the thin client:

Metric 1: Image Quality–NVIDIA nVector uses a lightweight agent on the VDI desktop and the client to measure image quality. These agents take multiple screens captures on

the VDI desktop and on the thin client for later comparison. The structural similarity (SSIM) of the screen capture taken on the client is computed by comparing it to the screen capture taken on the VDI desktop. When the two images are similar, the heatmap reflects more colors above the spectrum shown on the right with an SSIM value closer to 1.0 (see Figure 4). As the images become less similar, the heatmap reflects more colors down the spectrum with a value of 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 users.

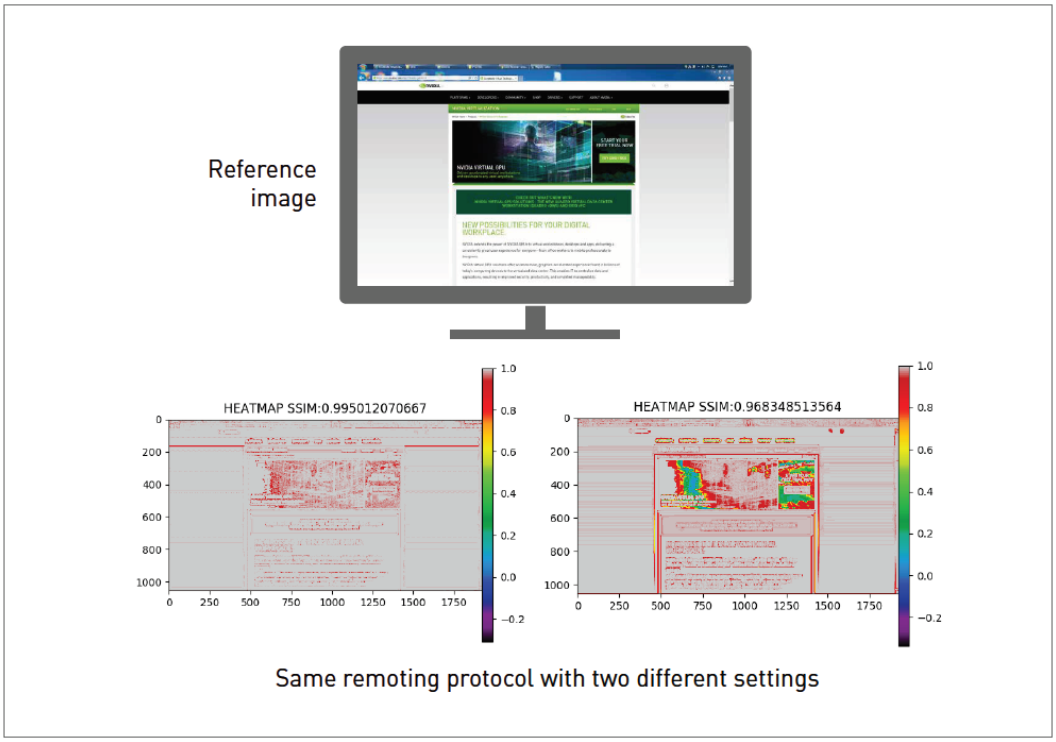

Figure 4. SSIM as a measure of image quality

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 duration of the workload, NVIDIA nVector collects data on the frames per second (fps) 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 for reporting. A larger fps indicates a more fluid user experience.

Metric 3: End-user latency−The end-user latency metric defines the level of response of a remote desktop or application. It measures the duration of any lag that an end-user experiences when interacting with a remote desktop or application.

Results summary

Overview

This section summarizes the SPECviewperf 13 benchmark scores obtained from the three test cases that we ran. A higher SPEC score indicates greater speed and better performance for the simulated graphics application. We performed the first two test cases by running four concurrent virtual workstations on the MX740c compute host and the third test case by running a single workstation on the MX740c compute host. In the first test case, we connected to one of the virtual workstations through a thin client and collected additional endpoint experience metrics with the NVIDIA nVector Lite tool. The three test cases that we performed were:

- Test Case 1: NVIDIA nVector SPECviewperf 13 + NVIDIA nVector Lite, Multi VM test results
- Test Case 2: NVIDIA nVector SPECviewperf 13 Multi VM test results
- Test Case 3: NVIDIA nVector SPECviewperf 13 Single VM test results

Test Case 1: NVIDIA nVector SPECviewperf 13 + nVector Lite, Multi VMs test results

For this test case, the MX740c GPU-enabled compute host was populated with four vGPU-enabled virtual workstation using the NVIDIA T4-8Q vGPU profile. Three of the endpoints used in this test were VMs, while the fourth endpoint was a Wyse 5070 Thin Client. We used the Wyse Thin Client to collect nVector Lite endpoint metrics, including image quality, frame rate, and end-user latency. Because the nVector Lite tool cannot collect endpoint metrics from virtual machine endpoints, we used the thin client option. You can compare SPEC scores from our performance testing with other published scores on the SPEC website. SPEC scores from our tests indicate an excellent graphics performance by the virtual workstations. The following table summarizes the SPEC scores for this test case (larger scores indicate greater speed):

Table 5. SPEC score summary for Test Case 1

VM name	3ds Max score	Maya score
Desktop-1	60.896	119.042
Desktop-2	62.154	129.522
Desktop-3	61.706	117.536
Desktop-4 (Thin Client)	71.19	118.94
Average	63.99	121.26

Test Case 2: NVIDIA nVector SPECviewperf 13, Multi VMs test results

For this test case, we populated the MX740c GPU-enabled compute host with four vGPU-enabled virtual workstations, each with an NVIDIA T4-8Q vGPU profile. The following table summarizes the SPEC scores. For host utilization metric graphs for this test case, see Appendix C: NVIDIA nVector, SPECviewperf 13, Multi VMs test.

Table 6. SPEC score summary for Test Case 2

VM name	3ds Max score	Maya score
Desktop-1	62.513	118.403
Desktop-2	62.822	114.088
Desktop-3	61.902	117.315
Desktop-4	60.703	115.133
Average	61.985	116.23475

Test Case 3: NVIDIA nVector SPECviewperf 13, Single VM test results

For this test case, we populated the MX740c GPU-enabled compute host with only one vGPU-enabled virtual workstation that had an NVIDIA T4-8Q vGPU profile. The following table summarizes the SPEC scores. For host utilization metric graphs for this test case, see Appendix D: NVIDIA nVector, SPECviewperf 13, Single VM test.

The single virtual workstation was able to consume more than its default allocated resources due to the “Best Effort” vGPU scheduler being used. Because there was only one VM per physical GPU, there was no contention for GPU resources. For additional information, see NVIDIA's virtual GPU software documentation.

Also, because there was no CPU oversubscription, the single virtual workstation got more CPU time slices. Because no concurrent VMs were running on the host, the single workstation was able to consume allocated resources in a way that delivered a better performance than VMs running concurrently on the host. As a result, you can see a relatively higher SPEC score in this test compared to the multi VM tests.

Table 7. SPEC score summary Test Case 3

VM name	3ds Max score	Maya score
Desktop-1	126.755	233.107

Note: You can find the results and raw data for the SPECperfview 13 benchmark testing here: https://dell.box.com

NVIDIA nVector SPECviewperf 13 + nVector Lite, 3ds Max, Multi VMs test

This section describes the details of Test Case 1, which we ran with a SPECviewperf 13 Autodesk 3ds Max workload. We analyze the SPEC scores, and host and endpoint resource utilization.

We provisioned four virtual workstation VMs on the MX740c compute host that was enabled with GPUs. NVIDIA T4-8Q vGPU profiles were configured for each virtual workstation. Three virtual workstations were connected to virtual machine endpoints running the NVIDIA nVector automated workload, and one of the virtual workstations was connected to the Wyse 5070 thin client. For the virtual workstation connected to the thin client, we ran the test manually and collected EUE metrics using the NVIDIA nVector Lite tool.

We tested the virtual workstations with the configuration shown in the following table:

Table 8. Workstation VM configuration

Configuration	Value
vCPU	6
vMemory	32768
HardDisk	120
GPU	grid_t4-8q
GPUDriverVersion	442.06
FRL	disabled
vSYNC	default
vDAVersion	7.10.1
DirectConnectVersion	7.10.1
CPUAffinity	Unset
Screen Resolution	1920 x 1080
QuantityMonitors	1

For this test, we sized virtual workstations to use the maximum available GPU frame buffer of 32 GB in the compute host with four virtual workstations running an 8 GB vGPU profile each. Two NVIDIA T4 GPUs, each with a frame buffer of 16 GB, were enabled per compute host. The following table shows the SPECviewperf 13 benchmark score for the test:

Table 9. 3ds Max SPEC score for test case 1

VM name	3ds Max score
Desktop-1	60.896
Desktop-2	62.154

VM name	3ds Max score
Desktop-3	61.706
Desktop-4 (Thin client)	71.19
Average	63.99

The following table shows the compute host utilization metrics during the test:

Table 10. Average host utilization metrics

Workload	Density per host	Average CPU %	Average GPU %	Average memory consumed	Average memory active	Average net Mbps per user
SPECviewperf 13 3ds Max	4	17%	77%	128 GB	128 GB	18.85 Mbps

The following graphs explain the host utilization metrics that we obtained when we ran the SPECviewperf 13 3ds Max workload and the end-user experience metrics, including image quality, frame rate, and end-user latency collected from the nVector Lite tool.

The following figure depicts the CPU core utilization on the compute host enabled with GPUs running four virtual workstation machines. The average CPU core utilization during the test was approximately 17 percent, with a peak utilization of 23 percent. The SPECviewperf 13 workload is more graphics-intensive than CPU-intensive and is designed to measure the GPU performance. For that reason, the low CPU utilization shown in the following graph is to be expected:

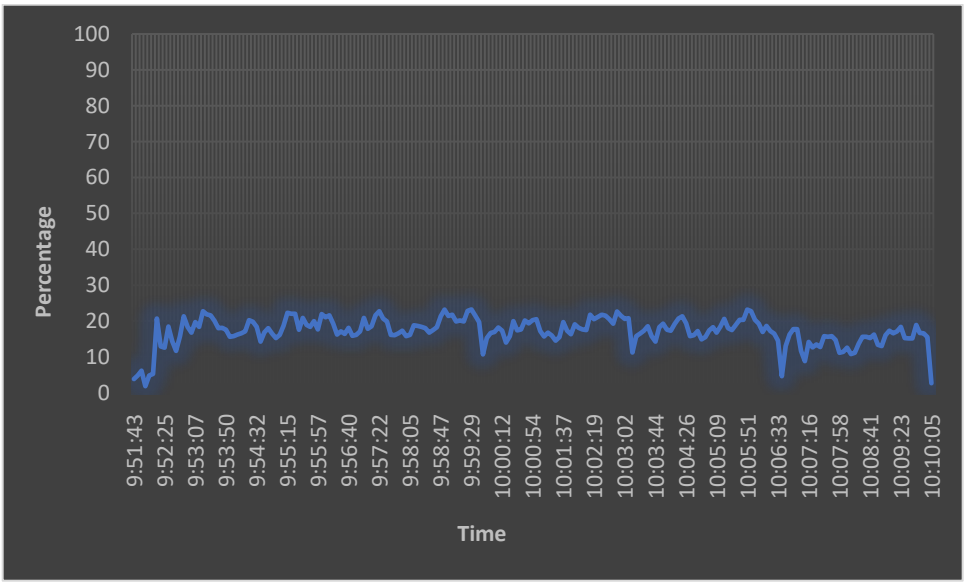

Figure 5. CPU core utilization

The following figure shows the GPU utilization for the system under test. GPU metrics were gathered through the NVIDIA System Management Interface on the VMware ESXi host. As shown in the graph, GPU utilization was high for this test and spiked to 100 percent during the testing. This high GPU utilization is expected because the SPECviewperf 13 workload is designed to stress the GPU to measure its performance.

