



HPE Reference Architecture for Remote Artist Workstations with Avid, Teradici, and Nvidia on HPE Proliant DL380 Server

Virtual Desktop Infrastructure to deliver Avid Media Composer via Teradici

CONTENTS

Executive Summary	3
Introduction.....	3
Product names and terminology	4
Solution overview.....	5
Solution components.....	6
Hardware.....	6
Software	9
Configuration guidance.....	11
Capacity and sizing	12
Vertical sizing.....	12
Test setup	14
Workload description.....	15
System performance data.....	19
Results	25
Analysis and recommendations	25
Summary.....	26
Resources and additional links	27



EXECUTIVE SUMMARY

The move to virtualized production environments has become standard practice with media organizations of all sizes over the past several years, providing discernable benefits in efficiency and cost savings, and providing the required performance to meet the demands of a modern media production environment. Industry thought leaders envisioned the industry expanding virtualization across the entire workflow to realize the benefits of VDI in a hyperconverged infrastructure (HCI). In an HCI environment, the deployment happens in software-defined IT infrastructure which virtualizes all the elements of the conventional hardware-defined systems. This includes the virtualized computing already described and extends this to software-defined storage and software-defined networking.

A recent need for remote work and business continuity have accelerated media companies' shift to hyperconverged infrastructure as part of an initiative to more aggressively deploying production in hybrid cloud scenarios. This is happening as a more distributed workforce needs access to media from anywhere and on any device, enabling more work to happen simultaneously while providing the performance and security demanded by modern media workflows. One of the many lessons learned during this shift is that technology exists to enable work from anywhere and still meet business requirements, and this has been proven as media companies were able to successfully pivot to work from home. The fact that these workflows have proven to be effective has opened up a new world of possibilities that will have both operational and economic impacts on what is now considered standard operating procedure. The way things used to be done has changed forever and media companies are building for their hybrid future.

To help address the challenges that a remote distributed workforce presents, enabling editing of high resolution, high bit rate content, from anywhere, this Reference Architecture is offered as a way to deliver low latency, high performance, secure, remote access to hybrid production environments.

This solution includes:

- HPE ProLiant DL380 server(s)
- NVIDIA® Quadro RTX GPUs
- VMware® vSphere® and VMware vSAN software
- Teradici Cloud Access software
- Avid Media Composer non-linear video editing software
- Avid NEXIS software-defined storage

Target audience: This Reference Architecture is designed for Technologists, Architects, and IT professionals who use, manage, or administer Avid production environments or workflows. This paper specifically focuses on Avid Media Composer in use cases that have the potential for high client counts, high-performance requirements, a smooth virtualized graphics experience, and that are secure, scalable, and low latency, even in challenging network conditions. Lastly, this information is intended for those who evaluate, recommend, or design high-performance architectures that support production environments in the post, news, and sports.

INTRODUCTION

The work required of remote artists using Avid Media Composer becomes more graphic and compute-intensive as time passes, and the cost of CPUs and GPUs is not decreasing. Despite the need for high-powered GPUs, end users do not require or use the entire GPU or CPU. End users also require the ability to collaborate wherever and whenever they can. The lack of mobility of a desktop equipped with a high-powered GPU is detrimental to the productivity of end-users who need the ability to work remotely.

The industry is also seeing an accelerating shift to as-a-service and op-ex models, and a major part of that shift is a move to virtualized infrastructures that can provide a hybrid approach for work from anywhere production methodologies. Virtualized environments made quickly standing up new channels economical for broadcasters and easier to implement in a shorter amount of time. They have also made supporting a distributed workforce easier to manage for the studios, which has paved the way for a growing shift to hyperconverged infrastructures that can provide the benefits of a software-defined strategy across the production pipeline. Taking advantage of standardized hardware for on-premises deployments, HCI allows for simplified management, lower overall costs, and increased resource utilization rates regardless of where the deployment is taking place (on-premises, private data center, or public cloud).



As many productions ramp back up at scale, the distributed nature of the workforce is predicted to remain, while the studios and broadcasters alike re-evaluate their operational strategies to take advantage of what is now possible. Media companies have the flexibility to broaden their ability to hire from an expanded talent pool regardless of their proximity to the on-prem facility, and talent can now reside where they prefer to live as opposed to being limited to specific locales where production is prominent. More workers able to collaborate more efficiently, working across geographies and time zones has accelerated an industry-wide adoption of technologies. This further enables more of the production pipeline to happen simultaneously, while delivering the performance and accessibility needed to efficiently collaborate and deliver content quickly.

Combined with concerns of user mobility and under- or overprovisioned resources, along with other general concerns such as ergonomics and data security, more businesses are continuing to move to virtualized desktops. To address the need for high-powered GPUs amongst video editors in the industry, a VDI solution with GPU acceleration can be used. This approach eliminates unnecessarily high costs for excessive amounts of hardware dedicated to a single user, allows reconfigurability of the hardware to meet changing project and performance requirements, and also addresses security concerns by allowing users' compute to be moved to a more secure location than the office floor.

HPE ProLiant DL380 servers and NVIDIA Quadro RTX GPUs are a great combination to effectively provide end-users with the performance they require. For virtualization purposes, the industry-leading compute virtualization platform VMware vSphere is used along with vSAN for storage virtualization. Finally, the virtual desktops are delivered via Teradici Cloud Access Software.

Hewlett Packard Enterprise servers deliver world-class results. The HPE ProLiant DL380 server powered by NVIDIA Quadro RTX GPUs and delivered with Teradici Cloud Access Software ensures the optimal solution for end users with high-performance and high-resolution needs such as video editors using Avid Media Composer.

PRODUCT NAMES AND TERMINOLOGY

To help avoid confusion around the different types of Media Composer editing clients and the additional technology involved with VDI the following are some descriptions of the products and terminology commonly used.

Media Composer I Ultimate: Avid Media Composer I Ultimate leverages the completely redesigned user interface, increased user-defined layout, and unified media engine enabling editing of SD, HD, UHD, HDR, and stereoscopic 3D content with compatibility with a growing number of codecs and formats. Includes the Media Composer I PhraseFind option to find content fast with dialog indexing, Media Composer I ScriptSync for finding the best takes to accelerate script-based editing also using dialog search, Media Composer I Symphony option enabling correct and color grade color and finishing capabilities, and the Media Composer I Newscutter option which adds MediaCentral I Newsroom management, iNews, and ENPS integrations into Media Composer. This comes with Media Composer I Ultimate software updates and supports providing access to all new releases and ExpertPlus support throughout the life of the subscription.

Media Composer | Cloud VM: Media Composer | Cloud VM is a product license add-on option available for [Media Composer I Ultimate](#)—this option provides a virtualized version of Media Composer, enabling anyone in your facility to access the software remotely using a workstation, laptop, or zero client with a high-speed network connection.

Avid INEXIS: Avid NEXIS software-defined storage, with tiered storage management capabilities, tuned to run in production workflows with options for full redundancy, no single point of failure, and easily scalable. Online and nearline tiers with integrated 3rd parties for archive tier available. Extends to the cloud using NEXIS I Cloudspaces to enable parking NEXIS workspaces into a cloud tier managed from within the NEXIS user management.

VDI: VDI stands for Virtual Desktop Infrastructure. While there are technically different types of VDI, this document will be referring to VDI where a workstation resource is running as a virtual machine in a data center that a user is accessing and operating over the network from an endpoint client.

Endpoint Client: An endpoint client is a hardware device that an operator uses to access a workstation or VM remotely. An endpoint client setup will usually consist of one or two monitors, a keyboard, mouse and desktop computer, laptop, or thin or zero clients.

Teradici Zero Client: A zero client is an endpoint hardware device that an operator uses to access a workstation or VM remotely. Zero clients have no operating system and run firmware for a specific remote display protocol. The chipset on a zero client is specifically designed to efficiently transmit command data and receive the desktop pixel data over the remote display protocol.



Teradici Thin client: A low-resource PC that often includes a simplified operating system (e.g.: Windows Embedded). Thin clients are computer terminals that are purpose-built for remoting into a server. The server is responsible for the majority of processing and data storage. The HP t740 is an example of a thin client well suited for this solution.

Teradici Thick client: Teradici software client is installed on traditional PCs or re-purposed PCs. Compared to thin client, the host was not purpose-built for remoting into a server. Teradici software application is running among other office applications.

Teradici PCoIP: “PC over IP” is a remote display protocol originally developed by Teradici to compress and decompress images and sound across local and wide area networks. This technology lowers the demand on the network while maintaining high-quality images and sound. Note that PCoIP is a brand name product for Teradici, but people will often use the term PCoIP generically referring to the fact that a desktop resource is being remotely accessed from an endpoint client over a LAN or WAN.

Teradici PCoIP Ultra: PCoIP Ultra is an enhancement to the original PCoIP protocol from Teradici. PCoIP Ultra contains expanded configuration options that enable you to choose the most efficient settings according to content characteristics. This flexibility results in a faster, more interactive experience for users of remote workstations working with high-resolution content, including applications for video editing, motion graphics, and animation. PCoIP Ultra is recommended for 4K UHD high frame rate content. Most newer software client’s devices with i5 processors or better have the compute resources required to run PCoIP Ultra. Please consult with Teradici for minimum requirements.

Teradici PCoIP Agent: This is the software component installed on the virtual machine that will be accessed and controlled remotely from an endpoint device. Teradici PCoIP Software Client: This is the software component installed on the endpoint device the operator physically interacts with to remotely connect to the virtual machine with Media Composer installed. The endpoint device hardware is typically a Zero Client or Software Client.

SOLUTION OVERVIEW

This Reference Architecture highlights a solution to create virtual workloads in which end users using Avid Media Composer can have the highest quality and most seamless experience possible. This is accomplished by using HPE ProLiant DL380 servers as the physical hosts with VMware ESXi™ serving as the virtualization layer to provide virtual machines (VMs) to end-users. The HPE ProLiant DL380 servers include NVIDIA Quadro RTX 6000 or RTX 8000 GPUs and NVIDIA vGPU software, which allow the physical GPUs to be shared between multiple virtual machines, each receiving a dedicated GPU frame buffer.

