3D Graphics for Virtual Desktops Smackdown

Author(s): Shawn Bass, Benny Tritsch and Ruben Spruijt

Version: 1.11

Date: May 2014

PQR

Eenvoud in ICT

CONTENTS

1. INTRODUCTION

Are you looking for an independent overview of desktop virtualization solutions and are curious about different strategies?

Are you interested in the use-cases and benefits of delivering 3D graphics? Do you want detailed information about the features and functions each vendor is offering? Do you want to know how you can enable and leverage 3D graphics for virtual desktops? If so this is the whitepaper you definitely must read!

In the current market there is an increasing demand for unbiased information about hardware accelerated graphics desktop virtualization solutions. This white paper is focused on solutions that are anticipated to have an important role in desktop virtualization deployments. An overview of available features of the various solutions has been created to provide a better understanding of capabilities and to assist you in understanding important differences between these technologies.

1.1 OBJECTIVES

The goals of this whitepaper are to:

- Provide an application and desktop delivery solutions overview
- Explain different desktop virtualization concepts
- Explain the pros and cons of desktop virtualization
- Highlight the use-cases and benefits of 3D graphics for virtual desktops
- Describe the vendors and their solutions to enable graphics within desktop virtualization
- Compare the features of the various 3D graphics for virtual desktops solutions

1.2 INTENDED AUDIENCE

This document is intended for IT Managers, architects, analysts, system administrators and IT-Pros in general who are responsible for and/or interested in designing, implementing and maintaining hardware accelerated graphics desktop virtualization Infrastructures.

1.3 VENDOR INVOLVEMENT

All major vendors whose products are covered such as AMD, Citrix, HP, NVIDIA, Microsoft, Teradici and VMware have been approached in advance to create awareness of this whitepaper and discuss their solutions functionality and features.

1.4 FEEDBACK

We try to provide accurate, clear, complete and usable information. We appreciate your feedback. If you have any comments, corrections or suggestions for improvements to this document we want to hear from you! Please send an email to team@teamRGE.com include the product name, version number and the title of the document you're reading in your message.

1.5 CONTACT

PQR; Tel: +31 (0)30 6629729

E-mail: info@pqr.nl ; www.pqr.com

Twitter: http://www.twitter.com/pqrnl

THIS DOCUMENT IS PROVIDED "AS IS"
WITHOUT WARRANTY OF ANY KIND
FOR REFERENCE PURPOSES ONLY

COPYRIGHT 2014 PQR

IT IS NOT ALLOWED TO (PARTIALLY) PUBLISH OR DISTRIBUTE CONTENT WITHOUT APPROVAL

2. ABOUT

2.1 ABOUT PQR

PQR is a professional ICT infrastructure company focusing on the availability of data, applications and workspaces, with optimized user experience, in a secure and manageable way. PQR provides its customers innovative ICT solutions, from on-premises to cloud management, without complicating matters. Simplicity in ICT, is what PQR stands for. PQR has verifiable references and a wide range of expertise in the field, proven by many of our high partner statuses and certifications. PQR is a Citrix Platinum Solution Advisor, HDS Tier 1 Platinum Partner, HP GOLD Preferred Partner, Microsoft Gold Partner, NetApp Star Partner, RES Platinum Reseller, VMware Premier Partner and VMware Gold Authorized Consultant Partner.

PQR's approach is based on four main pillars:

- Data & System Availability
- Application & Desktop Delivery
- Secure Access & Secure Networking
- Advanced IT Infrastructure & (Cloud) Management

PQR, founded in 1990, is headquartered in De Meern and has over 107 employees. In fiscal year 2011/2012 posted sales of € 94.9 million and a net after tax profit of € 4.6 million have been recorded..

2.2 ACKNOWLEDGEMENTS

Bernhard Tritsch, CTO at bluecue consulting

Dr. Benny Tritsch is the Chief Technology Officer with bluecue consulting in Germany. He is a business developer, principal consultant, market analyst, author, and all-around geek specializing in enterprise Windows remoting and virtualization solutions. His areas of expertise include: Remote Desktop Services, VDI, high-end graphics remoting, Hyper-V, enterprise Windows deployment tools and solutions, PowerShell

for IT Pros, and user workspace management. Benny speaks around the world at several conferences each year, including Microsoft TechEd, Citrix Synergy, VMware VMworld, BriForum and E2EVC. He has received the Microsoft Most Valuable Professional (MVP) award for RDS since 2004 and the Citrix Technology Professional (CTP) since 2006. If you want to know more about Benny's activities in the virtualization community, check out his website at www.DrTritsch.com or follow @drtritsch on Twitter

Shawn Bass, Solution Architect at shawnbass.com

Shawn Bass, an independent consultant based in the Chicago area, is a Citrix Technology Professional (CTP) since 2006, and a Microsoft Most Valuable Professional (MVP) since 2008. He's been working with SBC, VDI and Application Virtualization technologies since their inception. Shawn is a highly rated presenter at a variety of IT events like Citrix Synergy, BriForum, TechEd, VMworld, etc. Follow @shawnbass on Twitter or contact Shawn via email at shawn@shawnbass.com

Ruben Spruijt, CTO @ PQR

Ruben Spruijt (1975) is CTO and focuses primarily on Enterprise Mobility, Virtualization and Cloud Management. He is actively involved in determining PQR's vision and strategy. Ruben is a Microsoft Most Valuable Professional (MVP), Citrix Technology Professional (CTP) and VMware vExpert and is the only European with these three virtualization awards. He gives customers advice and has them benefit from his expertise; he motivates his colleagues and writes blogs, articles and opinion pieces on a regular basis. During presentations in several national and international congresses, Ruben shares his thoughts and knowledge on application and desktop delivery, and on virtualization solutions. To contact Ruben directly send an email to rsp@pqr.nl. Follow Ruben on twitter: @rspruijt

3. TEAM REMOTING GRAPHICS EXPERTS - TEAMRGE

Shawn Bass, Bernhard Tritsch and Ruben Spruijt are the founders of TeamRGE.

TeamRGE is a community group of experts with focus on Remoting Graphics for Virtual Desktops and Applications. The goal of this group of thought leaders is to share unbiased and independent knowledge via blog posts, white papers, videos and presentations at local and international events.

TeamWork - it's only through the effort and persistence of the 'Smackdown' community team that we achieved the goals, a big thanks to them!

Name	Position	Role	Twitter
Thomas Poppelgaard	Technology Evangelist and Blogger	Contributor	@_poppelgaard
Jits Langedijk	Sr. Consultant @PQR	Reviewer	@JitsLangedijk
Peter Sterk	Solution Architect @PQR	Reviewer	@PeterSterk

4. QUOTES

"This paper is great because who has the time to go research all these things one-by-one, let alone actually set them up and take measurements? (Well, these guys do I guess!) I love that they go into the background of how 3D graphics are used in desktop virtualization today, like how GPUs and APUs work, how the different types of GPU virtualization work, etc. They also cover the details of how the various remoting protocols work, like client-side versus host-side rendering, GDI remoting, etc. Really anyone who's delivering remote Windows desktops needs to read this paper."

Brian Madden, Speaker, author, blogger @brianmadden

"Most desktop virtualization projects start by standing up solutions in the lab, deciding they're good enough, and proceeding with a deployment. Only after scaling up the number of users and use cases does anyone really start to look under the hood to see how things work. This <white paper / book> gives you everything you need to know before you start, diving into all the combinations of desktop virtualization platforms (Citrix, VMware, MainFrame2, Microsoft, etc…) and protocols, exploring the different graphics engines (GDI, GDI+, OpenGL, DirectX, Flash, etc…), explaining the differences between graphics cards (AMD, Nvidia), and how everything (literally everything!) relates to each other.

The discussion around how to assess your 3D graphics needs and performance, not to mention the comparisons of different solutions is unparalleled in the industry, and everyone that reads this can take away useful information."

Gabe Knuth - Author, blogger and editor at TechTarget
@GabeKnuth

Shawn, Ruben, and Benny expertly paint a picture of a fast-moving landscape of remote 3D graphics. It's an instant reference guide and a must-read for anyone contemplating remote deployments of 3D graphics apps.

Nikola Bozinovic, CEO and Founder Mainframe2
www.mainframe2.com

"Virtualization industry experts Benny, Ruben and Shawn have produced a comprehensive and up-to-date document covering exactly what technologists need to know about '3D Graphics for Virtual Desktops'. It's a thorough, independent and well-researched technical deep dive into strategy, vendors, solutions, features, qualifying questions and much more. This will be an invaluable source of information for CTO's and technical architects looking to realize the many benefits of virtualizing and remoting 3D graphics applications."

Derek Thorslund, Director of Product Management, Citrix Systems @DerektCitrix

Shawn, Benny, and Ruben have created something special with their 3D Graphics for Virtual Desktops Smackdown whitepaper. This document is a deep dive into the technology and product landscape that is useful to newcomers and experienced practitioners alike. And they don't disappoint with the new 3D Graphics for Virtual Desktops Smackdown. Once just the domain of a subset of power users, 3D graphics are now a requirement for many different users and use cases. Shawn, Benny, and Ruben do an excellent job reviewing the technology behind 3D graphics on virtual desktops and providing an objective view of the product offerings. It's a must-read for any CIO or CTO considering virtual desktops.

Kit Colbert, CTO, End-User Computing, VMware @KitColbert

5. TOMORROW'S WORKSPACE

Introduction

Flexible Workstyles, Bring Your Own, consumerization of IT, Enterprise Mobility Management, Unified Communications, Mobile Devices, Applification, Cloud Computing, Social Enterprise and Application and Desktop Delivery are the main trends in 'Tommorow's Workspace'. 3D graphics for virtual desktops is part of the application and desktop delivery solutions stack. Virtualization is incorporated in several of these trends. Virtualization is nothing more than the decoupling of IT resources. The forms of virtualization that are most frequently applied include network, storage, server, application and desktop virtualization. Application and desktop delivery is a process which has the goal of offering applications independent of location and device, so the business consumers can work onsite, online, off site, offline, anywhere, with any (own) device and at any time. The dynamic delivery of applications is an essential functionality and part of a broader strategy of an optimized Desktop.

Application and Desktop Delivery

One of the fundamental questions in application and desktop delivery is the question "What is

the execution platform for the applications, and where is my data stored?"

Within the execution platform, system resources such as the CPU, GPU, memory, disk and network are used in order to execute the windows, web-architected, rich mobile and mobile web-applications. The most frequently used execution platforms include the following: tablet, smartphone, desktop, laptop and desktop virtualization with both Virtual Desktop Infrastructure (VDI) and Session Virtualization (Server Based Computing).

The choice of an execution platform is the most fundamental questions and the application and desktop delivery strategy. Applications are either executed locally on a device or centrally in a private or public datacenter. Each execution platform has its own characteristics. The theories: "Less is more", "Cut out the exceptions" and "Manage diversity" should always be in mind. An execution platform is great; but if there are no applications available on the platform, the platform is of no real value to the business consumer, the end-user. The other questions which needs to be answered is: "How will (Windows) applications get onto the execution platform?!" A number of solutions exist for making Windows, web-architected, rich internet and mobile web applications available on the platform. The forms most frequently used with windows applications include installation, application virtualization, and more recently, layering.

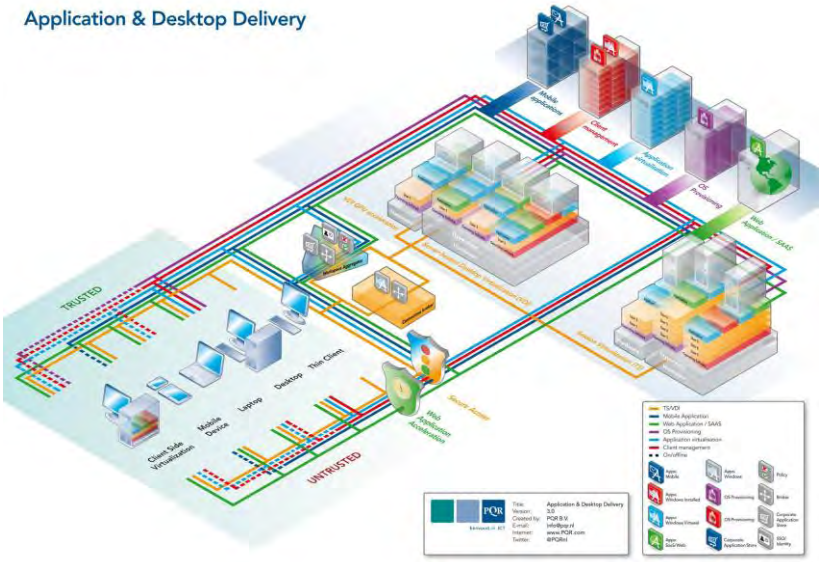
Other questions in the strategy are: "Do you need to managed the environment, what do you want to manage? (Device, application, data) from both an IT and Business Consumer perspective.

It's great to see more and more scenarios where high-end graphics,and resource intensive (Windows) applications within desktop virtualization solutions are being used. The goal of this document is to explain the use-cases, business benefits, various solutions and differences between the solutions.

Overview of application and Desktop Delivery

Before doing a 'deep dive' into 3D graphics for virtual desktops, we think it's important to have an overview of all the 'Application and Desktop Delivery' solutions before you proceed with 3D graphics for virtual desktops. PQR created the "Application & Desktop Delivery Solutions Overview" to provide an at-a-glance outline of the various

application and desktop delivery solutions. Reading this chapter will give you a complete outline of the diagram and all the application and desktop solutions that are included within it. This paragraph does not aim to describe all application scenarios or their technical advantages and disadvantages, but to give a general idea of the state of the union in the application and desktop delivery segment, independent of vendors.



Secure Access

Secure Access solutions provide secure access for (untrusted) devices to corporate IT resources. A Secure Access solution could be a full (SSL) VPN solution or a Gateway Services which is targeted for Server Hosted Desktops. Solutions that can be used to realize secure access scenarios include Cisco ISE, Citrix Netscaler Gateway, Juniper SSL VPN, Microsoft Remote Desktop Services Gateway, Microsoft UAG and VMware View Security Server.

A complete overview of Secure Access and Secure Networking solutions has been created and can be downloaded here.

Mobile Application Delivery

Rich Mobile applications running natively on Apple iOS, Google Android, Blackberry, Windows Phone and Modern Apps on RT or Windows 8.x are delivered by the Mobile Application delivery solution. In enterprise customer scenarios this function is incorporated in most of the Enterprise Mobility Management solutions but it can be a more consumer focused application store as well. The application store is the interface for application access, rich mobile application delivery and usage reporting functionality.

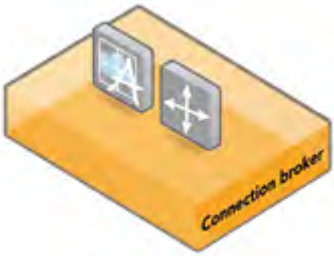
Web application acceleration

Web Application Acceleration appliances, or application delivery controllers accelerate and secure web-architected applications. All of us are encountering these solutions everyday. Large public facing internet applications, such as Amazon and eBay, all make use of these devices. Web Application Acceleration solutions are not just useful for large public organizations; you can also use them for your own web applications. Solutions that facilitate web

application acceleration and security include packages such as Citrix Netscaler and F5 BigIP. Functions of Web Application acceleration are

- Secure access to SaaS and Web resources, represented by the shield symbol.
- Fast and optimized access to web applications, presented by the accelerator symbol.

Connection Broker



A connection broker determines which hosted remote applications and desktop will be available to a user. When using a hosted virtual desktop infrastructure for this, it is possible to either designate dedicated desktops or a pool of remote desktops. The desktop broker can automatically create, remove or pause remote desktops. Functions of this object are:

- Hosted Applications and desktops, represented by the grey window with "A"pplication logo.
- Represented by the compass icon, which determines who will get access to which applications or desktops, both persistent and non-persistent.

Application Virtualization

Application virtualization can make applications available to desktops, laptops, server-hosted VDI and Remote Desktop Session Host (TS-Terminal Server) platforms. The applications are executed on the target platform, without needing to make any persistent modifications to the platform. The advantages of application virtualization include: installation, upgrade, roll-back, delivery speed and the ease of application support and management). Installation of applications is no longer necessary, eliminating the possibility of

conflicts. The result is a dynamic application delivery infrastructure. More unique information about application virtualization can be found here.

Application layering

Application layering has recently emerged as an alternative way to package and deliver applications separate from the operating system, without having to reinstall them on every desktop. Applications can be layered by running a standard setup procedure. The changes in files, directories, and Registry keys made by the installation procedure are captured as a "layer." When the layer is assigned to one or more desktops, its files and Registry keys are overlayed (layered) with the Windows operating system layer and all other application layers. Layered applications appear in Windows Add/Remove programs, and look as if they are natively installed to Windows and all other applications. The advantages of application layering include: simple installation, upgrade, and roll-back; ease and speed of application packaging; support for Boot 0 applications, device drivers, and other complex applications that cannot be virtualized with traditional application virtualization; and interoperability with all other applications and Windows itself. A key advantage of software layering versus application virtualization is typically a higher rate of application compatibility. The isolation introduced by Application Virtualization is both a blessing (for the apps that need it) and a curse (for those that don't).

OS Provisioning

OS Provisioning, or Machine Based Imaging, allows workstations to boot up and run from a central image. A single image can be used simultaneously by multiple workstations. The advantage of this is that complete operating systems, including applications and agents, can be made available quickly and securely. It is possible to make a single image available to multiple VDIs, RDSH, Client Side Virtualization and physical desktop environments without causing conflicts. As a result, it is possible to

upgrade or roll-back an OS quickly, simply, and without significant risks. When virtual desktops use OS streaming, (valuable) storage is saved and the management of virtual desktops becomes relatively simple. This means that virtual or physical machines using OS Provisioning can become - stateless devices..

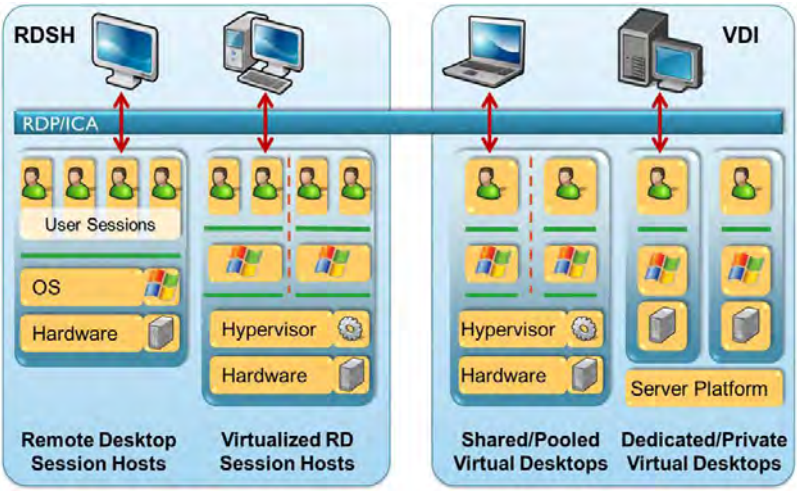
Server-Hosted Desktop Virtualization (VDI)

VDI with GPU acceleration, 3D graphics for Virtual Desktops

3D graphics for virtual desktops can be a server-hosted VDI solution. It provides each (virtual) machine with (GPU) graphic performance to run multimedia, 2D/3D, NextGen, design and engineering applications. The GPU can be shared, dedicated or virtualized for the Virtual Machine or Terminal Server environment. Display data is presented to the client device via an optimized remote display protocol. To ensure that the end-users experience the best possible performance, the bandwidth, latency, or local (software/hardware) components have to meet extra requirements.

Session Virtualization (RDSH)

With session virtualization every user has his or her unique (terminal server) session. Session Virtualization, also known as Terminal Services, is a solution for the remote access to desktops and applications that are run on a terminal server in a data center. Access to the desktop or application is not tied to a location or end-user machine, and programs are executed centrally on the terminal server. The data appears on the client screen through a remote display protocol such as Microsoft RDP/RemoteFX or Citrix ICA/HDX. Remote Desktop Services consists of various infrastructure components for management, load balancing, session control and support. It has the advantage that applications are made available quickly and securely, the TCO is low, and applications can be accessed irrespective of location or work place. While this whitepaper is not the proper place for a discussion of the pros and cons of RDSH vs VDI, it is important to note that RDSH does have a greater chance of limitations around application compatibility due to being based on a server operating system, whereas VDI is most often delivered via a client operating system.



Client-Side Desktop Virtualization

Client-side desktop virtualization is a solution where the virtual machines run locally on the client endpoint device. The client hypervisor ensures that each virtual machine is hardware-independent, and makes it possible to simultaneously use a number of virtual machines at the same workstation. The hypervisor plays an essential part in client-side VDI solutions while the management portion handles synchronization, policy, enforcement and management insights. The two types of Client-side Desktop Virtualization solution are:

- Client hosted hypervisor is installed and runs as an application on the operating system (be that Windows, Mac OS X or Linux) of the end device. This offers great flexibility of endpoint hardware and operating system compatibility at the expense of less performance.
- 'Bare-metal' client hypervisor acts as the device's base operating system and must be installed before other operating systems. This offers great performance at the expense of more limited hardware and operating system compatibility. Since a bare metal solution requires a low level installation

on the target device, it is rarely an acceptable solution for the modern BYO (Bring Your Own) model of compute.

The main differences between both types of Client-side desktop virtualization solutions are Usage in BYO scenarios, hardware support, performance, manageability and end-user experience.

Workspace Aggregation

The term Workspace Aggregator is used to describe software that unifies the delivery of multiple application or desktop types such as:

- Native mobile applications;
- Software as a Service (SaaS) applications;
- Mobile web applications;
- Windows Applications via application virtualization or installation;
- Desktops, local, server Hosted Virtual Desktops (VDI) or published Desktops in RDSH/Terminal Services..

A workspace aggregator ideally evaluates the user's device to determine which applications are available for this user on this particular device and at this particular time (context-aware access). In addition to application delivery, workspace aggregators provide secure file system (data) access from a broad range of devices. A workspace aggregator simplifies the deployment and life cycle management of applications. As the single point of access, (de)provisioning, auditing and monitoring is easily accomplished through security rules and policy enforcement. Workspace aggregators that are available at the market today also provide Single Sign On capabilities to applications..

Client management

Any self-respecting professional IT organization is bound to use a Client Management solution, as it is needed to facilitate things such as OS deployment, patch management, application and client deployment, asset management, service desk integration, and remote control.

Functions of Client Management (in the context of Application and Desktop Delivery) are:

- Deliver and manage Windows applications
- Deliver and manage Windows Operating systems

5.1 VENDOR MATRIX, WHO DELIVERS WHAT

Vendor	Product	Solution
Citrix	XenDesktop	VDI
Citrix	VDI-in-a-Box	VDI
Dell	vWorkspace	VDI
Ericom	PowerTerm WebConnect	VDI
Microsoft	VDI with RemoteFX	VDI
NVIDIA	VCA	VDI
NICE	DCV	VDI
RedHat	Enterprise Desktop Virtualization	VDI
Virtual Bridges	VERDE	VDI
VMware	View	VDI
Citrix	HDX-ICA	Remote Display Protocol
Dell	EOP/RDP	Remote Display Protocol

Ericom	RDP/Blaze	Remote Display Protocol
HP	Remote Graphics Solution	Remote Display Protocol
Microsoft	RemoteFX/RDP	Remote Display Protocol
NVIDIA	GRID	Remote Display Protocol
Oracle	ALP	Remote Display Protocol
RealVNC	RFB	
Virtual Bridges	SPICE	Remote Display Protocol
VMware/Teradici	PCoverIP	Remote Display Protocol
Citrix	XenDesktop HDX3D Pro	SH- Physical Desktops
Ericom	PowerTerm WebConnect	SH- Physical Desktops
HP	Remote Graphics Software	SH- Physical Desktops
VMware	View	SH- Physical Desktops
Teradici	PCoverIP	SH- Physical Desktops
Citrix	XenClient	CSV, baremetal - centrally managed
Citrix	XenClient Enterprise	CSV, baremetal - centrally managed
Microsoft	Windows 8 Client Hyper-V	CSV, baremetal - not managed

MokaFive		CSV, baremetal - centrally managed
Virtual Bridges	LEAF	CSV, baremetal - centrally managed
Bromium	vSentry	CSV, Security
Citrix	DesktopPlayer for Mac	CSV, Type-2 hypervisor - centrally managed
Intel	DeepSafe	CSV, Security
VMware	Player/Fusion/Workstation/View	CSV, Type-2 Hypervisor – View centrally managed, the rest not managed.
Citrix	Provisioning Services / MCS	OS Provisioning
Citrix	Personal vDisk (former RingCube)	OS Provisioning++
Double Take	Flex	OS Provisioning
Dell	Streaming Manager (WSM)	OS Provisioning
Dell	HyperDeploy	OS Provisioning
VMware	View Composer	OS Provisioning
VMware	Mirage	OS Provisioning++
Unidesk	Unidesk	OS Provisioning++
Virtual Bridges	Dynamic Gold Imaging	OS Provisioning
Microsoft	Windows Server 2008R2/2012/2012R2	Session Virtualization
Citrix	XenApp	Session Virtualization++
Dell/Quest	vWorkspace	Session Virtualization++

Cisco	ISE	Secure Access
Citrix	Netscaler Gateway	Secure Access
Ericom	Ericom Secure Gateway	Secure Access
Juniper	SA / MAG	Secure Access
Microsoft	Unified Access Gateway / RDG	Secure Access
VMware	View Security Server	Secure Access
Cameyo		Application virtualization
Numecent	Application Jukebox	Application virtualization
Microsoft	App-V	Application virtualization
Symantec	Workspace Virtualization	Application virtualization
Spoon	Spoon	Application virtualization
VMware	ThinApp	Application virtualization
CloudVolumes	CloudVolumes	Application layering
FSLogix	Apps	Not layering, but achieves similar goals
Unidesk	Unidesk	Application layering
VMware	Mirage	Image and Application management

IBM	BigFix	Client Management
LANDesk	Client Management Suite	Client Management
Microsoft	System Center Config Manager	Client Management
Novell	ZenWorks Configuration Mgr	Client Management
RES	Automation Manager	RunBook Automation
Symantec	Client Management Suite	Client Management
Citrix	AppController / Storefront	Workspace Aggregator
Centrix	Workspace Universal	Workspace Aggregator
VMware	Horizon Workspace	Workspace Aggregator
AMD	FirePro	GPU enabling technology for 3D Graphics DV
HP	Remote Graphics Software	Enabling technology for 3D Graphics DV
NVIDIA	GRID and Quadro	GPU enabling technology for 3D Graphics DV
Teradici	PCoverIP	Enabling technology for 3D Graphics DV
Mainframe2		3D Graphics DaaS

OTOY	Cloud Workstation	3D Graphics DaaS
VMware	Horizon View DaaS	3D Graphics DaaS (basic 3D graphics)

6. DESKTOP VIRTUALIZATION 101

The previous chapter gave an overview of 'Application and Desktop Delivery' solutions. This chapter describes 'desktop virtualization' in more detail.

6.1 SERVER HOSTED DESKTOP VIRTUALIZATION DIRECTIONS

The different server hosted desktop virtualization solutions are outlined in the following paragraphs.