Figure 6. GPU utilization

The following figure shows the free memory that is available on the compute host during the test. There were not many variations in memory usage throughout the test because all vGPU-enabled VM memory was reserved. There was no memory ballooning or swapping on either host, indicating that there were no memory constraints on the host.

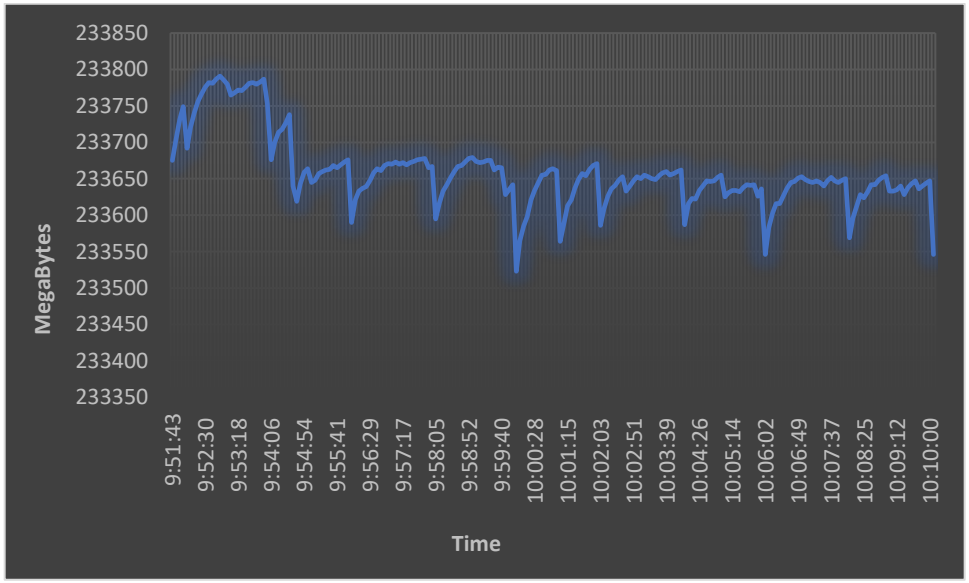

Figure 7. Free memory

The following figure shows the network utilization of the tested system for the duration of the test. A spike of 695 MBits was recorded. There was more than enough bandwidth available for the workload and the display protocol to accommodate the workload. Each server had a single 25 GbE network card (25,000 MBits) connected to it.

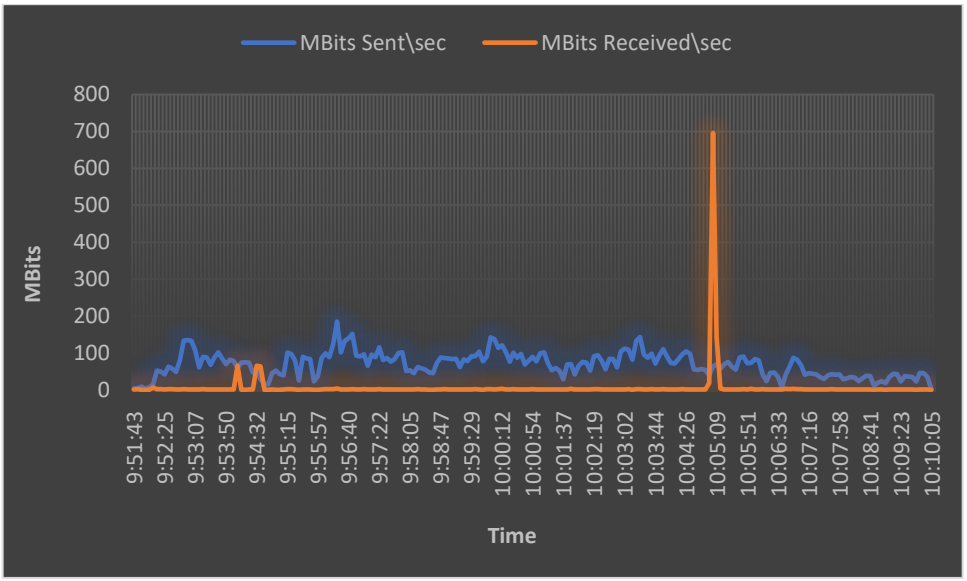

Figure 8. Network usage

nVector Lite Endpoint Metrics

The image quality on the Wyse 5070 thin client was captured using nVector Lite and the SSIM heatmap index. This testing takes screenshots on the endpoint (the Wyse Thin Client) and the virtual workstation and makes comparisons of how the display protocol is performing. As shown in the following figure, the image quality was approximately 94 percent. This SSIM measurement value shows that the image quality was not degraded significantly by the remoting protocol.

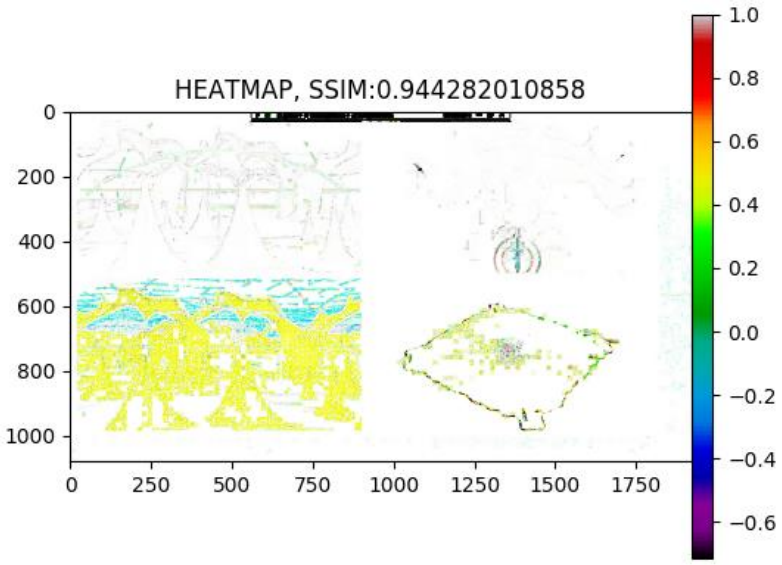

Figure 9. Image quality–Heatmap

The image frame rate was captured on the virtual workstations and the Wyse 5070 thin client. The frames that were measured at the thin client are those frames that the remote protocol displayed. The maximum remoted frames that the display protocol displays was set to 60 frames. The nVector tool measured FPS and Flip rate, which are defined as follows:

- FPS–The frames per second as measured from the GPU
- Flip rate–The remoted frames per second as measured from the Blast Extreme display protocol

Figure 10 shows a comparison of FPS and flip rate. The display protocol setting caps the flip rate at 60. For a good user experience, the maximum configured flip rate should be no lower than the frame rate and the maximum observed flip rate should be equal to the frame rate.

Figure 10 shows the flip rate to be close to the frame rate at all times, indicating a good user experience.

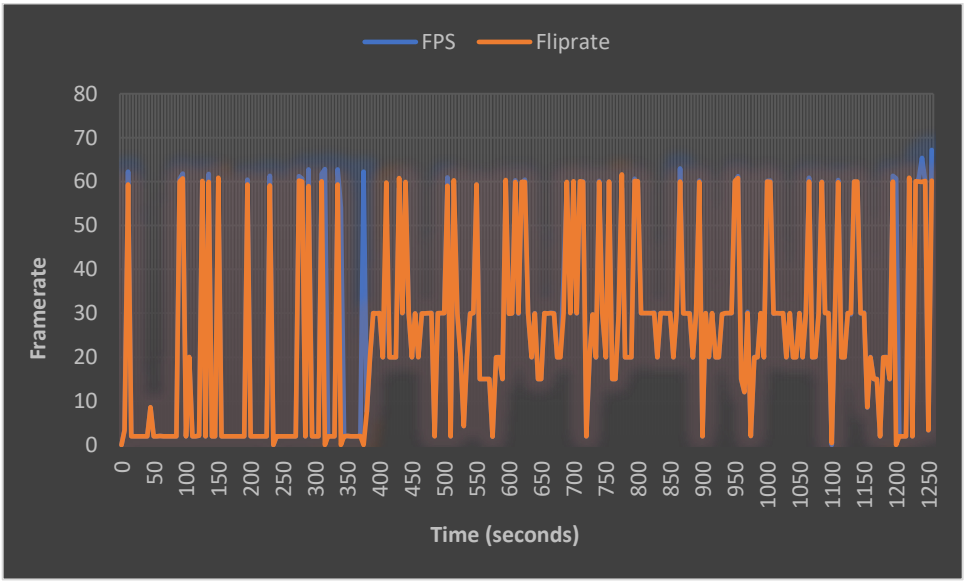

Figure 10. Image frame rate

The end-user latency metric gives an indication to VDI administrators of how responsive the VDI session is at the user's endpoint. The average end-user latency measured by the nVector Lite tool during the test was 85.75 ms. This low latency figure indicates that the remote session was very responsive.

NVIDIA nVector + nVector Lite, SPECviewperf 13, Maya, Multi VMs test

This section describes the details of Test Case 1, which we ran with the SPECviewperf 13 Autodesk Maya workload. We analyze the SPEC scores and the host and endpoint resource utilization.

We provisioned four virtual workstation VMs on the MX740c compute host that was enabled with GPUs. We configured NVIDIA T4-8Q vGPU profiles for each virtual workstation. Three virtual workstations were connected to virtual machine endpoints running the NVIDIA nVector automated workload, and one of the virtual workstations was connected to the Wyse 5070 thin client. For the virtual workstation connected to the thin client, we ran the test manually and collected EUE metrics using the NVIDIA nVector Lite tool.

The following table shows the virtual workstation configuration that we used for testing:

Table 11. Workstation VM configuration

Configuration	Value
vCPU	6
vMemory	32768
HardDisk	120
GPU	grid_t4-8q
GPUDriverVersion	442.06
FRL	disabled
vSYNC	default
vDAVersion	7.10.1
DirectConnectVersion	7.10.1
CPUAffinity	Unset
Screen Resolution	1920 x 1080
QuantityMonitors	1

For this test, we sized virtual workstations to use the maximum available GPU frame buffer of 32 GB on the compute host with four virtual workstations running an 8 GB profile each. We enabled two NVIDIA T4 GPUs, each having a frame buffer of 16 GB, per compute host. The following table gives the SPECviewperf 13 benchmark score for the testing. The performance of this system can be compared with the other benchmark scores available on the SPEC website.

Table 12. Maya SPEC score for Test Case 1

VM name	Maya score
Desktop-1	119.042
Desktop-2	129.522
Desktop-3	117.536
Desktop-4 (Thin Client)	118.94
Average	121.26

The following table shows the average compute host utilization metrics during the test:

Table 13. Average host utilization metrics

Workload	Density per host	Average CPU %	Average GPU %	Average memory consumed	Average memory active	Average net Mbps per user
SPECviewperf 13 Maya	4	18.77%	79.42%	114 GB	114 GB	9.63 Mbps

The following graphs show the host utilization metrics while running the SPECviewperf 13 Maya workload and the EUE metrics, including image quality, frame rate, and end-user latency collected from the nVector Lite tool.

The following figure shows the CPU core utilization on the compute host enabled with GPUs running four virtual workstation machines. The average CPU core utilization during the test was 18.76 percent, with a peak utilization of 24.72 percent. The SPECviewperf 13 workload is typically more graphics-intensive than CPU-intensive and is designed to measure the GPU performance. Therefore, the low CPU utilization shown is expected.