Figure 1 depicts the solution flow with the following steps:

1. End users, working either remotely from home or on-premises, request a VM desktop.
2. Teradici PCoIP Connection Manager validates and accepts the request. Connection Manager hands off the request to the connection broker.
3. A connection broker handles authentication and then assigns a desktop.
4. Within the Windows 10 image, Avid, NVIDIA, and Teradici agents request the appropriate licenses.
5. End users are connected directly to the remote VMs using the Teradici PCoIP Ultra protocol.



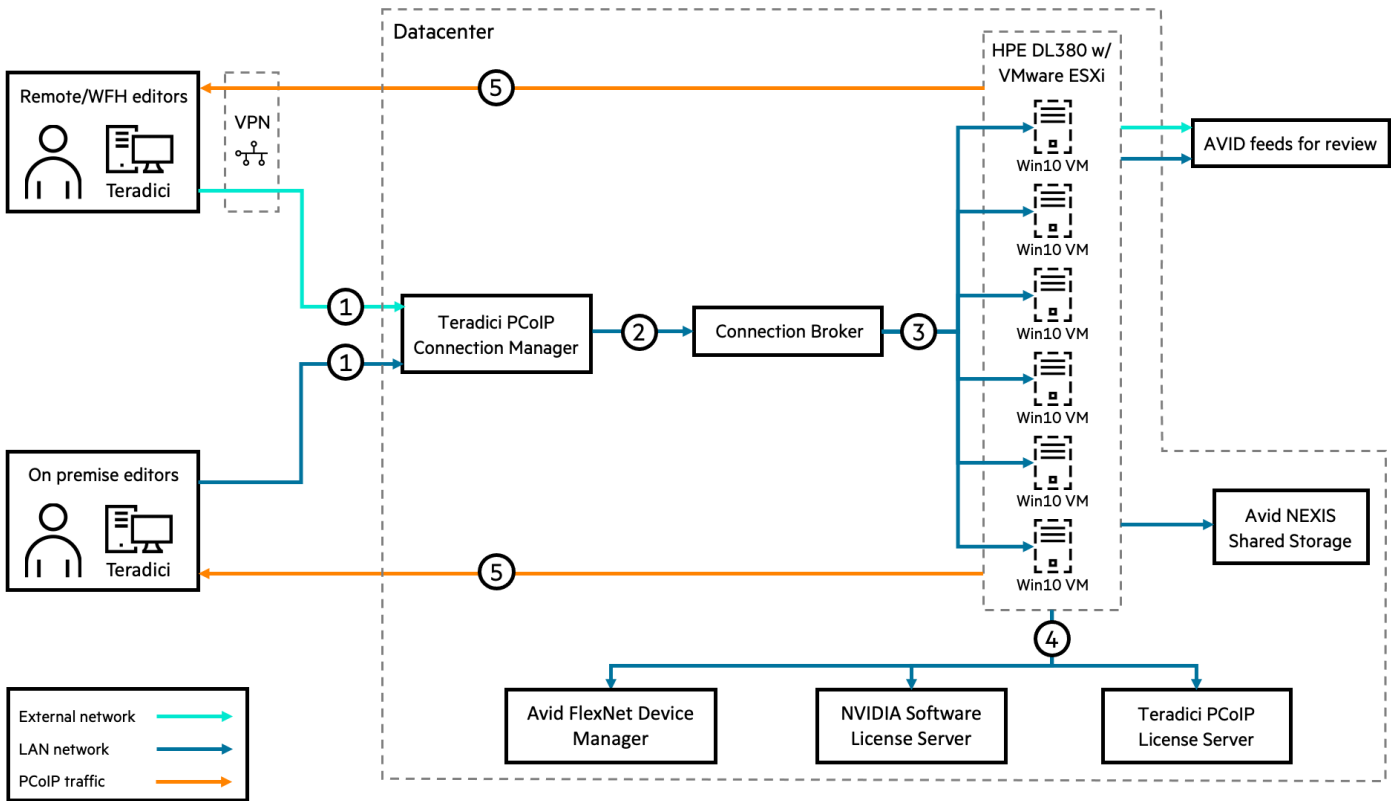


FIGURE 1. Solution flow

SOLUTION COMPONENTS

Hardware

HPE ProLiant DL380 Gen10 Server

The HPE ProLiant DL380 Gen10 server delivers the latest in security, performance, and expandability. It supports the Intel® Xeon® Processor Scalable Family supporting HPE 2933 MT/s DDR4 SmartMemory. The HPE ProLiant DL380 Gen10 server has an adaptable chassis, including new HPE modular drive bay configuration options with up to 30 SFF, up to 19 LFF, or up to 20 NVMe drive options along with support for up to 3 double wide GPU options. Along with an embedded 4x1GbE, there is a choice of HPE FlexibleLOM or PCIe standup adapters which offer a choice of networking bandwidth (1GbE to 40GbE) and fabric allowing customers to adapt and grow for changing business needs. The HPE ProLiant DL380 Gen10 server comes with a complete set of HPE Technology Services, delivering confidence, reducing risk, and helping customers realize agility and stability.



Table 1 describes the configuration of the HPE ProLiant DL380 Gen10 tested for this document.

TABLE 1. Components of HPE ProLiant DL380 Gen10 Server

Component	Quantity	Description
CPU	2	Intel® Xeon® Gold 6258R CPU @ 2.7GHz, 28-core, or Intel Xeon Gold 6248R CPU @ 3.0GHz, 24-core
Memory	24	12x HPE 32GB 2Rx4 PC4-2933Y-R Smart Kit 12x HPE 64GB 2Rx4 PC4-2933Y-R Smart Kit
GPU	2	NVIDIA Quadro RTX 8000, or NVIDIA Quadro RTX 6000
Storage Controller	2	1x HPE Smart Array E208i-p SR Gen10 (boot drives) 1x HPE Smart Array E816i-a SR Gen10 (vSAN)
Disks	8	2x HPE 240GB SATA Read Intensive SFF SSD (boot drives) 2x HPE 375GB NVMe x4 Write Intensive SFF SSD (vSAN cache tier) 4x HPE 3.2TB SAS Mixed Use SFF SSD (vSAN capacity tier)
Network Adapter	2	1x HPE 40Gb 2P 544+QSFP Adapter 1x HPE 40Gb 2P 544+FLR-QSFP Adapter

Figure 2 represents the physical layout of the HPE ProLiant DL380 Gen10 tested for this solution.

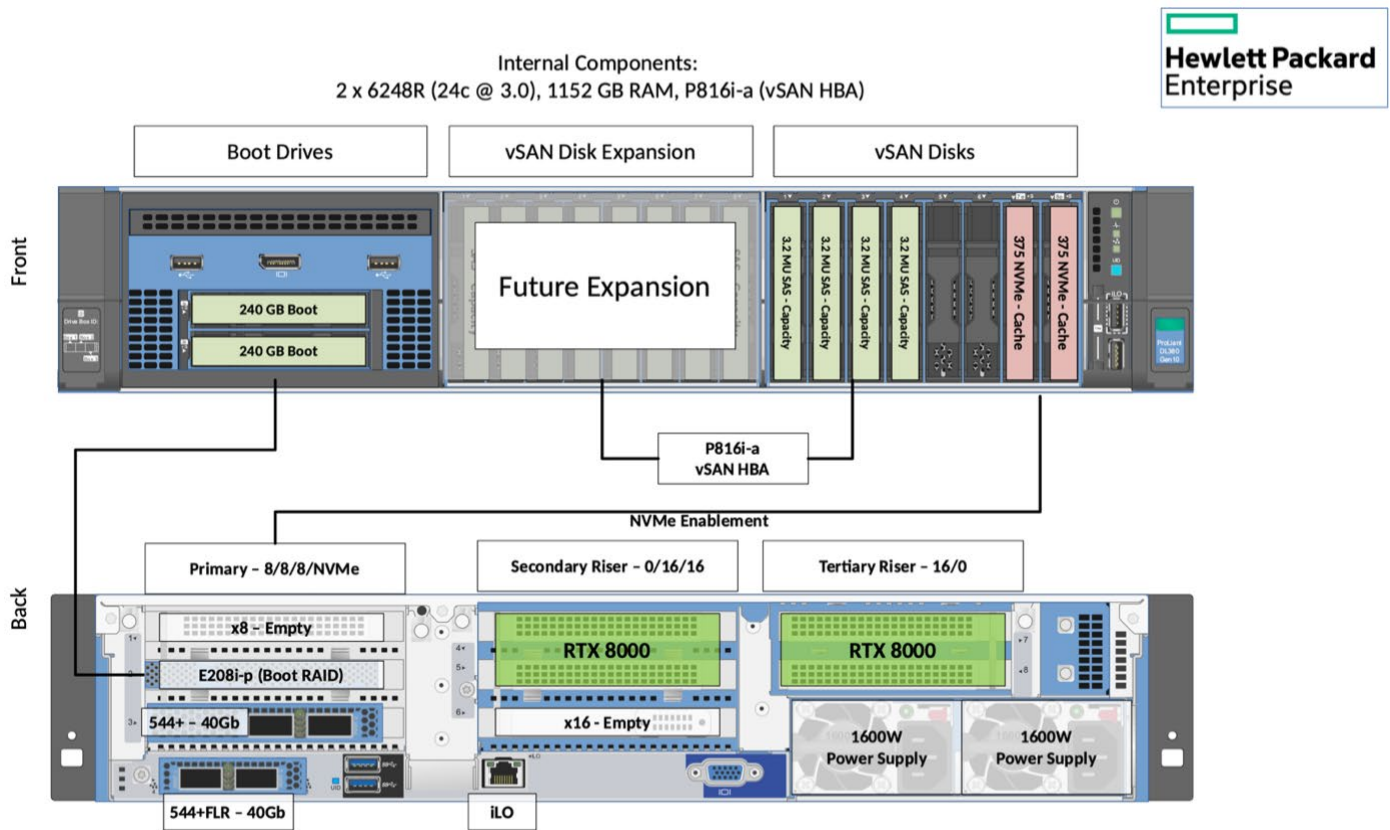


FIGURE 2. Physical layout of the HPE ProLiant DL380 Gen10 Server



NVIDIA Quadro RTX GPUs

NVIDIA GPUs can be used for accelerated graphics, as well as a general-purpose graphics processor unit (GPGPU) for high-performance computing and analytical applications. The NVIDIA Quadro RTX 6000 and RTX 8000 GPUs are both 295-watt PCIe form factor GPUs with 4608 CUDA cores. The Quadro RTX 6000 has 24GB of GDDR6 memory, and the Quadro RTX 8000 has 48GB of GDDR6 memory. Both can be split among multiple virtual machines with NVIDIA vGPU. The choice of which of these GPUs to use depends entirely on the GPU frame buffer requirements based on the user workload, since they are nearly identical other than the amount of GPU memory they have.