6.1.1 Server Hosted Virtual Desktop (VDI)

Server hosted Virtual Desktops (SH-VDI) is a solution for accessing Windows 7/8 or legacy Windows XP desktops that are executed remotely on a virtual machine in the datacenter. The Virtual Infrastructure ensures availability and manageability. Other frequently used terms for this type of desktop virtualization include: Virtual Desktop Infrastructure (VDI), Remote Desktop Services Virtualization Host.

6.1.2 Non-persistent, persistent and layered virtual desktops

Typically, server-hosted Virtual Desktops have come in 2 flavors: non-persistent (pooled) desktops, in which users are given a fresh / new virtual machine every time they login, and persistent (private) desktops, in which users are assigned a dedicated virtual machines that remains the same every time they login. Non-persistent (pooled) desktops are often called stateless desktops because they will always revert back to their original state.

Persistent (personal) desktops are often called stateful desktops because they give users the freedom to install software (rights permitting) and make workspace-related adjustments by saving changes i.e. the state of the workspace, between sessions. Keep in mind that there are different levels of user personalization. Some items like desktop icons, wallpaper, etc are items that can be

persisted in a user profile / workspace management product. You do not need to implement stateful desktops to achieve basic customization of the virtual desktop. However, if you want the user to be able to make deeper changes to the desktop (like having different installed desktop applications) then this is a use case for persistent VDI (or at a minimum require the addition of a layering solution to accomplish your goals).

6.1.3 Stateless (pooled) Desktops

The advantages of non-persistent (stateless) desktops are:

- Very simple to roll-out and fairly easy to update the base images (think applying monthly OS hotfixes).
- A virtual desktop is guaranteed to be 100% identical
- The user always has a clean desktop (no registry clutter over time)
- Less effort is required by the management team due to the standardization of images
- Less storage is required because a single base operating system image can be shared across many desktops.
- Image management may require a new method of maintenance vs existing PCLM tools

The major disadvantage of non-persistent desktops is that any customizations made by users and any applications that are delivered outside of the base image by IT are lost after each desktop reboot. In an effort to make stateless, non-persistent desktops act more like persistent desktops, VDI vendors have recommended profile management tools to restore user customizations each time users log in to their new virtual machines, and application virtualization tools to deliver different applications outside of the base gold image. User-installed applications and deep OS and application security settings that live outside of profiles cannot be captured and restored by profile management tools.

6.1.4 Stateful (personal) Desktops

The advantages of persistent (stateful) desktops are:

- Freedom to install software within the desktop (rights permitting);
- Maintaining all changes to the operating system between reboots.

- Maintain desktops using the exact same tools you manage your existing desktops with

Traditional Client Management solutions can be used to deploy, maintain and support personal desktops the same as laptop, ultrabook and desktop scenarios.

The major disadvantage of persistent desktops is the higher (storage) cost. Most enterprises cannot afford the high costs of storage required to implement full-sized, thick-provisioned virtual machines for every user. . Also, since every desktop is unique, this can introduce additional unique troubleshooting scenarios for applications, etc.

6.1.5 Layered Desktops

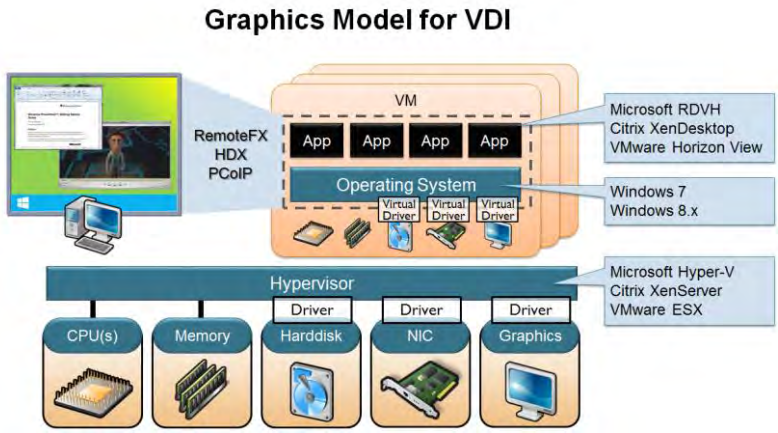
Another technology has recently emerged, that gives administrators a new virtual desktop provisioning and management option for VDI: "layering." Layering combines the benefits of stateful and stateless desktops in one solution. With layering, a non-persistent virtual machine is assigned to every user, guaranteeing that systems are consistent between reboots and allow IT a better way of managing image updates. Yet, unique departmental applications, user installed applications and all customizations will be remembered through reboots and base image updates. This makes the layered virtual desktops stateful, while still providing common base image updates shared across multiple systems.

The advantages of layered desktops are:

- Freedom to install software within the desktop (rights-permitting and IT policy approved)
- Maintaining all changes to the operating system between reboots (controlled via IT policy)
- The simple roll-out and update of basic images
- A virtual desktop is guaranteed to be 100% identical at the OS and App levels (in the base image layer). Individual departmental app layers and/or user layers can allow user/application deviation as needed.
- The user can always be reverted back to a clean desktop (controlled via IT policy)
- Less effort is required by the management team due to the standardization of images, simpler application packaging and ability to rollback or "undo" OS and application patches

- Less storage capacity is required because a single base operating system image and single image of common applications can be shared across many desktops.

A disadvantage of layering is that some of the vendors offer a solution supported in Server Hosted VDI environments only. However, the fact that there are now hundreds of real world customer implementations of layered desktops, many with over 1,000 users in production, indicates that VDI-only layering solutions are being generally accepted by the mid-market.



6.2 VDCRY?!

While reading the 'WhyVDI!?' paragraph a generic and better understanding of the benefits of VDI has been provided. Every solution has its weaknesses or challenges, as does VDI. Challenges doesn't mean it cannot be done, it just means that attention must be paid to ensure the solution outcome. The attention points for VDI are:

- End-user experience (UX) and their perceived performance of the solution
  - Voice, webcam, VOIP, conferencing solutions usability and supportability
  - The requirement for networking; bandwidth, QoS, latency and packet loss
  - Usability of rich media; audio/video, Adobe Flash and H.264 content

    - 2D/3D graphics in LAN and WAN
    - Delivery of hosted applications and desktops to mobile and touch/tablet oriented devices
- The overall cost reduction of the technology stack
- The overall complexity in the complete technology stack
- Streamline Management between VDI and classic clients
- Total Cost of Ownership and Return On Investment
- Licensing (SA, VDA, CDL, Intune, SPLA, RDS, corp. issued, and Microsoft Office). These are especially important in Bring Your Own scenarios because Microsoft licensing is a major challenge.
- The end user's user acceptance criteria. When does the solution meet the requirements for the user? Where are the requirements and who has contributed creating the requirements. If the solution does not satisfy the user, it cannot be successful.
- Does the IT team possess the skills and mindset to succeed in the project?
- The very serious impact of VDI on storage. When you don't know the impact of VDI on storage, seek help before attempting the project. IOPS, throughput and latency are only the beginning. Read more here, here and here

So in essence, User Experience, 3D Graphics, Storage and Microsoft Licensing are the main barriers to widespread adoption for VDI. Three of these challenges can be solved by setting the right expectations and using the proper solution.

7. 3D GRAPHICS FOR VIRTUAL DESKTOPS, STRATEGY

Virtual Desktops is a key component in an optimized desktop strategy. The transition to a dynamic and optimized desktop is causing many IT organizations to reevaluate traditional IT operations, deployment, delivery, packaging, support and management methods.

It's important to have a Vision and Strategy around 'Tomorrow's Workspace'. We see a lot organizations primarily focusing on products and vendors and lacking a clear vision and strategy. This approach is fine for point solutions, but a proper vision and strategy is crucial for a next generation optimized desktop platform. How can the vision and strategy be successful?Success = Vision x Execution x Adoption

The following discussions and corresponding topics should be part of the virtual desktop strategy. This includes 3D graphics for virtual desktops.

Generic:

- What are the use cases for Virtual Desktops? Does the use case require Desktop Virtualization?
- VD-why? What are you trying to achieve? Is this a business enabler? An overall cost of ownership (TCO) and cost reducer? Will it allow your business to be more agile in how you share information and collaborate on activities? Are you doing technology for technology's sake?
- What is the business-case? What do you expect as a ROI? This can be expressed in money saved, money not spent, or even time saved (which ultimately means a boost in productivity and/or a reduction of headcount).
- Are you investigating a tactical (point) or a strategic solution? What is your roadmap timeframe?
- Is work shifting a key driver for the optimized virtual desktop? How are the roaming/flexible and mobile users within

the organization facilitated? How do you take care of application and desktop delivery when the user has different access scenarios? Are you a company that insists your users aren't allowed to use DropBox, et al or are you actively seeking ways to enable your employees to work the way they want to work?

- How do you deliver applications to users in a Bring Your Own (BYO) or Choose Your Own (CYO) scenario?

Desktop strategy:

- What's your desktop delivery and migration strategy for Windows 7/Windows 8? Beyond staying secure, what is the business value to your organization in moving to Windows vNext?
- What is your application and desktop delivery model? Is delivery of applications/desktop focused on SaaS, Enterprise, SMB or the Consumer space? It's not uncommon for this to be a blend of technologies and that's perfectly alright.
- Is the VDI deployment targeted for small, medium or enterprise environment? Is the solution easy to deploy and easy to maintain? Don't be afraid to look to VDI as a solution for a silo of the business if it provides the ROI/TCO to satisfy your needs. Being strategic in the beginning doesn't always mean it's the best path to proving the technology at your organization. Some firms need small wins to become motivated to be strategic about things. (i.e. don't bite off more than you can chew…)
- Do you want to deliver windows and web applications to mobile devices such as tablets and smartphones via VDI? Is this what the user wants? Try spending a day in their shoes and see if you're satisfied with that solution. Do you have an enterprise mobility strategy?
- How does the desktop virtualization solution fit into existing deployment and management tools? Will you need to retrain your software packaging / deployment team? What about your desktop / application support teams? Finally, what impact does your decision on desktop and application delivery impact your high availability and disaster recovery needs?

- Is separation of Operating System - Application - and User Preferences inside and outside the virtual desktop part of the overall strategy?

User Experience:

- What's the user experience using Multimedia, Video/Voice, 2D/3D applications? Always try to look to the future with regards to this. There's nothing worse than deciding on a virtual desktop architecture and then having to re-architect it six months after it's rolled out because you were short-sighted.
- Is 3D graphics needed? What's the perceived end user experience expectation?
- Is Unified Communications and VoIP functionality within VDI needed? Is it supported by the VDI and UC-vendor? Don't forget how your mobility strategy is impacted by this. Also, UC/VoIP is very different inside the firewall vs outside the firewall. Do you have a list of tested VoIP/UC peripherals? Supported softphone headsets, speakerphones, etc. Is your UC/VoIP solution spanning long distances over poor WAN links? Have you tested these scenarios?
- What are the user expectations of the vDesktop? Are users involved in a proof of concept and pilot? What are their acceptation criteria?
- Never conduct a proof of concept or pilot on the easiest use case users in your environment. It will create a false sense of accomplishment within the project and may very well land you in hot water when you're asked to scale up to thousands of users across multiple regions with very complex requirements. Always seek out a mixture of users from different departments with different needs to ensure you've handled challenges from every corner of the business.
- What endpoints do we support and facilitate and what is the role of these devices in the end-user experience? Are the endpoints managed by the IT organization? It's very important to have endpoints and peripheral devices that your users are using in the environment so IT can see how it works. Even if you claim you don't support third party hardware in the environment, you always end up supporting more than you'd like depending upon how important the user is.

Infrastructure

- Secure Access and Secure networking; SASN how do users, with a variety of endpoints (rich, thin or zero clients and mobile devices) connect to the vDesktop?!
- What is the impact of Secure Access and Secure Networking solutions on mobile devices while connecting to the vDesktop? What is the user experience with these secure access solutions?
- What is the performance and bandwidth impact on the network infrastructure; LAN, WAN, WLAN, Mobile;

VDI Solution

- Does the VDI solution needs a client/agent component on the endpoint? Is there a supported agent for the OS/endpoint? What is the User Experience with this agent? What is the feature and future roadmap of these agents? Is agentless via HTML(5) included and important?
- Is image deployment and management part of the (virtual) Desktop Strategy?
- Do we need a vMachine based image management solution?
- How do we design, build and maintain the (golden) Image(s)?
- Does the end-user needs the ability to install and update applications? Is User Installed Applications functionality needed? Does the user have the correct privileges to install or update software?
- How are Windows applications delivered within the vDesktop? Unattended or manual installation, application virtualization or the applications are part of the (golden) image? What is the strategy?
- Do we focus on stateless (non-persistent) and/or stateful (hybrid/persistent) images? What is, for example, the impact on storage, manageability, security, legal and business-case?
- Linux or Windows as Guest OS?
- Windows 8 or Windows 7 as core guest OS platform? x64 or x86?
- How does the solution scale? What do we need from a scalability point of view?
- Is there a validated design?

- How do we size the vDesktop and corresponding infrastructure and what are the best-practices for optimizing the vDesktop?
- What is your site topology? Multi-site, multiple datacenters?
- Is the VDI solution as a whole highly available? Is that built-in or are additional planning and solutions needed to get high availability? Is a highly available vDesktop needed?
- Is the VDI vendor a financially healthy organization? Is this important in evaluation of the vendor and the solution?
- Is there a huge ecosystem with partners, consultancy, training and education around the VDI solution? Is this important for you?
- What is your public cloud computing strategy and how does VDI/DaaS fits in that strategy? What do you want to solve?
- Is the solution reliant on public cloud infrastructure (IaaS, e.g. Amazon EC2, Windows Azure, VMware vCHS)? Where is the data stored? Is US Safe Harbour a challenge? What is the exit strategy for your public cloud offering?

User Environment Management

- Do you need context awareness? Based on user/role, device, location and various settings is access to application resources controlled and enforced when needed?
- How do you design and build the user's profile and his 'workspace'? Does Application virtualization fit into this strategy? The 'User Environment Management' Smackdown can be helpful.
- Do you want to integrate and run local applications in the centralized desktop environment and present centralized and local applications in one single interface to the end-users?
- What is the impact of client management and user environment management solutions in a stateless (pooled) VDI scenario? Is it supported?

Application Readiness

- What is your application readiness assessment strategy? Are Windows 7, Windows 8, VDI, application virtualization and x64 included?
- What is the (business) applications vendor support policy for virtualized desktops?
- Do you have insights which applications are graphical,- and resource intensive?

Storage

- What is the performance and storage impact of application virtualization in combination with VDI? Is this important from a business-case or technology perspective?
- Do you need local or centralized storage? What storage optimization (IOPS/latency) technology is being used?
- What is the impact on storage (http://bit.ly/5HTajV) and how does it affect the business case?
- Licensing; VDI solution, guest and client operating system, Client Access Licenses and (Business) Applications.
- Do we need to backup (and restore) the vDesktops?
- Is Antivirus needed? Inside the VM or as layer on the Hypervisor? What is the real performance impact of Antivirus?

IT.org

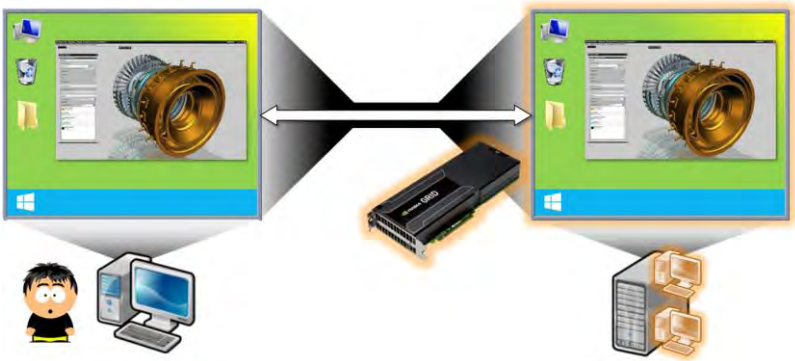
- Is the IT department able to adopt the technology with right knowledge and skills? What subject matter experts are needed to get and keep the VDI solution up and running in production environment? Is this expertise available? Who has overview of the complete VDI solution stack?

8. 3D GRAPHICS FOR VIRTUAL DESKTOPS

8.1 INTRODUCTION

Delivering 3D graphics for virtual desktops isn't a niche market. Millions of users can benefits from graphics with their virtual desktop. While technology evolves and competition increased the cost of adding 3D graphics to virtual desktops will be affordable for everyone.

Every PC has a GPU so why not for VDI?! Most new operating systems and application leverage GPU's while a larger group of people benefit from using high-end graphics for 2D, 3D, multimedia applications and HPC environments.



8.2 WHY 3D GRAPHICS FOR VIRTUAL DESKTOPS?

What are the business drivers for enabling 3D graphics for Virtual Desktops?

- Flexibility: Work isn't a place, it's something you do from anywhere.
- Access: Virtual desktop works independently of location, endpoint and network. Use the Virtual Desktop from any client endpoint; Work Anywhere, with Any Device in LAN, WiFi

and WAN scenarios. Support work from home, global product development teams, contractors and BYO scenarios.

- Application integration: more and more application integrate with each other, large files and datasets (sometimes TBs) are accessed by users all over the globe. File transfer and WAN acceleration are not the solution.
- Security and control: Information systems and data stays in the datacenter center; Protect Intellectual Property (IP).
- Freedom: Every user can have his own (virtual) desktop with administrator privileges when needed.
- BYO: enables delivery of applications and desktops for Bring Your Own Device or Corporate Owned Personally Enabled (COPE) scenarios;
- High-Performance: High-end graphics and flash based storage solutions delivers high-end 2D/3D graphics and resource intensive applications to any device while keeping the data central.
- Reduce Cost: Operating System, Applications and user environment is centrally managed. resources are shared where needed.
- Legacy: It is simple to offer legacy applications on a state-of-the-art end-user client computing platform;
- Cooling: Use thin clients with multiple screens in larger setups in the offices and keep heat generating workstations inside the cooler datacenter. Reduced noise helps to improve employee working conditions as well.
- Disaster Recovery: Multi-site and multi-datacenter around the globe is easier (not easy) to design because apps, data and desktops are centralized.

8.3 USE-CASES

The use-cases of 3D graphics for virtual desktops are huge and is increasing fast.

Deliver 3D graphics and resource intensive applications for Aerospace, Automotive, Construction, Energy, Film/media, Engineering, Hi Tech Electronics, Industrial Equipment, Medical Equipment, Oil/Gas exploration and HPC environments is possible today.

Various customer around the world in different segments are using the technology to meet their use-cases. We are happy to connect these clients to you when appropriate.

8.4 CLASSIFICATION

Perceived performance and user experience are critical elements in a 3D graphics for virtual desktop project. It's important to understand the user population, what applications or type of applications they use and what the performance and resource impact of these applications is. It doesn't matter how you classify your users and what names you use to group them.

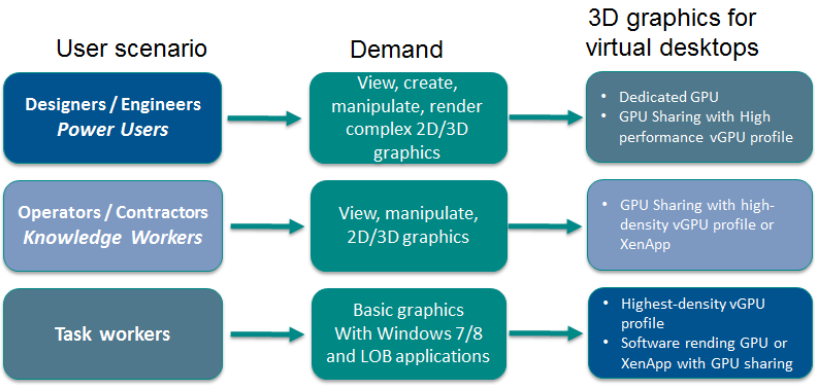
Classify the GPU requirements of the applications is also important. Does the application requires an GPU to operate, is the application GPU intensive or is the GPU just an assistant for the application. Below is an example to group the various 3D graphics users with examples of 3D graphics for virtual desktop solutions. The concepts of the different 3D graphics solutions will be explained further in the document.

Designers / Engineers - Power Users: Designers are the most demanding user group in the organization, they view, create, manipulate, render complex 2D/3D graphics. GPU acceleration of Direct3D, OpenGL, CUDA, OpenCL applications is required. Designer require a high-end virtual desktops with high performance storage and dedicated GPU for graphics acceleration. GPU pass-through with NVIDIA or AMD can be the solution for designers.

Engineers view, create, manipulate, 2D/3D graphics. Engineers require a higher-end virtual desktops with high performance storage and a high-end vGPU profile.

Operators / Contractors - Knowledge Workers: These power users need to view or edit 3D files, access graphical applications and workflows from anywhere. High-performance, SSD speed, storage to start and use applications is needed. Hardware accelerated graphics is needed while high-end graphics is overkill. High density NVIDIA vGPU profile, GPU sharing with Citrix XenApp can be suitable for power users.

Task workers: The segment of users in the organization that are not engaged in professional graphics design. Hardware accelerated graphics may or may not be required to deliver business graphics, such as the Windows 8 style apps, PowerPoint transitions in Office 2013, or perform light 2D and 3D work.



9. 3D GRAPHICS FOR VIRTUAL DESKTOP CONCEPTS

This chapter explains the different solutions to deliver 3D graphics for virtual desktops.

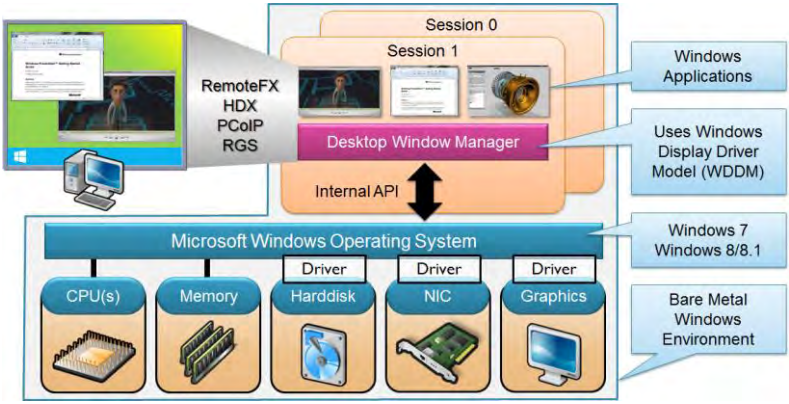
9.1 GPU AND APU

Delivering high-end graphics for virtual desktops can be accelerated by using physical Graphics Processing Units (GPUs). A GPU is a highly integrated electronic circuit used in personal computers, mobile phones, embedded systems, graphics workstations, game consoles and cloud platforms. It is designed to perform data manipulation at high performance and in parallel, with the intention to create digital images in a frame buffer and send the results to one or multiple displays. In contrast to a computer's Central Processing Unit (CPU) with only few cores designed for sequential serial processing, a GPU consists of thousands of smaller cores optimized for dealing with many tasks simultaneously. In a graphics workstation or a cloud server optimized for graphics remoting, GPUs are typically present on a separate video card.

An Accelerated Processing Unit (APU, also Advanced Processing Unit) combines a CPU with additional processing capability designed to accelerate certain types of computations. Such an APU may include a GPU used for general-purpose computing (GPGPU) and for graphics acceleration. In essence, an APU combines a CPU and a GPU on the same die. allowing for high data transfer rates between the two while keeping power consumption at a relatively low level. A good example of a state-of-the-art APU is Intel's Haswell CPU with integrated Iris Pro GPU. While a dedicated GPU on a separate video card may outperform an APU in many high-end graphics scenarios, lower overall power consumption and lower price are typically advantages on the APU side.

9.2 BARE METAL

This method is the (blade) PC/workstation remoting architecture where no virtualization is being used. The end user connects via a connection broker to the workstation via remote protocols such as Citrix HDX, Microsoft RDP, VMware Horizon view PCoIP, HP RGS or via Teradici hardware based PCoIP. HP's moonshot is an example of a solution to deliver a virtual desktop running on a bare metal PC-like hardware.



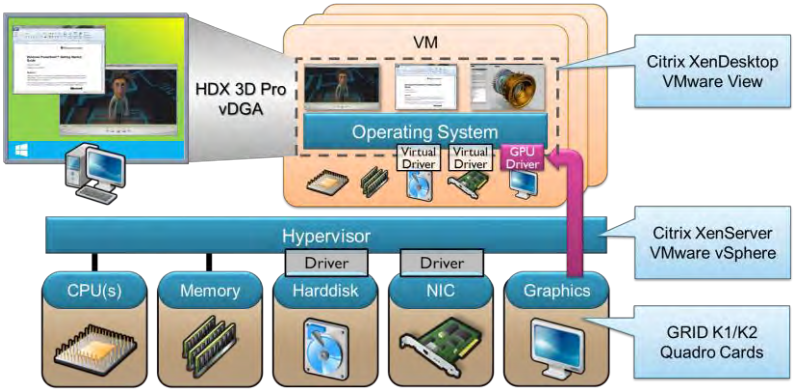
Benefits:

- High-end 3D Graphics and workstation class performance.
- Support for all technologies such as OpenGL, OpenCL, NVIDIA CUDA, DirectX 9/10/11/12
- Full application support, native GPU drivers are used.

Challenges:

- Hardware dependent
- Less cost effective
- No Virtualization benefits such as hardware independence, resource sharing, snapshots, Live Migrations etc.

9.3 PASS-THROUGH OR DIRECT ATTACHED GPU FOR VIRTUAL DESKTOPS - VDI



The Virtual Machine (VM) has a 1:1 relation with a GPU and has through the Hypervisor full and direct access to the GPU. The native NVIDIA/AMD graphics driver is installed in the Virtual Machine. VMware call this technology "vDGA - Direct Graphics Acceleration", Citrix calls this technology "GPU pass-through". Despite it's name the solution offers full GPU performance and high-end graphics needed for designers and engineers.

Citrix XenServer 6.x and VMware vSphere 5.1 or higher support the Pass-through or direct attached GPU for Virtual Desktops.

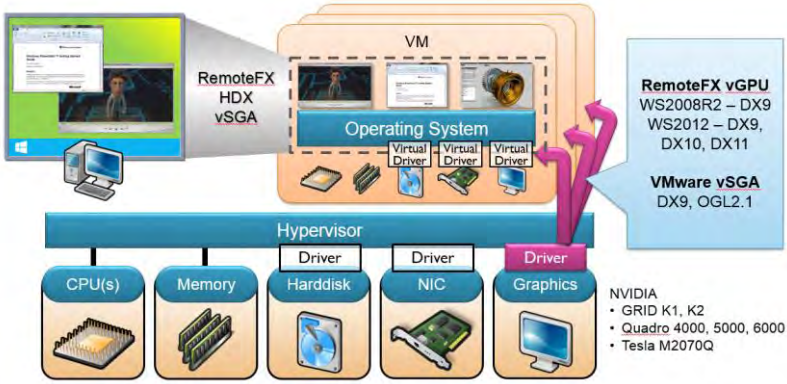
Benefits:

- High-end 3D Graphics and workstation class performance for virtual desktops.
- Support for all technologies such as OpenGL, OpenCL, NVIDIA CUDA, DirectX 9/10/11/12
- Full application support, native GPU drivers are used.
- Virtual machines with and without direct access to the GPU can be hosted on the platform.