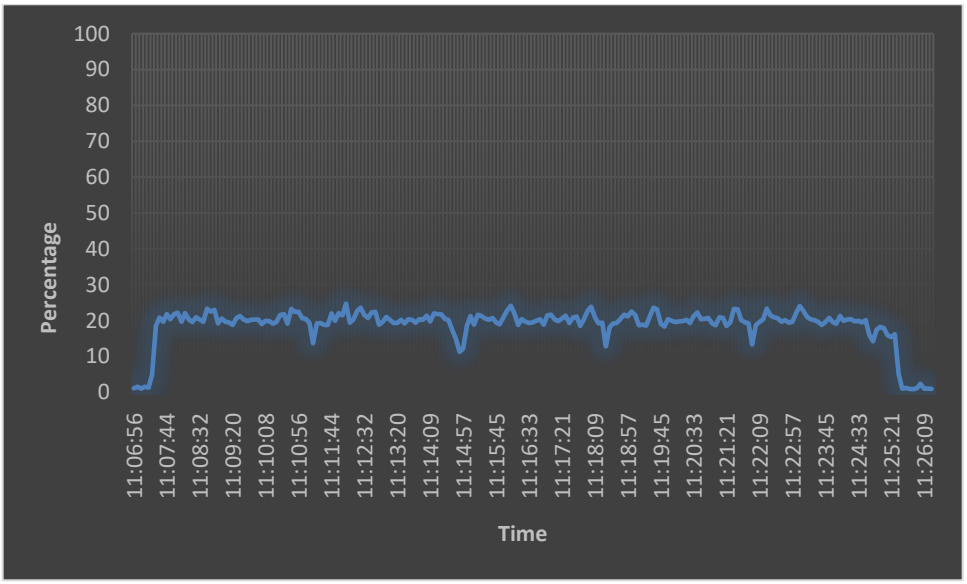

Figure 11. CPU core utilization

The following figure shows the GPU utilization for the system under test. We gathered GPU metrics through the NVIDIA System Management Interface on the VMware ESXi host. GPU utilization spiked to a maximum during the test. The high GPU utilization is expected because the SPECviewperf 13 workload is designed to measure the performance of the GPU. The average GPU utilization across the GPUs during testing was 79 percent.

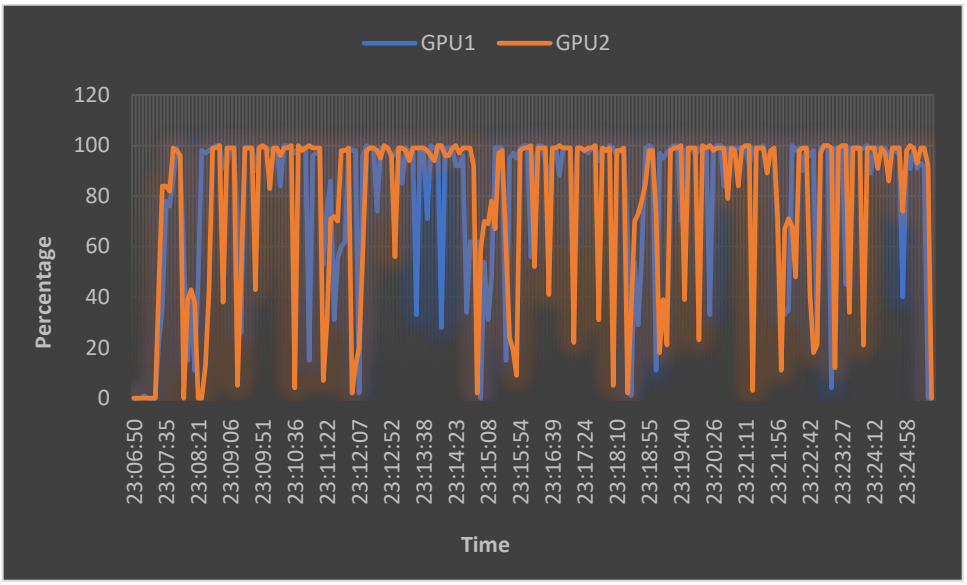

Figure 12. GPU utilization

The following figure shows the free memory that was available on the compute host during the test. There were not many variations in memory usage throughout the test because all vGPU-enabled VM memory was reserved. There was no memory ballooning or swapping on either host, indicating that there were no memory constraints on the host.

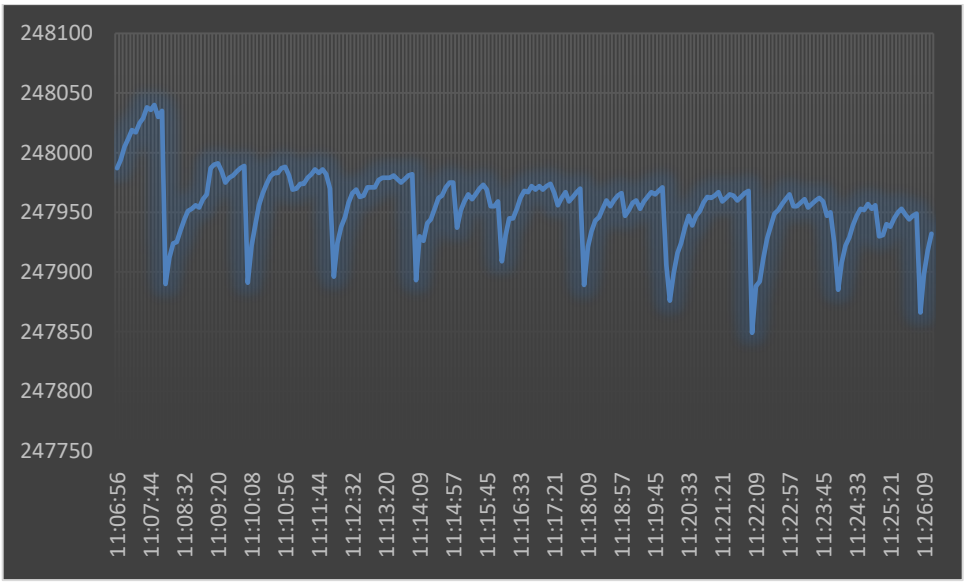

Figure 13. Free memory

The following figure shows the network utilization of the tested system for the duration of the test. The peak network utilization recorded during the test was 111 MBits. There was more than enough bandwidth available for the workload and the display protocol to accommodate the workload. Each server had a single 25 GbE (25,000 MBits) network card connected to it.

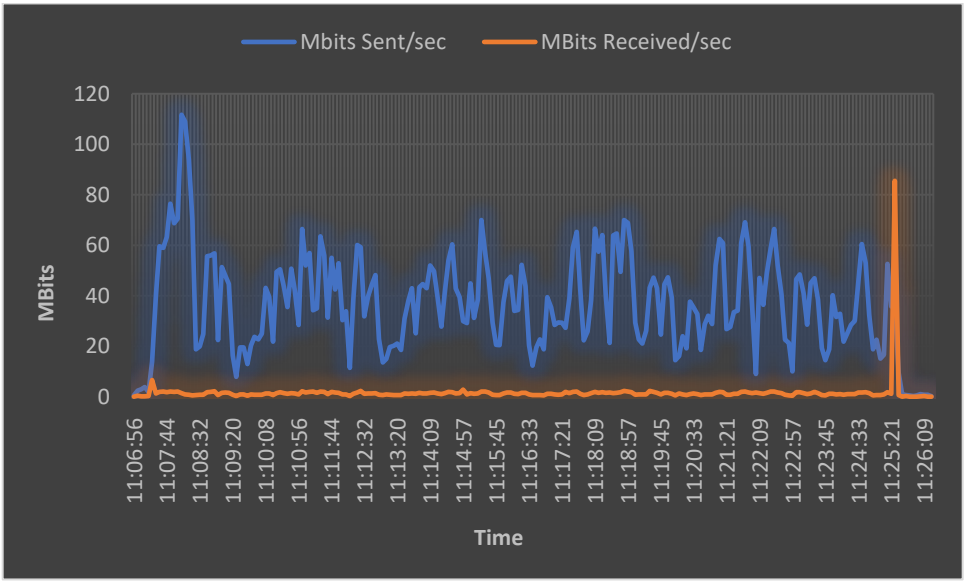

Figure 14. Network usage

nVector Lite Endpoint Metrics

The image quality on the Wyse 5070 thin client was captured using nVector Lite and the SSIM heatmap index. This testing takes screenshots on the endpoint (the Wyse Thin Client) and the virtual workstation and makes comparisons of how the display protocol is performing. The image quality recorded was approximately 91 percent, as shown in the following figure. This SSIM measurement value shows that the image quality was not degraded significantly by the remoting protocol.

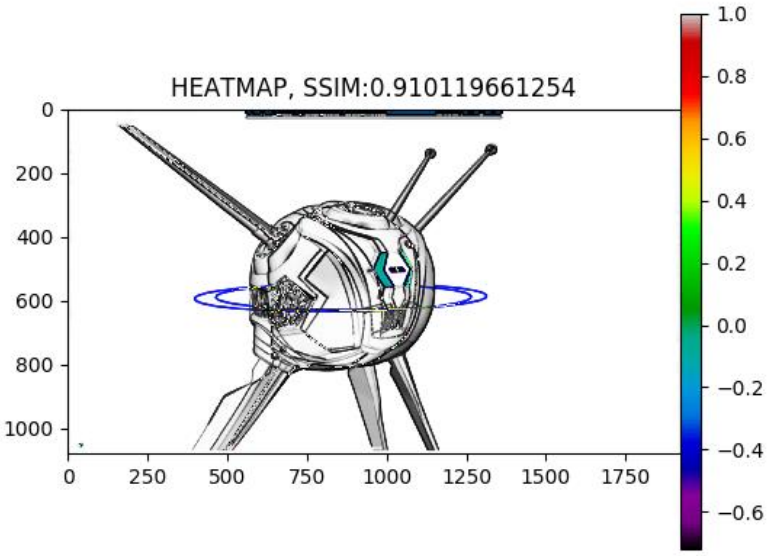

Figure 15. Image quality–Heatmap

The image frame rate was captured on the virtual workstations and the Wyse 5070 Thin Client. The frames that were measured at the thin client are those frames that the remote protocol displayed. The maximum remoted frames that the display protocol displays was set to 60 frames. The nVector tool measured FPS and flip rate, which are defined as follows:

- FPS–The frames per second as measured from the GPU
- Flip rate–The remoted frames per second as measured from the Blast Extreme display protocol

Figure 16 shows a comparison of FPS and flip rate. The display protocol setting caps the flip rate at 60. For a good use experience, ensure that the maximum configured flip rate is no lower than the frame rate and the maximum observed flip rate is equal to the frame rate.
Figure 16 shows the flip rate to be close to the frame rate at all times, indicating a good user experience.

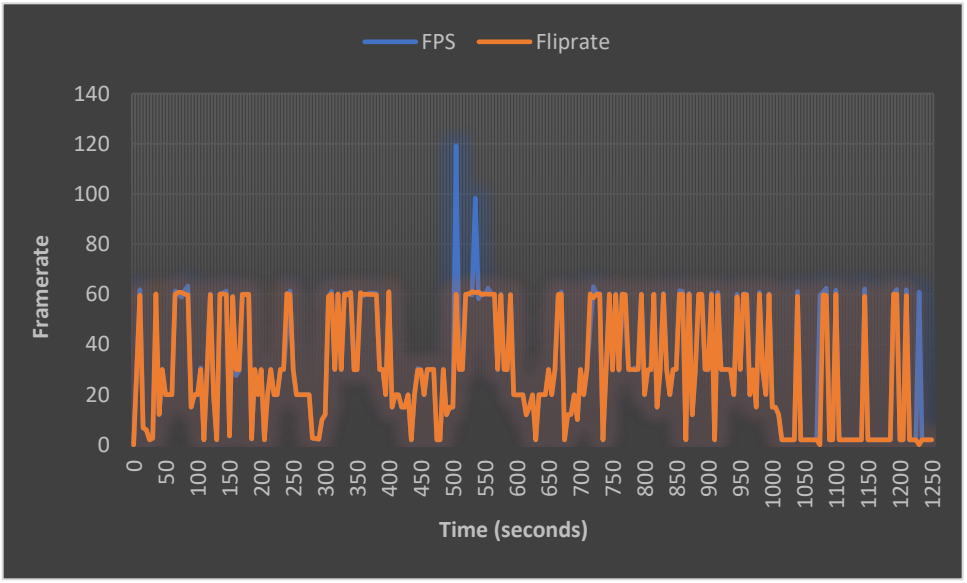

Figure 16. Image frame rate

The end-user latency metric gives an indication to VDI administrators of how responsive the VDI session is at the user's endpoint. The average end-user latency measured by the nVector Lite tool during the test was 140.48 ms. This value is higher than the value recorded during the 3ds Max test. However, this value was not high enough to impact the responsiveness in the endpoints.