Figure 3 shows the NVIDIA Quadro RTX GPU.



FIGURE 3. NVIDIA Quadro RTX 8000 GPU

Table 2 describes the specifications of the NVIDIA Quadro RTX 8000 GPU tested for this solution and the NVIDIA Quadro RTX 6000 GPU which may fit some workloads better.

TABLE 2. NVIDIA Quadro RTX 8000 GPU components

Component	Specification
CUDA cores	4608
Additional cores	576 NVIDIA Tensor Cores, 72 NVIDIA RT Cores
Memory	RTX 6000: 24 GB GDDR6 with ECC RTX 8000: 48 GB GDDR6 with ECC
vGPU profiles	1 GB, 2 GB, 3 GB, 4 GB, 6 GB, 8 GB, 12 GB, 16 GB, 24 GB, 48 GB
System interface	X16 PCIe Gen3
Form factor	Dual slot, full-height
Power	Total board power: 295W Total graphics power: 260W
Thermal solution	Active

Client-side hardware

When driving multiple high-resolution monitors as part of a remote computing solution, the choice of the thin client can have a major impact on apparent system performance to the end-user. Importantly for media workloads, the performance of client devices can also make a difference in audio synchronization. They must also be able to support the desired user peripherals, such as Wacom tablets.



A number of thin clients were evaluated for this environment, including the HP t740, powered by a quad-core AMD processor with Radeon Vega 8 graphics, and supporting up to four 4K monitors. The HP t740 delivers an excellent end-user experience as a thin client, even for demanding remote workloads, and has been validated by Teradici to work well with the PCoIP protocol. Other devices evaluated included an Intel NUC, IGEL UD3, an LG 38" curved all-in-one, and various Mac computers. Note that the PCoIP software client for macOS is not at the full feature and functionality parity with the Windows client, and therefore Linux® and Windows thin clients, like the HP t740, are recommended over Mac clients.

Software

This Reference Architecture uses virtualization infrastructure tested on both ESXi 6.7 and 7.0, using vCenter Server® Appliance™ (VCSA) 7.0 and Teradici Cloud Access Software 2021.03.

Table 3 provides the solution software components used in this Reference Architecture.

TABLE 3. Solution software components

Component	Version
VMware	
VMware vCenter	7.0.2
VMware ESXi	6.7 U3 and 7.0.1
NVIDIA	
NVIDIA License Manager	2020.05.0
NVIDIA vGPU Manager	460.73.02
NVIDIA Software Graphics Driver	462.31
Teradici	
PCoIP Graphics Agent	21.03.0
Teradici Cloud Access Connector	2021.3
Teradici PCoIP Client	21.03.1
Avid	
Avid Media Composer	2021.3.0

VMware vSphere and VMware vSAN

VMware vSphere is the industry-leading compute virtualization platform. ESXi is the bare-metal hypervisor that is a part of vSphere. This solution has been tested on ESXi 6.7 U3 and ESXi 7.0.

VMware vSAN is storage virtualization software that works together with vSphere to allow management of both compute and storage through a single platform. Instead of traditional, external storage, the solution becomes hyperconverged by using physical storage on the same server that is providing virtualization compute.

NVIDIA vGPU

NVIDIA vGPU, or virtual GPU, virtualizes data center GPUs and allows them to be shared across multiple virtual machines. This can allow for significant performance improvements when compared against CPU-only virtual machines, especially on graphics-accelerated applications, such as media workloads. Specifically for this workload, NVIDIA Quadro vDWS, or Quadro Virtual Data Center Workstation delivers the highest performance and full capabilities of NVIDIA vGPUs.



Figure 4 shows how an NVIDIA RTX GPU can deliver multiple vGPUs to separate virtual machines, through the vSphere hypervisor using the NVIDIA vGPU Manager.

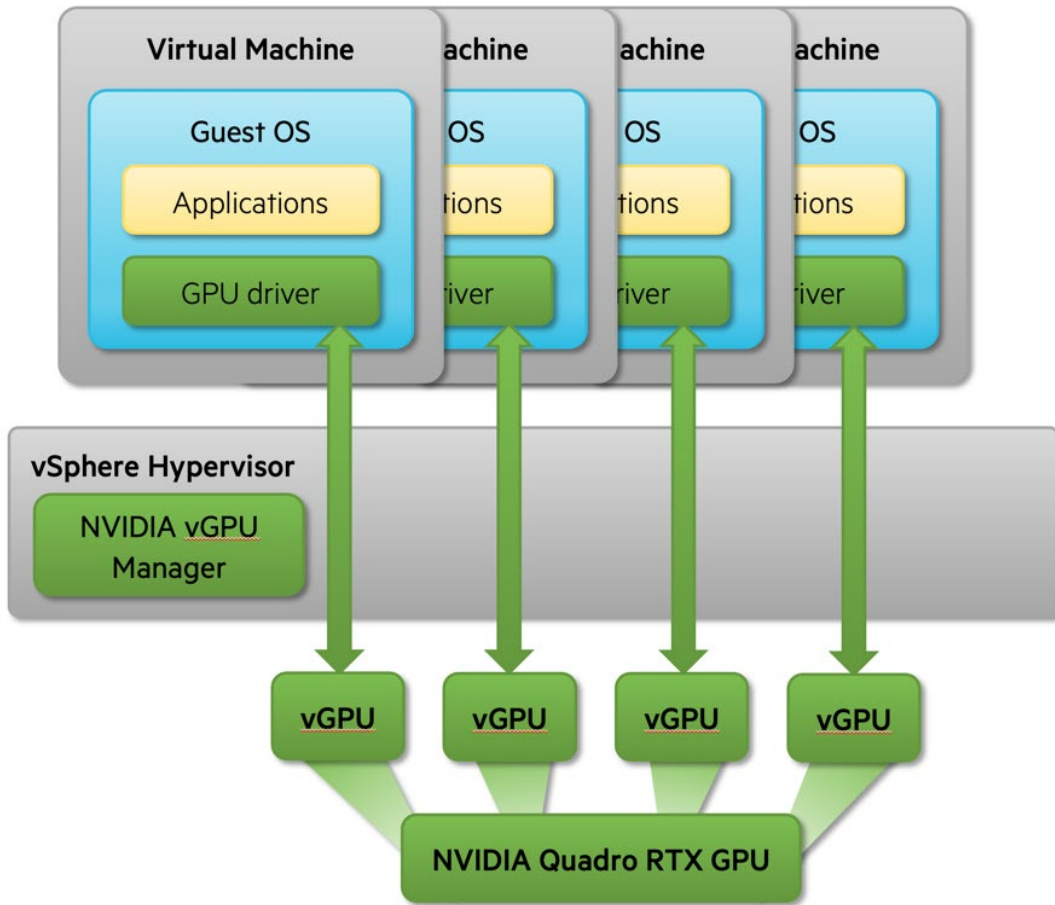


FIGURE 4. NVIDIA Quadro RTX vGPU

Teradici Cloud Access Software

Teradici is the creator of the industry-leading PCoIP remoting protocol technology and Cloud Access Software. Teradici Cloud Access Software enables enterprises to securely deliver high-performance graphics-intensive applications and workstations from private data centers, public clouds, or hybrid environments with crisp text clarity, true color accuracy, and lossless image quality to any endpoint, anywhere.



Figure 5 shows the Teradici PCoIP Ultra with NVIDIA RTX Server that can provide virtual machines to multiple artists resulting in virtual machines that are indistinguishable from physical workstations. Artists can enjoy workspaces set up on the latest hardware, and work with confidence in high fidelity with steady frame rates.