Challenges:

- Live migration of VM with pass-through devices is not supported
- Hardware dependent.

9.4 GPU Sharing for Virtual Desktops, API intercept



GPU Sharing for Virtual Desktops is a 3D graphics solution suitable for both VMware Horizon view (VDI) and Session Virtualization with Microsoft Remote Desktop Services Virtualization Host. The shared GPU model is also called an API intercept model where the GPU is managed and owned by Microsoft Hyper-V with RemoteFX or VMware vSphere hypervisor with Horizon View vSGA (Shared Graphics Acceleration). The graphics API requests generated from the VMs are intercepted via the capture driver in the virtual machine. The requests and redirected and executed on the mentioned hypervisors. After processing the information is send back to the VM. The VM has no direct access to the GPU, as with GPU Virtualization for VDI.

The advantage of GPU sharing is it's scalability, where 10s of VMs or sessions can provide GPU access to knowledge workers and task workers who are using the Line of Business applications, Windows 7 and 8.x which and requires graphic acceleration.

Benefits:

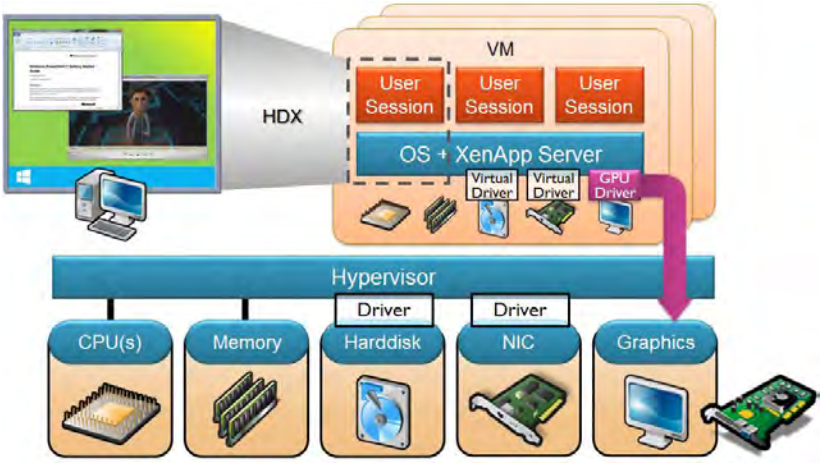
- Cost effective solution for knowledge users and task workers who requires graphic acceleration
- Scalable solutions, 10s of VMs or sessions, of course depending on application and usage.

- VMware vSGA implementation can switch between GPU sharing and Software 3D renderer. Soft3D is a VMware WDDM drivers which is installed in the VM.
- Allows each user to have power user performance with enhanced support for DirectX 3D and Windows® Aero
- Live migration of virtual machines VM's with VMware vSGA supported

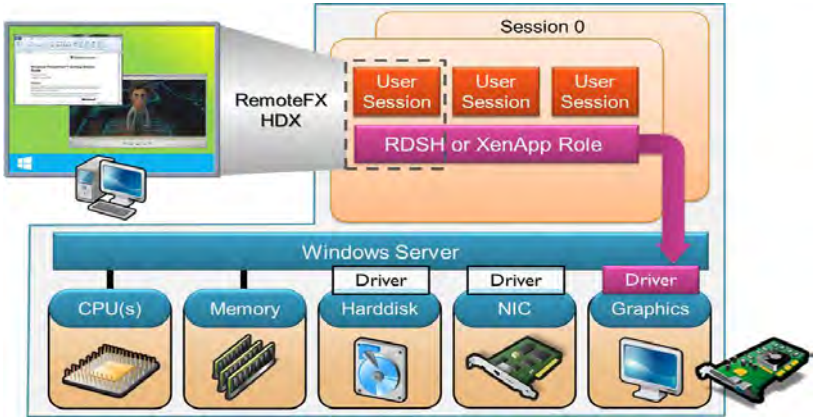
Challenges:

- The solution doesn't meet the requirements for high-end graphics where users will view, create, manipulate, render complex 2D/3D graphics
- The GPU can become a performance bottleneck as many users will share the same GPU.
- Application compatibility issues due to limitation of 3D API's supported
    - OpenGL supported versions limited
    - DirectX supported versions limited to DirectX 9 in some cases

GPU Sharing for Virtual Citrix XenApp



GPU Sharing for physical Microsoft RDSH and Citrix XenApp



In this model the 3D applications are installed in a multi-user Operating System such as Windows Server 2008R2/2012R2 and published as hosted shared desktop or seamless application. The Citrix XenApp or Microsoft RDSH server need to have access to a GPU, which can be the scenario with bare-metal or a virtual machine with GPU passthrough. NOTE: In a pure Microsoft stack, RDSH only supports bare metal GPU access since Hyper-V does not support PCI device passthrough on the hypervisor. RDSH hosted on alternative hypervisors would provide PCI device passthrough access to execute as a VM. In order to leverage a hardware graphics adapter for RDSH, you must enable a GPO setting named "Use the hardware default graphics adapter for all Remote Desktop Services sessions" for Server 2012/2012R2.

Benefits:

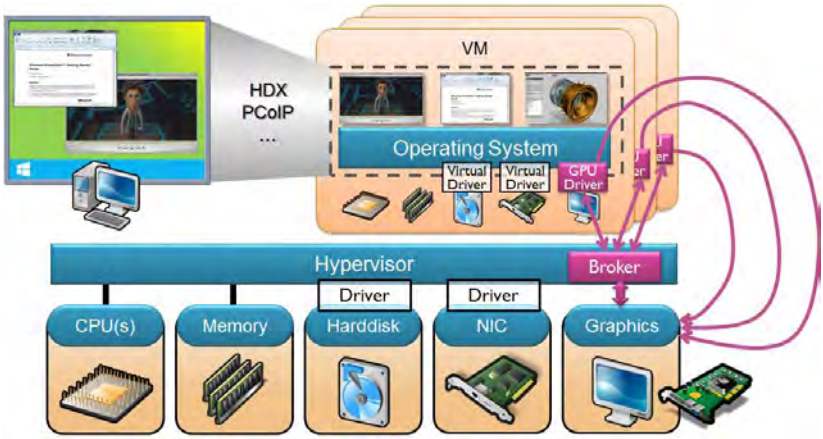
- Cost effective platform to deliver 3D graphics for Virtual Desktops/Apps.

Challenges:

- Some 3D application may not work in a multi-user environment.
- The solution doesn't meet the requirements for high-end graphics where users will view, create, manipulate, render complex 2D/3D graphics

- The GPU can become a performance bottleneck as many users will share the same GPU.

9.5 GPU VIRTUALIZATION FOR VDI - vGPU



GPU Virtualization for VDI or hardware virtualized GPU offers the benefit of GPU sharing and gives the virtual machine via de native graphic driver access to the GPU. Each Intel or NVIDIA GPU is shared by multiple VMs.

With NVIDIA GRID each VM is configured with a vGPU profile with a configurable static vRAM size and the VM will share GPU processing power. Currently Citrix Xenserver 6.2 support the only hardware virtualization solution in the market, NVIDIA GRID vGPU.

Benefits:

- Application support for 3D technologies such as OpenGL and DirectX.
- Depending on the vGPU profile the solution can deliver high-end performance and highest-density solution for knowledge workers or operators.

Challenges:

- Lower overall VM density per GPU as compared to GPU Sharing for Virtual Desktops.

- Little competition, only NVIDIA and Citrix provide this technology today
- No OpenCL and CUDA API support today.

9.6 APPLICATION VENDOR SUPPORT

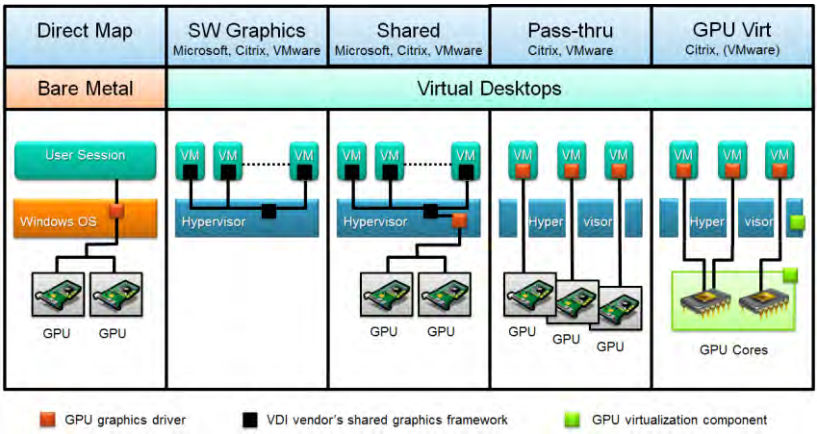
GPU Virtualization,- and Pass-through GPU for VDI allow applications to use GPU vendors drivers. The ISVs such as Autodesk, Solidworks and PTC certify their solutions with specific GPU drivers. Application certification for NVIDIA GRID can be found here:

http://www.nvidia.com/content/cloud-computing/pdf/GRID_Certifications.pdf

Application certification for AMD FirePro can be found here:: http://support.amd.com/en-us/download/workstation/certified

9.7 3D GRAPHICS FOR VIRTUAL DESKTOP CONCEPTS SUMMARY

The five different concepts to leverage 3D graphics for virtual desktops are:



9.8 How to choose the right 3D graphics for virtual desktop solution?!

What is the right 3D graphics for virtual desktop solution and how do you make the choice for that solution? First of all, when you don't know the requirements from end-user, application or IT perspective the easiest and best way to determine the right solution is flipping the coin.

It's key to classify application usage, perceived performance and IT,- and use acceptance.

Chapter 'Enabling 3D graphics for virtual desktops solutions' will explain some of the solutions which can give insights in graphics application usage. The diagram displays the high-lights of the different 3D graphics concepts and its characteristics.

3D graphics for Virtual Desktops	High performance	3D API support (openGL, DirectX etc.)	Cost effective	Application compatibility
Software 3D Graphics	★	★	★ ★	★
Bare metal OS	★ ★ ★ ★	★ ★ ★ ★	★	★ ★ ★ ★
Pass through GPU	★ ★ ★ ★	★ ★ ★ ★	★	★ ★ ★ ★
GPU Virtualization - vGPU	★ ★ ★	★ ★ ★	★ ★ ★	★ ★ ★
GPU Sharing for VDI	★ ★	★ ★	★ ★ ★ ★	★ ★
GPU Sharing for XenApp/RDSH	★ ★	★ ★	★ ★ ★ ★	★ ★

9.9 Platform accessing 3D graphics

In today's world of consumerization of IT, Bring Your Own and Corporate Ows Personally Enabled (COPE) scenarios there isn't one platform anymore. Today's reality is Laptops, hybrid laptop/tablet devices, ultrabooks, zero clients, thin clients, desktops and workstations running Windows, Apple OSX, Android or ChromeOS.

On the mobility side it's Android, iOS, Windows Phone, Blackberry, MozillaOS as native mobile platform and HTML5 as web application delivery platform.

What is the best platform for accessing 3D graphics applications and/or desktops? Is it a Google ChromeBook with HTML5? a Zero client with smartcard integration?, a Windows 8 tablet with touch?, a MacBook Pro accessing AutoDesk Revit, a Windows 7 corporate laptop accessing PTC Creo? The best platform for accessing 3D graphics applications and/or desktops doesn't exist.It all depends on perceived performance, user experience, required functionality such as USB support, monitor monitor support, Client printing, network scenarios etc. etc.So the main question here is do you know what your business consumers, your partners and co-workers expect from the solution and do you know the impact on endpoint choice and platform accessing 3D graphics applications?

9.10 GUEST OS FOR 3D GRAPHICS

The Guest OS is the platform being used for running Windows or Linux 3D graphics applications. This OS is running inside the virtual machine, physical server or professional workstation and is responsible for running the 3D graphic application.

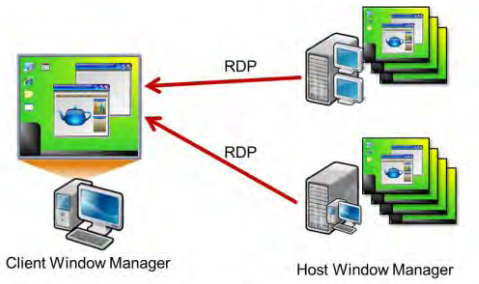
What is the best guest OS for delivering 3D graphics applications? The main question to answer is what are the application requirements? Does the application requires a Linux or Windows Client OS, can the application leverage x64 bits? Is VDI or RDSH being used? Do we want to deliver the windows application as a service through a service provider and what's the impact on licensing.In such as case a Windows Server OS used in a personal can be a good solution. The majority of 3D graphics for virtual desktop solutions are focused on delivering Windows applications and desktops to users. Citrix, Microsoft, NVIDIA and VMware are examples. Remoting protocols and it's corresponding solutions from HP RGS and Nice DCV, do support Windows and Linux guest OSes.

10. REMOTING PROTOCOL TURNED INSIDE OUT

This chapter highlights the concepts behind remoting protocols.

10.1 GRAPHICS REMOTING FUNDAMENTALS

User experience remoting has been around since the 60s of the last century, initially focusing on text-only remoting across serial lines. Modern graphics remoting goes far beyond such a basic model. It allows rich Windows applications and their graphical output device to be separated by a remoting boundary. It facilitates user interaction with a remote computer system by using a remoting protocol to transfer graphics display data from a host system to the user sitting in front of a client system. User input is transported from the client to the host and replayed there.



Modern versions of remoting protocols, such as Microsoft RemoteFX, Citrix HDX or Teradici PCoIP, try to improve performance by taking advantage of physical Graphics Processing Units (GPUs) on the remoting host. In addition, the remoting protocols identify the client capabilities at connection time and constantly analyze the network conditions throughout the entire remoting session time in order to adjust communication settings dynamically. This all helps to make the desktop remoting environment self-adaptable and provide the best performance possible.

When looking at the impact network conditions have on the performance of remoting protocols, most people believe that bandwidth is the most significant factor. While this is true for networks with a low bandwidth profile, it is different for networks with more than 2Mbps per remote user session. In such cases, latency and packet loss become the limiting factors. User experience

will typically degrade when latency is more than 50ms and it will be very challenging when latency exceeds the 200ms threshold. Packet loss should be below 1% for a good user experience. However, there are new remoting protocol variants that were specifically designed to perform well in low bandwidth and high latency/packet loss scenarios.

Note: Remoting protocols typically don't limit themselves with respect to available network resources. In other words, if one remote user session requires high bandwidth for rich multimedia content, it may well consume up to 100Mbps if such bandwidth is available.

10.2 REMOTING PROTOCOL FEATURES

Remoting protocols run on top of the Internet Protocol (IP), using Transmission Control Protocol (TCP), User Datagram Protocol (UDP) or a combination a TCP and UDP for different aspects of remoting. While older remoting protocols only used TCP, the modern ones use UDP for the graphics remoting aspect.

TCP is a connection-oriented protocol providing high reliability through error checking, congestion control and a built-in mechanism that rearranges data packets in the order specified. It also guarantees that all data remains intact in the packets transferred. But all this makes TCP relatively heavy-weight, significantly reducing graphics remoting performance on low bandwidth and high latency/packet loss networks.

UDP is a connectionless protocol, allowing a program to send individual data packets to another program. There is no error checking and no guarantee that all packets were delivered in the right order. Only if a data packet arrives at its destination it is checked for integrity, This makes UDP light-weight and much faster than TCP, but at the risk of losing chunks of data. The receiver program needs to be prepared to deal with this kind of data loss.

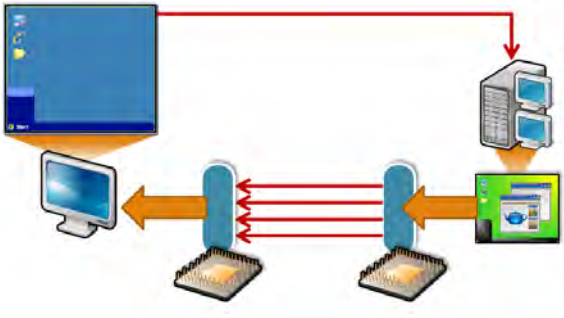
But there is more in a remoting protocol, in particular when it comes to extensibility. The concept of virtual channels provides a way to establish separate streams of data communication while taking advantage of the remote session communication already established. Many remoting protocols use virtual channels to add functions that allow a strict separation from the core features or are not yet

specified in the protocol. They represent a platform that future developments can be based on without having to modify the communication methods between host and clients. Examples for virtual channel use cases are joint client and server clipboards or redirecting print jobs to local client printers.

Other notable remoting protocol features include bi-directional audio transmission, client side caching mechanisms, session reconnect after a temporary loss of connectivity, device redirection, multi-monitor support, location awareness and support of Unified Communications. They all are relevant for an acceptable user experience.

10.3 Client Side Rendering versus Host Side Rendering

In a graphics remoting environment, the Windows desktop including its applications is rendered in a different way compared to traditional, local PCs. But what exactly is "rendering"? The most common definition is as follows: If an image described in an abstract model using vector graphics and bitmap primitives is converted into a raster image for output on a screen the term rendering is used. In graphics remoting environments, rendering may happen either on the client side or on the host side.



When graphics primitives are transmitted from the host to the client and rendered on the receiver side, this is referred to as client side rendering. When multimedia data streams are involved, the preferred term is multimedia redirection (MMR). The advantage of this remoting method is that it can greatly reduce host side CPU impact. In addition, redirecting media typically works well in constrained network and due to the nature of the transmitted objects it has a lower bandwidth profile than host side rendering.

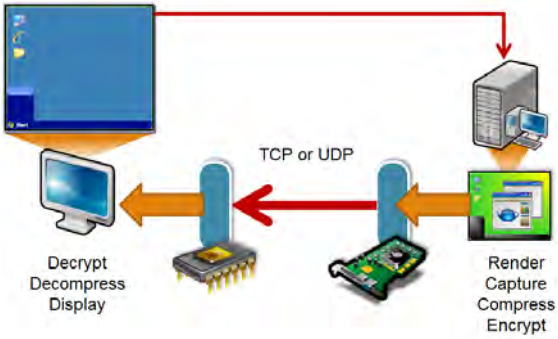
But client side rendering also has many disadvantages: The breadth of features can be very dependent of the client operating system, some forms of media may fall back to server side rendering, client rendered media may require some buffering time, and client side resource requirements may be significantly increased. In addition to that, client side rendering introduces audio/video codec dependencies and possible security vulnerabilities through outdated client side components.

Another significant challenge of client side rendering is the synchronized playback of audio and video in film or animation sequences. This is also referred to as "lip sync", which is a common term for matching lip movements with spoken vocals. Dedicated codecs and processing pipelines for the separate audio and video data streams in combination with network delays may result in A/V sequences that are out of sync.

Until recently, client side rendering was regarded as the preferred Windows desktop remoting method, providing superior performance and user experience. But modern Windows desktops and applications may use various graphics formats which makes it hard to implement associated rendering algorithms across all popular client platforms. As a result, modern remoting protocols are typically not based on client side rendering anymore.

When graphics primitives are rendered on the host device instead of on the client device, this is called host side rendering. As a result, all graphics types can be by transmitted as a highly compressed bitmap images to the endpoint device in an adaptive manner. Ideally, the host computer takes advantage of one or multiple physical GPUs to improve graphics performance.

The advantage of host side rendering is that it generally is client OS independent and also independent of client hardware. More advantages are that any form of mixed media content can be



rendered, media content playback begins immediately, only fairly low client side resources are required and audio/video playback is provided regardless of client side codecs.

Disadvantages are that host side rendering may generate a very high host CPU impact, it can have A/V sync issues in constrained networks (depending on the protocol's implementation) and it typically requires a higher bandwidth profile if compared to client or hybrid approach.

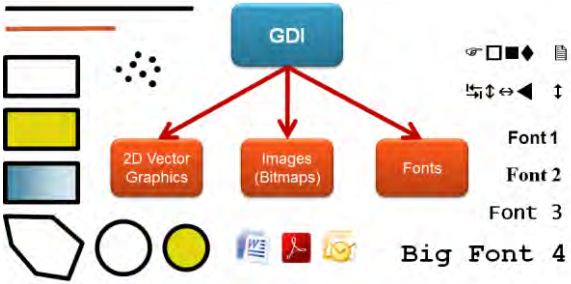
As mentioned earlier, Windows desktops and applications are not only built on top of one single graphics format. As a matter of fact, a typical modern Windows system uses multiple graphics formats side by side for composing the desktop. It is important to understand nature and purpose of the individual graphics formats as some of them are treated differently in the different remoting protocols. The following paragraphs introduce the most important graphics and multimedia formats used for Windows desktop composition.

10.4 GDI REMOTING

Graphics Device Interface (GDI) is an application programming interface (API) that was developed by Microsoft in the 80s and 90s of the last century. It is responsible for defining and rendering graphical objects on Microsoft Windows. GDI rendering or rasterization is the task of taking an image described through vector shapes or primitives and converting it into a raster image consisting of pixels or dots for output on a GDI-compatible video display or printer. Such a rasterized image can also be stored in a standard bitmap file format.

Typical GDI primitives are lines, rectangles, ellipses, arcs, Bezier splines, filled areas, bitmaps and fonts. A significant capability of GDI is its indirect method of accessing the underlying hardware by drawing on a Device Context (DC). A DC represents an abstraction layer that can be mapped to multiple physical target devices, such as screens and printers. A GDI applications sends its graphics output to the DC and then the Windows system sends the DC content to the target device.

It is important to note that GDI was designed in such a way that it can be hardware accelerated in a graphics card. Every modern PC graphics card supports the full GDI function set in hardware. During the Windows initialization phase, GDI hardware acceleration is registered in the system, significantly reducing CPU load generated by graphics output during system runtime.



In GDI there is no mechanism of synchronizing the DC with the graphics card frame buffer. As a result, GDI is not good at animating graphics primitives. There is no built-in method for double buffering animated output. In addition to that, GDI is a 2D graphics system, so it lacks the capability of rendering 3D primitives. The only 3D capability of GDI is the z-order of a window indicating its position in a stack of potentially overlapping windows. A window in the z-order list overlaps all other windows that are closer to the bottom of the z-order.

Most traditional Windows applications are built on top of GDI. The application windows and dialog boxes consist of basic GDI window objects, such as borders, title bars, captions, control boxes, scroll bars, menu bars and icons. These window objects are built of even more elementary GDI primitives, creating a hierarchical system of GDI primitives and window objects. A message pipeline allows an application to send graphical output to the client area of one of its windows or dialog boxes.

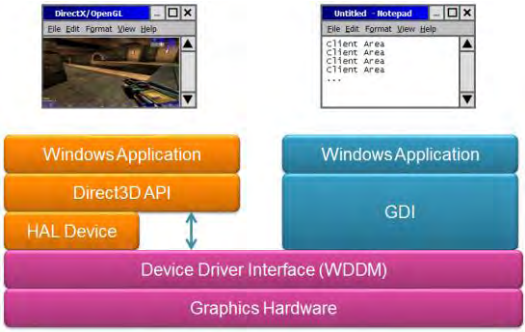
When Windows XP was introduced, GDI was extended by the GDI+ subsystem. GDI+ added more 2D graphics features and the support of modern graphics file formats. The Microsoft .NET Framework includes a managed GDI+ interface through the System.Drawing namespace. Substantial portions of GDI+ are not hardware accelerated.

The first designs of the Microsoft Remote Desktop Protocol (RDP) adopted many concepts of GDI and GDI+. The general idea was that

remoting client and host negotiated their common set of GDI capabilities during connection, allowing for an advanced level of client side rendering. If the resulting session properties did not include the support of GDI elements like bitmaps, pointers, fonts, brushes, glyphs and standard windows, the communication fell back to pixels. This was referred to as "screen scraping", where all pixels of a rendered desktop were "scraped" from the host video memory and then transmitted to the client.

Today, both GDI and GDI+ are not as relevant as they used to be in the Windows XP times. They were replaced by more modern graphics formats, such as DirectX and Windows Presentation Foundation. As GDI falls in popularity, so does the relevance of client side rendering.

10.5 DIRECTX REMOTING



DirectX is a collection of application programming interfaces (APIs), covering different aspects of multimedia formats. These include Direct3D (D3D) for 3D graphics, DirectDraw (D2D) for 2D graphics and audio/video APIs called DirectMusic, DirectPlay and DirectSound. DirectX was introduced to compensate for the shortcomings of the Graphics Device Interface (GDI), in particular in Windows games. It is directly connected to display drivers and gets better results at rendering than GDI.

When Windows applications are using Direct3D they write to a 3D surface. In Windows Vista and later versions, Direct3D uses the Windows Display Driver Model (WDDM) to share the 3D surface with the Desktop Window Manager (DWM). DWM then uses this surface directly and maps it on to the window desktop. One of the central element of the WDDM graphics architecture is the new video memory manager. It supports the virtualization of graphics hardware

for a range of applications and services like the Desktop Window Manager.

If an adequate graphics card is available, Direct3D can use it for hardware acceleration of the entire 3D rendering pipeline or part of it. The advanced graphics capabilities of the underlying 3D graphics hardware are exposed through Direct3D. Combining such capabilities with the other DirectX technologies allows for advanced multimedia scenarios.

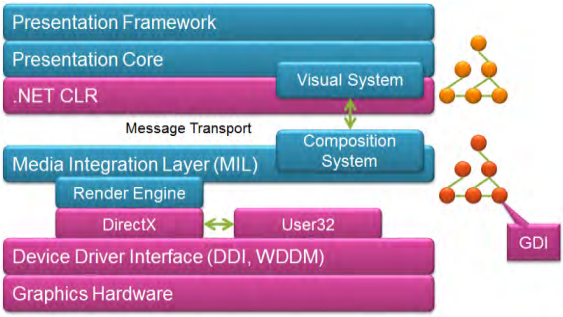
Over the years, Direct3D became a de-facto standard for many software vendors when developing software applications for visualization, games and other high-end graphics tasks. Direct3D 9, Direct3D 10 and Direct3D 11 are only available on Windows Vista and later due to the fact that they require the Windows Display Driver Model.

Note: Microsoft released Direct3D in 1995, which became a competitor of OpenGL. In December 1997, Microsoft and Silicon Graphics initiated a joint effort with the goal of unifying the OpenGL and Direct3D interfaces.

Remoting DirectX is typically done on the host side, resulting in very poor performance when using the older remoting protocols. However, there are forms of client-side DX remoting that are enabled by both Microsoft (Aero Glass Remoting in RDP) as well as Citrix XenDesktop (Desktop Composition Redirection). These forms of DX remoting even work well in decent WAN conditions, but they are designed to support basic DX remoting and are not designed to support high-end DirectX graphics remoting. By redirecting DirectX calls to the client device, you can reduce the server load on the hosted desktop platform thereby scaling more users. The downside of this approach is a more limited set of client hardware / OS support (though fallback to server side rendering occurs in those cases).