The user experience was measured in a subjective manner by logging into a user session and performing a series of tasks that indicate responsiveness. Some of these tasks include opening multiple applications and observing the launch time, drawing a figure 8 in a graphics editor, and dragging windows around the desktop to observe smoothness. These tasks performed as expected and the experience was determined to be excellent with no observable latency or delays.

Guidance for configuring VMware Horizon virtual workstations on MX7000

The Dell Technologies Ready Solutions for VDI team recommends the virtual workstation configuration in the following table for running 3ds Max and Maya applications in an MX7000 modular environment. This configuration provides the optimal performance for running applications such as 3ds Max and Maya for the MX7000 hardware configuration that is listed in Table 2.

Table 14. Virtual workstation configuration for MX7000

Users per MX740c node	vCPUs	ESXi memory configured	Hard disk	vGPU profile	Screen resolution	Operating system
4	6	32 GB	> 250 GB	T4-8Q	1920 X 1080	Windows 10 Enterprise 64-bit

We recommend that you consider the following factors when configuring a VMware Horizon virtual workstation on Dell EMC hardware:

MX740C Host Configuration:

1. On the MX740c compute host, set the **System Profile** option in **BIOS > System Profile Settings** to **Workstation Performance**.

2. If the ambient temperature is not cool enough, consider changing the iDRAC **Thermal Profile Optimization** to **Maximum Performance**. This setting reduces the probability of thermal throttling of the CPU and GPUs. For more information, see the Integrated Dell Remote Access Controller 9 Version 3.15.15.15 User's Guide.

3. Size the virtual CPUs of the workstation VMs within the NUMA node boundaries of the physical processors. For more information, see Virtual Machine vCPU and vNUMA Rightsizing–Rules of Thumb.

4. Choose and set the appropriate GPU scheduler for your workload and use case. The available GPU scheduler options are:

   - Best Effort (Default)
   - Equal Share
   - Fixed Share

   For more information, see NVIDIA's virtual GPU software documentation.

Workstation VM Configuration:

1. Set the VMware Horizon Blast **Max Frame Rate** to **60** using the registry key or GPO. For more information, see VMware Blast Policy Settings.

2. Using the VMware Horizon Performance Tracker in the Virtual Workstation VM, ensure that the Protocol is **Blast** and the Encoder Name is **NVIDIA NvEnc H264**. The recommended protocol for running graphics workloads in a VMware Horizon environment is Blast Extreme.

3. Use the latest version of the VMware Horizon Client. At the time of publication of this white paper, the latest version was 5.4.2.

Configuration for Knowledge Worker virtual desktops on MX7000

The Dell Technologies Ready Solutions for VDI team recommends using the virtual workstation configuration in the following table for the high-end Knowledge Worker use cases, which benefit from graphics acceleration. This configuration provides the optimal performance for the MX7000 hardware configuration that is listed in Table 2.

Table 15. Virtual workstation configuration for MX7000

Users per MX740c node	vCPUs	ESXi memory configured	Hard disk	vGPU profile	Screen resolution	Operating system
32	4	8 GB	60 GB	T4-1B	1920 X 1080	Windows 10 Enterprise 64-bit

Network guidance for an MX7000 VDI environment

The MX scalable fabric architecture ties multiple MX7000 chassis into a single network domain that behaves like a single logical chassis from a networking perspective. A scalable fabric consists of two main components: a pair of MX9116n Fabric Switching Engines (FSE) and additional pairs of MX7116n Fabric Expander Modules (FEM) that are used to connect remote chassis to the FSEs. This is a hardware-enabled architecture and it applies if the switch is running in either Full Switch or Fabric modes.

A total of ten MX7000 chassis are supported in a scalable fabric. The following figure shows the networking configuration of the MX7000 chassis used in this solution. Fabric slots A1/A2 are used for VDI compute infrastructure networking. The B1/B2 fabric slots are used by Amulet Hotkey CoreModules that contain the NVIDIA T4 GPUs. Fabric slots C1/C2 are designated for optional Fiber Channel switches. Fiber Channel options leveraging slots C1/C2 for external block storage are possible but are not within the scope of this document.

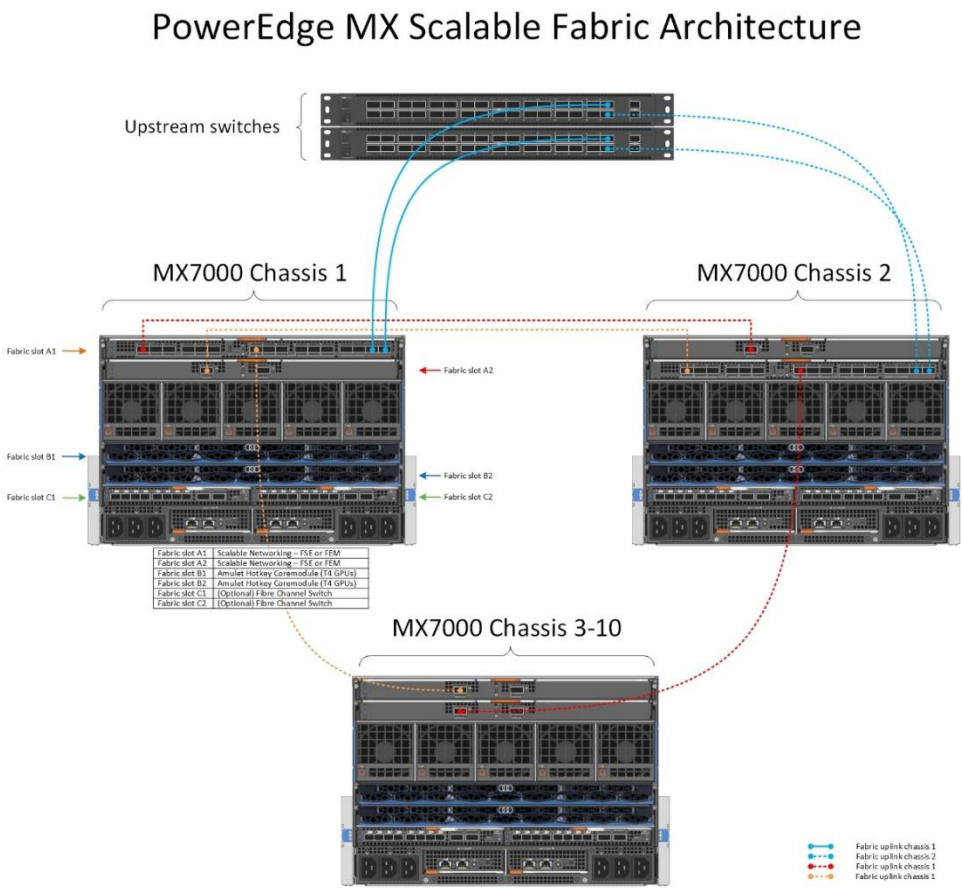

Figure 17. MX7000 network diagram

Amulet Hotkey CoreModules use the fabric B1 and B2 slots in the MX7000 chassis, meaning that they are not available for networking purposes. For this reason, the networking architecture for the MX7000 chassis hosting the VDI compute nodes must use a single fabric design leveraging slots A1 and A2. For more information about networking configurations, see the PowerEdge MX I/O Guide.

Conclusion

Summary

The VDI virtual workstation solution on an MX7000 modular chassis that is accelerated by Amulet Hotkey CoreModules equipped with NVIDIA T4 GPUs offers an excellent platform to run professional graphics applications, including Autodesk 3ds Max and Maya. The high-performing CPU and GPU acceleration that the solution provides is well suited for the application workloads that are used in the MEG industry. These virtual workstations provide an experience that is on a par with more expensive and less flexible physical workstations. The solution simplifies the provisioning and management of virtual workstations and lets your IT team focus on innovation.

We have provided SPEC benchmark scores for the Autodesk 3dxmax and Maya applications that are used widely in the MEG industry. SPEC scores for our virtual workstations were excellent during our testing. You can compare these scores with other scores that are available on the SPEC website. We observed that application performance and user experience were excellent. In addition, the monitored host utilization and endpoint metrics from the testing were under the threshold.

The configuration in this white paper shows the preferred way to integrate an MX7000 modular chassis running NVIDIA T4 GPUs and Horizon 7 virtual workstations powered by NVIDIA Quadro vDWS. The Dell Technologies Ready Solutions for VDI team recommends this solution configuration for running the professional graphics applications that are typically used in the MEG industry. However, we recommend that you perform pilot tests to correctly size the virtual workstation environment according to your workload requirements.

References

Dell Technologies

The following Dell Technologies documentation provides additional information:

- *Dell PowerEdge MX7000 Technical Guide*
- *Dell PowerEdge MX I/O Guide*

VMware

The following documentation provides additional information:

VMware Horizon 7 Documentation

NVIDIA

The following documentation provides additional information:

NVIDIA Virtual GPU Software Documentation

Amulet Hotkey

The following documentation provides additional information:

Amulet Hotkey CoreModules

Appendix A: Technical specifications of MX7000 modular platform

Table 16. Technical specifications of MX7000

Feature	Specifications
Form factor	7U rack
Sled support	Up to 8 standard-height, or 4 double-wide, or 16 half-height sleds
I/O modules	3+3 (A/B/C). A and B direct I/O Ethernet interconnect. C is midplane-based Fiber Channel, or SAS interconnect
Power modules	6 x 3 kW
KVM	Integrated on the Management Module

Appendix B: Amulet Hotkey Modular Workstation–Data Center Optimized

Amulet Hotkey develops unique modular workstation solutions that provide powerful performance and availability, including the Amulet Hotkey CoreModule, which was developed in partnership with Dell Technologies OEM Embedded and Edge Solutions.

This Amulet Hotkey-delivered solution combines the powerful GPU acceleration of the Amulet Hotkey CoreModule with the density and efficiency benefits of the Dell PowerEdge MX modular architecture. The solution is supported by the mission-critical expertise of the Amulet Hotkey Global Support organization.

The Amulet Hotkey modular workstation solution allows professional users to connect to graphics- and compute-intensive applications any time, anywhere, using remote access devices.

The following section provides system configuration information and benchmark results obtained using SPECworkstation 3 and OTOY Octanebench. The SPECworkstation 3 tests compare remote access client scenarios using a commercially available mobile client and an Amulet Hotkey DXZ4 PCoIP zero client. Similar results would be expected if a Dell Wyse 5030 Zero Client or an Amulet Hotkey DXZC PCoIP zero client were used instead of the DXZ4 device.