HPE DL380 RTX Server with Teradici

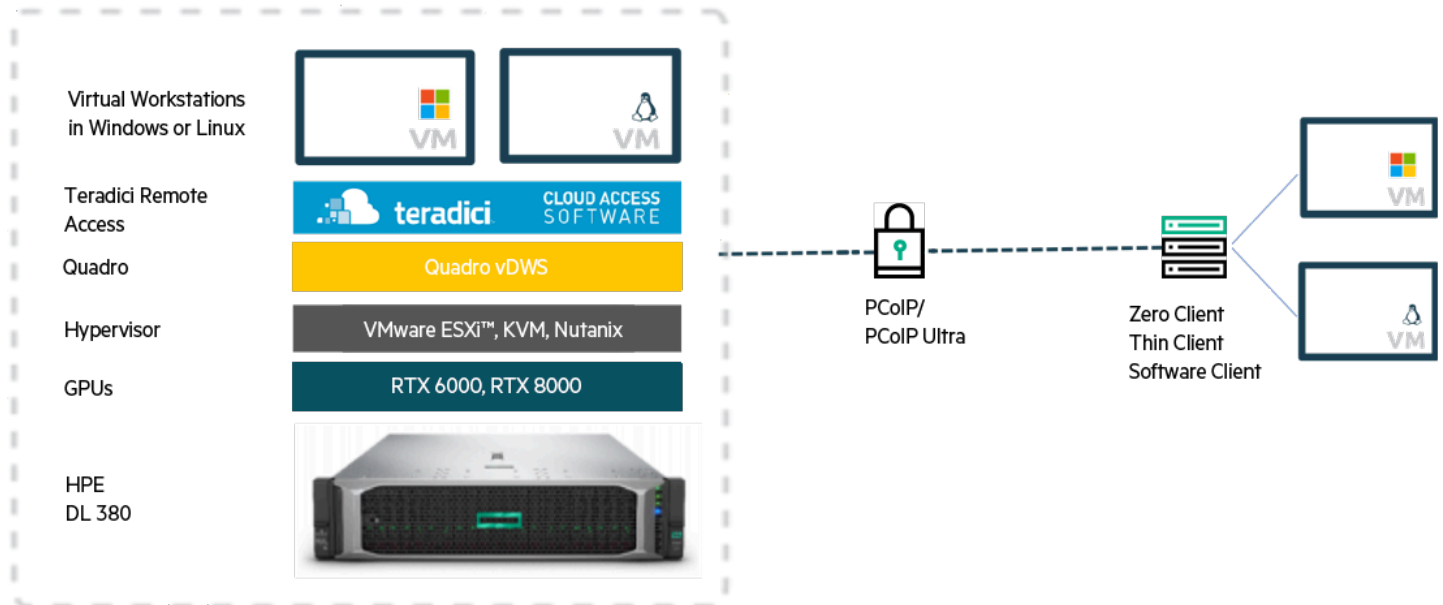


FIGURE 5. HPE ProLiant DL380 Server with Teradici

Connection Broker

There are multiple options for connection brokers to use with the solution. Customers have successfully deployed this solution as specified, but with Leostream Connection Broker added on. It is also possible to use the connection broker in Teradici Cloud Access Software. The choice here will not impact user performance and is dependent on the environment and how users expect to access their systems.

Avid Media Composer

Avid first delivered Media Composer NLE over 30 years ago and it has helped revolutionize how content is edited across television, film, broadcast, commercials, digital and more. A standard in Hollywood, Media Composer has evolved to become a world-class offline and finishing toolset. Recently redesigned in 2019, Media Composer not only enhanced its user interface but delivered significant changes under the hood with a new universal media engine and color floating-point capabilities extending the capabilities of the platform. From the first cut to final finishing, Media Composer delivers the performance, reliability, and feature set post-production teams demand to deliver content across more platforms while enabling editors to focus on creativity and not technology. Options that extend Media Composer to MediaCentral | Cloud UX for fast access to content in your production and media asset management libraries and tuned to run on Avid | NEXIS, Media Composer continues to set the bar for usability and performance.

CONFIGURATION GUIDANCE

Though deployment methodologies will vary greatly depending on the environment in which a solution is actually deployed, this section provides an example high-level overview of the steps to stand up this solution from hardware to user-accessible and graphics-accelerated virtual machines. For further detail, including step-by-step guides for these high-level actions, refer to the accompanying Deployment Guide.

1. Rack, stack, and cable physical hardware.
2. Provision networking at switches.
3. Image all servers with ESXi 6.7 U3 or ESXi 7.0.



4. Configure networking information.
5. Install vCenter and add servers.
6. Install and configure NVIDIA vGPU software license server.
7. Install and configure NVIDIA vGPU Manager for ESXi.
8. Install and configure vSAN.
9. Provision a Windows 10 VM.
10. Configure the VM and OS settings.
11. Install NVIDIA drivers and point to the NVIDIA license server.
12. Install desired applications.
13. Convert the VM to a template.
14. Test VM performance in Avid Media Composer.

CAPACITY AND SIZING

Vertical sizing

Vertical sizing, or compute resource allocation, is the term used to describe how many CPU cores, how much memory, which vGPU profile, and how much disk space will be allocated to a virtual machine to perform a certain workload or function.

Two traditional editing profiles can typically be defined, specifically in the context of Avid Media Composer:

Light editing: Editing of a relatively simple sequence timeline with minimal video streams, such as for news or sports highlights. This is usually for broadcast companies.

Advanced editing: Editing of more complex sequences, like for episodic television or feature films with extensive video stream counts. This is usually for post-production companies.

It is important to remember that editing requirements differ from customer to customer. A customer environment may be different than the profiles described above. It is recommended to use those profiles as starting points, and resize the virtual machine to match the necessary specification for a specific workload.

Solution architects should consider additional variables. Some of them are:

- The codec used will also play a role. Codecs like XDCAM and Avid DNxHD can playback more streams per core because they require less compute power to encode and decode, versus more dense codecs like AVC-Intra, AVC LongG, or XAVC.
- Lower framerates like 24 or 25 fps are more forgiving than 30 fps projects when it comes to screen refresh rates involved with the remote desktop display (PCoIP) process.
- The Video Quality Playback Setting used in the Media Composer Timeline. As a reminder, Media Composer has three settings for playback quality: Draft quality, Performance, and Full Quality. Performance mode would potentially allow more streams without dropping frames. Full quality on the contrary would decrease the number of streams.

CPU

The majority of Media Composer workflows require the virtual machine to have between 6 and 12 cores assigned. For lighter editing workloads, the virtual to physical core allocation ratio can safely be calculated and provisioned at a rate of 1.5:1. For more intensive editing where a timeline will often playback more than three streams simultaneously with a lot of effects, the virtual to physical core allocation ratio should be reduced to 1.2:1. These virtual to physical core allocation ratios have been tested and validated by Avid directly for this solution, as well as for other media virtual machine deployments. Some CPU cores should be reserved as overhead for the ESXi hypervisor to avoid overprovisioning the processors.

Hyperthreading may be enabled on the VMware host but CPU sizing should be based on physical cores and not on logical processors.



Table 4 shows density calculation for light editing workloads, using 6 cores per VM, on both the Intel Xeon Gold 6258R and 6248R.

TABLE 4. Density calculation for lighting editing workloads using 6 cores

Component	6258R, 2.7GHz, 28-core	6248R, 3.0GHz, 24-core
Total physical cores	56 cores	48 cores
Hypervisor overhead	-8 cores	-8 cores
Workload physical cores	48 cores	40 cores
Virtual to physical core ratio	1.5x	1.5x
Assignable virtual cores	72 vCPU	60 vCPU
vCPU per VM	6 vCPU	6 vCPU
Density (VMs per host)	Up to 12 VMs per host	Up to 10 VMs per host

Table 5 shows density calculation for advanced editing workloads, using 12 cores per VM, on both the Intel Xeon Gold 6258R and 6248R.

TABLE 5. Density calculation for advanced editing workloads using 12 cores

Component	6258R, 2.7GHz, 28-core	6248R, 3.0GHz, 24-core
Total physical cores	56 cores	48 cores
Hypervisor overhead	-8 cores	-8 cores
Workload physical cores	48 cores	40 cores
Virtual to physical core ratio	1.2x	1.2x
Assignable virtual cores	58 vCPU	48 vCPU
vCPU per VM	12 vCPU	12 vCPU
Density (VMs per host)	Up to 5 VMs per host	Up to 4 VMs per host

Memory

Calculating memory allocation for a VM running Media Composer is nearly identical to deciding on memory specifications for a physical workstation. For smaller projects, allocating 32GB of memory to each VM is recommended. Larger projects that have large bins and bins with complex sequences open use more memory, and therefore should have 48GB – 64GB of memory assigned to them.

GPU

Media Composer VMs require a supported NVIDIA graphics card that supports NVIDIA vGPU to be installed in the host server node so that the virtual machine can be assigned a vGPU profile. When using a GPU card with virtualization, the vGPU profile represents the size of the frame buffer that is reserved for the virtual machine. The same profile must be used by all virtual machines sharing a single GPU card. However, different profiles can be assigned to VMs sharing a different physical GPU. All HD projects should use at least the 4Q profile which reserves 4GB of the frame buffer. In testing, the 4Q profile provided sufficient frame buffer to support two 1920x1080 monitors, several streams of commonly used HD codecs, and the remote display data. For UHD, 4Q profile can still be used but it is recommended to configure a virtual machine with an 8Q profile and higher.

While these requirements apply for Avid Media Composer, it is common for the media editors who are using these VMs to have other applications as part of their workflows, and they often require a larger frame buffer. Applications such as Autodesk Maya and 3ds Max, Adobe Premiere Pro, Foundry Nuke, and others can utilize more GPU memory, and so it is common for customers to deploy with vGPU profile sizes up to 16GB.

Media Composer GPU effects processing is disabled by default in Media Composer v8.8.5 (and higher) when installed on a virtual machine. Avid has observed issues with the GPU resource assignment in a multi-client VM environment that impacts overall performance. For this reason, Avid recommends keeping the GPU effects processing disabled for all Media Composer virtual machines.



Table 6 shows some likely possibilities for physical GPU layouts and vGPU provisioning for a virtual media editor environment. Note that it is also possible to have 48GB of video memory with a single RTX 8000 GPU or two RTX 6000 GPUs, but the two RTX 6000 configurations will have double the GPU processing power compared with a single RTX 8000 GPU.

TABLE 6. Physical GPU layouts and vGPU provisioning for a virtual media editor

vGPU and user profile	1x RTX 6000 (24GB)	2x RTX 6000 (48GB)	2x RTX 8000 (96GB)
4Q (4GB) HD/ UHD editing	Up to 6 users GPU limited	Up to 12 users CPU dependent	
8Q (8GB) UHD+ editing	Up to 3 users GPU limited	Up to 6 users CPU dependent	
16Q (16GB) UHD editing plus additional applications		Up to 3 users GPU limited	Up to 6 users CPU dependent

Disk

A virtual machine running Avid Media Composer does not require a lot of OS disk space. The Windows 10 OS will use about 20GB of space. Media Composer, Avid NEXIS Client, and the Avid Workgroup Client apps will use another 5GB of space. Any other third-party applications that might be installed, as well as some space for Windows updates to download, must be accounted for. Most use cases can safely configure the OS virtual disk with 80–200GB of space for the Media Composer VM.