10.6 WPF REMOTING

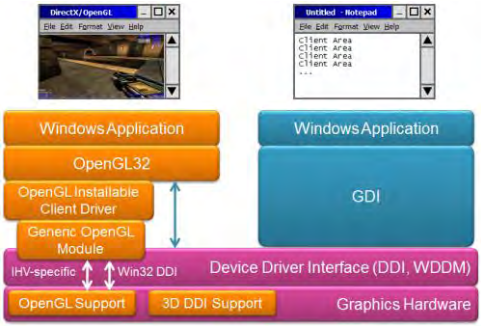
Windows Presentation Foundation (WPF) was introduced with Windows Vista as part of the .NET Framework 3.0. It was designed to be the successor of GDI for the standard graphical user interface of Windows desktops and applications.



WPF is completely built on top of DirectX, enabling modern UI features like transparency, gradients and transforms. One of the major WPF design goals was to provide a consistent programming model for building applications. In particular, it separates the user interface from business logic by providing a markup language (XAML) to define UI elements and relationships with other UI elements.

WPF applications allow for a variety of runtime options. They can either be deployed as stand alone desktop programs or hosted as an embedded object in a website. Independently of the deployment method, WPF supports a basic set of common user interface elements for 2D and 3D rendering, including vector graphics, runtime animation, and pre-rendered media. These elements can then be linked and manipulated based on various events, user interactions, and data bindings. One such element is the 2D surface that is generated by all open GDI applications. This GDI surface is mapped on the WPF-based Desktop Window Manager. As a result this concept for desktop composition allows to combine 2D and 3D elements seamlessly, including moving bitmaps, transparency and anti-aliasing. One of the biggest advantages for GDI applications when used in the WPF context is that the application logic doesn't need to re-render graphical content when the application window comes to the foreground after it was covered by an other window before. Remoting WPF is identical to remoting DirectX, which means that it is typically rendered on the host side and can be GPU accelerated.

10.7 OpenGL/WebGL Remoting



Open Graphics Library (OpenGL) is a multi-platform application programming interface (API) widely used in 2D and 3D applications. It was developed by Silicon Graphics Inc. in the 90s and is managed by a non-profit technology consortium, the Khronos Group, since 2006. The OpenGL API consists of over 250 different function calls to draw complex three-dimensional scenes from simple primitives. It is capable of interacting with a physical GPU for hardware-accelerated graphics rendering. Version 1.0 of OpenGL was released in January 1992. New versions incorporate a number of extensions, defining a new set of features which all conforming graphics cards must support. OpenGL 1.1 was released in 1997 and is the supported version in Remote Desktop Protocol (RDP) 8.1 including RemoteFX. OpenGL 1.2, 1.3, 1.4 and 1.5 followed in 1998, 2001, 2002 and 2003 respectively. OpenGL 2.0 introduced a Shading Language in 2004 and was succeeded by version 2.1 in 2006. OpenGL 2.1 is the supported version in VMware vSGA.

OpenGL 3.0 was released in 2008 and it introduced a deprecation mechanism for API revisions. OpenGL 3.1, 3.2 and 3.3 followed in the years until 2010. HDX 3D Pro is fully compatible to OpenGL 3.3 and earlier. The design goal of OpenGL 4.0 was to support DirectX 11 compatible hardware. It was also released in 2010 alongside version 3.3. New NVIDIA and AMD graphics cards led to versions 4.1, 4.2, 4.3 and finally 4.4. In general, OpenGL releases are backwards compatible. Most graphics cards released after the release date of a particular OpenGL version support those version features, and all earlier features.

In remoting scenarios, OpenGL is typically rendered on the host side, and it can greatly benefit from GPU acceleration.

WebGL is a cross-platform web standard for a low-level 3D graphics API. It is based on OpenGL ES 2.0 and exposed in Apple Safari,

Google Chrome, Mozilla Firefox and Opera browsers through the HTML5 Canvas element as Document Object Model interfaces.

10.8 OpenCL Remoting

Open Computing Language (OpenCL) is an open, cross-platform standard for writing gaming, entertainment, scientific and medical software. It was designed for parallel programming of CPUs, GPUs, digital signal processors and field-programmable gate arrays while it interoperates with OpenGL and other graphics APIs. Like OpenGL, OpenCL is maintained by the non-profit technology consortium Khronos Group. It has been adopted by vendors such as NVIDIA, AMD, Intel, Apple, Qualcomm, Samsung and ARM Holdings.

Version 1.0 of OpenCL was released in December 2008, allowing applications to tap into multi-core CPUs and GPUs. OpenCL 1.1 and 1.2 followed in June 2010 and November 2011 respectively, adding functionality for enhanced parallel programming and improved OpenGL and DirectX interoperability. OpenCL 2.0 is the latest significant version. It was released in July 2013 and is designed to further simplify cross-platform parallel programming.

Remoting OpenCL is pretty much identical to remoting OpenGL. Rendering typically happens on the sender side.

10.9 Flash Remoting

Adobe Flash is a popular multimedia platform to create Web advertisements, interactive games and Rich Internet Applications (RIAs), allowing the combination of still images, animation, vector graphics, audio and video. Viewing and interacting with the interactive multimedia content of a Flash application requires the Adobe Flash Player which supports the Flash Video Format (FLV). From a playback perspective, the Flash Player has a lot in common with a custom-developed video player using a proprietary multimedia format. But in contrast to a video player, Flash allows user interaction through mouse, keyboard, microphone and camera. This includes bidirectional streaming of audio and video data. A clear advantage of Flash is that the Flash Player is available free of charge as a plug-in for all major Web browsers and for many operating systems, smartphones and tablet devices.

Creating interactive Flash applications is based on an object-oriented programming language called ActionScript. Professional development tools are required to build such application. Since a couple of years, the introduction of new or updated Flash applications is declining on websites. However, there are still many Flash applications in use.

Remoting Flash is typically done on the host side, resulting in very poor performance when using the older remoting protocols. Only Citrix and Dell implemented a technology called Flash Redirection which relies on an existing Flash Player on the client side. Flash Redirection improves usability and reduces bandwidth requirements significantly, but also requires a client that is capable of installing and maintaining a Flash Player. There are some negative aspects of performing client side Flash remoting such as a dependency of keeping the server and client Flash versions up to date as well as some security related risks when redirecting Flash to the client system (i.e. a server hosted Flash control could infect a client PC since the Flash content is rendered on the client).

10.10 SILVERLIGHT REMOTING

Microsoft released Silverlight in 2007 with the ambition to replace Adobe Flash as the most popular application framework for interactive multimedia applications. Now in its fifth version, the Silverlight runtime environment is available as a plug-in for Web browsers running under Windows and Mac OS X. In addition, Silverlight is one of the Windows Phone development platforms, even though the resulting applications can only run natively on Windows Phone but not from Web pages in Internet Explorer on Windows Phone, Windows Mobile or Windows RT. Like with Adobe Flash, Silverlight remoting is done on the host side. There is no technology available today that allows redirecting Silverlight content even if a Silverlight runtime is installed on the client side. As a result, Silverlight performance is degraded when using the older remoting protocols.

10.11 AUDIO/VIDEO REMOTING

Playing back audio and video streams is very easy on Windows systems. The built-in Media Player supports a number of multimedia

file formats, such as WAV for audio and WMV or AVI for video. RealVideo and Apple QuickTime can be seen as competitors.

Most video files consist of two or more types of files objects. One is the media container which describes the overall structure while another represents the compression/decompression (codec) algorithm used inside the container.

Here is a short overview of the most popular file formats and codecs: MPEG-1 and MPEG-2 files are based on a compression format standardized by the Moving Picture Experts Group. MP4 is a video format that enhances the MPEG standard by the support of video/audio "objects", 3D content, low bitrate encoding and Digital Rights Management. In addition, it allows the separation of audio and video tracks. Video is compressed with MPEG-4 and audio uses AAC compression. Audio/Video Interleave (AVI) was developed by Microsoft and most commonly contains MPEG or DivX codecs even though it can contain almost any codec. The Windows Media Video format (WMV) was also developed by Microsoft. Initially designed for Internet streaming, it is now a common format used for video playback. QuickTime was developed by Apple. The format contains one or multiple tracks containing video, audio and other objects. The RealMedia format was created by RealNetworks, allowing media streaming of both audio and video data. The RealVideo format was developed by Real Media with a low bandwidth use case in mind.

The most popular video codec today is H.264/MPEG-4 Part 10 or AVC (Advanced Video Coding), released in May 2003. It is a block-oriented motion-compensation-based video compression standard, also used as one of the video encoding standards for Blu-ray discs. H.264 can be implemented in hardware which allows modern GPUs to accelerate video compression and decompression on hardware level. By default, H.264 compression is lossy, but it can be configured to be lossless which is an important aspect for use cases such as medical imaging.

Many remoting protocols try to redirect audio and video data streams to the client if on session connection the required codec packages can be found there. If a video stream is decoded in host video memory and reencoded for remoting the content, this comes with substantial resource requirements.

Note: Remoting protocols implementing host side rendering typically use H.264 to encode and compress the Windows desktop into video frames before sending it down the wire. Taking advantage of the H.264 hardware encoder on the GPU improves remoting performance significantly.

10.12 HTML5 FOR REMOTING

Hypertext Markup Language version 5 (HTML5 for short) is the advanced version of HTML, used for presenting text and multimedia content in the World Wide Web. An important new feature in HTML5 is the way it embeds graphics, audio, video and interactive documents, allowing to create Web applications. New syntactic features provide a simpler way to manipulate multimedia and graphical content without the need to install additional browser plugins. The latest versions of all modern browsers, such as Internet Explorer, Firefox, Chrome, Safari and Opera, support HTML5 to a certain extend. A public HTML5 test site (http://html5test.com/) creates a HTML5 compatibility score.

An interesting aspect of HTML5 is its capability to deliver remote desktops via native browser components. This requires WebSockets to create a TCP connection channel for continuous transmission of data throughout the entire remote session lifespan as opposed to the connectionless communication via HTTP. In addition, a native HTML5 element called Canvas is needed that has the ability to control single pixels discretely and output complete raster images at high speed. In combination, WebSockets and Canvas provide the necessary mechanisms to connect to a remoting host from a browser and render the 2D graphics output dynamically. As a result, the browser receives the remoting data through WebSockets and uses the Canvas element to draw the desktop. A gateway component or driver between the host and the browser is required to render the content received through the remoting protocol and convert it into a binary data stream of desktop frames retransmitted through WebSockets.

Among industry experts and market analysts, remoting clients based on HTML5 are regarded as a top candidate for the preferred future technology. But at this stage it is too early for declaring HTML5 as the successor of native remoting clients.

10.13 CUDA IN GRAPHICS REMOTING ENVIRONMENTS

The Compute Unified Device Architecture (CUDA) is NVIDIA's parallel computing platform and programming model. CUDA was designed to be used with NVIDIA GPUs, giving developers direct access to instruction set and memory of the graphics chip. Use cases are not only limited to high-end graphics only. CUDA GPUs can also be used for general purpose processing, an approach called GPGPU. The advantage a GPU has over a CPU is that GPUs are optimized for running a big number of concurrent processes or threads side by side, resulting in superior overall performance. This makes GPGPU extremely powerful for scientific calculations and cryptography.

NVIDIA offers CUDA-accelerated libraries and extensions to standard programming languages such as C and C++, allowing developers to take advantage of the GPU's capabilities. Vendors like Adobe or Autodesk use such features to accelerate their 3D CAD/CAM programs.

CUDA-enabled applications only run if there is a physical NVIDIA GPU present on the system. Remoting such applications also benefits of the hardware acceleration on the GPU.

11. 3D GRAPHICS FOR VIRTUAL DESKTOPS VENDORS AND THEIR SOLUTIONS

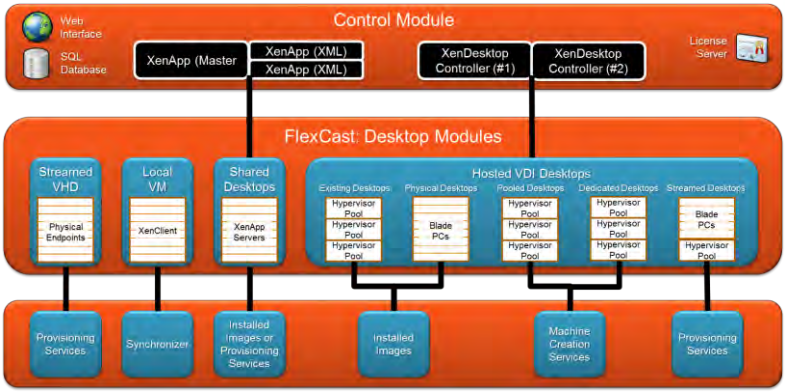
11.1 INTRODUCTION

This chapter will give you an overview of the major vendors and their solutions to deliver 3D graphics for virtual desktop.

11.2 CITRIX

Introduction

Citrix has a long history in Server Based Computing with its XenApp product. To expand the market, Citrix focused on using a broad set of virtualization technologies to create XenDesktop, an end-to-end solution for providing virtual desktops and apps for a broad set of use cases. XenDesktop includes the ability to host personal, pooled, and dedicated virtual machines (traditionally known as VDI,) but also enables cost-effective and scalable hosted-shared environments with powerful personalization tools, locally executed desktops with XenClient (a Type-1 hypervisor) and streamed VHD (real-time streamed OS at boot). Finally XenDesktop also enables GPU enabled PC blades and GPU assisted Hypervisors for providing rich, pixel-perfect 3D professional graphics support. All these different options are all part of the Flexcast Management Architecture.



However, not every user needs a full virtual desktop. Increasingly, the demands for extending Windows apps to tablet and smartphone form factors or the need to provision virtual apps on-demand to BYO laptops have increased the need for an integrated application virtualization and delivery infrastructure. XenDesktop includes these capabilities to support enterprise workers who may roam between device types and networks but need a consistent work environment that is easy to use, centrally managed and secure.

Architecture

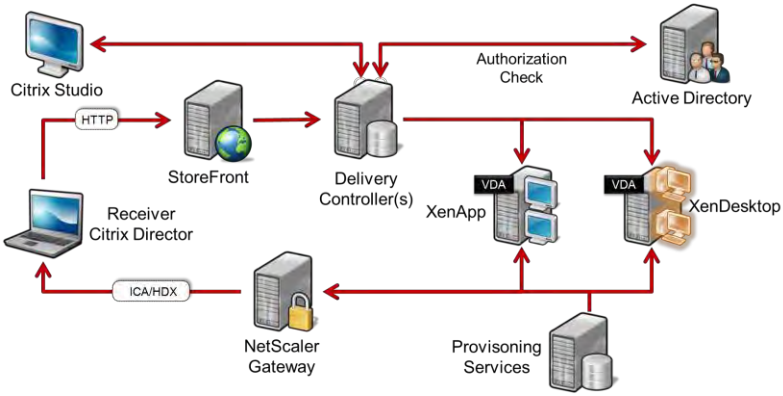
The architecture of XenDesktop, today, consists of multiple components. This makes the initial setup of XenDesktop more complex than some other solutions, but is designed for flexibility and scale. The tasks of the connection broker or the "Delivery Controller" are:

- Authenticates users
- Manages assembling the user's desktop environment
- Brokers connections between the user and his virtual desktop

Other components of the XenDesktop infrastructure are: Virtual Infrastructure (XenServer, Microsoft Hyper-V or VMware vSphere), Licensing server, database, Provisioning server, Storefront, Director and NetScaler Gateway for secure remote access.

Introduced in XenDesktop 7 is a single architecture to host virtual machines, remote pc and hosted shared desktops. XenApp 7.5 delivers shared desktops through a Remote Desktop Services based

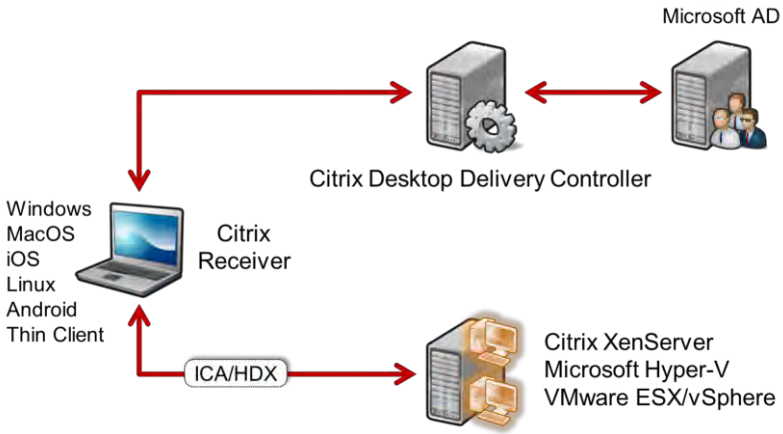
Windows Server OS. The XenApp and XenDesktop infrastructure is hereby capable to include access to remote or streamed applications on a Windows Server OS.



(Citrix XenDesktop and XenApp 7.5 overview)

A XenDesktop agent is required in the guest VM or Blade PC. The agent enables direct connections between the endpoint and the user's virtual desktop. On the endpoint the Citrix Receiver or Receiver for HTML5 can be used to provide simple access from any device, anywhere, to virtual desktops and applications. The Citrix Receiver provides the most comprehensive experience, while the Receiver for HTML5 has less endpoint requirements as well as capabilities..

The protocol used for XenDesktop (and XenApp) is the ICA protocol and has some special features. The ICA protocol performs better on limited bandwidth and higher latency than the RDP/RemoteFX protocol.Citrix HDX Technologies were introduced in XenDesktop 3. HDX includes several elements that improve user experience, such as VoIP and webcam support, 3D support, enhanced audio and optimization for use over WAN.



(Citrix XenDesktop overview)

Licensing

Each license type includes, next to the XenDesktop Controller (the broker) the right to use certain components and features of Citrix XenDesktop: VDI desktops, Hosted Shared Desktops, XenServer, Provisioning Server, NetScaler Gateway, EdgeSight, etc. VDI edition is the most basic version with a partial but still strong feature set for pure VDI deployments. Platinum is the most advanced edition and delivers the complete feature set. NOTE: With respect to 3D graphics support, you must have the Enterprise or Platinum Editions of XenApp/XenDesktop to leverage the HDX3DPro feature set.

An overview of the different XenApp and XenDesktop versions and its features can be found here.

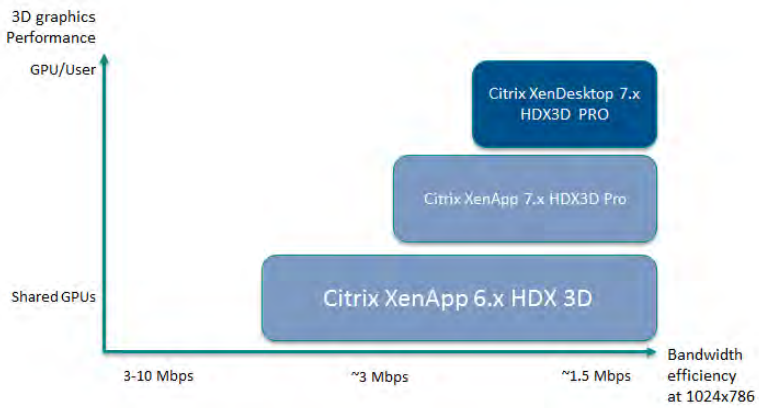
	XenDesktop Editions			XenApp Editions		
Feature	VDI	Enterprise	Platinum	Advanced	Enterprise*	Platinum*
Application Virtualization and Hosting (RDS)		✓	✓	✓	✓	✓
VM Hosted Apps (16 bit, 32 bit, or 64 bit Windows apps)		✓	✓		✓	✓
App-V delivery for offline apps		✓	✓		✓	✓
Hosted Shared Desktop (RDS)		✓	✓	✓	✓	✓
Local - Type 1 client hypervisor (XenClient Enterprise)		✓	✓			
Pooled VDI Desktop	✓	✓	✓			
Dedicated VDI Desktop	✓	✓	✓			
Physical Desktop (Blades, racked workstations)		✓	✓			
Remote PC access to office-based PCs (Wake on LAN)		✓	✓			
HDX Seamless Local Apps			✓			✓
Persistent, personalized VDI desktops with single image management (Personal vDisk technology)	✓	✓	✓			

A complete list of the different functions within various editions can be found here:
https://www.citrix.com/go/products/xendesktop/feature-matrix

Citrix and Hardware Accelerated Graphics for Desktop Virtualization

Citrix's XenDesktop and XenApp support for 3D graphics for Virtual desktops timeline:

- 2009: Availability of XenDesktop HDX 3D Pro with Deep Compression
- 2011: XenServer 6.0 supports GPU passthrough
- 2012: HDX3DPro increased fps via NVIDIA GRID. Improved compression and bandwidth reduction
- 2013: XenDesktop 7.0+ GPU sharing via GRID K1/K2 interfaces. XenApp 6.5+ GPU sharing (OpenGL/WebGL), NVIDIA vGPU released and integrated in XenServer and XenDesktop



Citrix XenDesktop or XenApp

What are the benefits of using 3D graphics with Citrix XenDesktop and Citrix XenApp?

Citrix XenDesktop is able to deliver a high-performance 3D graphics solution for the most demanding users such as design engineers who

view, create, manipulate, render complex 2D/3D graphics. XenDesktop It provides GPU acceleration for Direct3D, OpenGL, CUDA, OpenCL applications with H.264 based deep compression. This assures that the least amount of bandwidth is consumed. High-performance 3D graphics can be achieved because every Virtual Machine is connected to a GPU. 3D Spacemouse support is available. Microsoft licensing for VDI, Virtual Desktop Access, can be a challenge. Especially in a Bring Your Own Device or work from anywhere scenario.Both Citrix XenServer 6.2+ and VMware vSphere 5.1+ can support Citrix XenDesktop with GPU pass-through. The GPU virtualization technology of NVIDIA's with vGPU is only supported with Citrix XenServer 6.2+.

Citrix XenApp is able to deliver 3D graphics solutions to operators and contractors who view, manipulate, 2D/3D graphics. XenApp provides GPU support acceleration for Direct3D, OpenGL, CUDA, OpenCL. XenApp is installed on a multi-user Windows Server OS, applications needs to be compatible with Citrix XenApp/RDSH. 3D Spacemouse support is currently not available yet. An Microsoft Virtual Desktop Access (VDA) License isn't needed, instead a RDS Client Access License is needed. Citrix XenApp with 3D graphics support is cost effective solutions because of GPU sharing. The downside is that 3D graphics performance isn't as high as Citrix XenDesktop with GPU Pass-Through or the high-performance vGPU profile.

Citrix XenServer

Citrix XenServer is an open source virtualization platform for managing cloud, server and desktop virtual infrastructures. XenServer is designed to optimize private datacenters and cloud. Citrix Xenserver support both GPU pass-through and GPU virtualization via NVIDIA GRID vGPU. In XenServer the NVIDIA driver, the vGPU Manager, is installed in the control domain and the NVIDIA vGPU Windows driver is installed in the guest OS of the virtual machine. Keep in mind that XenServer only support pass-through of one GPU/VM. Each physical GPU can be split into 2, 4 or even 8 vGPUs when using NVIDIA GRID K1/K2.

Citrix, 3D graphical for virtual desktops

Vendor support for different 3D graphical solutions	Software 3D graphics*	Bare metal Client OS	Pass-through GPU	GPU Virtualization - vGPU	GPU sharing for VDI (API Intercept)	GPU sharing for XenApp/RDSH**
Citrix XenApp – Server OS (RDSH)	√	N/A	√	√	N/A	√
Citrix XenDesktop – Client OS (VDI)	√	√	√	√	X	N/A

11.3 MAINFRAME2

Introduction

Mainframe 2 is a startup that aims to run thousands of heavy-duty non-web Windows and Linux applications on HTML5-supporting browsers without plug-ins. Users can simply sign in to Mainframe2 from any HTML5 capable browser and have instant access to their apps. No plugins, such as Java or Flash, are required. Applications are streamed directly from the the public cloud as a video stream. Founded in summer 2012, Mainframe2 is based in Menlo Park, CA.

The Mainframe2 team has deep expertise in video streaming via an earlier organization, MotionDSP. While video streaming using H.264 is a key component of their technology, the Mainframe2 solution includes a variety of other capabilities ranging from cloud infrastructure orchestration to dynamic quality of service adjustment to mobile device support.

Today's GPU virtualization, faster networks and elastic clouds are turning all of our state-of-the-art devices back into terminals like in the early Mainframe days.

The vision of Mainframe2 is to make any application, including 3D graphics, available on demand and available to everyone with an Internet connection. It's delivering applications on-demand, present them seamless to end-users wherever they are leveraging high-performance and scalable public cloud infrastructures.

Solution

With Mainframe2 every users has its own Virtual Machine, or when needed every application can have its own Virtual Machine. The focus of Mainframe2 is application delivery and not desktop delivery. Currently that Virtual Machine is running Windows Server 2012 R2. The applications are currently only running on Amazon EC2,

leveraging flash based storage, GPUs and CPUs of EC2 GPU instances. The technology can run in theory on-premises but right now the implementation is running on Amazon. Mainframe2 can use various Amazon instances, c3.large, c3.xlarge, and g2.2c large are the 3 instances Mainframe2 currently offers to their customers. The last one is with a powerful GPU. While the application runs in the datacenter, each video frame is converted to H.264 video and the application is seamlessly presented to any endpoint running a HTML5 browser or Mainframe2 client. WebSockets are being used, No WebRTC. Also there is a native client for Windows, OSX, iOS and Android available.

The remoting protocol is tuned to consume as little bandwidth as possible, dropping to a few kbps and 1fps when the content is still and then instantly jumping to 10Mbps+ and 60fps when needed. This makes the solution suitable to deliver applications in LAN, WAN, 3G, 4G scenarios.

Multiple applications can be installed in the same virtual machine. Administrators can easily setup different users with access to different apps or groups of apps. Software vendors don't have to rebuild their native Windows applications for the web in order to run them on Mainframe2. The applications, MSI, exe or ISO, are installed via a quick installation process into the platform. This can be done by the admin.

It's key to emphasize that there are other services that Mainframe2 provide on top -- like integration with cloud storage (e.g., dropbox, box, or your own..), integration with SSO (e.g., Okta), metering, billing, and user management. Customers pay a single bill to Mainframe2 and it includes EC2 and all other infrastructure related costs.

Availability

The plan is to make the platform available for public Q2-2014.

Awesome demos and preview of the solution can be found at www.Mainframe2.com

11.4 MICROSOFT REMOTE DESKTOP SERVICES

Introduction

With the technology originally called Terminal Services, a user is connected with a session on a Windows Server which allows for providing a full desktop session or individual programs (called RemoteApp) on the user's client device. With the launch of Windows Server 2008 R2 in October 2009, the name "Terminal Services" was replaced by "Remote Desktop Session Host" (RDSH). Windows Server 2008 R2 also introduced a new role, namely "Remote Desktop Virtualization Host". By adding this role to Microsoft's hardware virtualization platform, called Hyper-V, Remote Desktop Virtualization Host now also allows for providing users with a dedicated virtual desktop running a Windows Client operating system. With SP1 for Windows Server 2008 R2, Microsoft added RemoteFX to Remote Desktop Services. The RemoteFX technology allows for a better user experience when using Remote Desktop Services. RemoteFX allows USB redirection, high-definition video and using 3D applications over Microsoft's Remote Desktop Protocol. RemoteFX works for virtual desktops as well as session based desktops.