The following table shows the system configuration for the Amulet Hotkey modular workstation:

Table 17. Amulet Hotkey modular workstation system configuration

Hardware component	Workstation	Zero client	Mobile client
System	Dell EMC PowerEdge MX740c	Amulet Hotkey DXZ4 PCoIP Zero Client	Microsoft Surface Book 2
CPU	2 x Intel Xeon Gold 6244 GPU @ 3.6 GHz	Teradici Tera2140	Intel Core i7-8650U @ 1.9 GHz
Memory	12 x 32 GB 2933MT DIMMs	512 MB DDR3 SDRAM	8 GB 1866 MHz LPDDR
Controller	PERC H730P MX	N/A	Embedded
Capacity SSDs	3 x 1.92 TB MU SSDs (RAID 5)	N/A	256 GB SSD
Boot device	BOSS card	32 MB Flash	N/A
Network	Qlogic 41232 (2 x 25 GE)	Gigabit Ethernet (RJ45)	Gigabit Ethernet (RJ45)
GPU	2 x NVIDIA T4 GPU	N/A	Intel UHD Graphics 620

Note: While the MX740c compute sled is configured for dual 8C CPUs, the SPECWorkstation 3.0.2 test runs were set to use only eight physical cores.

The following table shows the software components:

Table 18. Software components

Software component	Workstation	Zero client	Mobile client
Operating system	Windows 10 Enterprise Version 1909	Green Hills ThreadX RTOS	Microsoft Windows 10 Version 1909
Remote Access	Teradici Cloud Access Software Plus (CAS+) Graphics Agent	Teradici Tera2 Firmware	Teradici Cloud Access Software Client

SPECworkstation 3 with Remote Access

The SPECworkstation 3 benchmark measures key aspects of workstation performance based on diverse professional applications. The SPECworkstation 3 benchmark introduced in October 2018 includes updates such as:

- Workloads that reflect changes in updated versions of Blender, Handbrake, Python, and Luxrender applications
- GPU-accelerated workloads based on the Luxrender and Caffe applications
- Graphics workloads from the SPECviewperf 13 benchmark, including new viewsets representing Autodesk Maya, PTC Creo, and medical applications

The workstation benchmarks cover both 2K and 4K SPECworkstation tests with three remote access scenarios including:

- No remote access “baseline”
- Amulet Hotkey DXZ4 PCoIP Zero Client
- Commercially available mobile client

The remote access client devices used in this test were connected to the Amulet Hotkey modular workstation (based on a PowerEdge MX740c compute sled) across a WAN network link with up to 30 ms of network latency.

The following graphs show the results we obtained:

SPECworkstation 3: 2K Results

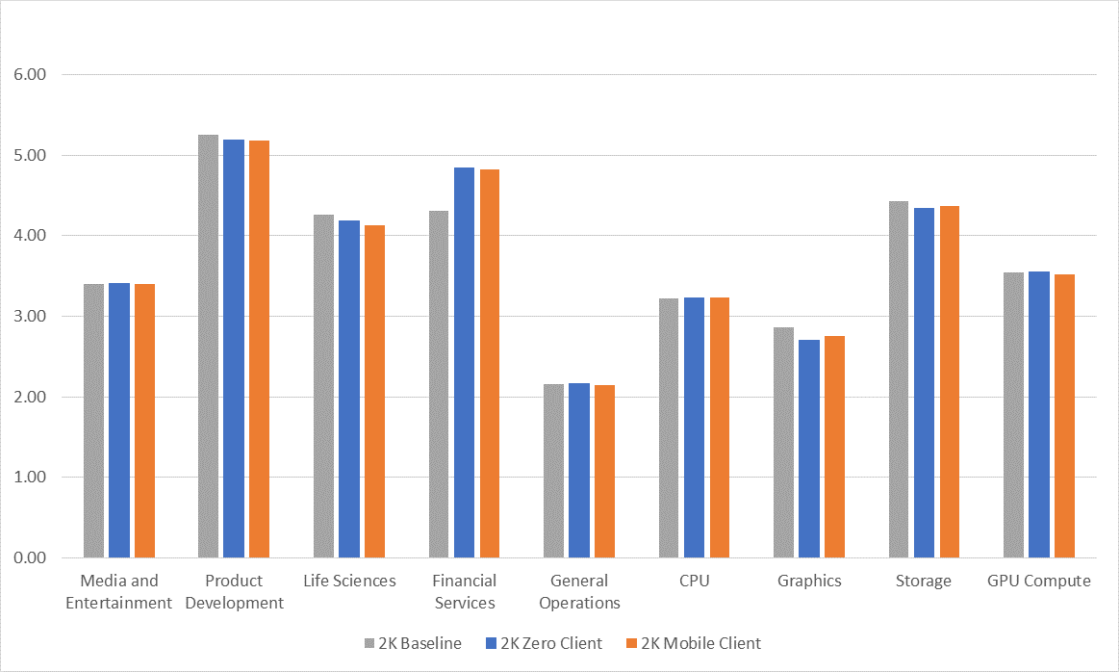

Figure 18. SPECWorkstation 3: MX740c 2K results

The following table describes the SPECWorkstation 3 2K results:

Table 19. SPECWorkstation 3 2K results

SPECworkstation 3.0.2	2K baseline	2K zero client	2K mobile client
Workstation	MX740c	MX740c	MX740c
Remote Access Software	N/A	Teradici Cloud Access Plus	Teradici Cloud Access Plus
Remote Access Client	No Remote Access	DXZ4 PCoIP Zero Client	Surface Pro PCoIP Software Client
SPECworkstation CPU Configuration	8 core test	8 core test	8 core test
SPECworkstation Render Resolution	1900 x 1060	1900 x 1060	1900 x 1060
Display Resolution	1920 x 1080	1920 x 1080	1920 x 1080
Media and Entertainment	3.40	3.41	3.40
Product Development	5.26	5.19	5.18
Life Sciences	4.26	4.18	4.13
Financial Services	4.31	4.84	4.82
General Operations	2.16	2.16	2.15
CPU	3.22	3.20	3.23

SPECworkstation 3.0.2	2K baseline	2K zero client	2K mobile client
Graphics	2.86	2.70	2.75
Storage	4.43	4.34	4.37
GPU Compute OpenCL	3.55	3.55	3.52

SPECworkstation 3: 4K Results

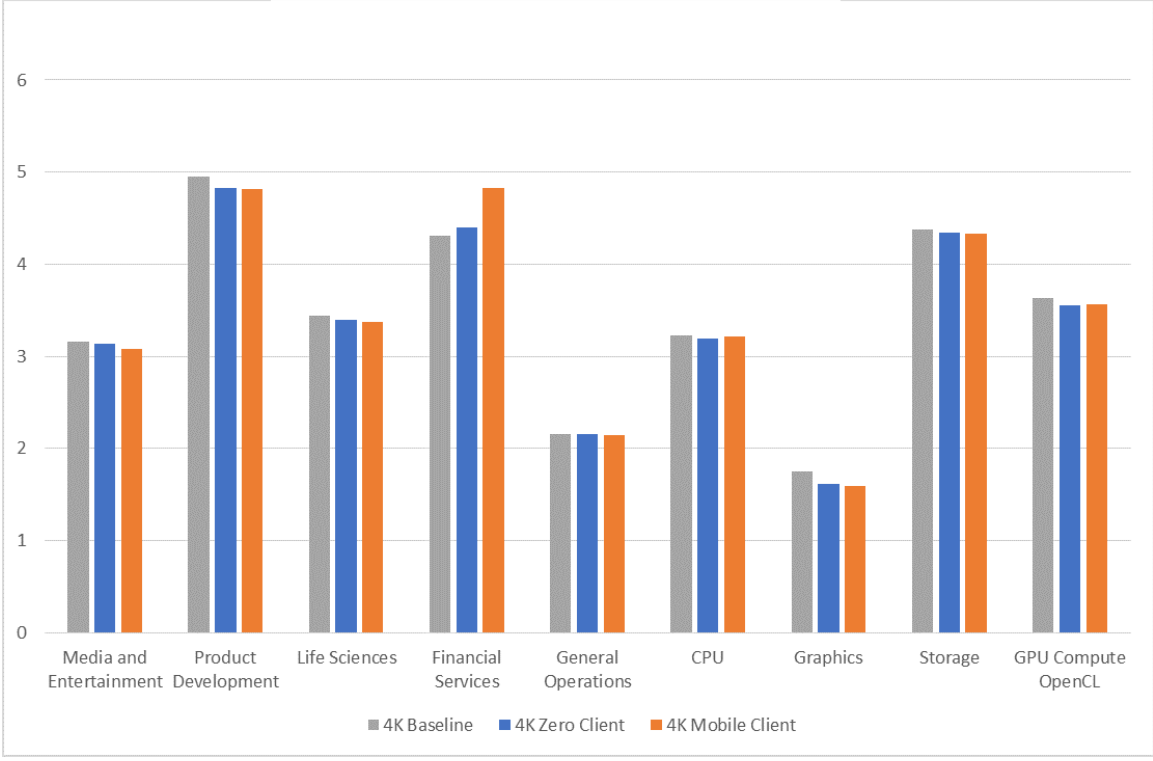

Figure 19. SPECWorkstation 3: MX740c 4K results

The following table describes the SPECWorkstation 3 4K results:

Table 20. SPECWorkstation 3 4K results

SPECworkstation 3.0.2	4K baseline	4K zero client	4K mobile client
Workstation	MX740c	MX740c	MX740c
Remote Access Software	N/A	Teradici Cloud Access Plus	Teradici Cloud Access Plus
Remote Access Client	No Remote Access	DXZ4 PCoIP Zero Client	Surface Pro PCoIP Software Client
SPECworkstation CPU Configuration	8 core test	8 core test	8 core test
SPECworkstation Render Resolution	3800 x 2120	3800 x 2120	3800 x 2120
Display Resolution	3840 x 2160	3840 x 2160	3840 x 2160

SPECworkstation 3.0.2	4K baseline	4K zero client	4K mobile client
Media and Entertainment	3.16	3.14	3.08
Product Development	4.95	4.83	4.82
Life Sciences	3.44	3.40	3.37
Financial Services	4.31	4.40	4.83
General Operations	2.16	2.15	2.14
CPU	3.23	3.19	3.22
Graphics	1.75	1.62	1.59
Storage	4.38	4.34	4.33
GPU Compute OpenCL	3.63	3.55	3.57

OctaneBench GPU Render Benchmark

The Amulet Hotkey CoreModule with NVIDIA T4 GPUs and PowerEdge MX modular architecture supports graphics-intensive acceleration that enhances the performance-user experience.

The OctaneBench benchmark tool provides GPU and multi-GPU benchmarking using OctaneRender by OTOY, which is a spectrally correct GPU render engine that is designed to deliver quality and speed. OctaneRender incorporates advanced graphics technologies with machine learning optimizations, out-of-core geometry support, and RTX raytracing GPU hardware acceleration.

The following table shows the OctaneBench MX740c and NVIDIA T4 results:

Table 21. OctaneBench–MX740c and NVIDIA T4 results

OctaneBench v4	GPU	Multi-GPU
Configuration	1 x NVIDIA T4 Score	2 x NVIDIA T4 Score
Total Score	145.20	289.20

The following table shows the OctaneBench scenes and scores (MX740c and NVIDIA T4):

Table 22. OctaneBench scenes and scores–MX740c and NVIDIA T4

Scene	GPU	Multi-GPU
Configuration	1 x NVIDIA T4 Score	2 x NVIDIA T4 Score
Interior (by Julia Lynen)	87.28	175.99
Interior (by Julia Lynen)	29.89	60.24
Interior (by Julia Lynen)	13.40	26.57
Idea (by Julio Cayetano)	100.20	201.51
Idea (by Julio Cayetano)	28.28	56.26

Scene	GPU	Multi-GPU
Idea (by Julio Cayetano)	25.33	50.43
ATV (by Jurgen Aleksejev)	56.90	112.28
ATV (by Jurgen Aleksejev)	21.75	43.20
ATV (by Jurgen Aleksejev)	18.26	36.34
Box (by Enrico Cerica)	96.00	190.69
Box (by Enrico Cerica)	19.09	37.86
Box (by Enrico Cerica)	19.29	38.42

The Octane Bench benchmark tool can be found at:

https://render.otoy.com/octanebench/

Test Result Data

The raw test data results can be found here:

https://resources.amulethotkey.com/download/20200409_Modular_Workstation_Test_Data.pdf

Appendix C: NVIDIA nVector, SPECviewperf 13, Multi VMs test

NVIDIA nVector, SPECviewperf 13, 3ds Max–4 users

This section looks at the four virtual workstations that ran the NVIDIA nVector automated SPECviewperf 13 3ds Max workload. All the virtual workstations were connected to virtual machine endpoints.