It is recommended to add a dedicated disk to the virtual machine to store Media Composer projects and corresponding media files during import. Disk space will depend on the codecs used and hypervisor datastore capacity. If a customer uses Nexis storage, there is no need for an additional disk. All projects and media will be directly stored in the Nexis storage and accessed over the network.

Recommendation

Remember that it is possible to overprovision CPU cores. With virtualization, the hypervisor manages CPU calls through CPU scheduling, and assigning too many cores to a single VM can actually decrease the performance due to an increase in CPU ready or wait times. It is easier to understand the VM resource and performance by starting with the minimum required resources and increasing them as needed.

On the other hand, memory, vGPU profiles, and disk space cannot be over-provisioned. If more memory than needed is allocated, it will not negatively impact the performance of the virtual machine. The physical GPU will consume all of its memory with fewer users if those users have larger vGPU profiles, but with GPUs like the NVIDIA Quadro RTX 8000 with 48GB of memory, this is not a configuration issue. And finally, more disk space will not negatively impact performance.

Test setup

The following baseline tests and calculations were based on the tested HPE ProLiant DL380 server specification in Table 7.

TABLE 7. HPE ProLiant DL380 server specification

Component	Specification
CPU	2x Xeon Gold 6258R, 28-core, 2.7GHz or 2x Xeon Gold 6248R, 24-core, 3.0GHz
Memory	768GB
GPU	2x RTX 8000 (96GB total vRAM)
Disk	5.2TB
OS	ESXi 7.1, Windows 10



Table 8 describes the two types of virtual machine templates that were tested with the HPE ProLiant DL380 Gen10 server with two NVIDIA RTX 8000 GPUs. Each VM was running Teradici Cloud Access Software and was connected to two monitors running at 1080p resolution using HP t310 quad-port thin clients.

TABLE 8. Two types of virtual machine templates

Component	Light editing	Advanced editing
vCPU	6	12
Memory	32GB	64GB
vGPU profile	4Q	8Q
OS disk	200GB	200GB

Table 9 describes the user densities for two types of virtual machine templates that were deployed on the HPE ProLiant DL380 Gen10 server. Each VM was connected to two monitors running at 1080p resolution using HP t310 quad-port thin clients.

TABLE 9. User density

Processor	Light editing	Advanced editing
Intel Xeon Gold 6258R	12 workstation VMs	5 workstation VMs
Intel Xeon Gold 6248R	10 workstation VMs	4 workstation VMs

Two media resolutions were tested: HD and UHD. DNxHR SQ was chosen as the video codec to perform the baseline test.

Workload description

Once all the workstations were built on the hypervisor, baseline tests were performed on each virtual machine.

These steps are designed to provide a logical and efficient process to establish a baseline performance with Media Composer VDI clients. Executing these steps will help determine that the setup and configuration will meet the performance standards necessary to achieve both operational and business goals, and importantly a proper end-user experience.

If possible, configure remote access to one or more physical Media Composer workstations using the Teradici PCoIP client, just like the VM. This way if there are any issues, the same steps and tests can be performed with the workstation and the results compared to the Media Composer VM. This will help determine if the issue has more to do with the PCoIP experience, or if it is a problem inherent to the performance of the VM.

Table 10 shows data on the number of video streams per virtual machine tested by Avid for each workload.

TABLE 10. Number of video streams

Stream resolution	Light editing	Advanced editing
HD(1080p59.94 DNxHR SQ)	6 streams (draft quality mode)	12 streams (draft quality mode)
UHD(3840xx2160 30p DNxHR SQ)	3 streams (draft quality mode)	7 streams (draft quality mode)

Media Composer VM playback test setup

The following steps were performed for testing Media Composer VM playback:

1. Create a project that is 1920x1080, 59.94 for HD, and 3840x2160 for UHD.
2. Create media using the DNxHR SQ Codec to perform this baseline test.
3. Using the DNxHR SQ media create a simple single track, cuts the only sequence using several different clips.



- Open the Media Composer Console window (from the Tools Menu). In the field at the bottom type and enter the command `sℓ`. This command will provide information on whether any frames were late or dropped during playback. This command will persist until you quit Media Composer.

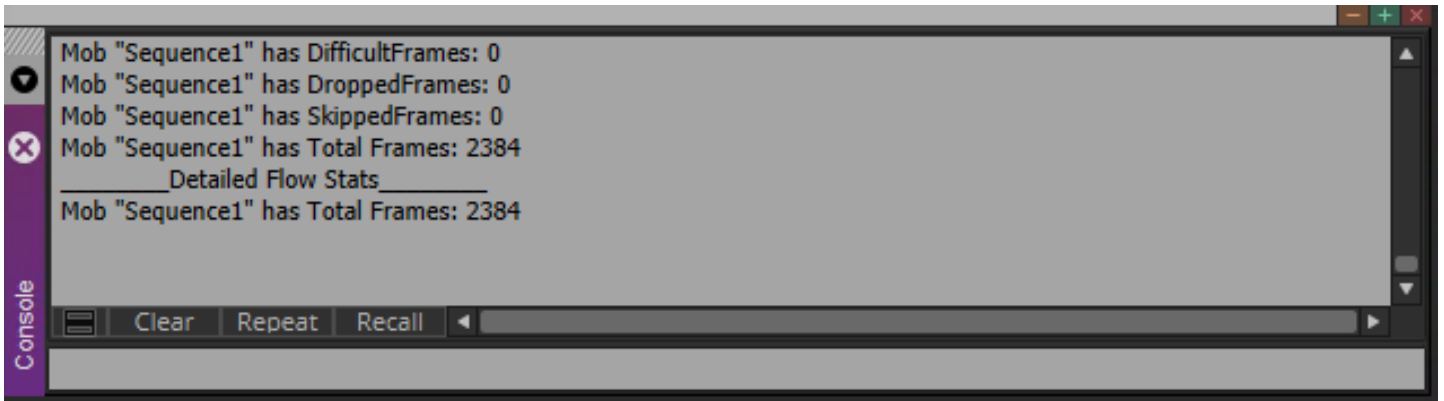


FIGURE 6. Media Composer Console window

- Play the sequence with one video track. Check the Console window to verify there were no dropped frames. Playback should not drop any frames with a single track. If the console does indicate frames were dropped, the source of the problem needs to be troubleshooted. If no frames were dropped proceed with the next steps to build a sequence with the PIP effect to test playback with multiple streams.
- Make the sequence to be about 50 seconds long. Add four more tracks of different video clips to your sequence and add a PIP effect to V2 - V5. Make the IN-point for each new layer after the previous layer plays for 10 seconds.

Figure 7 shows the timeline that looks like the following screenshot.

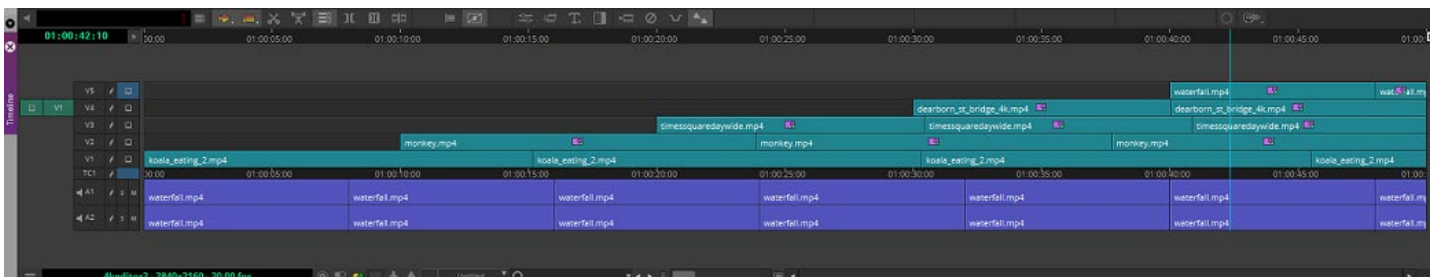


FIGURE 7. Timeline



- 7. Shrink the PIP effect on V2 to about 1/8 the size of the playback window, reposition it to the top left corner as shown in Figure 8. Repeat the process with the PIP effect on V3 - V5 positioning each track so that a grid is created that looks like the screenshots below at the end of the sequence when monitoring all seven tracks.



FIGURE 8. PIP effect

- 8. Make sure the monitor icon is on the 5th track as shown in Figure 9, so the video for each track appears when the sequence is played back.

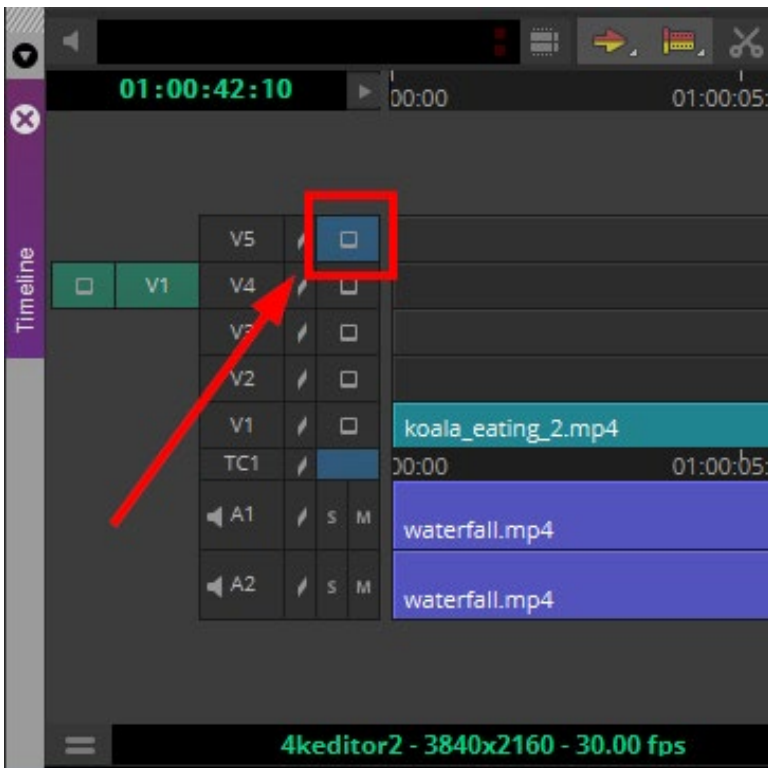


FIGURE 9. Media editor – Monitor icon



9. Play the sequence back and check the console for dropped frames. If there are no dropped frames loop play the sequence for several minutes and check the console again for dropped frames. Check Performance under Task Manager (in Windows) to see the graph and percentage values of CPU utilization for the VM. The graph should show an increase in CPU utilization with each track added and then drop back down when the sequence finishes and restarts at the beginning again with just one track playing as shown in Figure 10.