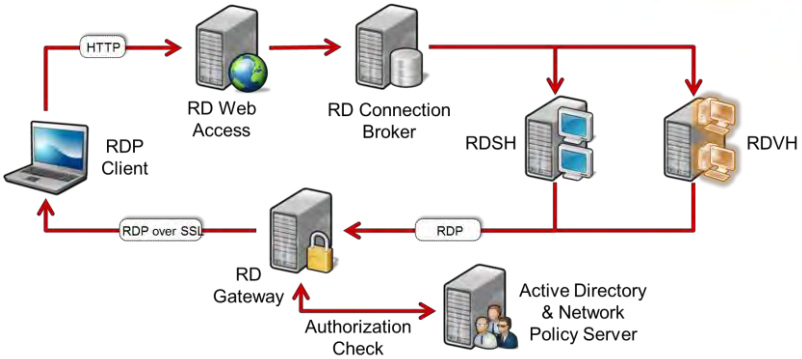
With the Windows Server 2012 release Microsoft redesigned the complete Remote Desktop Services stack to improve the ease of deployment and management. RemoteFX features are also improved and broadened to support a new generation of Windows (8) devices. RemoteFX is optimized for WAN deployments and touch-based devices. Microsoft also delivers a built-in profile management solution for virtual desktops and session based desktops.

Architecture

Microsoft utilizes its own Remote Desktop Protocol (RDP) with RemoteFX (RFX) enhancements to provide the remote desktop or remote application to the user. For the best experience and functionality a client device should run Windows 7 SP1 or Windows 8 with RDP8. The Microsoft Virtual Desktop Infrastructure consists of the following Windows Server 2012 roles:

- Remote Desktop Gateway (RDG). This is an optional role to provide secure access to the Microsoft Virtual Desktop Infrastructure from internet-based clients.

- Remote Desktop Web Access (RDWA). This role provides access to the desktops and/or remote applications available for a specific user. After the user browses to the Web Access URL and authenticates, Web Access provides a webpage displaying the shortcuts to the resources available to this user. If the client device is running Windows 7/8 and is on the corporate LAN the shortcuts can be also integrated in the user's Start Menu.
- Remote Desktop Connection Broker (RDCB). The Connection Broker tells Web Access which resources are available to the user. The RDCB role is the broker which connects the client to the correct resource selected by the user in Web Access. The Connection Broker also contains the Remote Desktop Management Service. The Remote Desktop Management Service maintains a database with the static configuration of the deployed RDG, RDWA, RDCB, RDSH and RDVH roles, and dynamic session information of the managed RDSH and RDVH servers.
- Remote Desktop Session Host (RDSH). Formerly known as a Terminal Server, RDSH provides server hosted desktops or remote applications to the client. The RDSH role is not required for a Windows Server 2012 virtual desktop infrastructure, but could be added to provide a hybrid solution.
- Remote Desktop Virtualization Host (RDVH). A Virtualization Host is a Microsoft Hyper-V host with the Virtualization Host agent service installed. RDVH provides virtual desktops or remote applications to the client. The Virtualization Host agent service manages the starting of the virtual machines or remote applications (in a virtual machine) when a user wants to connect.



3D Graphics for Virtual Desktops with RemoteFX

Microsoft Remote RemoteFX delivered shared GPU access to virtual machines with Microsoft Hyper-V as a hypervisor. A DirectX 11 driver needs to be installed in the parent partition of Windows Server 2012 R2 Hyper-V to leverage shared GPU access within the virtual machines. The shared GPU functionality is also known as virtualized GPU with API intercept technology. RemoteFX in Windows Server 2012 R2 is primarily designed for fluid user experience with Line of Business applications and Windows graphical functions such as Aero. The shared GPU provided optimized graphics for DirectX enabled applications in the virtual machine. The solution isn't targeted for applications which requires OpenGL 1.1+, OpenCL, CUDA and users who need high-end 3D graphic performance.

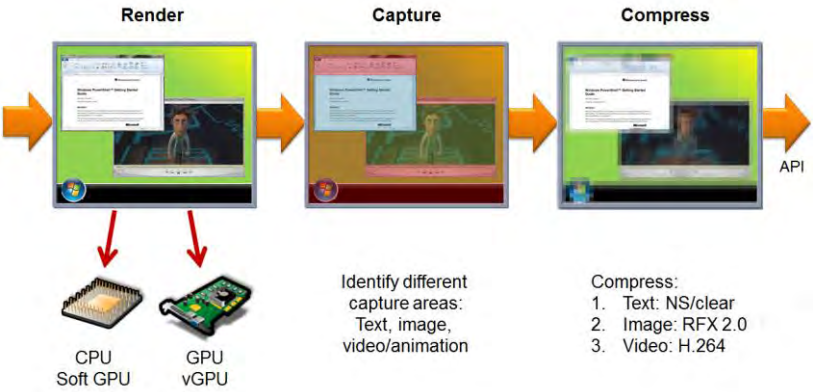
The NVIDIA/AMD GPU has a dedicated amount of video RAM. Each virtual machine consumes a specific amount of video RAM. The amount is based on the maximum amount of monitors and the resolution configured for each virtual machine. The combination of monitors and resolution determines the maximum number of virtual machines per physical NVIDIA/AMD GPU.

Maximum resolution	Maximum number of monitors in virtual machine setting							
	1 monitor	2 monitors	3 monitors	4 monitors	5 monitors	6 monitors	7 monitors	8 monitors
1024 x 768	48 MB	51 MB	54 MB	57 MB	60 MB	63 MB	66 MB	69 MB
1280 x 1024	80 MB	85 MB	90 MB	95 MB	100 MB	105 MB	110 MB	115 MB
1600 x 1200	118 MB	125 MB	132 MB	140 MB	N/A	N/A	N/A	N/A
1920 x 1200	141 MB	150 MB	159 MB	167 MB	N/A	N/A	N/A	N/A
2560 x 1600	250 MB	266 MB	N/A	N/A	N/A	N/A	N/A	N/A

Starting with Windows Server 2012 R2, the default resolution that virtual machines with RemoteFX virtual GPU support is 1920 x 1200. This prevents manually changing the default resolution for full-screen sessions in most cases. In addition, RemoteFX virtual GPU also supports a higher maximum video RAM allocation. The maximum amount of VRAM for a Windows 8.1 VM on Windows Server 2012 R2 RDVH is 256MB dedicated VRAM plus 1GB shared VRAM. But only virtual machines with a minimum of 2,512MB of system memory will get the maximum of 1GB shared VRAM, which results in a total of 1,280MB dedicated + shared VRAM per VM. The VRAM configuration can be observed by running the command dxdiag within a virtual machine's command prompt and clicking one of the Display tabs to view the Approx. Total Memory value.

RemoteFX Under the Covers

When Microsoft introduced RemoteFX version 1 in Windows Server 2008 R2 SP1, it could only be used in VDI scenarios when a compatible physical GPU was present in the server. Starting with Windows Server 2012 this was changed in such a way that RemoteFX can be used with or without physical GPU. When no GPU is present, RemoteFX version 2 uses a CPU-emulated GPU (called SoftGPU) for rendering graphics. When, however, one or more GPUs are present in the server, the synthetic Hyper-V graphics driver in the VM intercepts and forwards a defined subset of graphics API calls to the physical GPU. This situation is referred to as RemoteFX vGPU - which should not be confused with Citrix XenServer vGPU. The advantage of the RemoteFX v2 architecture is that independent of the existence of a physical GPU in the server the same algorithms can be used for remoting the screen buffer content. Requirements for Windows Server 2012 (R2) RemoteFX vGPU are that all VMs must be running on Hyper-V, the server CPU must have Second Level Address Translation (SLAT) enabled and the graphics card(s) installed on the Hyper-V server must be DirectX 11 capable. Examples for RemoteFX-compatible GPUs are NVIDIA GRID K1 and K2 as well as AMD FirePro S10000, S9000 and S7000. A compatible GPU which was virtualized and shared by Hyper-V is available as an assigned GPU resource in the VM, almost like a physical PC with a physical GPU.

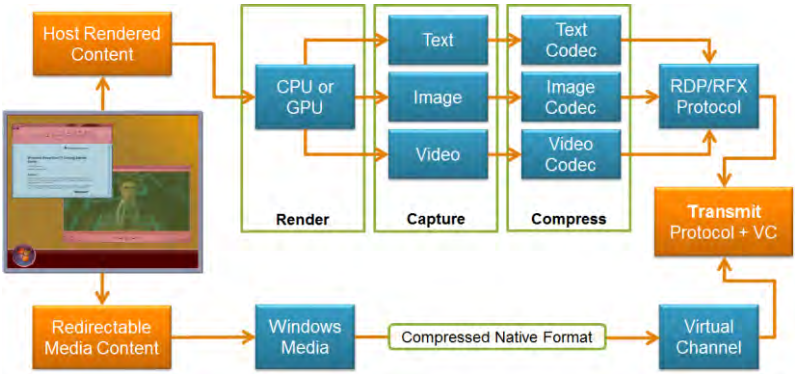


But RemoteFX v2 is not only about rendering, it's also about capturing and compressing screen frames. The RemoteFX capture component analyzes what was rendered in the previous frame and what's been rendered in the current frame. It dynamically divides the screen content into three different capture areas, representing text, still images and video/animation. Microsoft calls this mechanism Intelligent Screen Capture, which is responsible for checking screen content changes in the three capture areas and forwarding the changed bits for encoding. In addition, the capture component continuously communicates with the remoting protocol stack, allowing it to track network speed and understand which buffers and frames have been acknowledged by the client. As a result. the transmission of screen content can be adjusted dynamically according to the available bandwidth.

The final component is the RemoteFX encoder which is responsible for compressing the screen data before it is sent to the network protocol stack. Encoding can be done on the CPU, on the GPU or on dedicated hardware. In other words, the compress cycle may use both the GPU and CPU. Compression of the text area uses an NS/clear algorithm which provides high compression ratio and high quality without any noticeable time delay. Image compression uses either a RemoteFX-specific algorithm or AVC/H.264 in an adaptive manner, meaning that the image output quality continually improves within a couple of seconds. Video and animation content that is not redirected also uses H.264 for re-encoding. From the capture area encoder output the compressed RemoteFX content is passed on to the RDP stack where it will go through the "standard" processing,

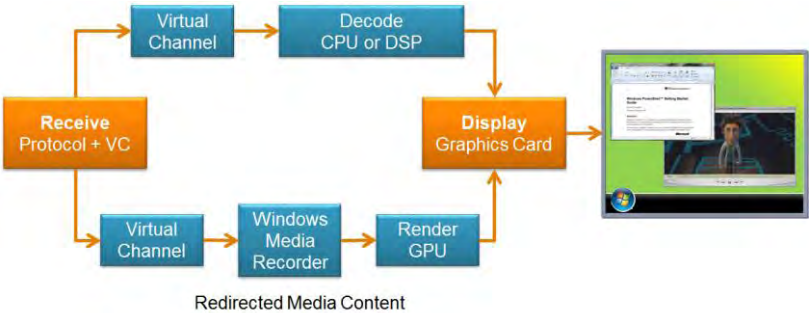
including bulk compression, encryption and prioritization. As much as possible, compression is offloaded to the GPU.

Typically, the entire screen goes through the RemoteFX render-capture-compress pipeline at 30 frames per second, except for redirected multimedia, such as WMV streams. Redirected media content is directly transmitted to the client in its compressed native format using an RDP virtual channel.



(RemoteFX v2 - Host side)

On the client side, the Remote Desktop Connection application receives the screen data from the remoting protocol network stack. It decodes the RemoteFX data using CPU, GPU or Digital Signal Processor (DSP), re-composes the entire desktop and displays the received frames on the local screen. Only redirected media content is processed in a data stream separated from the RemoteFX receiver pipeline. It is sent directly to the associated media player, typically supported by the GPU.



(RemoteFX v2 - Client side)

Microsoft RDS, 3D graphics for virtual desktops

Vendor support for different 3D graphical solutions	Software 3D graphics*	Bare metal Client OS	Pass-through GPU	GPU Virtualization - vGPU	GPU sharing for VDI (API intercept)	GPU sharing for XenApp/RDSH**
Microsoft RDSH – Server OS (RDSH)	√	X	X	X	N/A	√**
Microsoft RDVH - Client OS (VDI)	√	X	X	X	√	N/A

11.5 NICE DCV

Introduction

NICE vision is to provide high performance access for the most demanding applications, including high end 3D and computing tools, from virtually any kind of device and network. NICE been pioneers in Technical Cloud deployments at leading companies and institutions, in many vertical markets including Oil&Gas, Industrial Manufacturing, Life sciences, Government, Electronic Design, Financial Services and Research / Academic.

NICE is a software and consulting Services Company that is dedicated to improving ease of use and productivity, and achieving major hardware cost savings for companies running high performance technical applications. NICE developed the world's first HPC web portal, EnginFrame, some 15 years ago and this is widely used by engineers, geologists and others to ease the process of submitting HPC computations. EnginFrame is application aware, has a rich set of data management features and provides feedback to the user while the job is running. More recently, NICE acquired DCV from IBM, after many years of development, and has since further developed this software to the point where is it has a broader set of capabilities than any other software of its kind. NICE Desktop Cloud Visualisation (DCV) enables remote visualisation, and workstation virtualization for a mix of Windows and Linux HPC applications, all sharing graphics cards (GPUs).

Solution

NICE remote access software for 3D and HPC applications is built on two pillars: the EnginFrame portal and the Desktop Cloud Visualization (DCV) protocol.

EnginFrame provides a flexible Web-based framework to publish interactive, batch and full desktop services, a comprehensive set of data transfer and management features, coupled with a proven integration with mainstream HPC management software. It offers a user-friendly experience for any Private and Public Cloud computing need, from simple remote desktop brokering to the most advanced engineering workflows.

DCV delivers the highly responsive and network optimized experience needed for remote desktop and remote GUI access for high end applications based on OpenGL or DirectX. With its support for the latest GPU virtualization technologies from NVIDIA, real-time collaboration capabilities, hypervisor independence and compatibility with Linux and Windows OS, it allows users to be enjoy running high end 3D applications from any device and over most network conditions.

Providing both high-performance computing (HPC) and visualization tools is becoming more difficult than ever. The rising cost of workstation management, larger file sizes, more remote workers, and the resulting need for collaboration are requiring IT professionals to reexamine how they deliver services.

NICE EnginFrame, NICE's HPC portal offering, is an advanced portal that provides access to grid-enabled infrastructures, HPC clusters, data, licenses, and interactive applications. It can be accessed by any authorized user with a standard web browser. EnginFrame handles computationally-intensive and sometimes parallel job submission, control, and monitoring. EnginFrame is based on standard protocols that facilitate the deployment of engineer-friendly portals to create, discover, and explore more efficiently. It provides for encrypted access and file transfers, protecting intellectual property and infrastructures.

EnginFrame provides HPC capabilities for large calculations, while Desktop Cloud Visualization provides the 3D modeling tools needed for complex, detailed models. Used together, they meet the computing needs for technical end users while delivering services in an intuitive, user-friendly interface.

Architecture

The NICE DCV architecture consist of:

- DCV Server, equipped with one or more GPUs, used for OpenGL rendering
- One or more DCV End Stations, running on "thin clients", only used for visualization
- Heterogeneous networking infrastructures (like LAN, WAN and VPN), optimized balancing quality vs frame rate

3D application delivery requires servers / blades equipped with NVIDIA GPUs, including many Quadro and Tesla models but best performing with the latest GRID cards, like the K2 or the gaming cards provided by Amazon G2 instances. A free accelerated endpoint software is provided for Linux, Windows and Mac OS/X, running on a broad range of devices including many thin clients or tablets, and other clients are also available on iOS, Android and other devices via our VNC compatibility mode. Thanks to its adaptive compression, DCV can accommodate high frame rates over LAN and WAN connections even on very high trans-oceanic latency and low bandwidth scenarios.

Besides a Web server layer for the portal publishing, the HPC application delivery leverages existing HPC clusters running leading job schedulers (including Torque/Moab, LSF, GridEngine, PBS/Pro, etc.) and mainstream batch applications, and requires only a Web browser to enjoy the full HPC Cloud experience.

The NICE DCV Technical Cloud software consists of three main components:

1. **Self-service portal**: The self-service portal enables engineers and scientists to access the applications and data in a web browser–based setting. It also provides security, monitoring, and management to ensure that users cannot leak company data and that IT managers can track usage. Engineers and scientists access applications and data directly from their web browsers, with no need for a separate software installation on their local client.
2. **Resource control and abstraction layer:** The resource control and abstraction layer lies underneath the portal, not

visible to end users. It handles job scheduling, remote visualization, resource provisioning, interactive workloads, and distributed data management without detracting from the user experience. This layer translates the user request from the browser and facilitates the delivery of resources needed to complete the visualization or HPC tasks. This layer has a scalable architecture to work on a single cluster or server, as well as a multi-site WAN implementation.

3. **Computational and storage resources:** The technical cloud's horsepower includes the company's existing or newly provisioned industry-standard resources, such as servers, HPC schedulers, memory, graphical processing units (GPUs), and visualization servers, as well as the required storage to host application binaries, models and intermediate results. These are all accessed through the web-based portal via the resource control and abstraction layer and are provisioned according to the end user's needs by the middle layer.

Technical cloud software is built on common technology standards. The software adapts to network infrastructures so that an enterprise can create its own secure engineering cloud without major network upgrades. The software also secures data, removing the need to transfer it and stage it on the workstation, since both technical applications and data stay in the cloud. These solutions feature the best characteristics of cloud computing—simple, self-service, dynamic, and scalable, while still being powerful enough to provide 3D visualization as well as HPC capabilities to end users, regardless of their location.

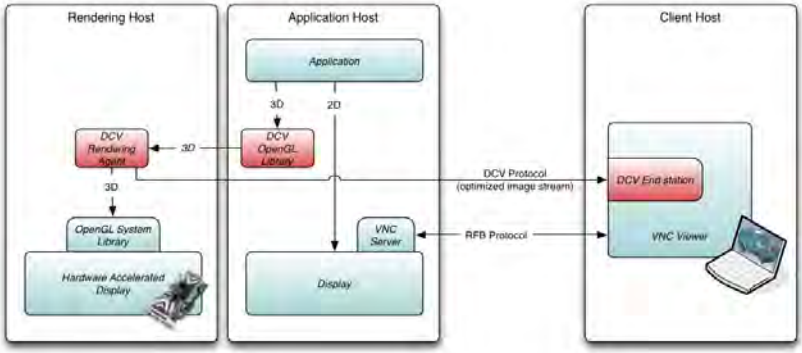
The technical cloud software solution can be specialized and optimized in two primary ways:

- HPC portal
- Remote 3D "Virtual Workstations". >> the focus of this whitepaper is VDI

Remote 3D "Virtual Workstations"

This model is ideal for 3D visualization and connecting remote users to OpenGL applications that run in a data center. Users access full desktops - like in common VDI (Virtual Desktop Infrastructure) – or

individual applications and data through a web browser, but the 3D performance is much better than a traditional VDI. The "Virtual Workstation" model of technical cloud software primarily provides visualization capabilities. It includes visualization servers equipped with one or more GPUs that may or may not be shared among users. These GPUs provide OpenGL acceleration required by technical applications that can be installed and run in the cloud without modifications. The end user works from a thin client, such as a laptop equipped with a web browser, that is only used to show pixels and requires no data or application logic locally. Therefore, even tablets can therefore be used to work or collaborate. Data and applications are kept in the data center, but remote desktop access allows engineers and scientists to use them and be productive from remote locations. Technical cloud software can be designed and installed by in-house IT staff; however, technical cloud consulting firms can provide support throughout the implementation processes. Thus managers and architects can focus on delivering the required HPC and visualization capabilities while minimizing disruption to business operations.



The OpenGL application runs in the Application Host, normally a VirtualMachine. All the OpenGL calls performed by the application are intercepted by the NICE DCV OpenGL interposing library and are redirected to an external rendering host.

The external rendering host must run the Linux OS and must be equipped with one or more (NVIDIA) 3D accelerated graphic adapters.For maximum performances, the advice is to use the application host and rendering host on the same physical host. The

rendering host could be a Linux box running KVM hosting multiple Windows VMs acting as application hosts, or a Linux VM with GPU passthrough running on the same physical host as one or more Windows VMs using Citrix XenServer or VMware ESX.

In addition to sharing a single GPU among multiple Windows VMs, OpenGL applications can run on the Linux rendering host and you can share a GPU among multiple OSes, applications running in the Windows VMs and applications running in the Linux host.

Another technology NICE developed is the ability to share the same GPU among multiple users on the same physical Linux host, or Linux VM with GPU passthrough.

In this scenario multiple users are all running on the same Linux instance leveraging the multi-user capability of Linux each one with his own desktop instance with full hardware accelerated 3D capabilities.

NICE DCV highlights

- Enables high performance remote access to interactive 2D/3D software applications on low bandwidth/high latency
- Supports multiple heterogeneous OS (Windows, Linux)
- Enables GPU sharing
- Supports 3D acceleration for OpenGL applications running on Virtual Machines
- Supports multiple user collaboration via session sharing
- Enables attractive Return-on-Investment through resource sharing and consolidation to data centers (GPU, memory, CPU, ...)
- Keeps the data secure in the data center, reducing data load and save time
- Enables right sizing of system allocation based on user's dynamic needs
- Facilitates application deployment: all applications, updates and patches are instantly available to everyone, without any changes to original code

NICE supports both Windows as Linux with Oracle Linux, SUSE Linux Enterprise Server and RedHat Enterprise Linux guest OS instances.

Both DCV and EnginFrame are licensed on concurrent usage: DCV based on the number of desktop sessions running on the servers,

EnginFrame on the number of browsers connected to the portal. Licenses can be annual subscriptions or permanent with annual maintenance contracts to get access to the new releases and technical support.

NICE DCV, 3D graphics for virtual desktops

Vendor support for different 3D graphical solutions	Software 3D graphics*	Bare metal Client OS	Pass-through GPU	GPU Virtualization - vGPU	GPU sharing for VDI (API intercept)	GPU sharing for XenApp/RDSH**
NICE DCV	X	√	√	√	√	N/A

11.6 NVIDIA - Visual Computing Appliance - VCA

NVIDIA Visual Computing Appliance (VCA) is a datacenter appliance which delivers a full 3D graphics for Virtual Desktops solution. The solution contains CPUs, GPUs, Memory, Storage, Hypervisor, eight Windows 7 VMs and one management VM, the GRID remoting protocol and software clients for various endpoints. It uses is the same protocol (H.264 hardware encoding) as NVIDIA is using for their GRID gaming solution. The NVIDIA VCA currently is using Xen (XCP) hypervisor, Citrix XenServer will be used in the future. The VCA appliance is designed for high-end 3D graphics applications providing pass-through access to NVIDIA GRID GPUs.Currently there are two VCA models. One model has 8 GPUs and 192GB of RAM and the other model has 16 GPUs and 384GB of RAM. Both models are 4U. One of the unique capabilities is the ability to put up to 4 GPUs into a single VM. Note that this is multiple GPUs per user not the opposite. One of the goals of NVidia with Visual Computing Appliance is to find new use-cases at customers and inspire others in the industry with new developments.

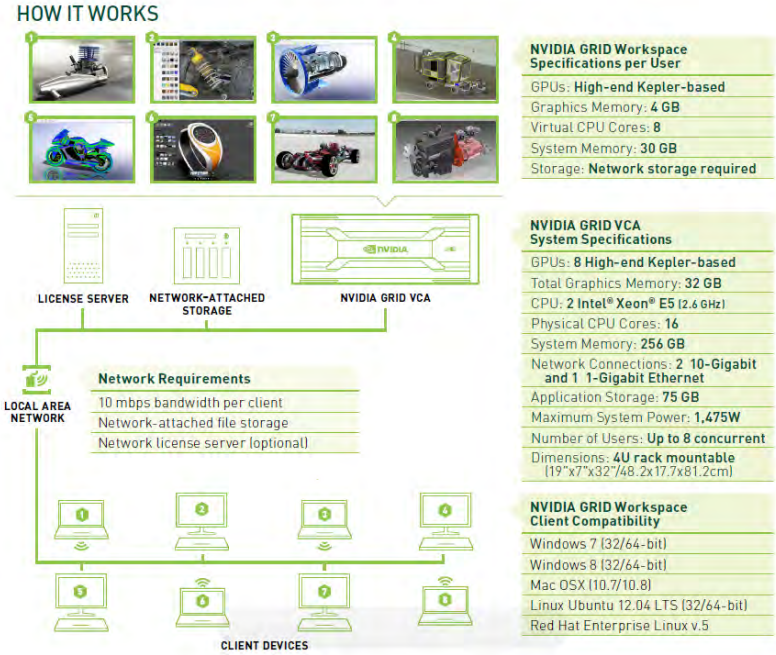
Benefits of the VCA are:

- Turnkey platform for 3D Graphics, simple installation and management
- Workstation class hardware, based on SuperMicro solution
- Powered by NVIDIA GRID K2 boards
- Includes 500GB SSD

- Up to eight concurrent users and up to 30GB of RAM/Virtual Machine
- 4GB VRAM per Virtual Machine
- Windows, Mac and Linux client support

Challenges:

- Focus is Small and Medium customers therefore lacks enterprise features like Active Directory integration and large scale Virtual Desktop management tools
- The GRID remoting protocol is targeted for LAN usage (5-30 Mbps)
- Secure access delivery of 3D graphics applications over public network not available
- Delivery of the solution via NVIDIA partners only is very limited (ÜSA only)
- No, or very limited mobile mobile client support
- Deliver a stateless environment only, Client templates on VCA are being used



(Source: NVIDIA)

NVIDIA VCA, 3D graphics for virtual desktop

Vendor support for different 3D graphical solutions	Software 3D graphics*	Bare metal Client OS	Pass-through GPU	GPU Virtualization - vGPU	GPU sharing for VDI (API intercept)	GPU sharing for XenApp/RDSH**
NVIDIA VCA	X	X	√	X	X	N/A

11.7 OTOY

OTOY is a Los Angeles-based cloud rendering company. Their aim is to deliver real-time, cinema-quality 3D graphics to filmmakers, 3D designers, animators, game developers, and end users by using the rendering power of the cloud, consisting of clusters of multiple GPUs, and streaming it to the client through a web browser.

OTOY launched the ORBX and OctaneCloud Amazon Machine Images (AMIs) on the AWS Marketplace in early November 2013. The use cases have ranged from HTML5 cloud gaming to high-end, plugin-free application streaming. OTOY wants to provide a dependable, royalty-free platform for delivering next gen video, cloud gaming and app streaming to any HTML5 browser that will make the open web the platform of choice for content creators and developers worldwide. Run any app in a browser -- no plugins required.

Solutions overview

ORBX.js

ORBX.js is a pure JavaScript framework developed in partnership with Mozilla and Autodesk to make the defacto standard for consuming high-performance cloud graphics. ORBX.js enables state of the art 1080p60 cloud streaming to all HTML5 browsers, without the use of plug-ins, browser-specific video codecs, thin client-installations or native code dependencies.

HTML client includes applet to launch native client from JavaScript (on PC/OSX/Linux/iOS) and run the session as a secure fullscreen or windowed process with much stronger encryption than a browser based https/wss stream.