The following table shows the configuration of the virtual workstations that we tested:

Table 23. Virtual workstations configuration

Configuration	Value
vCPU	6
vMemory	32768
HardDisk	120
GPU	grid_t4-8q
GPUDriverVersion	442.06
FRL	Disabled
vSYNC	Default
vDAVersion	7.10.1
DirectConnectVersion	7.10.1
CPUAffinity	Unset
Screen resolution	1920 x 1080
QuantityMonitors	1

The following table shows the average host utilization metrics for this test:

Table 24. Average host utilization metrics

Workload	Density per host	Average CPU %	Average GPU %	Average free memory	Average net Mbps per user
SPECviewperf 13 3ds Max	4	17%	77%	232 GB	11.27 Mbps

The following figure shows the CPU core utilization for the system under test. CPU core utilization remained relatively low, with the highest point reaching around 33 percent.

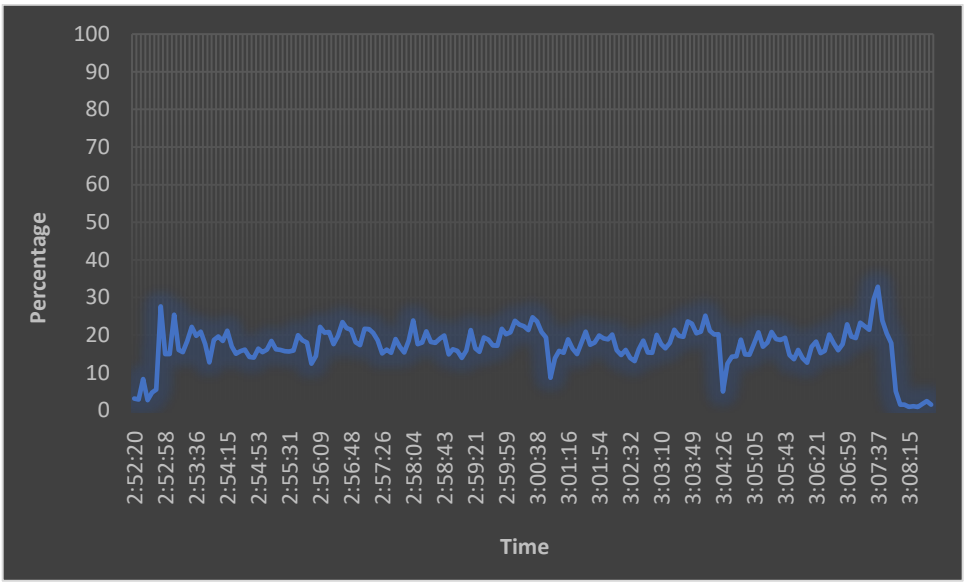

Figure 20. CPU core utilization

The following figure shows the GPU utilization for the system under test. GPU metrics were gathered using NVIDIA-SMI on the VMware ESXi host. GPU utilization was quite high for this test, but it is to be expected for the SPECviewperf 13 workload as the test is largely a measure of GPU performance.

Figure 21. GPU utilization

The following figure shows the free memory in the system during the test. There were no variations in memory usage throughout the test because all vGPU-enabled VM memory was reserved. There was no memory ballooning or swapping on either host.

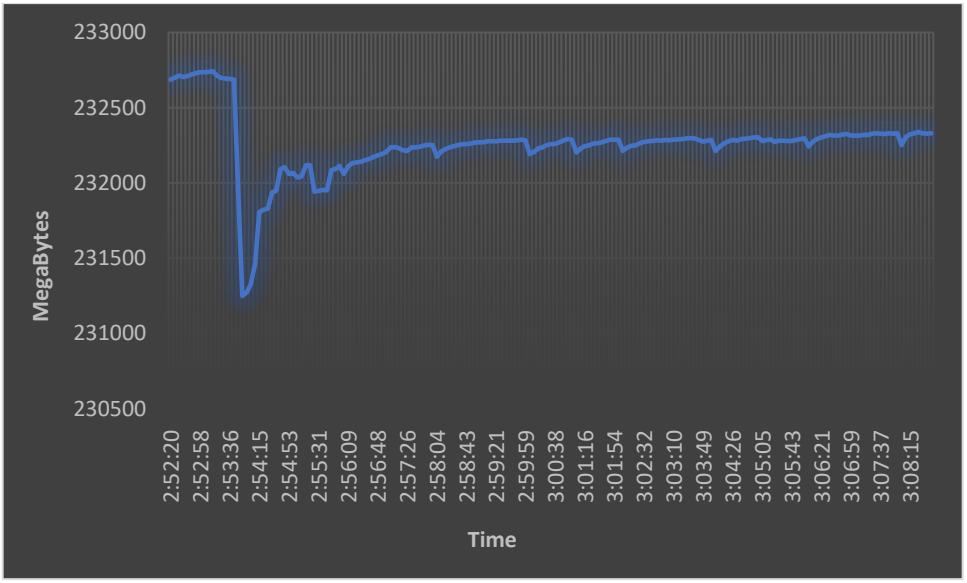

Figure 22. Free memory

The following figure shows the network utilization of the system under test for the duration of the test. There was more than enough bandwidth available for the workload and the display protocol to accommodate the workload. Each server had a single 25 GbE (25,000 MBits) connected to it.

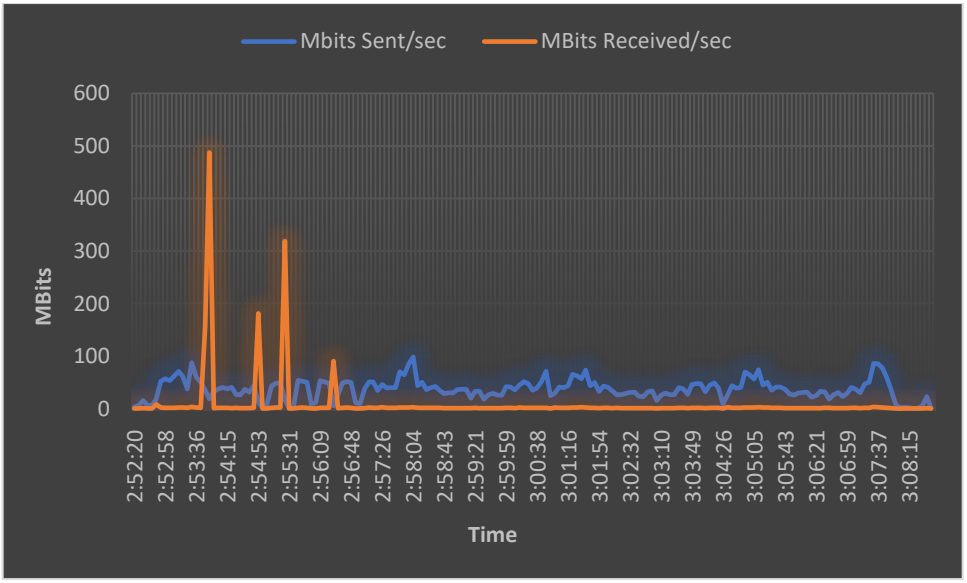

Figure 23. Network usage

NVIDIA nVector, SPECviewperf 13, Maya–4 users

This section describes the four virtual workstations that ran the NVIDIA nVector automated SPECviewperf 13 Maya workload. All the virtual workstations were connected to virtual machine endpoints.

The following table shows the configuration of the virtual workstations that we tested:

Table 25. Virtual workstations configuration

Configuration	Value
vCPU	6
vMemory	32768
HardDisk	120
GPU	grid_t4-8q
GPUDriverVersion	442.06
FRL	Disabled
vSYNC	Default
vDAVersion	7.10.1
DirectConnectVersion	7.10.1
CPUAffinity	Unset
Screen Resolution	1920x1080
QuantityMonitors	1

The following table gives the average host utilization metrics for this test:

Table 26. Average host utilization metrics

Workload	Density per host	Average CPU	Average GPU	Average free memory	Average net Mbps per user
SPECviewperf 13 Maya	4	19%	74%	232 GB	9.24 Mbps

The following figure shows the CPU core utilization for the system under test. CPU core utilization remained low, with the highest point reaching 26.4 percent.

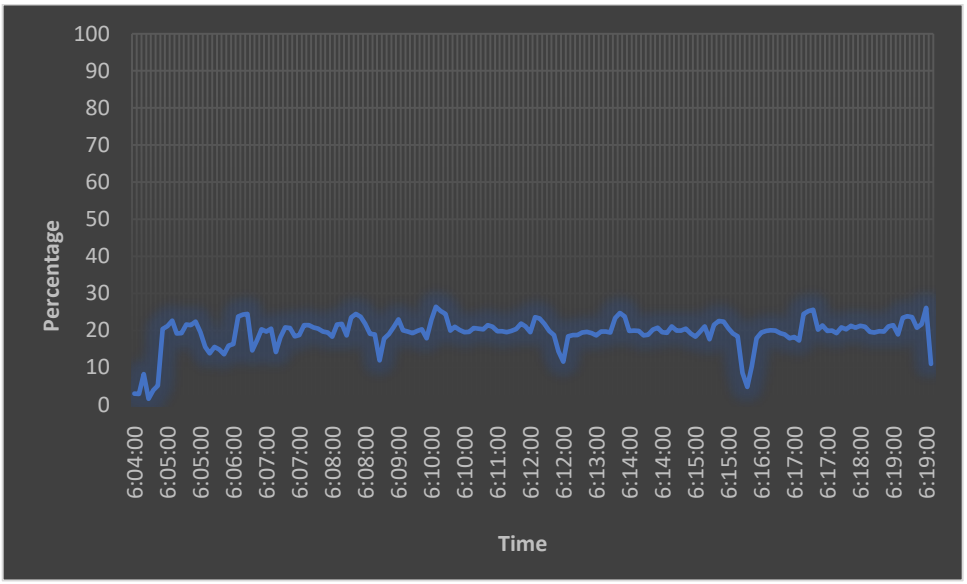

Figure 24. CPU core utilization

The following figure shows the GPU utilization for the system under test. We gathered GPU metrics through NVIDIA-SMI on the VMware ESXi host. GPU utilization was quite high for this test, but it is to be expected for this workload as the test is largely a measure of GPU performance.