CPU

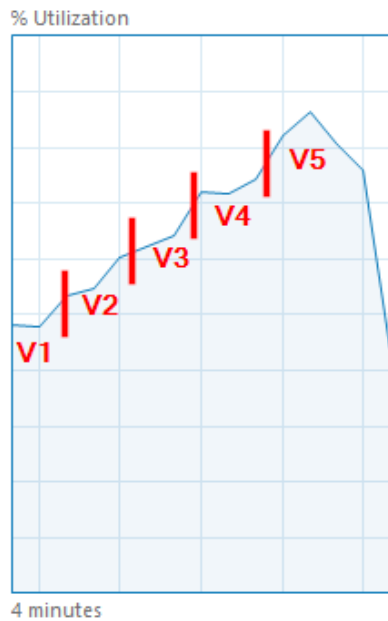


FIGURE 10. CPU utilization

10. Continue to add another track with a PIP effect. Playback the sequence with each added layer and check for dropped frames and continue to monitor the CPU and GPU utilization of the VM with each playback. There shouldn't be dropped frames until the CPU utilization exceeds 90%.
11. Once the sequence has six or more tracks, continue to make changes, and repeat the tests using the same sequence by just changing which track the monitor icon is placed on. If there is a sequence with 10 video tracks but the monitor is placed on V6, then only the 6 layers will be playing and monitored.

Testing multiple VMs and documenting baseline performance data

Perform the following steps to test multiple VMs:

1. Deploy additional VMs on the same hypervisor until the number of the system was designed to host is reached. Repeat all the steps above on each VM so a sequence can be loop played from all the VMs simultaneously.
2. While all the VMs are playing their sequence, make a note of how much CPU is being used by each VM and the overall CPU utilization of the host.
3. Continue to perform tests using this method but build sequences that will be more similar to those in a real-world project.



4. Media Composer has three settings for Playback Quality. New projects will default to Timeline Settings using Draft Quality Playback. Changing this setting could yield more streams without dropping frames in Performance Mode or less streams in Full Quality Mode.

	Draft Quality (Default)	Processes and plays a subsample of the full image raster for the project that uses 1/4 of the image information. Uses a bit depth of 8 bits. This option subsamples 50% of the raster width. For interlaced projects, this option uses one field. For progressive projects, this option uses 50% of the scan lines.
	Performance	Processes and plays a subsample of the full image raster for the project that uses 1/16 of the image information. Uses a bit depth of 8 bits. This option subsamples 25% of the raster width. For interlaced projects, this option uses 50% of the lines in one field. For progressive projects, this option uses 25% of the scan lines.
	Full Quality	Processes and plays the full image raster for the project. Uses a bit depth of 8 bits. This option provides the highest video playback quality by processing every image pixel. In interlaced projects, this option processes the full width of every line in both fields. In progressive projects, this option processes the full width of every scan line.

FIGURE 11. Media Composer settings

System performance data

Each graph shows one VM performance data in an advanced and light editing environment:

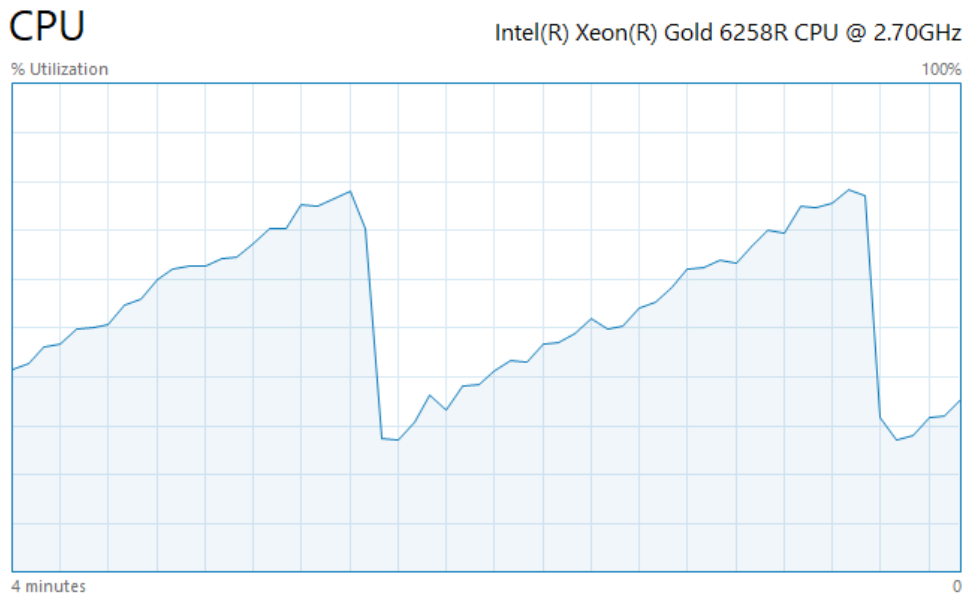


FIGURE 12. Advanced Editing 1080p – CPU utilization





FIGURE 13. Advanced Editing 1080p – RAM utilization

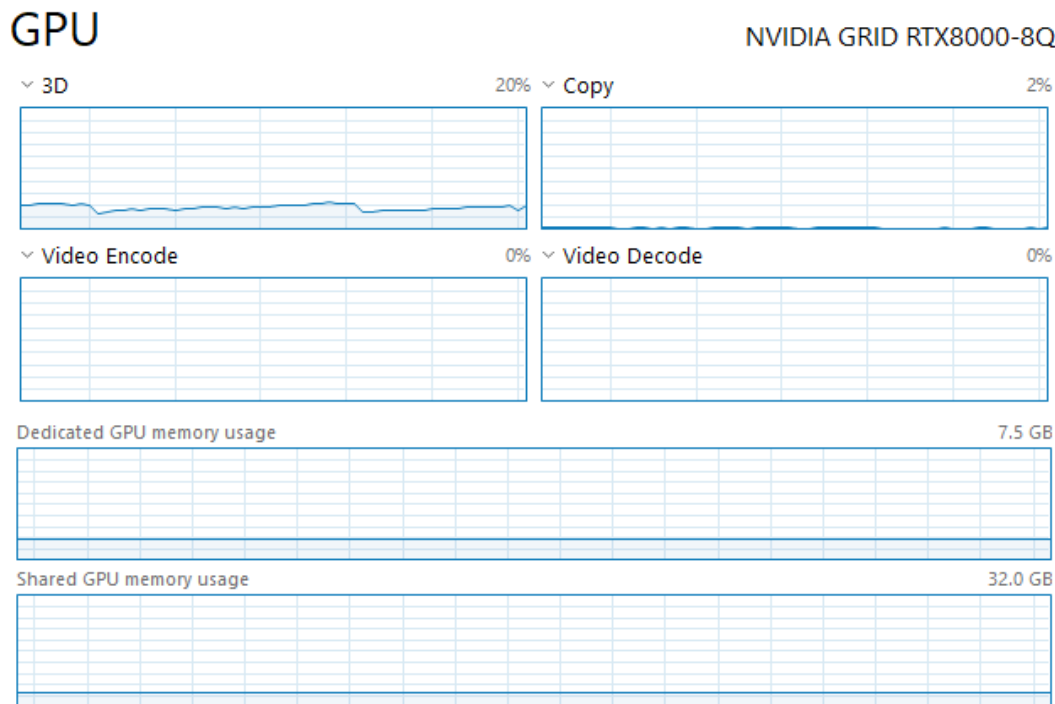


FIGURE 14. Advanced Editing 1080p – GPU utilization



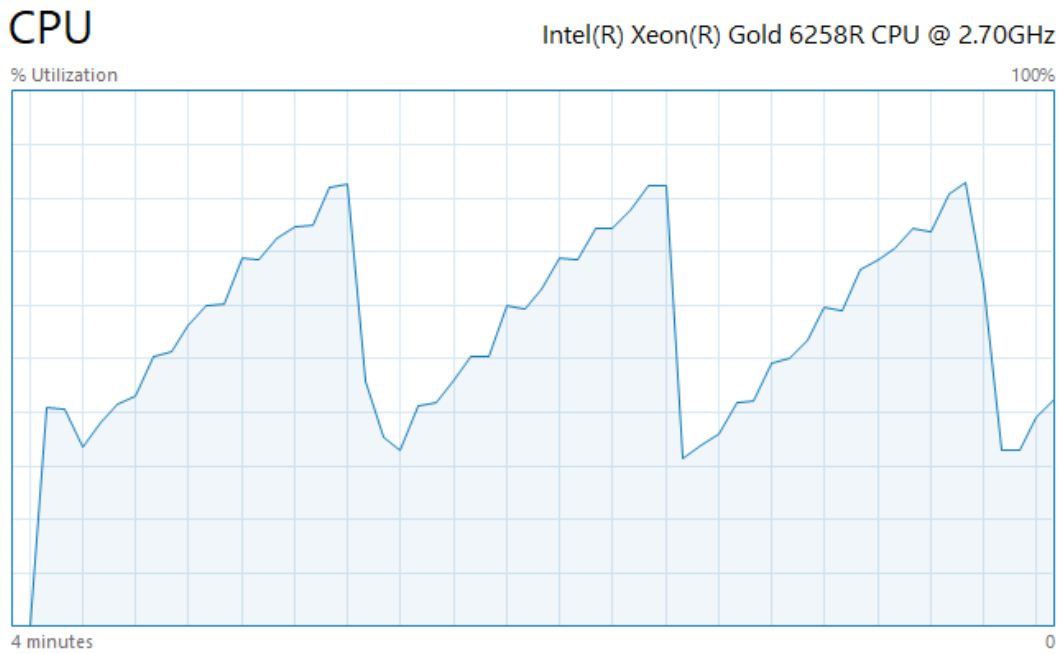


FIGURE 15. Advanced Editing 3840p – CPU utilization

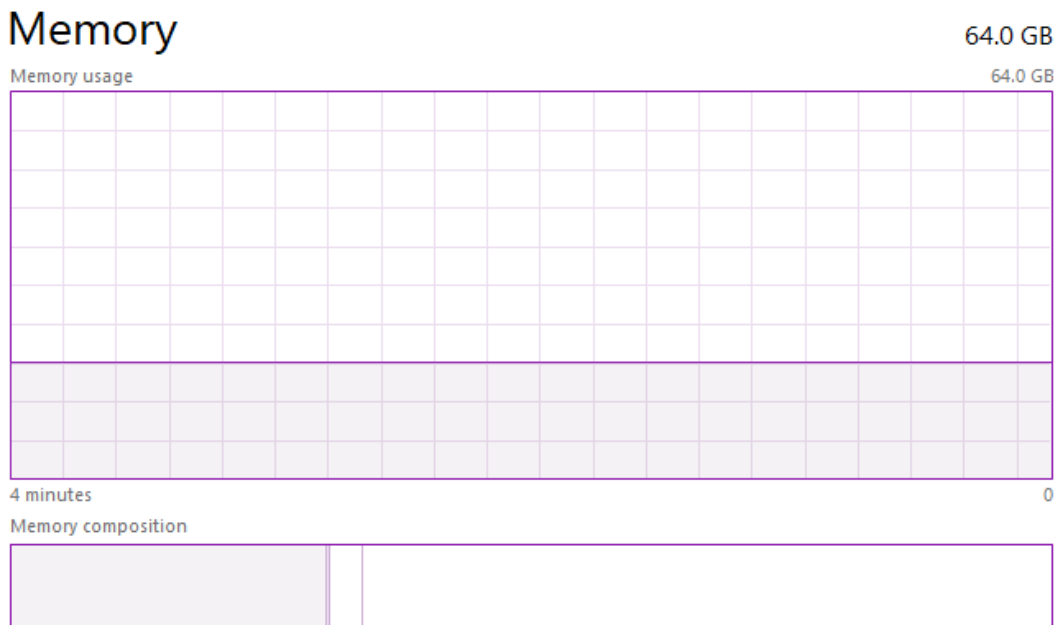


FIGURE 16. Advanced Editing 3840p – RAM utilization



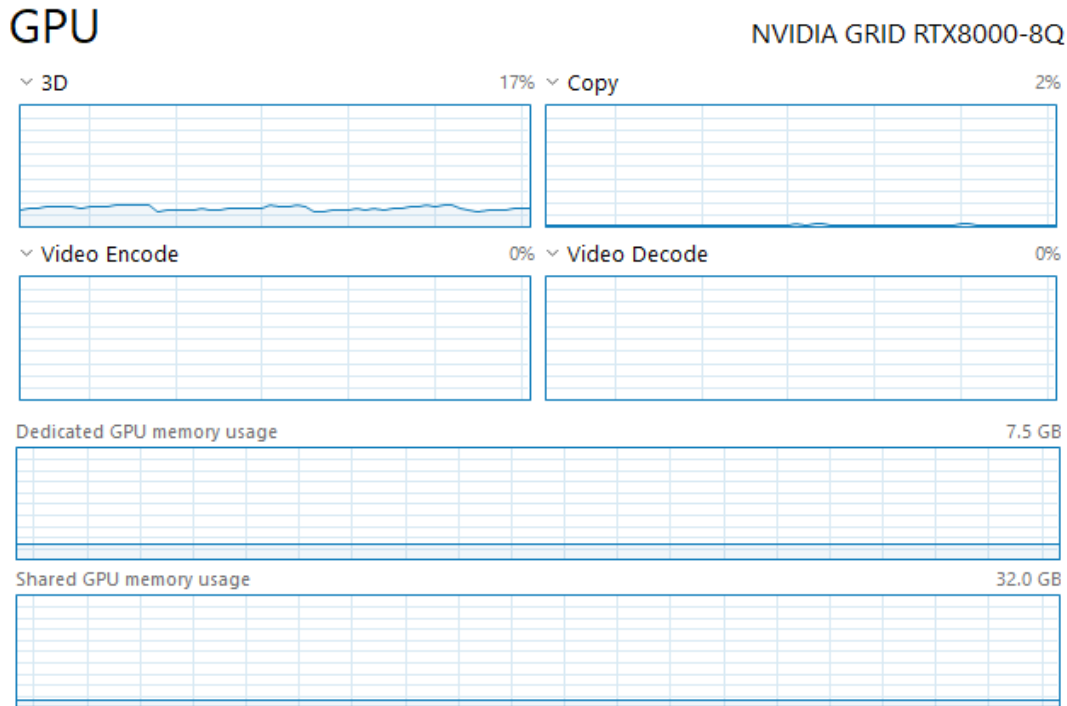


FIGURE 17. Advanced Editing 3840p – GPU utilization

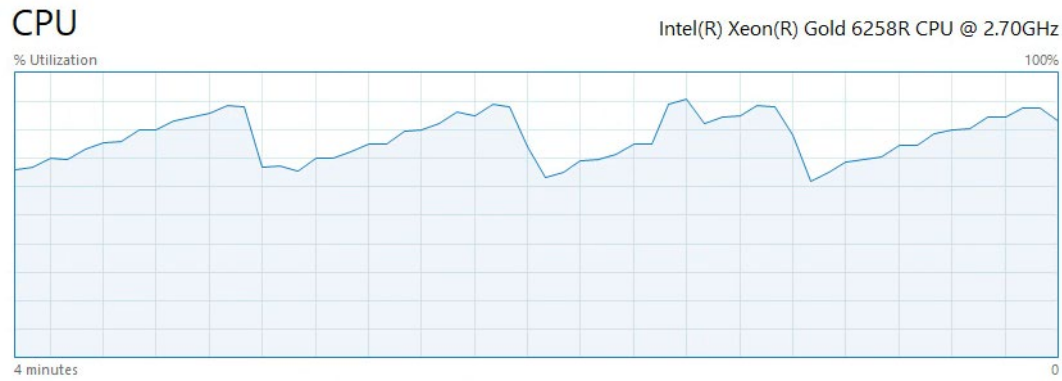


FIGURE 18. Light Editing 1080p – CPU utilization





FIGURE 19. Light Editing 1080p – RAM utilization

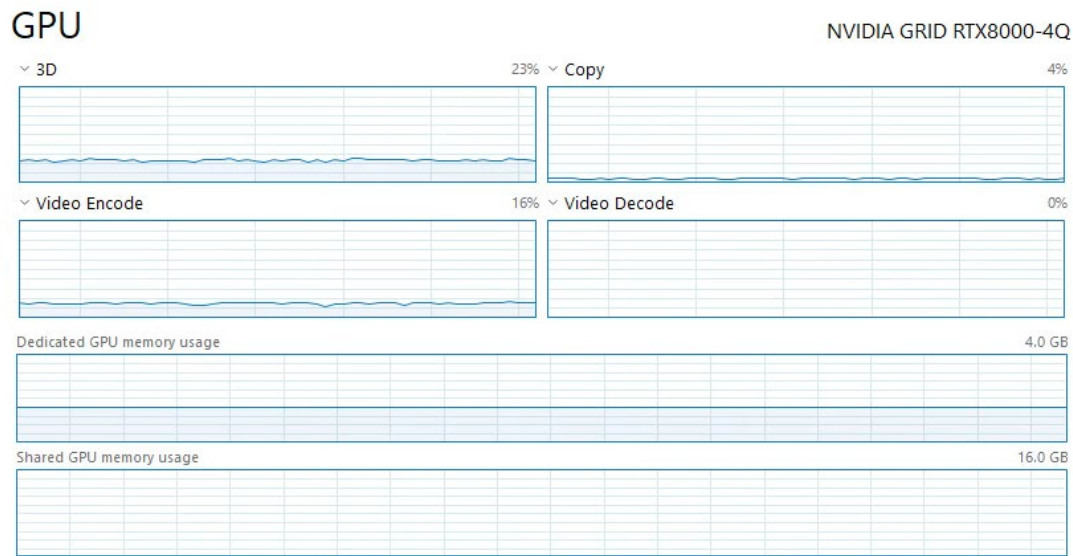


FIGURE 20. Light Editing 1080p – GPU utilization



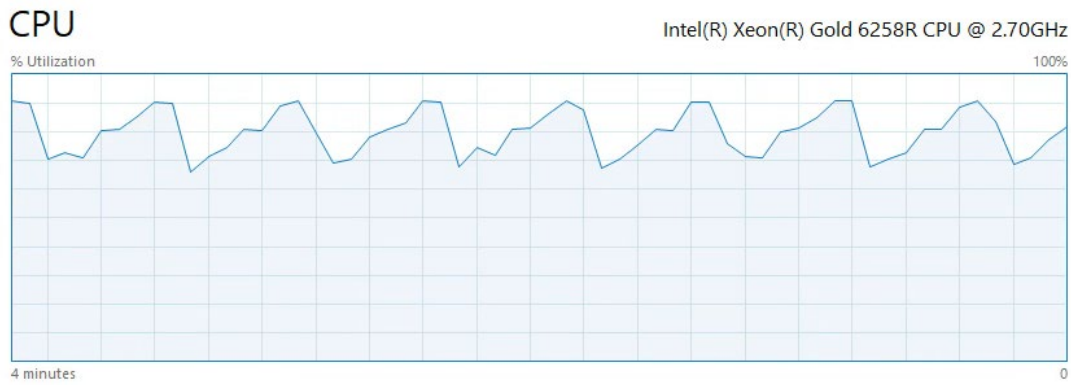


FIGURE 21. Light Editing 3840p – CPU utilization

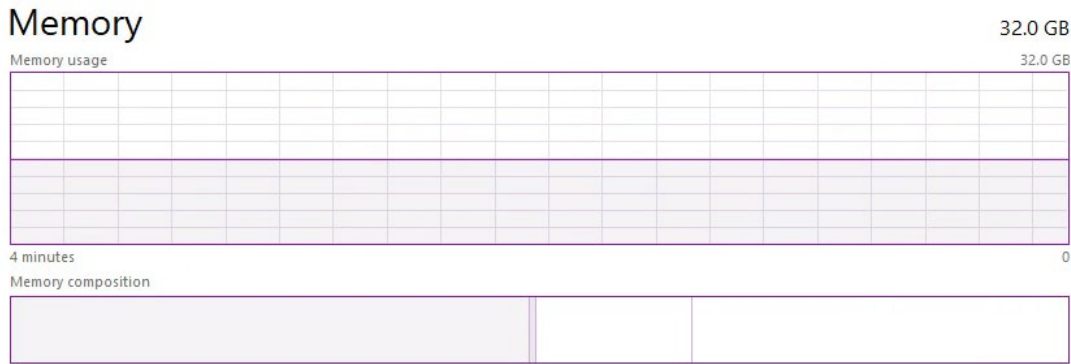


FIGURE 22. Light Editing 3840p – RAM utilization



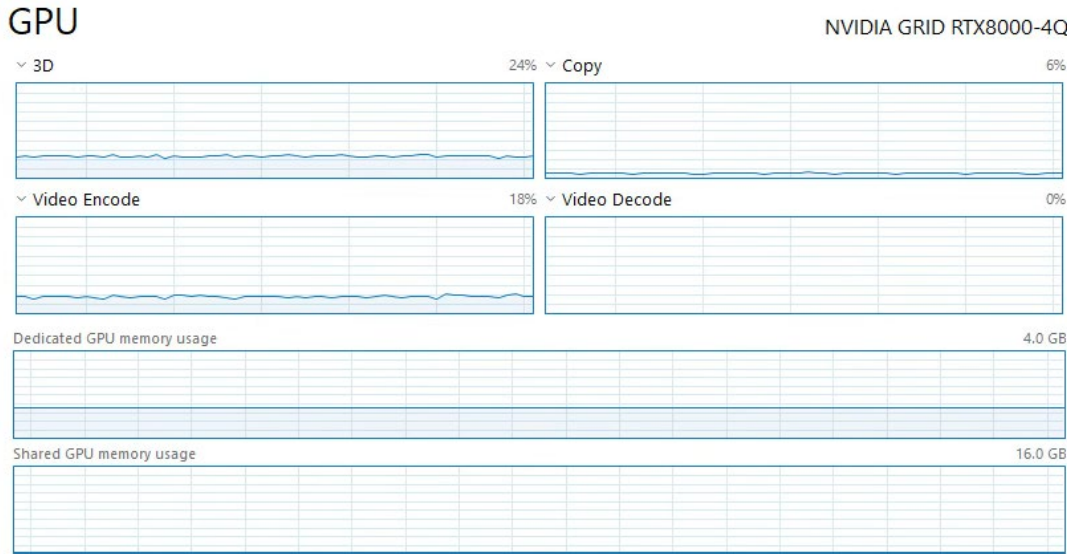


FIGURE 23. Light Editing 3840p – GPU utilization

Results

The virtual to physical core allocation ratios of 1.5 for light editing and 1.2 for advanced editing was confirmed by the results of the tests above. CPU utilization was pushed to 80-90% for the tested scenarios and densities based on those allocation ratios. It is best to leave CPU utilization below 90% for best performance and to leave room for short-term spikes that may drive usage slightly higher.

Table 11 shows the workload description in addition to stream counts, each column shows the peak CPU usage of the VM for each resolution during the tests.

TABLE 11. CPU usage of the VM

Stream resolution	Light editing	Advanced editing
HD(1080p59.94 DNxHR SQ)	6 streams (draft quality mode) 88% VM CPU	12 streams (draft quality mode) 78% VM CPU
UHD(3840xx2160 30p DNxHR SQ)	3 streams (draft quality mode) 90% VM CPU	7 streams (draft quality mode) 82% VM CPU

Analysis and recommendations

Total CPU core count is the primary limiting factor for density, though a high base clock speed is important as well. The best CPU choice for a media editor virtual workstation environment depends on several factors. For maximum density, the Intel Xeon Gold 6258R is recommended, because it can host an additional advanced editing VM or two additional light editing VMs per server. However, it is necessary to take into account the utilization of other system resources to determine the best option for a given environment. Some other factors to consider include:

GPU memory - The 6258R CPU choice, with five advanced editor users as tested would leave some GPU memory ‘stranded,’ or not utilized, whether using a single RTX 8000 or two RTX 6000s. Underutilization of GPU memory can be avoided by assigning different vGPU profiles across different cards as needed (for example, two 12Q users on one RTX6000 and three 8Q users on another), but this is only possible with multiple GPU cards, as all vGPU profiles on the same card must be identical, and it also introduces additional complexity for user mobility and disaster recovery, as only like profiles can be moved onto a single GPU. On the other hand, the 6248R CPU choice, with four advanced editor users as tested, divides GPU memory nicely in either a single or double GPU configuration.

System memory - Intel processors with six memory channels have different balanced memory configurations than in previous generations, with the best performance coming from configurations with either six or twelve DIMMs per CPU. From the light editor example with the Intel Xeon



Gold 6248R, 10 users each with 32GB of RAM would require 320GB of RAM, plus hypervisor overhead. This fits well into a 12x 32GB DIMM configuration of 384GB. On the other hand, from the light editor example with the Intel Xeon Gold 6258R, 12 users each with 32GB of RAM would require 384GB of RAM plus hypervisor overhead, which would require moving to the next balanced configuration of memory and some amount of overprovisioning. The specific breakpoints for the number of users and required memory depend on user requirements in each unique environment.