There is a beta HTML5 client for mobile available for Safari iOS, FireFox OS, Android Browsers (Chrome, Opera, Firefox, Maxthon) There is also a HTML 5 desktop beta client for Windows available

ORBX enabled AMIs on Amazon EC2 provide a complete Windows or Linux desktop or workstation replacement in the cloud, capable of installing and running software or services designed for traditional PCs. This means that PC software and games can be easily whitelisted to run remotely from Amazon EC2 without modification, including day and date deployment of games and apps through digital delivery platforms such as Valve's Steam OS or Ubisoft's Uplay.

Octane Cloud Workstation

Octane Cloud Workstation is the world's first turn-key high performance cloud desktop solution specifically designed for streaming high-end remote graphics. Use this AMI to stream a Windows based virtual desktop, hosted in the cloud, to a web browser anywhere in the world. - Access your performance intensive Windows-based applications and data from any device, regardless of your operating system or device performance. - Your applications seamlessly render with clarity in HD due to OTOY's next generation ORBX Video Codec.

Stream your applications using OTOY's native client application (available for Windows, Linux and iOS), or go plugin-free with ORBX.js on any modern browser. The solution includes OTOY WebCL remote graphics driver, the only OpenCL 1.2 GPU runtime for NVIDIA GRID. Demo video.

Highlights:

- ORBX.js delivers 60 hz 'zero client' HD cloud desktops to any HTML5 web browser using pure JavaScript. No plugins required!
- Access to a fully licensed version of Octane Render Cloud Edition - enabling real time GPU rendering on EC2.
- Supports GPU graphics applications (DirectX 9/10/11, OpenGL 4.x, CUDA, OpenCL) on G2 instances. OTOY WDDM Aero virtual desktop resolution up to 2048 x 1536. Validated to work with high performance 2D graphics and media applications on all supported instance types

Host GPU driver support:

- ORBX GPU Desktop Graphics: OTOY WDDM Aero virtual desktop (up to 2048x1536), DirectX 11, OpenGL 4.x, DXGI 1.1+ Application support

- ORBX Render: Built in OTOY Octane Render with integration across all major ADSK apps: 3DS Max, Maya, XSI, AutoCAD, Revit, Inventor.
- ORBX Compute: Built in OTOY OpenCL 1.2 driver/runtime/compiler for NVIDIA Kepler. Built on top of NVIDIA CUDA driver, compiled OpenCL directly to PTX, used for encoding/rendering (3rd party SDK being developed in partnership w/ NVIDIA)

Virtual Machine specifications

- Base OS: Windows, Windows Server 2008 R2 2008R2 x64
- Memory 15 GB
- CPU 22 EC2 Compute Units (8 virtual cores), plus 1 NVIDIA GK104 GPU
- Storage 1 x 60 GB SSD
- Amazon EC2 GPU G2.2XL instance

Host Device I/O support:

- Remote audio driver: speaker output, microphone input
- Remote USB HID device driver (requires clients that support via raw input, could be extended to support 3D Mouse, Wacom tablet, Oculus Rift, Leap Motion, etc.)
- Printer driver: supports remote image based printing and large format printer support
- Xinput driver: analog, gamepad, battery level, rumble pack support
- Clipboard driver: text, images, metafiles, OLE objects used by AutoCAD, Microsoft Office, etc.
- Droptarget process monitor: Drag and drop file and folder support (bi-directional on desktop clients)
- Webcam driver (client->host, not available on all clients, in beta)

11.8 VMWARE HORIZON VIEW

VMware was founded in 1998 and in the early years focused on providing a platform to run Windows- and Linux-based machines virtually. In 1999 this resulted in VMware Workstation. After the

launch of ESX in 2001, VMware made a name in server virtualization and quickly became the market leader in this segment.

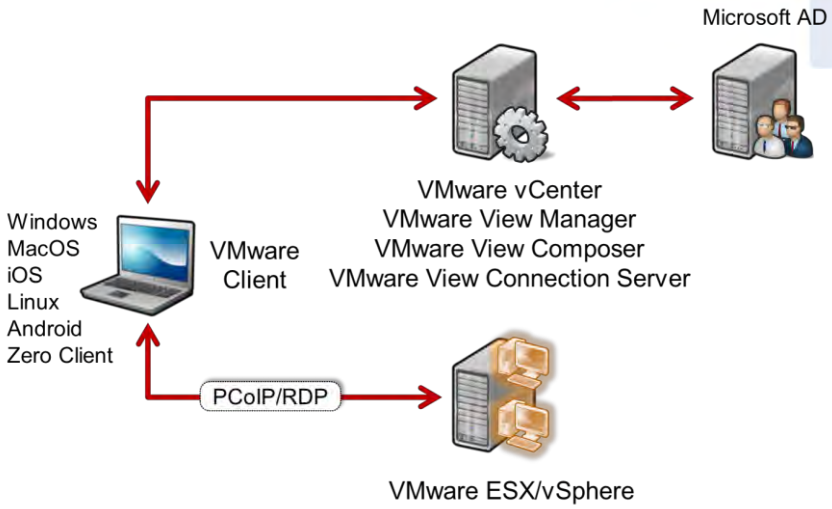
In May 2007, VMware acquired UK-based ISV Propero. In early 2008 VMware used some of the Propero software assets to enter the VDI market with the launch of VMware VDM 2.0. VMware VDM gave users a centrally hosted desktop. The name of VDM evolved into a new product suite: VMware View.

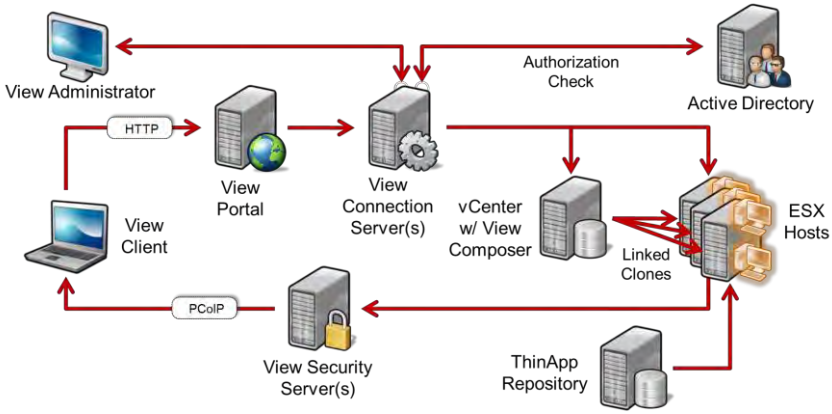
In 2009 VMware partnered with Teradici for their display protocol PC-over-IP (PCoIP) which was specifically developed for server hosted (virtual) desktops and Blade PCs. This partnership allowed VMware to make a change to the remote display capacity that resulted in a major enhancement to the user experience. PCoIP technology is delivered in both hardware and software implementations. VMware View 4.x and above are based on the software version of PCoIP.

It is important to note that the PCoIP remoting protocol is only responsible for the graphics portion of remote desktops. Other functionality, such as MMR, Unified Communications, Webcam Redirection, USB, were added by the VMware Horizon View team..

Architecture

VMware View's architecture is shown below. Apart from the VMware Virtual infrastructure, you only need the View Connection Server to be up and running. To use the linked-clones technology, the VMware View Composer is installed on the VMware vCenter server. With this feature it is possible to use one (snapshot of a) golden image to deploy virtual desktop VMs and save disk space because the VMs use the same golden image and an additional Delta file of changes. The protocols used with VMware View 4.x and above are RDP and PCoIP.





(VMware Horizon View overview)

Licensing

Horizon View includes Persona Management (User Profile Management), View Composer (linked clones, single disk-image provisioning), View Client with Local Mode, VMware vShield Endpoint (antivirus component) and VMware ThinApp (Application Virtualization). There are two packages available to Horizon View: Bundle and Add-on. Horizon View Bundle includes all components

needed for the virtual desktop infrastructure: vSphere Desktop, vCenter Server Standard Desktop, View Manager 5 and related edition components. Horizon View Add-On only includes desktop components and customers must already have vSphere in their environment or purchase vSphere Desktop separately. Horizon View is also available as part of the Horizon Suite, which includes Horizon Workspace and Horizon Mirage.

VMware Horizon View, 3D graphics for virtual desktop

VMware offers various solutions for 3D graphics.

The name of 'software 3D graphics' for VMware is: SVGA, Soft3D and VMware SVGA 3D. It's a VMware Windows Display Driver installed in the Windows guest OS.

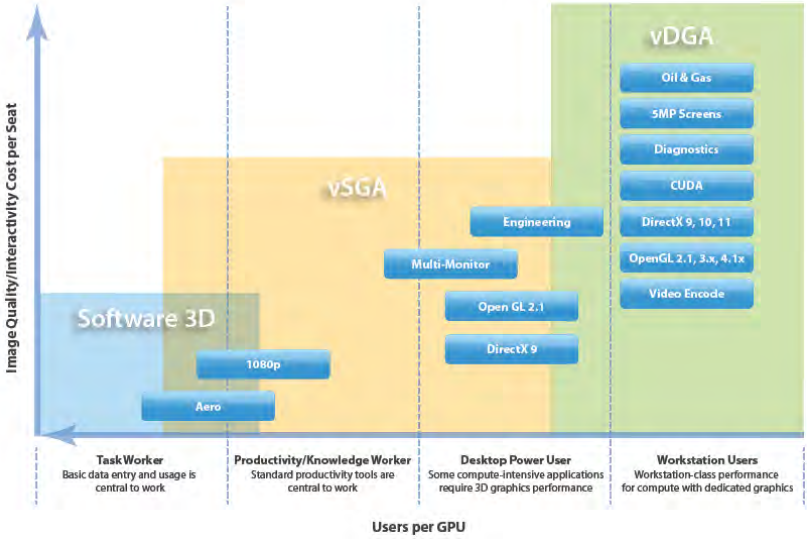
Pass-through GPU is VMware vDGA, Virtual Dedicated Graphics Acceleration. GPU access provided by a real NVidia driver where high performance 3D graphics is needed.

GPU sharing for VDI is VMware vSGA, Virtual Shared Graphics Acceleration. Multiple virtual machines leverage physical GPUs installed in the ESXi host and provide hardware accelerated 3D graphics to virtual machines. In each virtual machine a VMware display driver is installed that provides emulated access to the underlying GPU hardware. Applications designed or certified for NVIDIA or AMD GPUs need to be recertified against the VMware driver. vSGA is limited to DirectX 9.0c and openGL 2.1.

VMware's Hardware accelerated graphics for Desktop Virtualization timeline

- 2008: VMware supports Teradici with view Software/Hardware1:1
- 2009: Availability of PCI Passthrough
- 2011: VMware View 5.0 support for soft3D with CPU rendering
- 2012: Support migration of VM's with vSGA. vDGA in vSphere 5.1 and introduction to vSGA in Horizon View 5.1
- 2013: vSGA vSphere 5.5 support for AMD GPUs, vSphere 5.5 support for vDGA, GPU Pass-through support for Horizon View 5.3 and Citrix XenDesktop/XenApp.

- 2014: VMware announced vGPU support for VMware vSphere and Horizon View at NVIDIA GTC 2014.. Release date 2015
- vDGA is released and supported in vSphere 5.1+
- VMware Horizon View 5.3+ supports vDGA with all versions of OpenGL and DirectX
- VMware supports Citrix XenDesktop and XenApp with vDGA on vSphere 5.5
- vSGA in vSphere 5.5 supports AMD GPUs



(Source: VMware)

VMware Horizon View, 3D graphics for virtual desktop

Vendor support for different 3D graphical solutions	Software 3D graphics*	Bare metal Client OS	Pass-through GPU	GPU Virtualization - vGPU	GPU sharing for VDI (API intercept)	GPU sharing for XenApp/RDSH**
VMware Horizon View	√	√	√	X	√	N/A

11.9 3D GRAPHICS FOR VIRTUAL DESKTOPS VENDOR SOLUTIONS AT A GLANCE

Vendor support for different 3D graphical solutions	Software 3D graphics*	Bare metal Client OS	Pass-through GPU	GPU Virtualization - vGPU	GPU sharing for VDI (API intercept)	GPU sharing for XenApp/RDSH**
Citrix XenApp – Server OS (RDSH)	√	N/A	√	√	N/A	√
Citrix XenDesktop – Client OS (VDI)	√	√	√	√	X	N/A
NICE DCV	X	√	√	√	√	N/A
NVIDIA VCA	X	X	√	X	X	N/A
Microsoft RDSH – Server OS (RDSH)	√	X	X	X	N/A	√**
Microsoft RDVH - Client OS (VDI)	√	X	X	X	√	N/A
VMware Horizon View	√	√	√	X	√	N/A

The diagram below gives a complete overview of the 3D graphical capabilities of the different 3D graphics for virtual desktops solutions.

*) Level of capabilities dependent on OS and 3D APIs

**) Only BareMetal GPU on RDSH brings 3D API support. No GPU Passthrough in Hyper-V Virtual Machine scenario. Rendering content leverages GPU capabilities. RDP/RemoteFX Capture and Compress cannot take advantage of GPU.

12. 3D GRAPHICS FOR VIRTUAL DESKTOPS ENABLING TECHNOLOGY - GPU, CPU

12.1 AMD

Founded in 1969, AMD designs and integrate technology that powers intelligent devices.

AMD acquired ATI Technologies Inc in 2006. ATI was a semiconductor technology corporation based in Markham, Ontario, Canada, that specialized in the development of graphics processing units and chipsets. The acquisition of ATI in 2006 was important to AMD's continual development of GPU related technology and system technology like HSA (Heterogeneous System Architecture). Since 2010, the ATI brand name is no longer used for any of AMD's graphics processor products. The branded GPU product ranges are Radeon for consumer, and Firepro, for professional graphics. AMD combined system (CPU+GPU) technology drives the leading gaming console solutions. AMD believes that providing pixel rendering and compute via GPU are essential for delivering high quality 3D, multimedia and high performance compute experiences on next generation cloud based services.

SKY Enabled solutions from AMD deliver graphics rendering solutions today for:

- Enterprise (On Premises VDI and Remote Graphics for application or desktop delivery)
- Cloud Service Providers (Gaming, Consumer, DaaS and 3D/Pixel Rendering dependent services)

STREAM Enabled Solutions from AMD deliver solutions for Compute leveraging a vendor neutral industry standard programming environment, OpenCL. OpenCL is a primary feature of the latest Apple MacPro systems that utilized FirePro graphics.

AMD supports the S series GPU boards for both SKY and STREAM use cases. The same boards for VDI can be used for compute.

Radeon Series

ATI launched the Radeon line in 2000 as their consumer 3D accelerator add-in cards, its flagship product line and the direct competitor to Nvidia's GeForce. Mobility Radeon is a series of power optimized versions of Radeon graphics chips for use in laptops.

AMD CrossFireX was ATI's response to NVIDIA's SLI platform. It allowed, by using a secondary video card and a dual PCI-E motherboard based on an ATI Crossfire-compatible chipset, the ability to combine the power of the two, three or four video cards to increase performance through a variety of different rendering options. There is an option for additional PCI-E video card plugging into the third PCI-E slot for gaming physics, or another option to do physics on the second video card.

FirePro – FirePro – Launched in 2001, following ATI's acquisition of FireGL Graphics from Diamond Multimedia. Workstation CAD/CAM video card, based on the Radeon series. Since that time the software and hardware development changed dramatically with latest generation FirePro products providing modern drivers with stability and real world application performance in mind. With SPECviewperf 12, AMD FirePro products are some of the higher performing solutions across the professional graphics market.

The role of AMDs in 3D graphics for virtual desktops

- Driving Knowledge Worker and High Performance Workstation Virtualized Solutions with VMware.
- Engaged with Citrix on Application and Session delivery for remote users
- Engaged with Microsoft RemoteFX/Hyper-V product teams, AMD FirePro is the best performing GPU for Microsoft RemoteFX

AMD Firepro and VMware Horizon View, shared graphics with vSGA.

- AMD GPU is attached to ESXi with the AMD GPU driver, VMware software virtualizes the GPU.

- Uses VMware Graphics driver in each virtual machine providing DirectX 9 and Software OpenGL. Leverages shared S Series GPU for rendering DirectX 9
- S Series GPU's supported for shared graphics, Ideal for Knowledge Worker Segment
- vSGA Driver in Pre-release now
    - Certification expected by end of March
    - vSGA driver is certified by GPU vendor, no OEM/SI Certification needed.
- Ideal Use Cases
    - 3D Viewers/Model –Component Viewing
    - GIS/Mapping Users/Viewers
    - Education/Training Environments
    - Enterprise Power Users (Exec/Management)

AMD Firepro and VMware Horizon View, Direct Graphics with vDGA

- AMD GPU's direct mapped to individual virtual machines on ESXi
- Uses AMD FirePro Driver to ensure full performance and workstation ISV support
  along with full OpenCL support
- FirePro S and W Series products supported
- VMware and AMD developing enhanced support for vDGA over the next few months.
- Ideal Use Cases
    - MCAD/CAE/PLM Midrange Users
    - Power GIS Users
    - Application Classroom/Training
    - Content Creation

AMD Firepro and Microsoft RemoteFX with Hyper-V

- AMD GPU Attached to Windows Server with AMD driver, Microsoft virtualizes the GPU
- Scalable support for remote VDI users with access to DirectX/Direct3D applications and content
- S Series GPU's supported for shared graphics
- Ideal for Knowledge Worker Segment
- Ideal Use Cases

- 3D Viewers/Model – Component Viewing (DX)
- GIS/Mapping Users/Viewers (DX)
- Education/Training Environments
- Enterprise Power Users (Exec/Management)

AMD Firepro and Citrix XenApp Session Virtualization

- Direct Connected GPU to Windows Server
  or Direct Pass Thru via Hypervisor provides GPU support to Windows Server session running XenApp or XenDesktop Session Services
- Supports any existing content that can be delivered with HDX 3D
- Ideal for "packaging" applications for delivery to mobility users needing access to
  content for viewing/validation
- S Series GPU's supported
- Ideal Use Cases
  - Remote Workstation Users
  - Tablet Delivery for Factory Floor CAD access
  - Packaging of applications to users needing dataset/CAD viewing only

AMD FirePro solutions

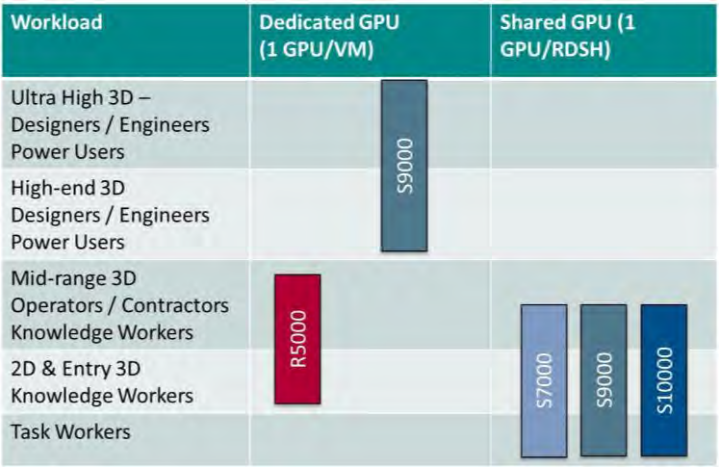
AMD FirePro R5000 - Remote Workstation

The AMD FirePro W5000 class GPU is combined with a Teradici PCoIP Hardware encoder for highest performance.

- 2 x Local Mini DisplayPorts
- 2 x Remote Display 2560x1600 or 4 x Remote display 1920x1200
- All PCoverIP transmissions are encrypted, only pixels transmitted
- Benefits:
  - GPU + Teradici Host card combined
  - Just 1 PCIe slot instead of 2
  - No External cabling
  - Local monitor support

AMD FirePro S10000, S9000, S7000 and R5000

	FirePro S10000	FirePro S9000	FirePro S7000	FirePro R5000
GPU	Tahiti x2	Tahiti	Pitcairn	Pitcairn
Stream Processors	3584	1792	1280	1280
Memory Size / type	6GB GDDR5 (3GB/GPU)	6GB GDDR5	4GB GDDR5	2GB GDDR5
Memory interfaces	384-bit	384-bit	256-bit	356-bit
Memory bandwidth	480 GB/s	264 GB/s	154 GB/s	102 GB/s
PCI express	3.0	3.0	3.0	3.0
ECC Memory	Internal and External	Internal and External	n/a	
Form Factor	FH/FL Dual Slot	FH/FL Dual Slot	FH/FL Single Slot	FH/FL Single Slot
Core Clock Speed	825 MHz	900 MHz	950 MHz	825 MHz
Max Power	375W	225W	150W	150W
Peak Double Precision	1.48 TFLOPS	806 GFLOPS	152 GFLOPS	
Peak Single Precision	5.91 TFLOPS	3.23 TFLOPS	2.4 TFLOPS	
Teradici Processor	n/a	n/a	n/a	Tera 2240
Notes	Workstation chassis solution			



Link to AMD FirePro Server OEMs

http://www.amd.com/us/products/workstation/graphics/firepro-remote-graphics/Pages/dell-and-amd-firepro.aspx

12.2 INTEL

Intel VT

Intel Virtualization Technology (Intel VT) provides hardware support that simplifies processor and platform virtualization, enabling reductions in virtual machine monitor (VMM) complexity and providing considerable improvements in performance over software-only VM hosting. VT is an "umbrella" term and the overall capabilities are often categorized as VT-x, which refers to processor capabilities, VT-d relating to chipset capabilities and VT-c referring to network and connectivity capabilities. Note that VT-d and VT-c are not dependent on VT-x. For example a VT-x enabled system can operate without VT-d, or without VT-d enabled or configured. You simply miss the benefits of the feature.

Intel's family of Xeon server processors provides support for hardware-based technologies enabling Desktop and Applications virtualization and security. The following section of the will cover specifically the following technologies: Intel VT, VT-x, VT-d, and VMCS Shadowing

VTX-support;

- A new CPU operating mode - VT Root Operation. An O/S expects to be able to run at the most privileged CPU ring, ring 0. However, when a VMM is running it has to "de-privilege" the guest OSs so they can simulate running at ring 0 which introduces considerable complexity and CPU overhead. With VT-x the VMM can now run in VT Root Mode and guest VMs can run at Ring 0 as they expect.
- Virtual processor IDs (VPID) to reduce the overhead of the VMM managing transitions,
- Extended page table (EPT) allow the VMM to efficiently and securely manage memory for each guest VM.
- Technically, VT-d is the term for a collection of Intel chipset technologies that enhance the performance of the I/O subsystem. One key problem is how to isolate device access so that one resource cannot access a device being managed by another resource. Some of the key VT-d capabilities are;
- Direct Device Attach (DDA) commonly referred to as "pass-through" where a particular guest VM can have an I/O device (such as the display) assigned. Performance improves

and all of the device capability (such as 3D rendering) is available to the guest.

- Direct Memory Address (DMA) and Interrupt remapping. Supports address translations for device DMA data transfers and VM routing and isolation of device interrupts. VT-d also improves security against DMA attacks.
- More information about Hardware-Assisted Virtualization technology can be found here and here

Intel VMCS Shadowing Technology

Intel VMCS Shadowing reduces the frequency with which the guest VMM must access the root VMM in a nested environment. This technology is needed because new usage models are emerging that would require two or more Virtual Machine Monitors (VMMs) to be hosted on the same client system, both physical and virtual. With Intel VMCS Shadowing, the root VMM is able to define a shadow VMCS in hardware. A guest VMM can access this shadow VMCS directly, without interrupting the root VMM. Since the shadow VMCS is implemented in hardware, required accesses can be completed nearly as fast as in a non-nested environment.

Security solutions such as Bromium and Intel DeepSAFE can leverage Intel VMCS delivering secure solutions inside a 3D graphics for virtual desktops environment.A nice blog post about Bromium inside VDI can be found here.

Intel GPU (GVT)

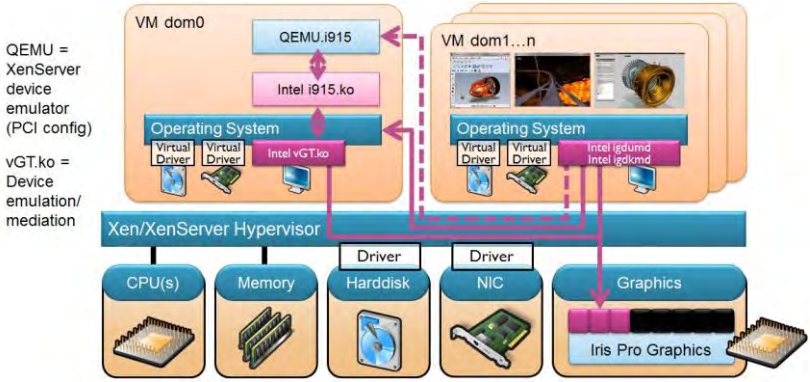
Intel offers integrated Processor Graphics technology in their Xeon E3 line of CPUs. A variety of processor configurations are available based on the GT2 and GT3e graphics components.

In September 2013, Intel announced their Haswell CPUs with four models of integrated GPUs. The high-end model is called GT3e or Iris Pro with 40 execution units, which is equivalent to 40 GPU cores. In addition to the 40 GPU cores, the Iris Pro model comes with 128 MB of embedded DRAM acting as a Level 4 cache. It is shared dynamically between the GPU and CPU, allowing for high data transfer rates.

Right after the Haswell announcement, Intel also announced XenGT (recently renamed to Intel Graphics Virtualization Technology or Intel GVT) as a GPU virtualization solution with mediated graphics

pass-through. As the name indicates, Intel GVT combines Xen hypervisor and Haswell Iris Pro GPU cores (GT3e). A virtual GPU is assigned to a guest VM, with resource assignment brokered and managed by a "mediator" component running in the Dom0 session. Controlled by the mediator, Intel's native graphics driver runs inside each guest VM, communicating directly with the assigned GPU cores and preventing hypervisor intervention in performance critical paths. On the hardware side, Haswell's Iris Graphics is Intel's main focus regarding hardware-accelerated graphics remoting along with next-generation Broadwell processors.

Intel GVT-g – GPU Virtualization



12.3 NVIDIA

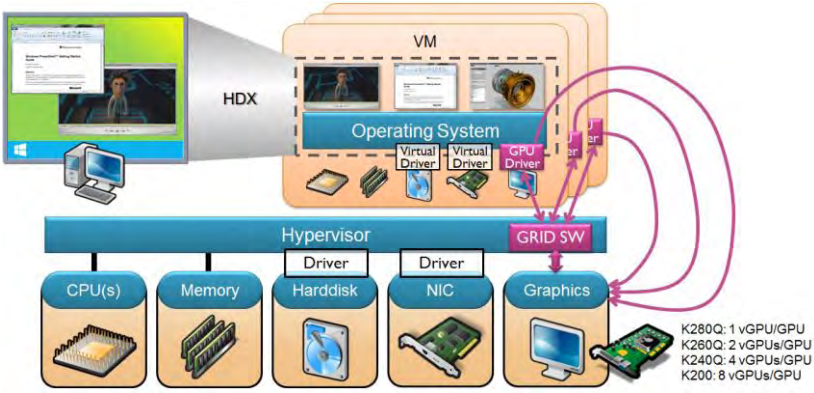
Nvidia is an American global technology company based in Santa Clara, California. Nvidia manufactures graphics processing units (GPUs), as well as having a significant stake in manufacture of system-on-a-chip units (SOCs) for the mobile computing market. Nvidia also joined the gaming industry with its handheld Nvidia Shield, as well as the tablet market with the Tegra Note. In addition to GPU manufacturing, Nvidia provides parallel processing capabilities to researchers and scientists that allow them to efficiently run high-performance applications. They are deployed in supercomputing sites around the world. More recently, Nvidia has moved into the mobile computing market, where it produces Tegra mobile processors for smartphones and tablets, as well as vehicle

infotainment systems. In addition to Advanced Micro Devices, its other competitors include Intel and Qualcomm.