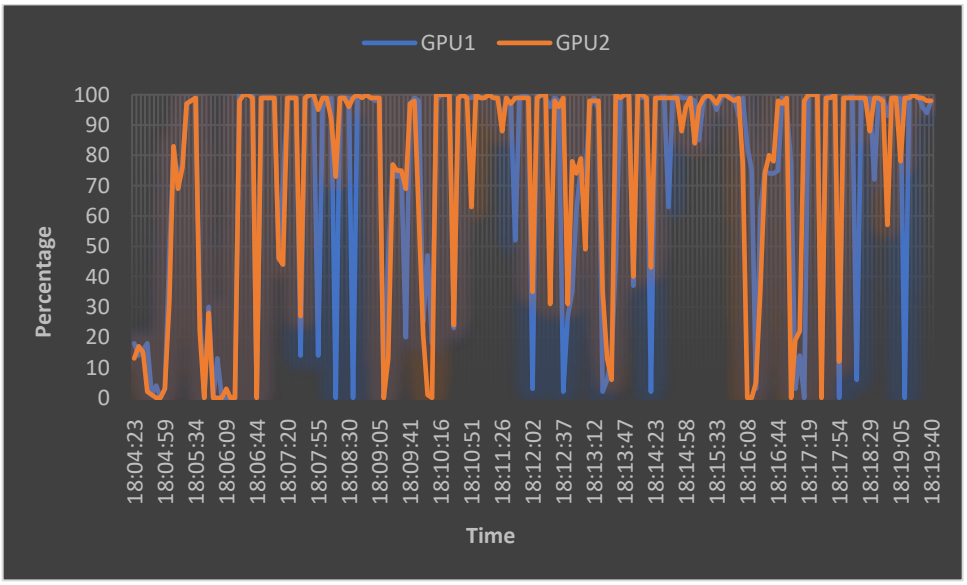

Figure 25. GPU utilization

The following figure shows the free memory in the system during the test. There were no variations in memory usage throughout the test because all vGPU-enabled VM memory was reserved. There was no memory ballooning or swapping on either host.

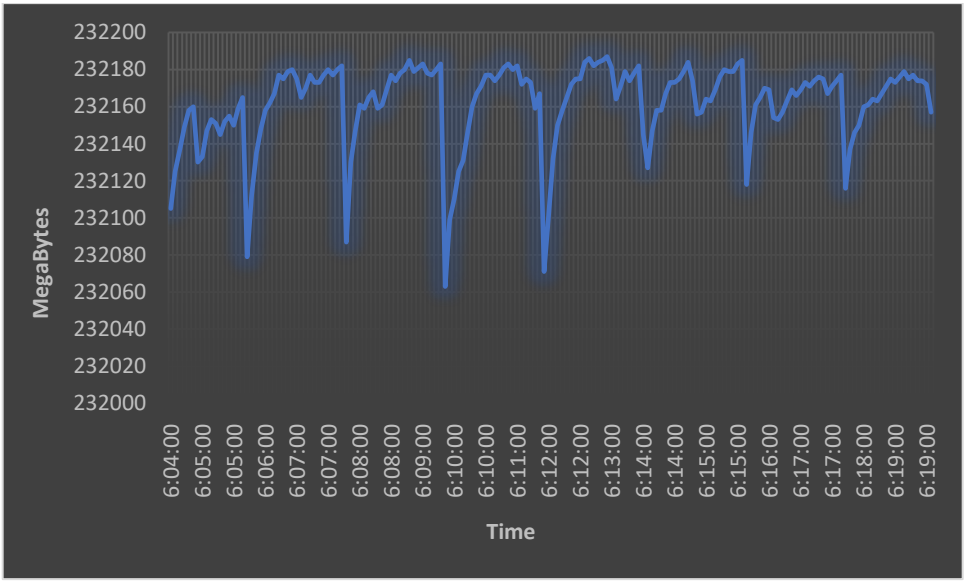

Figure 26. Free memory

The following figure shows the network utilization of the system during the test. There was more than enough bandwidth available for the workload and the display protocol to accommodate the workload. Each server had a single 25 GbE (25,000 MBits) network card connected to it.

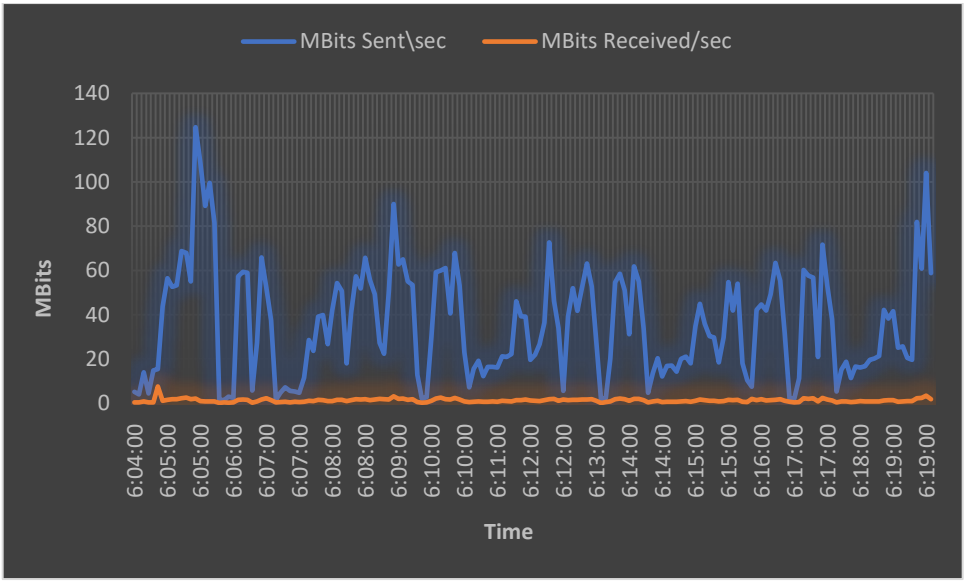

Figure 27. Network usage

Appendix D: NVIDIA nVector, SPECviewperf 13, Single VM test

NVIDIA nVector, SPECviewperf 13, 3ds Max–Single VM test

This section describes a single workstation that ran the NVIDIA nVector automated workload. The test was automated and data was captured by NVIDIA nVector. The single workstation was able to consume more than its default allocated resources due to the "Best Effort" vGPU scheduler being used. Because there was only a single workstation VM per GPU, there was no contention for GPU resources. For additional information, see the NVIDIA vGPU documentation.

The following table shows the single Virtual Workstation VM configuration:

Table 27. Virtual workstation VM configuration

Configuration	Value
vCPU	6
vMemory	32768
HardDisk	120
GPU	grid_t4-8q
GPUDriverVersion	442.06
FRL	Disabled
vSYNC	Default
vDAVersion	7.10.1
DirectConnectVersion	7.10.1
CPUAffinity	Unset
Screen Resolution	1920 x 1080
QuantityMonitors	1

The following table shows the average host utilization metrics for this test:

Table 28. Average host utilization metrics

Workload	Density per host	Average CPU	Average GPU	Average free memory	Average net Mbps per user
SPECviewperf 13 3ds Max	1	5%	39%	333 GB	13.68 Mbps

The following figure shows the CPU core utilization for the system under test. CPU core utilization remained relatively low with the highest point reaching approximately 8 percent.

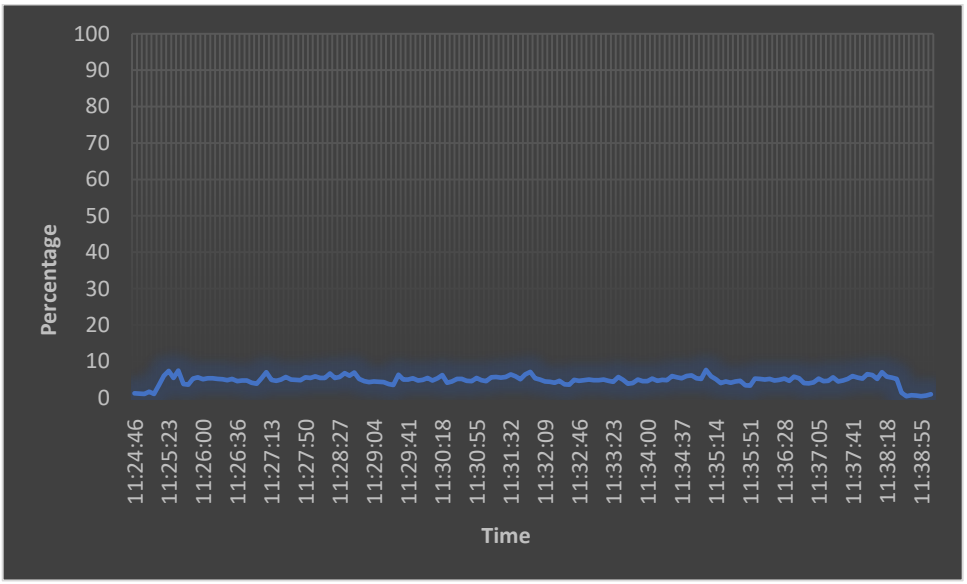

Figure 28. CPU core utilization

The following figure shows the GPU utilization for the system under test. GPU metrics were gathered by using NVIDIA-SMI on the VMware ESXi host. The utilization for one of the GPUs was quite high for this test, while the other remained unused. This is expected for this workload with a single workstation VM.

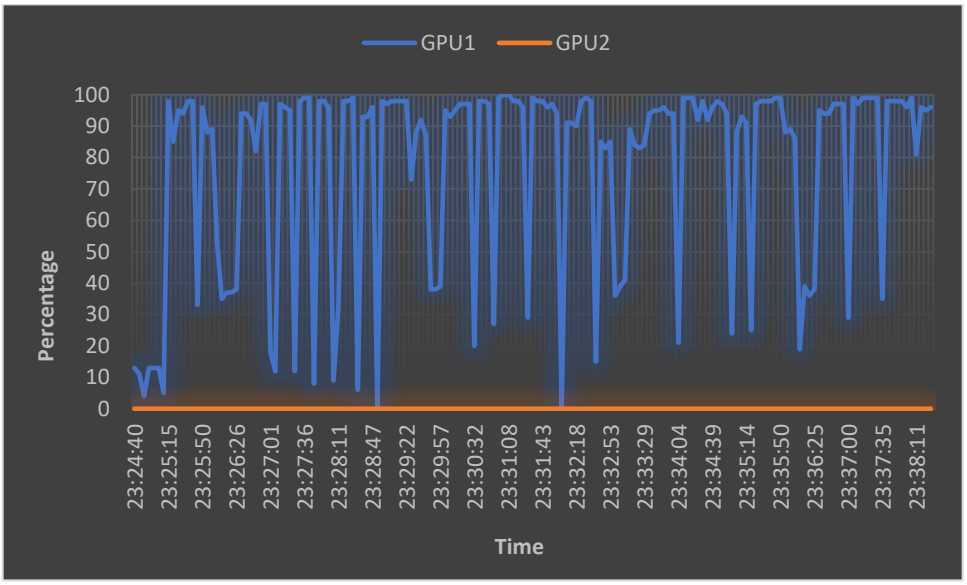

Figure 29. GPU utilization

The following figure shows the free memory in the system during the test. There were no variations in memory usage throughout the test because all vGPU-enabled VM memory was reserved. There was no memory ballooning or swapping on either host.

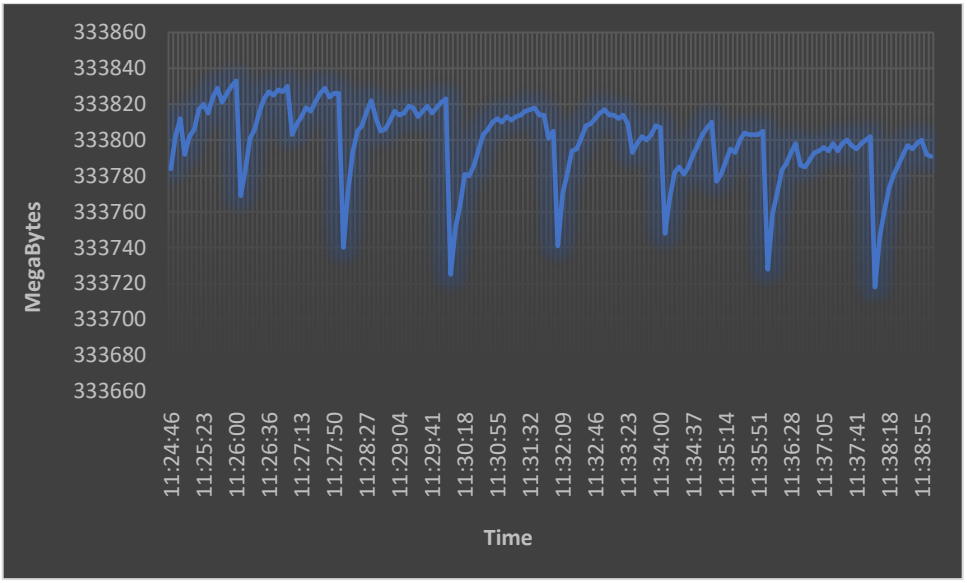

Figure 30. Free memory

The following figure shows the network utilization of the system during the test. There was more than enough bandwidth available for the workload and the display protocol to accommodate the workload. Each server had a single 25 GbE (25,000 MBits) network card connected to it.