It is important to remember that the pre-defined 'light' and 'advanced' editors tested for this solution are sizing starting points, and are not indicative of every user's workflow. If the intended editor users for the system fall in between these pre-defined profiles, it could be best to deploy an intermediate number of users that can divide GPU resources evenly, such as six or eight editor users per server. It is important to conduct user testing with actual workflows to validate that the solution sizing presented here will translate to an excellent end-user experience in a given environment.

GPU is not a limitation for the Avid workflow, but other applications used by media editors could require larger GPU frame buffers.

SUMMARY

Media editors require high-end desktops with fast processors and graphics acceleration to facilitate working with multiple full HD and 4K streams simultaneously in editing programs such as Avid Media Composer. The traditional desktside workstation model is cumbersome for many IT organizations, and they are searching for ways to improve data security, simplify desktop management and facilitate user mobility without the constant battle of never-ending hardware refreshes.

The HPE ProLiant DL380 Gen10 equipped with NVIDIA Quadro RTX 6000 or RTX 8000 GPUs can address these requirements, providing high-end graphics-accelerated virtual desktops, hosted securely in the data center. Media editors get an excellent editing experience on a desktop and they can access it from anywhere, their IT policies allow through the use of Teradici Cloud Access Software. The solution is easily configurable for light and advanced levels of editing, both now and in the future, with virtual machines using NVIDIA vGPU software.

The HPE ProLiant DL380 Gen10 equipped with NVIDIA Quadro RTX 8000 GPUs has been tested with Avid Media Composer for virtualized media editor workloads, to improve resource utilization as compared to dedicated desktside workstations. The key takeaways include:

- Support for up to 12 light editors or 5 advanced editors per HPE ProLiant DL380 Gen10 server, with optional CPUs to best fit the specific deployment and density requirements of different environments.
- Excellent end-user performance from a highly manageable VDI solution that brings many advantages over desktside workstations.
- GPU flexibility with NVIDIA vGPU, allowing for frame buffers to be deployed to match workloads focused solely on Avid Media Composer, or to meet the increased GPU needs of workloads that also include other media applications.



RESOURCES AND ADDITIONAL LINKS

HPE Reference Architectures, hpe.com/info/ra

HPE Servers, hpe.com/servers

HPE Storage, hpe.com/storage

HPE Networking, hpe.com/networking

HPE Technology Consulting Services, hpe.com/us/en/services/consulting.html

NVIDIA vGPU Software User Guide, <https://docs.nvidia.com/grid/latest/grid-vgpu-user-guide/index.html>

To help us improve our documents, please provide feedback at hpe.com/contact/feedback.

© Copyright 2021-2024 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Avid product specifications are subject to change without notice and do not represent a commitment on the part of Avid Technology, Inc. Use of Avid software is subject to the terms and conditions of a software license agreement provided with the software. Avid products may be protected by one or more U.S. and non-U.S. patents. Details are available at www.avid.com/patents. Avid, Avid DNxHR, Avid DNxHD, Avid NEXIS, Interplay, MediaCentral, Media Composer and all related product names and logos, are registered or unregistered trademarks of Avid Technology, Inc. in the United States and/or other countries. The Interplay name is used with the permission of the Interplay Entertainment Corp. which bears no responsibility for Avid products. All other trademarks are the property of their respective owners.

Microsoft and Windows are U.S. registered trademarks of Microsoft Corporation. NVIDIA is a registered trademark of NVIDIA Corporation in the U.S. and other countries. VMware is a registered trademark of VMware, Inc. in the United States and/or other jurisdictions. vSphere, VMware vCenter Server are the registered trademarks of VMware, Inc. in the United States and/or other jurisdictions. VMware ESXi, VMware vSAN are the trademarks of VMware, Inc. in the United States and/or other jurisdictions. Intel and Xeon are trademarks of Intel Corporation in the U.S. and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries. All third-party marks are property of their respective owners.