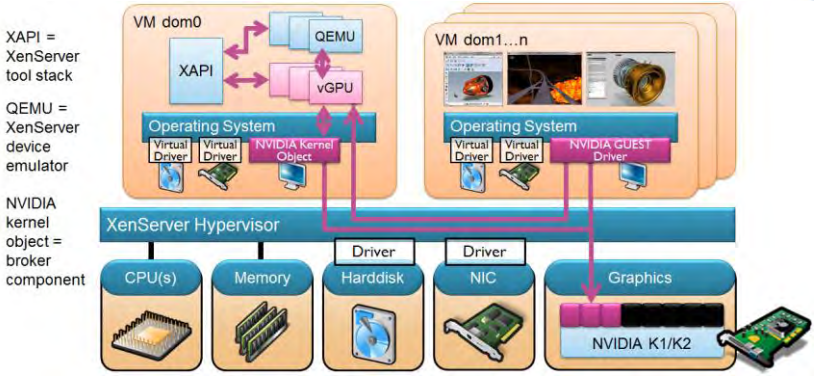
The Nvidia GRID portfolio of technologies leverages the power of the GPU and graphics applications to deliver GPU-accelerated applications and games over the network to any user. Nvidia GRID GPUs are based on the Nvidia Kepler GPU architecture. Nvidia GRID software is a complete stack of GPU virtualization, remoting and session-management libraries that allows multiple users to experience graphics-intensive desktops, applications and games using GPUs.

Nvidia GRID technology offers the ability to offload graphics processing from the CPU to the GPU in 3D graphics for virtual desktop environments, allowing the data center manager to deliver true PC graphics-rich experiences to more users for the first time.

NVIDIA vGPU – GPU Virtualization



NVIDIA/Citrix XenServer vGPU – GPU Virtualization



Nvidia GRID is available in two models:

- The GRID K1 card is for general purpose VDI sessions. It has four Nvidia Kepler GPUs (consisting of 768 CUDA cores and 16GB of DDR3 RAM), and Nvidia figures it's good for 100 VDI sessions on the host. The GRID K1 list price is about 2500$. The Server OEM sets the pricing.
- The GRID K2 card is a special purpose card designed for graphics-intensive environments. It only has 8GB of GDDR5 RAM and two GPUs (instead of four), though they're higher end with 3072 CUDA cores. The GRID K2 list price is about 4500$. The Server OEM sets the pricing.

The first implementation of the GRID vGPU technology will support up to 8 users per physical GPU, which equals 32 users on the quad GPU GRID K1 board (Servers typically support two boards, so 64 users per server). Future versions of vGPU will support higher scalability with just a software update. Of course scalability depends on application usage.

NVIDIA GRID Graphics Board	vGPU Profile	Graphics Memory	CUDA core/GPU	Max Displays Per User	Max Resolution Per Display	No. of vGPUs per GPU	Max Users Per Graphics Board	User Classification
GRID K2 (2 physical GPUs)	K280 (Pass-through)	4,096 MB	1542	4	2560x1600	1	2	Designers/Engineers–Power Users
	K260Q	2,048 MB	786	4	2560x1600	2	4	Designers/Engineers–Power Users
	K240Q	1,024 MB	384	2	2560x1600	4	8	Designer/Power User
	K220Q	512 MB	96	2	2560x1600	8	16	Operators/Contractors-Knowledge Workers
	K200	256 MB	96	2	1900x1200	8	16	Operators/Contractors-Knowledge Workers
GRID K1 (4 physical GPUs)	Pass-through	4,096 MB	192	2	2560x1600	4	4	Designer/Power User
	K140Q	1,024 MB	48	2	2560x1600	4	16	Operators/Contractors-Knowledge Workers
	K120Q	512 MB	24	2	2560x1600	8	32	Task workers
	K100	256 MB	24	2	1900x1200	8	32	Task workers

The Q in e.g. K260Q refers to Quadro. The higher vGPU profiles go through Quadro certification from an ISV perspective.

NVIDIA GRID K1 or K2?

If the 3D graphics application uses more GPU memory then the GRID K1 with 1GB memory and 786 CUDA core is the preferred choice. If the 3D graphics application is consuming more GPU power compared to GPU memory then the NVIDIA GRID K2 with 8GB memory and 3072 CUDA cores is the preferred choice. Keep in mind when using GPU passthrough that the K1 has 4 GPU's on one board while the K2 has 2 more powerful GPU's on one interface. NOTE: You can use Process Explorer v15.3 (or later) or GPU-Z to get an approximate understanding of how much GPU core processing power or frame buffer is consumed by an application in order to help you determine where to best place the application/user for vGPU profile or which GRID board to use. Simply run one of these utilities on physical workstation before you begin your complex 3D application and take note of the current GPU % utilization and GPU system memory before you begin your complex application. Then during

the execution of the complex graphical application you can want the allocation of memory and GPU core utilization. This will help you decide on whether the application is more bound by GPU core compute or GPU memory allocation.

The performance of one GRID K1 GPU in pass-through mode is comparable with a K600 with more VRAM. The performance of one GRID K2 GPU in pass-through mode is comparable with a K5000.

NVIDIA GRID and Quadro GPU specification

	GRID K1	GRID K2	Quadro K6000	Quadro K5000	Quadro K4000
GPU	4 Kepler GPU (K600-class	2 high end kepler GPUs (K5000-class)	Kepler	Kepler	Kepler
GPU Clock	850 MHz	745 MHz	902 MHz	706 MHz	810 MHz
CUDA cores	786 (192/GPU	3072 (1536/GPU)	2880	1536	786
Memory size	16GB GDDR3 (4GB/GPU)	8GB GDDR5 (4GB/GPU)	12GB GDDR5	4GB GDDR5	3GB GDDR5
Memory interface	128-bit	256-bit x2	384-bit	256-bit	192-bit
Memory bandwidth	Unknown	160GBs x2	288GBs	173GBs	134GBs
Max Power	130W	225W	225W	122W	80W
Form Factor	Dual Slot ATX, 10.5	Dual Slot ATX, 10.5	Double-width	Double-width	Single-width
Aux power req.	6 pin	8 pin	2 x 6 pin	6 pin	6 pin
PCIe	x16	x16	x16	x16	x16
PCIe generation	Gen3 (Gen2 compatible)	Gen3 (Gen2 compatible)	Gen3	Gen2	Gen2
Cooling solution	Passive	Passive	Active	Active	Active
# Max. users	4-100	2-64	1	1	1
OpenGL	4.4	4.4	4.4	4.4	4.4
OpenCL	1.1	1.1	1.1	1.1	1.1
DirectX	9/10/11	9/10/11	9/10/11	9/10/11	9/10/11
Shader model	5.0	5.0	5.0	5.0	5.0
Price (approx.)	4140USD	5199USD	5265USD	2499USD	1269 USD

GRID enabled OEM platforms

It's key that the combination of server hardware and software is supported.

The combination of server hardware with NVIDIA GRID, the amount of required GRID card, amount of network,- and storage controller interfaces, teradici PCoverIP server offload card or PCIe flash based storage controllers such as FusionIO needs to be supported by hardware vendors such as HP or Cisco. This is including the virtual desktop including hypervisor software. A full list of GRID certified servers can be found here: http://www.nvidia.com/object/enterprise-virtualization-where-to-buy.html

13. 3D GRAPHICS FOR VIRTUAL DESKTOPS ENABLING TECHNOLOGY - IAAS

13.1 AMAZON EC2 - GPU INSTANCES

Amazon provides an EC2 GPU instance, specific G2 and CG1, which provides high parallel processing capability via NVIDIA GPUs.

Many scientific, engineering, and rendering applications can leveraging the Compute Unified Device Architecture (CUDA) or OpenCL parallel computing frameworks via Amazon EC2 GPU instance. Also 3D graphics for virtual desktops, game streaming, 3-D application streaming, and other graphics workloads can run on EC2 GPU instance.

The G2 instances provide access to NVIDIA GRID GPUs ("Kepler" GK104) each with 1,536 CUDA cores and 4GB of video memory. The initial NVIDIA driver release provides support for OpenGL 4.3, DirectX 9, 10, and 11, CUDA 5.5, OpenCL 1.1, and GRID SDK.

CG1 instances provide access to NVIDIA Tesla M2050 GPUs ("Fermi" GF100) each with 448 CUDA cores and 3GB of video memory. The latest driver release provides support for CUDA 5.5, OpenCL 1.1, and DirectCompute.

The EC2 GPU instances run as Hardware Virtual Machines (HVM-based instances) it uses hardware-assist technology provided by the AWS platform.

With HVM virtualization, the guest VM runs as if it were on a native hardware platform, except that it still uses paravirtual (PV) network and storage drivers for improved performance.

This enables Amazon EC2 to provide dedicated access to one or more discrete GPUs in each GPU instance. GPU instances can be clustered and placed into a cluster placement group. Cluster placement groups provide low latency and high-bandwidth connectivity between the instances within a single Availability Zone.

NVIDIA provides Amazon Machine Images (AMIs) for GPU instances for Amazon Linux and Windows. These reference AMIs include the NVIDIA driver, which enables full functionality and performance of the NVIDIA GPUs. Keep in mind with Amazon EC2 GPU instances

- Aren't available in every region.
- Must be launched from HVM AMIs.
- Can't access the GPU unless the NVIDIA drivers are installed.
- Aren't available for use with Amazon DevPay.

A complete overview of of the Amazon EC2 instances can be found here.

14. 3D GRAPHICS FOR VIRTUAL DESKTOPS ENABLING TECHNOLOGY - REMOTE GRAPHICS

14.1 HP RGS - REMOTE GRAPHICS SOFTWARE

Introduction

Remote workstations are breaking free of network limitations with HP (RGS). HP Remote Graphics Software (RGS) is the collaboration and remote desktop solutions for serious virtual desktop or workstation users and their most demanding applications. All applications run natively on the remote workstation or inside virtual desktops and take full advantage of its graphics resources. The desktop of the remote workstation is transmitted over a standard network to a window on a local computer using advanced image compression technology specifically designed for digital imagery, text and high frame rate video applications. A local keyboard and mouse are supported as well as redirection of most USB devices to provide an interactive, high performance workstation experience.

With the just released HP RGS 7, HP is defining the future of mobile access to workstation applications. HP Remote Graphics Software (RGS) 7.0 is the first solution delivering workstation-class productivity from Windows 8 tablets by leveraging new touch controls along with HP's industry-leading software.

With RGS 7.0, HP turns swipes into hot keys, gestures into control and pinches into zooms, along with a precision onscreen mouse that uses the entire screen as a trackpad. These enhancements increase productivity for remote users of touch PCs or Windows tablets, including the HP ElitePad.

HP RGS opens up a completely different way for workstation users to get the job done, enabling real-time collaboration between multiple users simultaneously and providing remote access to a workstation with little to no loss of performance. RGS enables animators, designers and engineers to work and collaborate more efficiently

across geographies by connecting them to their high powered workstations, where ever they may be.

New with HP RGS 7.0, users can:

- Replace the computer mouse with simple taps, presses and swipes
- Replace the keyboard by mapping gestures to the hot keys most used
- Enable quick zoom in and zoom out on the touch-enabled tablet
- Increase the responsiveness of remote sessions over long distances with network issues using the upgraded HP Velocity 2.1

HP RGS and Windows 8 Tablets.

The lack of a keyboard, mouse, screen real-estate and a good network connection can limit a tablet's usefulness for workstation class applications. HP RGS 7 gives you the power to overcome these obstacles.

- No mouse, no problem. Replace the mouse with simple taps, presses and swipes.
- Fat fingers, no problem. Tab with precision with the HP Virtual Mouse. Use the entire screen as a trackpad for precise control of the onscreen mouse icon
- No keyboard, no problem. Map gestures to the hot keys you use most. Easily program your Windows 7 applications to recognize Windows 8 gestures.
- Small screen, no problem. Zoom in to use small menus. Zoom out to see multiple displays on a tablet screen
- Poor network connection, no problem. The network is the experience; fix it with HP Velocity Extreme. Increase the responsiveness of your remote session over long distances with network issues.

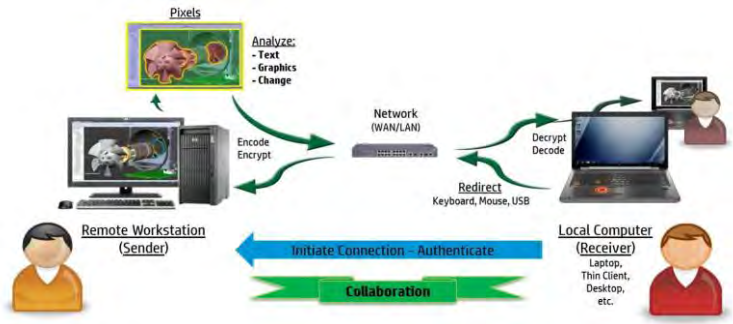
Unique functionality for RGS 7.0

- 10+ display support
- Linux applications and Linux sender
- Full interactive collaboration
- Build in WAN optimization, via HP Velocity
- > 30 fps

- Sender GPU or CPU encoding. With GPU encoding NVIDIA Quadro 2000+ required
- Mathematical and visual lossless
- 256 encryption
- OpenGL and DirectX 3D graphics support
- Connection broker support for Leostream., Dell/Quest, Mechdyne

Architecture

The HP Remote Graphics Software architecture:



(Source: HP)

Sender and Receiver support

The following platforms for sender and receiver are supported

	Windows 7 Professional, Enterprise	Windows 8.1 Pro, Enterprise	RHEL V5.9, V6.5	SLED 11.3
Sender:				
HP Z Workstations	✓	✓	✓	✓
HP WS460c (Gen6)	64-bit Bare Metal & Citrix XenServer v6		64-bit Bare Metal & Citrix XenServer v6	
HP WS460c (Gen8)				
HP SL390 (Gen7)				
HP SL250 (Gen8)				
HP DL380p (Gen8)				
HP Moonshot M700	✓			
Receiver	✓	✓	✓	✓
Thin Client Receiver	**Windows Embedded**		**Embedded Linux**	
HP t610	WES7, WES8		HP ThinPro 4.3, 4.4	
HP t620	WES7, WES8		HP ThinPro 4.4	
HP t820	WES7, WES8			
HP mt40/41	WES7			

(Source: HP)

Licensing

One HP RGS 7 license is required for each RGS sender system.HP RGS Floating license can be used where licenses are "checked out" on a concurrent use basis up to the number of licenses purchased. In this scenario the use of the license server is required.

With HP RGS there are no monthly fees, the Receiver is a free download, and the Sender is free on HP Z Workstations..

14.2 TERADICI

Teradici is a privately held software company and was founded in 2004. It has its headquartered in British Columbia and Santa Clara, CA. Between 2004 and 2007, PCoverIP is developed. PCoverIP is a Remote Display protocol available in hardware silicon and software. It's purpose is delivering applications and desktops to PCoverIP enabled endpoints. These endpoint can be equipped with a PCoverIP interface, such as a Zero client, or they can be software based. In 2007, the first generation PCoverIP processors and workstation products were shipped. In 2008, a strategic licensing and co-development with VMware produces a rich, remote desktop experience. VMware View 4.0 launched with PCoverIP features. In 2011, the Teradici hardware accelerator for VMware View is offered under the name APEX 2800. In 2012, the next generation PCoverIP takes off named TERA2. In 2013, Amazon selected Teradici PCoverIP for Amazon Workspace services, a Desktop as a Service offering.

PCoverIP

PCoverIP is the foundation for VMware Horizon View and Teradici Remote Workstation solutions. It is an UDP based protocol that is host rendered, multi-codec and dynamically adaptive. Images rendered on the server are captured as pixels, compressed and encoded and then sent to the client for decryption and decompression.

Depending on the image, different codecs are used to encode the pixels sent since techniques to compress video images differ in effectiveness compared to those for text.

The protocol also dynamically adapts its encoding based on the available bandwidth. In low bandwidth environments it utilizes lossy

compression where a highly compressed image is quickly delivered, followed by additional data to refine that image, a process termed "build to perceptually lossless".

Benefits

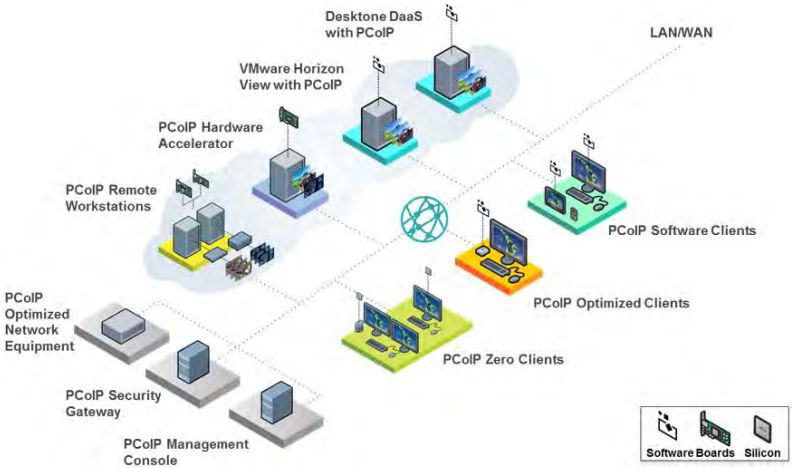
- Host rendering: Provides a rich user experience
- Optimized multi-codec: optimized bandwidth and image quality
- Dynamic Network Adaptation; automatically delivers best possible user experience under changing network conditions
- Only encrypted pixels are transmitted: data stays secure

Solutions

- PCoverIP ZeroClient System on a Chip (SOC); Is a System on a Chip for OEM's to use zeroclients with Tera2 silicon, which implement the PCoverIP remote display protocol. The SOC is delivered via partners and OEM products.
- PCoverIP workstation interface. It's a host card for 1:1 connection to a powerful (blade) workstation. This means one host system to one remote user, it is not virtualized or shared and can capture the output from a GPU for full HD remoting along with redirecting audio and USB peripherals. The endpoint can be a PCoverIP zeroclient or a software client.
- PCoIP software clients: Software implementation of the PCoIP protocol for flexible client device support. Select OEM's include this in their products for x86 and ARM based thin clients.
- Teradici APEX 2800 server offload card: A PCIe coprocessor board to offload the compression and encryption of graphics and audio to and from PCoIP based clients.
- PCoIP management console: A web-based management tool for administrative control of PCoIP devices from a central console.

Architecture

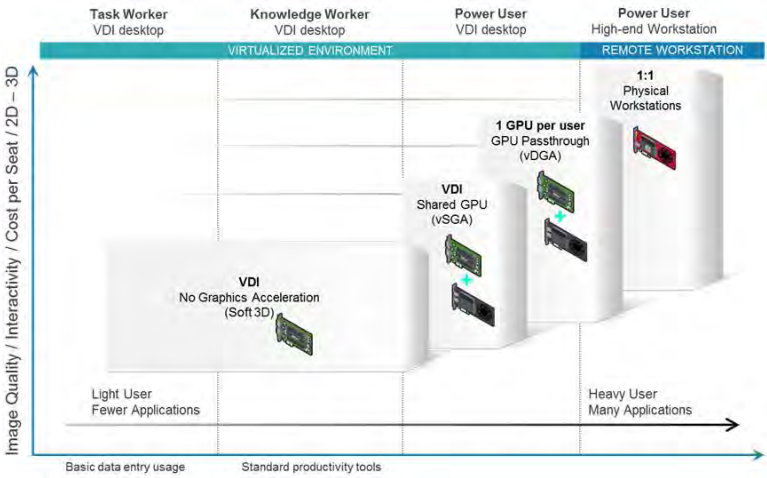
A complete overview of the PCoverIP hardware and software solutions:



(Source: Teradici)

Supporting 3D graphics in VMware Horizon View and workstation scenarios

How does PCoverIP support various 3D graphics for virtual desktop scenarios:



(Source: Teradici)

15. 3D GRAPHICS FOR VIRTUAL DESKTOPS ENABLING TECHNOLOGY - 3RD PARTY SOFTWARE SOLUTIONS

15.1 LAKESIDE SOFTWARE SYSTRACK

The first step toward leveraging hardware accelerated graphics for desktop virtualization is understanding the performance demands of applications in conjunction with user interaction.

SysTrack enables the detailed analysis of graphical demands to help build a portfolio that can be analyzed quickly to identify which applications would benefit the most from hardware accelerated graphics for desktop virtualization.

By sizing the supporting systems appropriately based on the actual application behavior observed in the environment it becomes much easier to guarantee optimal end-user experience. Through the continual capture of key performance aspects like compute and GPU resource consumption, memory, disk usage, and network utilization a complete, time correlated picture of the expected load in an environment can be created.

Once the solution has been fully developed and deployed, it's critical to keep track of the end-user experience to ensure that everything is working as anticipated.

This is another area where SysTrack can help, both in capturing the initial end-user experience metrics to benchmark the success of a transformation project and in steady state in the new environment to ensure continued quality.

Using proprietary API access provided by NVIDIA, the quality of the end-user experience can now be even more completely characterized with GPU utilization. This makes it much easier to pinpoint where oversubscription of resources may be causing issues, and guarantees a basic quality of service to keep productivity at peak levels.

Application Graphics Benchmarking

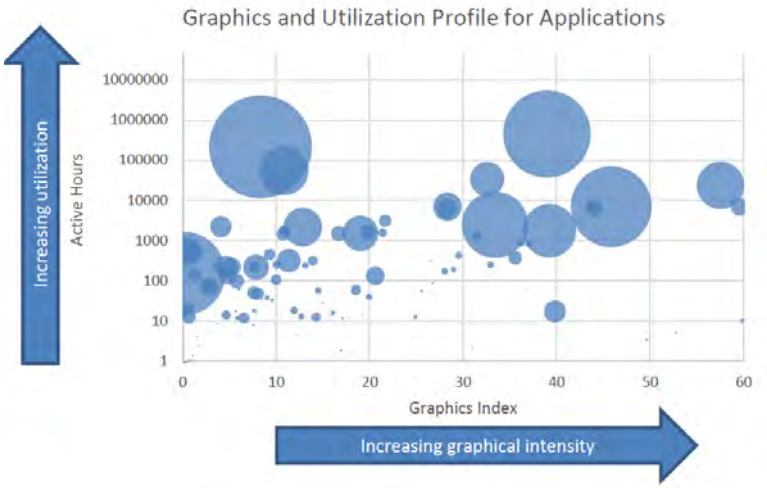
SysTrack continually collects information about software packages as they're used and normalizes all data points for cross platform comparison. One of the key performance parameters that's identified in this process is a graphical intensity measure (Graphics Index) that provides a way to identify those applications in the portfolio that have higher GPU demands than others.

With this information it becomes possible to segment the portfolio into groupings based on their requirements for specific resources.

By tying a general sense of which applications have peak demand to total length of usage it becomes easier to start developing a portfolio made up of different combinations of usage styles.

The figure below displays this relationship in a bubble chart format, this format groups applications based on their similar characteristics presenting clusters of similar applications in larger bubbles. The vast majority of applications exist in the "low graphics demand – Low Time Active" area in the bottom left, while only a select few have either high graphics demand or high time active.

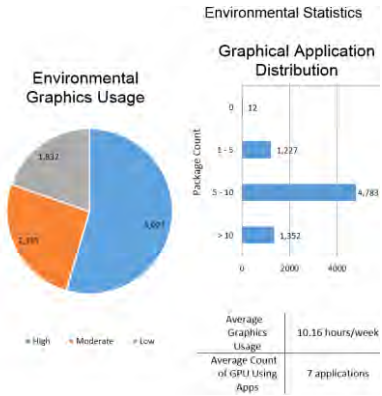
A natural expansion of this is grouping users into distinct workload types to understand how best to configure the profile types and GPU assignments for users. Once the target applications and users have



been characterized and a plan has been developed it's critical to begin the process of sizing the environment.

Resource Modeling and Capacity Planning

With a complete portfolio plan it now becomes possible to move into the next phase and start creating a model for what resources will be required for a complete environment. Because each of the users have been fully characterized throughout the assessment data collection interval it's possible to use SysTrack's Virtual Machine Planner (VMP) for mathematical analysis to provide insight into infrastructure provisioning.



The NVIDIA MarketPlace report from VMP outlines the number of users that fall into the various use cases (e.g. "high" for a designer or higher end power user), making it much easier to forecast how many users per NVIDIA board can be allocated and in turn how many total boards may be needed. This information creates an easy to use design for a set of user profiles, both for the actual desktop delivery and for the vGPU assignment.

Another key feature is the result of close collaboration with NVIDIA to leverage APIs presented in the guest operating system. This allows the capture of detailed GPU performance metrics to correlate vGPU consumption to end-user service quality. Specifically, with NVIDIA drivers present in the guest OS or on a physical system, the GPU utilization and key metrics from the graphics card can be captured and analyzed in the same way as CPU or other system metrics are currently in SysTrack. Metrics which are captured are:

- GPU % Usage
- Graphics Memory % Usage
- Number of GPU Driving Apps
- GPU Temperature and FAN RPMS
- Frame Buffer Usage
- Bus Usage

- Video Usage
- DeviceID

15.2 UBERAGENT

uberAgent takes Windows monitoring a step further. It does not just collect data - it gives you the information that matters. Other monitoring products rely on the performance counters built into Windows. uberAgent has its own metrics, covering key aspects of user experience and application performance.

Benefits and functionality of the solution

uberAgent stands out from the crowd because it is not another me-too product, but gives administrators the information they actually need to improve user experience. uberAgent's high-quality metrics include detailed logon duration and GPU usage, to name just two. But that is not all: uberAgent is built on a powerful platform, Splunk, which makes it possible to enrich uberAgent's data with information from arbitrary other sources (think NetApp, NetScaler, anything) and generate true operational intelligence.