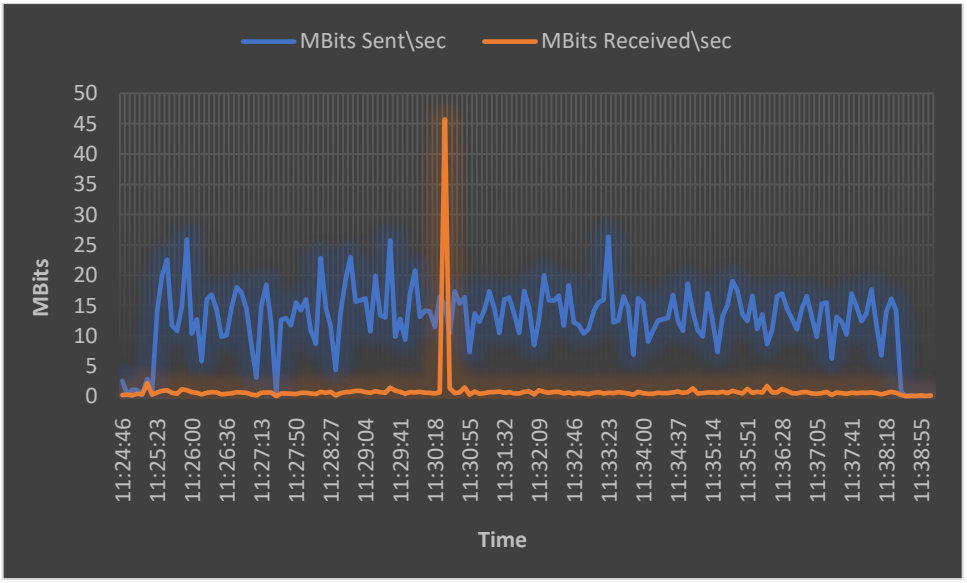

Figure 31. Network usage

NVIDIA nVector, SPECviewperf 13, Maya–Single VM Test

This section describes a single workstation that ran the NVIDIA nVector automated workload. The test was automated and data was captured by NVIDIA nVector. The single workstation VM was able to consume more than its default allocated resources due to the "Best Effort" vGPU scheduler being used. Because there was only a single VM per GPU, there was no contention for GPU resources. For additional information, see the NVIDIA vGPU documentation.

The following table shows the single Virtual Workstation VM configuration:

Table 29. Virtual Workstation VM configuration

Configuration	Value
vCPU	6
vMemory	32768
HardDisk	120
GPU	grid_t4-8q
GPUDriverVersion	442.06
FRL	Disabled
vSYNC	Default
vDAVersion	7.10.1
DirectConnectVersion	7.10.1
CPUAffinity	Unset
Screen Resolution	1920 x 1080
QuantityMonitors	1

The following table shows the average host utilization metrics for this test:

Table 30. Average host utilization metrics

Workload	Density per host	Average CPU	Average GPU	Average free memory	Average net Mbps per user
SPECviewperf 13 Maya	1	5%	34%	334 GB	10.08 Mbps

The following figure shows the CPU core utilization for the system under test. CPU core utilization remained relatively low, with the highest point reaching 7.24 percent.

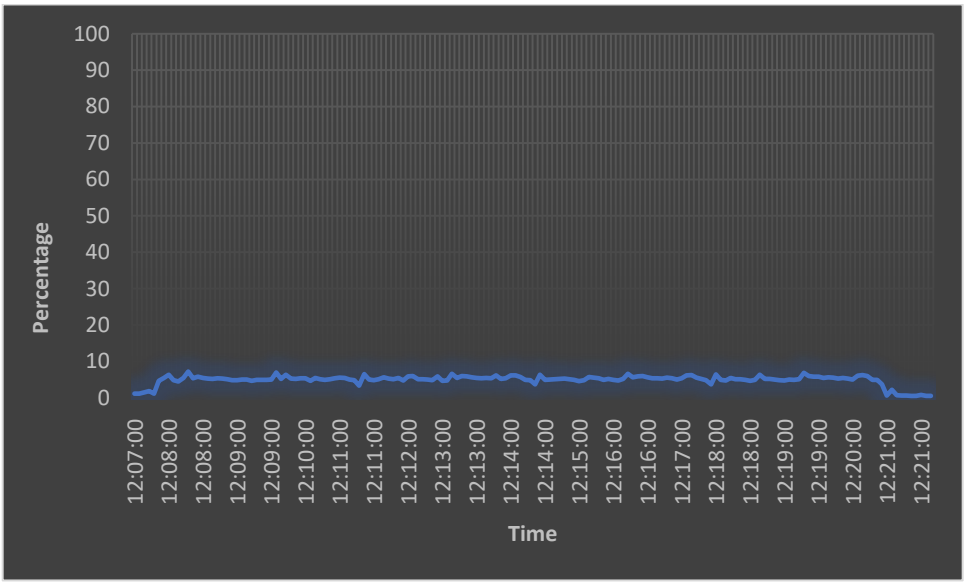

Figure 32. CPU core utilization

The following figure shows the GPU utilization for the system under test. GPU metrics were gathered using NVIDIA-SMI on the VMware ESXi host. GPU utilization for one of the GPUs was quite high for this test, while the other remained unused. This result is expected for this workload with a single VM.

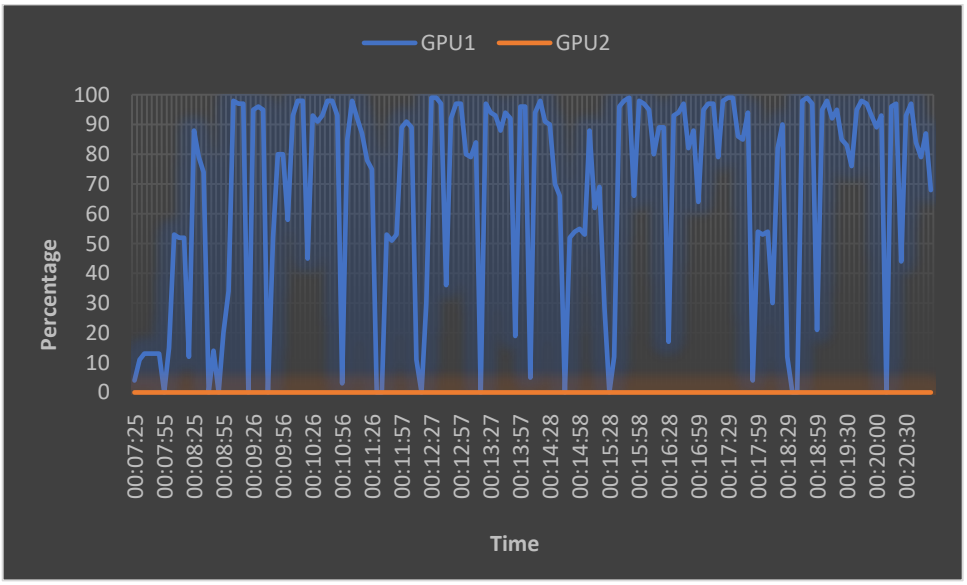

Figure 33. GPU utilization

The following figure shows the free memory in the system during the test. There were no variations in memory usage throughout the test because all vGPU-enabled workstation memory was reserved. There was no memory ballooning or swapping on either host.

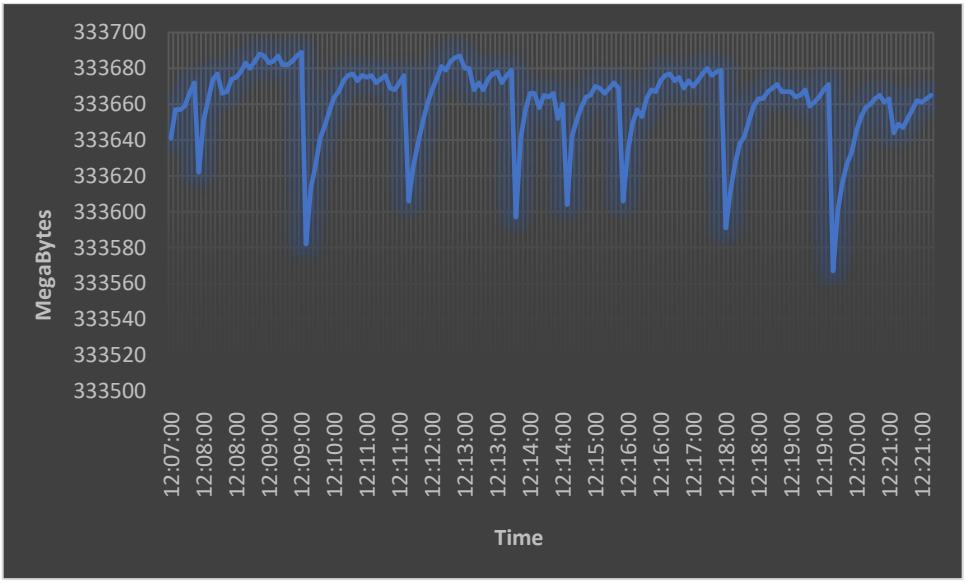

Figure 34. Free memory

The following figure shows the network utilization of the system during the test. More than enough bandwidth was available for the workload and the display protocol to accommodate the workload. Each server had a single 25 GbE network card (25,000 MBits) connected to it.

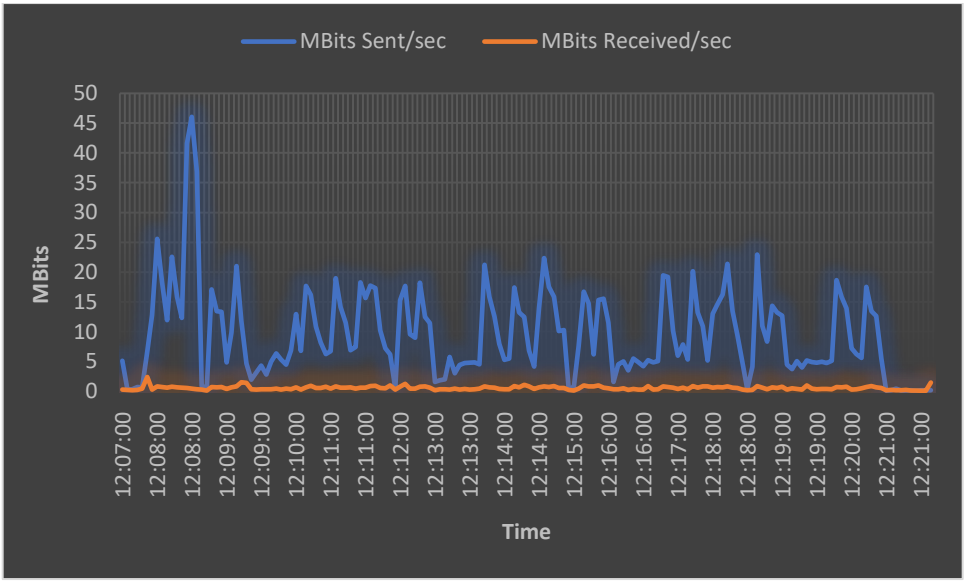

Figure 35. Network usage