Monitoring GPU usage

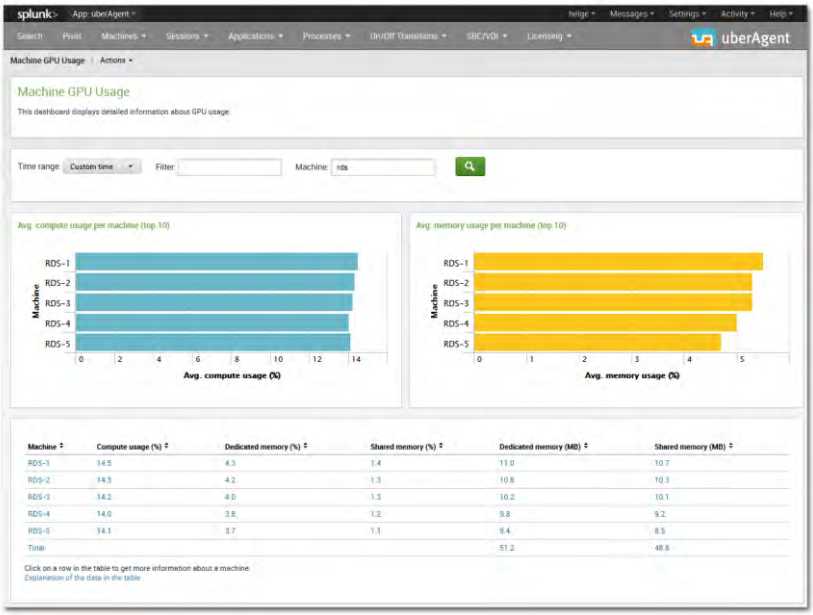
As it becomes more and more common to utilize the GPU for effects, video decoding and even general-purpose computing administrators need a tool that helps them understand exactly how their applications make use of GPU acceleration in order to optimally size the hardware for the workload. uberAgent is that tool.

GPUs have multiple engines, each serving a different purpose like 2D acceleration, 3D acceleration, video decoding, and so on. uberAgent determines the utilization of each engine individually. It also reports on shared and dedicated GPU memory usage separately. With that data it is already possible to gauge a machine's GPU usage with great accuracy. uberAgent even reports GPU usage per process, enabling administrators to find out where exactly a GPU's utilization is coming from and what purpose it might serve.

uberAgent measures:

- GPU memory and compute usage per process
- GPU memory and compute usage per machine

- Show memory usage for shared, main system RAM, and dedicated RAM on the GPU
- Shows compute usage per GPU engine e.g. 3D acceleration, video coding, 2D acceleration.



SLA Compliance Monitoring

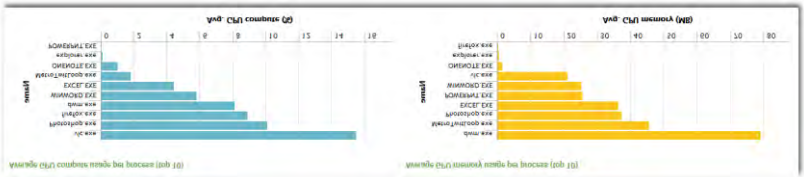
uberAgent helps you monitor SLA compliance by telling you what management wants to know: unnecessary pauses and interruptions in the workflow by IT systems that cannot keep up. It helps answer questions like:

- How long does it take a PC or laptop to boot?
- What is the duration of a logon?
- How long do end users need to wait for applications to start?

Sizing and Capacity Planning

Planning and sizing a new environment is impossible without proper data on user activities. uberAgent makes the whole process a lot simpler by:

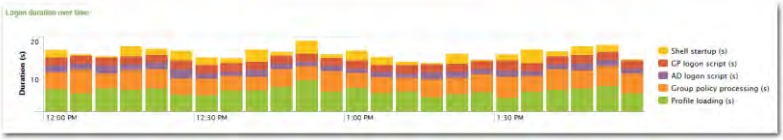
- Showing you the footprint of the average user session (RAM, IOPS, network throughput, CPU usage)
- Showing you the resource usage per application
- Showing you the footprint individual websites have on your systems' resources
- Showing you the GPU usage per individual process



Operations

IT operations involves keeping an eye on a huge number of different systems and ensuring all of them run smoothly.

- uberAgent shows you what is going on: how many sessions, how many users, how many machines, etc.
- Want to know more? Response times per machine, per application and even per session.
- Even more? Logon duration, group policy processing and user profile loading duration in great detail.



Requirements and architecture

uberAgent consists of two components: the technology add-on (TA) collects the data and sends it to Splunk for indexing. The dashboards

contained in the uberAgent app present the data stored in Splunk to the user.

The uberAgent TA is a lightweight component that does not require installation. It is executed from Splunk's Universal Forwarder, a kind of generic agent that is installed on the same computer. The uberAgent TA works on all client and server versions of Windows beginning with Vista / Server 2008, including TS/RDS/XenApp/XenDesktop and VDI.

The uberAgent app, on the other hand, contains the dashboards that present the data stored in Splunk. All dashboards are browser-based (no console installation required) and can be modified easily if the need arises.

The entire architecture is extremely flexible and scales from a single server to a globally distributed environment. It can optionally be clustered for redundancy and availability and supports multi-tenancy. uberAgent works with versions 5.x and 6.x of Splunk Free and Enterprise.

Licensing, pricing

uberAgent is licensed per monitored computer. There are two types of licenses available: client and server licenses. Both types are available as perpetual, term and service provider licenses. Prices start at €19 per client and €375 per server for perpetual licenses and include one year of maintenance. Volume discounts are available. In addition to the uberAgent license a Splunk license is required for uberAgent to work.

16. 3D GRAPHICS SOLUTIONS FOR VIRTUAL DESKTOPS-FEATURE COMPARISON

16.1 INTRODUCTION

It's important to understand that comparing features is the last step in the decision tree. Vision, Strategy and Technology are the former steps. Each solution has its own functionality and feature-set. This chapter describes some of the features and more importantly the qualifying questions to determine the right solution.

16.2 QUALIFYING QUESTIONS

Understanding the strategic questions before you start the 3D graphics for virtual desktop project is important. When the use-cases. Business-case, future direction, roadmap, requirements both from end-user and IT and the action plan is clear the next step is important.

The next step is to determine which solution, or solution stack, meets the requirements.

What are the key Qualifying Questions to differentiate the 3D graphics for virtual desktops solutions from each other?

Generic

- What is the customer focus? SMB, Enterprise, ISV, ISP
- Does the solution requires public IaaS, such Amazon GPU instance?
- Can the solution run on-premises, Private IaaS?
- How is the solution licensed?
- Is the solution optimized for LAN, WAN, WiFi, Mobile networks?
- Is the solution suitable for Server Based Computing?
- Is the solution suitable for Virtual Desktop Infrastructure?
- Does the solution work in a multi-site enterprise environment?

- Is the solution proven? What is your definition and what are the proof points? Are there reference customers you can contact?
- Is there a large ecosystem with OEMs, ISV, thought leaders, partners, community?

Application related

- Which version of DirectX, OpenGL, CUDA is supported? is this supported with vGPU, Pass-through, shared and Soft-GPU?
- Is the solution certified by the ISV?
- Are there Requirement for multiple GPU's to one physical or virtual machine?

User Experience

- Is Client drives and client printers supported?
- Is USB device redirection supported?
- is Audio and Microphone supported?
- Is 3D space mouse supported?
- Is 3D Glasses support needed?
-

EndPoint and Guest OS

- Is Linux as guest OS supported?
- Does the solution includes a connection broker?
- Do you need support for 4K screens ?
- What is the maximum amount of monitors per user (2,4,8+)? What is the impact on FPS, bandwidth?
- What is the maximum resolution?
- Is span/extend, clone and pivot of multi-monitor supported
- Is there support for HTML5, Apple OSX, Apple iOS, Windows7 / 8, Android, ChromeOS, ThinOS
- Understand the limitations with thin clients that you want to use with remote graphics. The amount of monitors, screen size, decoding capacity.

16.3 FEATURE COMPARE MATRIX

Used legend:

√*=Applicable;* **X***=Not applicable;* **--***Not needed;* A green √ or red **X** has nothing to do with advantage or disadvantage of a solution. It just presents the availability and support of the functionality.

	XD7.1 VDA SoftGPU	XD7/XA6.5 Shared GPU	XD7 HDX 3D Pro Pass-Through GPU	XD7.1 vGPU (Shared)	RDVH 2012 R2 Shared GPU	WS 2012R2 RDSH on Bare Metal	View 5.3 vDGA Pass-Through GPU	View 5.3 vSGA Shared GPU
DirectX 9	√	√	√	√	√	√	√	√
DirectX 10	X	√	√	√	√	√	√	X
DirectX 11	X	√	√	√	√	√	√	X
OpenGL 1.1	√	√	√	√	√	√	√	√
OpenGL 2.0	√	√	√	√	X	X	√	√
OpenGL 3.0	X	√	√	√	X	X	√	X
OpenGL 3.3	X	√	√	√	X	X	√	X
OpenGL 4.0	X	√	√	√	X	X	√	X
OpenGL 4.3	X	√	√	√	X	X	√	X
OpenGL 4.4	X	√	√	√	X	X	√	X
OpenCL 1.1	X	√	√	X	X	X	√	X
CUDA 3.0	X	X	√	X	X	X	√	X
CUDA 4.0	X	X	√	X	X	X	√	X
CUDA 5.0	X	X	√	X	X	X	√	X
CUDA 6.0	X	X	√	X	X	X	√	X

17. APPLICATION SUPPORTABILITY

#WorkInprogress here.

When you have a link to ISV application and their support statements please share.

18. TESTLAB

Investigating, analyzing and writing about 3D graphics for virtual desktops needs to be combined with real world customer experiences and R&D in lab environments.

Ruben Spruijt / PQR his R&D environment consists of different test environments, the current primary environment consists of:

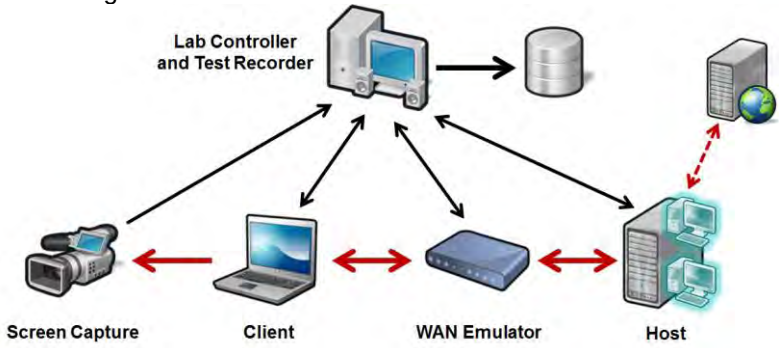
- **Host 1**
  - Cisco C240M3 server with 96GB RAM and a couple of 146GB SAS drives,
  - FusionIO 1.4TB interface (used with VMware vSphere)
  - NVIDIA GRID K1 and K2 cards
- **Platforms**
  - Host OS platform:
    - Citrix XenServer
    - VMware vSphere
    - Microsoft Windows Server 2012 R2 - HyperV
  - Guest OS platform:
    - Windows 8.1 both x64 and x86
    - Windows 7 both x64 and x86
  - Virtual Desktop connection brokers
    - Citrix XenDesktop 7.x
    - VMware Horizon View 5.3
    - Microsoft RDS Server 2012R2
- **WAN Emulator**
  - Apposite Linktropy mini2
- **Screen Capture**
  - Epiphan DVI2USB 3.0
- **Clients**
  - Apple MacBook Air 13
  - Windows Surface Pro
  - Lenovo X1 Carbon
  - Dell Wyse Z90
  - Apple iPad

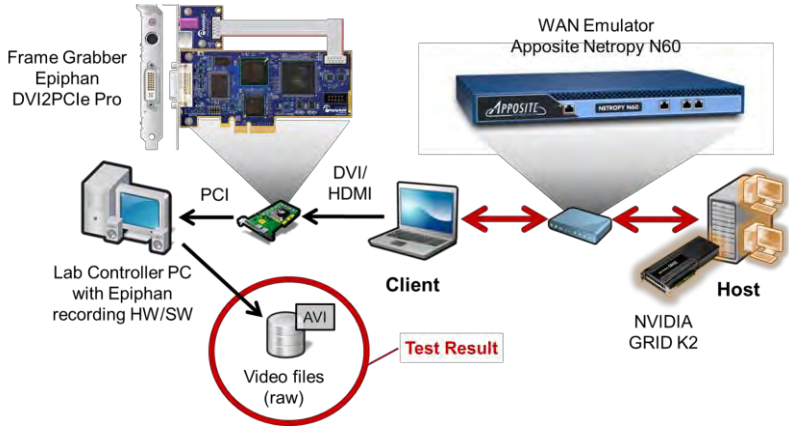
19. SMACKDOWN 3D GRAPHICS SOLUTIONS, VIDEO RECORDINGS

Reading the 3D graphics for virtual desktop smackdown is a great start for anyone who wants to start with 3D graphics for virtual desktops and also for the people who have knowledge and want to expand their knowledge. See different solutions in action, compare the performance differences and determine that impact for your own environment is where video recordings can support. Multiple 1000s of video are recorded by Shawn Bass and Bernhard Tritsch and new videos are created to show 3D graphics for virtual desktops in action over LAN, WAN, Mobile scenarios. Some of these videos will be used by Ruben, Shawn and Benny on stage and some will be posted to the TeamRGE and DarcoLabs YouTube channels.

The testing methodology behinds these videos is rather simple. We take a stock specification host system hardware and leverage multiple hotswap SSDs in order to create a system that allows us to easily switch between baremetal Server OS, baremetal client OS, or hosted hypervisor solutions from Microsoft (Hyper-V), VMware (vSphere/ESXi) and Citrix (XenServer). We use this host hardware to install the latest offerings from Microsoft, Citrix and VMware and compare the technologies against a set of known graphics solutions focused on 2D graphics (GDI, Flash, Shockwave, Silverlight, etc) as well as 3D (DX9-11, OpenGL 1.x-4.x, etc). In each case, the videos are recorded from the same client system to same host hardware. We'll often use this to compare how the various remoting protocol solutions differ when compared side-by-side given a different set of LAN/WAN conditions, a different set of virtual/physical graphics, etc. It gives the viewer an easy way of understanding the differences between the technologies. All test scenarios are automated using AutoIt scripts so we can compare the scenarios as closely as possible while eliminating as much human error as possible. In the future, we are hoping to provide synchronized video starts as well as overlayed graphs of CPU/GPU consumption, LAN/WAN bandwidth, etc. We also leverage network emulation hardware from Apposite

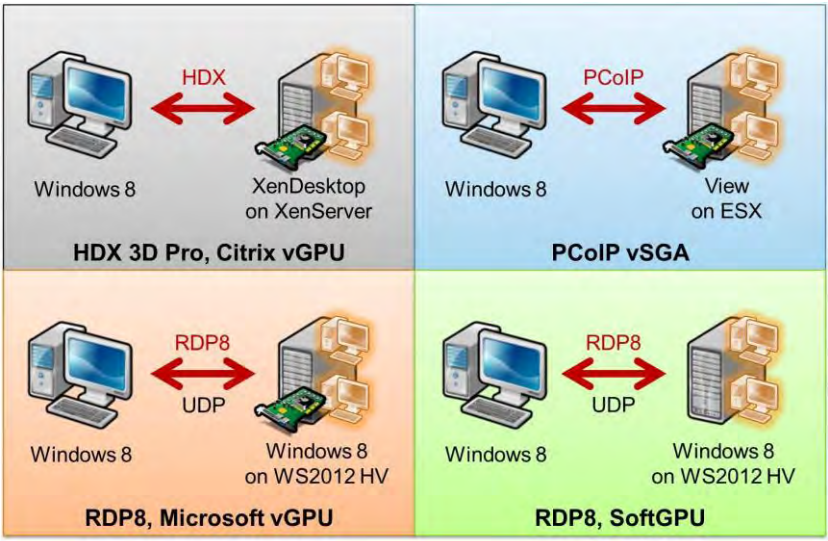
Technologies to ensure consistent WAN simulation scenarios.





(Setup for video recording detailed)

An example of how our 4-up SyncPlayer looks is shown below:



20. NOTES FROM THE FIELD

Working with 3D graphics for virtual desktop solutions at customers and in R&D environments is great. Every day you learn new things. This paragraph will share our notes from the field in random order.

Hardware related

- Performance isn't only about GPU. Understand the impact of CPU ClockSpeed, CPU cores and CPU cache size. Some 3D graphic are much more CPU bound than GPU bound. We have seen in practice that certain applications runs much much faster with 3.5GHZ, 4 cores CPUs than 2.8GHZ, 12 cores CPUs.
- Performance isn't only about GPU. Understand the performance impact of Flash-based storage solutions, both central and local storage solutions such as FusionIO.
- Triple check Hardware vendor support when using 3D graphics for virtual desktops especially the hardware combination such as CPUs, GPUs, PCIe storage controllers such as FusionIO, PCoverIP offloading interfaces. Most often, a server manufacturer will require a minimum wattage PSU configuration to be used with high end GPUs in the server. This PSU will usually be 1100 Watts, though in limited cases a 750Watt PSU will be adequate for a single card. Many server manufacturers will have a separate part number SKU for a GPU kit which may include additional fan shrouds, power connectors and possibly PCI riser cards. You will need to check with your specific supplier to determine if such a GPU kit is required.
- To support GPU virtualization, you will need to have a system BIOS that supports proper memory mapping for the GPU. This may be called a number of different things. It may be referred to as "Above 4G Decoding", "64-bit BAR Support" or "64-bit MMIO". It is imperative that your BIOS supports this feature setting in order for vGPU to work properly.
- Is the hardware combination on the HCL for Citrix, Microsoft, VMware?

Software related

- Does the 3D graphics application vendor support the virtual desktop solutions. Our experience is that the majority of applications aren't supported on the latest (and even one version earlier) virtual desktop solution. We understand that testing, QA, reporting for ISV isn't that easy but this is a fast and evolving market. Please speed up.
- Does snapshot with Citrix XenServer work with GPU cards?!
- Does snapshot with Citrix XenServer work with GPU cards?!
- Use to be issues with older versions, but I have good experience with XS62SP1 w. XenDesktop 7.1, that was a requirement with vGPU and MCS. Strange Citrix tells that.
- Citrix XenServer – Setting more than one vCPU per VM to improve application performance.
- Understand Citrix XenServer GPU Groups
- Understand the NUMA, vCPU pinning. Citrix XenServer – Setting more than one vCPU per VM to improve application performance.
- Enable extended P-states, TurboBoost mode in the BIOS.
- Set XenServer frequency governor to performance mode
- Turn on "hardware acceleration" in applications such as Autodeskhttp://blogs.citrix.com/2014/03/11/citrix-xenserver-setting-more-than-one-vcpu-per-vm-to-improve-application-performance-and-server-consolidation-e-g-for-cad3-d-graphical-applications/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+CitrixBlogs+%28Citrix+Blogs%29

21. 3D GRAPHICS BENCHMARK,- APPLICATIONS AND VARIOUS TOOLS

21.1 BENCHMARK APPLICATIONS

- GPU-Z: http://www.techpowerup.com/gpuz/
- GPU Caps Viewer: http://www.ozone3d.net/gpu_caps_viewer/
- Microsoft Process Explorer with GPU monitoring: http://technet.microsoft.com/en-us/sysinternals/bb896653.aspx
- Redway3d: Turbine http://www.redway3d.com/downloads/public/demos/turbine/turbineDemo.exe and Watch: http://www.redway3d.com/downloads/public/demos/watch/watchDemo.exe
- Siemens Teamcenter Vis Application: https://www.plm.automation.siemens.com/en_us/products/teamcenter/lifecycle-visualization/jt2go/index.shtml

  Models: https://www.plm.automation.siemens.com/en_us/products/teamcenter/lifecycle-visualization/jt2go/downloads/index.shtml#lightview-close
- 3DMark by Futuremark: http://www.3dmark.com/ (DirectX), comparison on http://www.futuremark.com
- Unigine: http://unigine.com/ (DirectX, OpenGL)
- SPEC ViewPerf 11/12: http://www.spec.org/ http://www.spec.org/benchmarks.html: 8 important Workstation programs (Catia, Ensight, Lightwave, Maya, Pro Engineer, Solidworks, Siemens Teamcenter, Siemens NX)

  http://www.spec.org/gwpg/gpc.static/vp11info.html
- Fraps: http://www.fraps.com/ (Framerate FPS for DirectX and OpenGL)

- Luxmark by Luxrender: http://www.luxrender.net/wiki/LuxMark (OpenCL)
- Furmark Freeware by Geeks3D.com: http://www.ozone3d.net/benchmarks/fur/ (OpenGL)
- Shader ToyMark: http://www.ozone3d.net/benchmarks/shadertoymark/ (OpenGL)
- Maxon CineBench: http://www.maxon.net/?id=1258 (OpenGL)
- eDrawings Viewer for SolidWorks: http://www.edrawingsviewer.com/
- CPU Benchmark by PassMark Software: http://www.cpubenchmark.net, http://www.videocardbenchmark.net/ (
- DaCapo Benchmark Suite, Open Source Java benchmarking: http://www.dacapobench.org/
- AutoDesk Revit Benchmark tools, RFOBenchmark: http://www.revitforum.org/hardware-infrastructure/1063-rfobenchmark.html
- Real-time info about GPU usage and HDX info overlaid on the screen http://virtexperience.com/2014/05/12/gpuperf-realtime-gpu-and-hdx-info/

If you have other awesome and handy benchmarking applications to share, please send an email to team@teamRGE.com so we can included them.

21.2 FREE 3D APPLICATIONS APPLICATIONS

- Google SketchUp (http://sketchup.google.com/) - To build models in SketchUp, you draw edges and faces using a few simple tools that you can learn in a short time. With Push/Pull tool you can extrude any flat surface into a 3D form. Furthermore, it works together with Google Earth, that you can import a scaled aerial photograph directly from Google Earth, or use SketchUp to build models which can be seen in Google Earth.
- 3DCrafter (http://www.amabilis.com/products.htm) - 3Dcrafter is a real-time 3D modeling and animation tool that incorporates an intuitive drag-and-drop approach to 3D modeling. The standard version of 3DCrafter is freeware.
- 3Dtin (http://www.3dtin.com/) - The simplest 3D software. You can draw directly from your browser.

- Anim8or (http://www.anim8or.com/main/) - Anim8or is a 3D modeling and character animation program.
- Art of Illusion (http://www.artofillusion.org/) - Art of Illusion is a free, open source 3D modelling and rendering studio. Art of Illusion is more as a 3D design system for animated computer graphics than as an engineering CAD tool.
- Blender (http://www.blender.org/) - Blender is the free open source 3D content creation suite, available for all major operating systems under the GNU General Public License. Blender was developed as an in-house application by the Dutch animation studio NeoGeo and Not a Number Technologies (NaN). It is a powerful program contains features that are characteristic of high-end 3D software.
- BRL-CAD (http://brlcad.org/) - BRL-CAD is a powerful cross-platform open source solid modeling system that includes interactive geometry editing, high-performance ray-tracing for rendering and geometric analysis, image and signal-processing tools, a system performance analysis benchmark suite, libraries for robust geometric representation. BRL-CAD has been the primary tri-service solid modeling CAD system used by the U.S. military to model weapons systems for vulnerability and lethality analyses for more than 20 years. It became an open source project on 21 December 2004.
- Creo Elements/Direct (http://www.ptc.com/products/creo-elements-direct/) - formerly CoCreate - is a complete design environment that offers direct 3D CAD modeler, along with 2D CAD, CAE and integrated product data management (PDM).
- DrawPlus Starter Edition (http://www.serif.com/free-graphic-design-software/?MC=FSSDRAWPLUS) - 100% free and simple, with Accurate vector drawing program, realistic brush, pen, and pencil tools, text on a path, blend modes for advance artistic effects.
- Ensight Free (http://www.ceisoftware.com/ensight-free/)
- FreeCAD (http://sourceforge.net/apps/mediawiki/free-cad/index.php?title=Main_Page) - FreeCAD is a general purpose Open Source 3D CAD/MCAD/CAx/CAE/PLM modeler, aimed directly at mechanical engineering and product design but also fits in architecture or other engineering specialties.
- GIMP (http://www.gimp.org/) – GNU Image Manipulation Program. GIMP is a versatile graphics manipulation package.

- GLC Player (http://www.glc-player.net/download.php) - GLC player is a free application used to view 3d models (COLLADA 3DXML OBJ 3DS STL OFF COFF Format) and to navigate easily in these models. With the album management, capture and multi-capture capabilities, html export and navigation possibilities GLC_Player is the accurate tools to review a lot of 3D models and to create illustrations. GLC_Player is a cross-platform application (Mac, Linux and Windows). It is lighter than regular modelling software so very handy.
- Lego Digital Designer (http://ldd.lego.com/) – 3D program that allows users to build models using virtual LEGO bricks.
- LeoCAD (http://www.leocad.org/trac) - LeoCAD is a CAD program that can be used to create virtual LEGO models. It has an easy to use interface and currently features over 3000 different types of pieces created by the LDaw community.
- Netfabb Studio Basic (http://www.netfabb.com/basic.php) - Netfabb Studio Basic provides mesh edit, repair and analysis capabilities. Its compact size of only a few megabytes allows a quick download, an easy installation and the handling of STL and slice files within seconds.
- K-3D (http://www.k-3d.org/wiki/Main_Page) - K-3D is free-as-in-freedom 3D modeling and animation software. It features a plugin-oriented procedural engine for all of its content, making K-3D a very versatile and powerful package. K-3D excels at polygonal modeling, and includes basic tools for NURBS, patches, curves and animation.
- OpenSCAD (http://www.openscad.org/) - OpenSCAD is a software for creating solid 3D CAD objects. It is free software and available for Linux/UNIX, MS Windows and Mac OS X. it does not focus on the artistic aspects of 3D modelling but instead on the CAD aspects.
- Osamu Mizuno's Metasequoia (http://metaseq.net/en/) – 3D Polygon modeler.
- Tinkercad (http://tinkercad.com/) - Tinkercad is a new and faster way of creating designs for your 3D printer. With only three basic tools you can create a wide range of useful things. Once your project is ready simply download the STL file and start your 3D print.
- Wings 3D (http://www.wings3d.com/) - Wings 3D is a subdivision modeler. It has been developed since 2001. Wings 3D

offers a wide range of modeling tools, a customizable interface, support for lights and materials, and a built-in AutoUV mapping facility. There is no support in Wings for animation.

- 3D Slicer is a free, comprehensive software platform for medical image analysis and visualization developed with NIH. supporhttp://www.slicer.org/publications/item/view/2331 and http://download.slicer.org/
-

If you have other free 3D graphics applications to share, please send an email to team@teamRGE.com so we can included them.

21.3 WEBGL DEMO WEBSITES

Advanced web API's, such as WebGL, delivers enhanced content, rich video, 3D graphics etc.. Next years more productivity tools and even fully featured applications will be delivered completely over the internet via WebGL. 3D graphics for virtual desktops is needed to receive a fluid user experience. Some nice WebGL demo websites:

- http://alteredqualia.com/three/examples/webgl_materials_skin.html
- http://webglsamples.googlecode.com/hg/aquarium/aquarium.html
- http://middle-earth.thehobbit.com/
- http://www.chromeexperiments.com/webgl/
- http://www.zygotebody.com/
- http://helloracer.com/webgl/
- http://madebyevan.com/webgl-water/

If you have other awesome WebGL websites to share, please send an email to team@teamRGE.com so we can included them.

22. CONCLUSION

The aim of this whitepaper is to educate the market about trends in virtual desktop/application delivery, graphics formats used by applications and the various products/solutions available from different vendors to assist in delivery of rich and complex graphical applications.

This whitepaper is by no means a complete coverage of all offerings in this space, but over time it will be enhanced and will ideally become the reference document for high end graphics for virtual desktop and application delivery. Currently, the most complete on-premises desktop and application delivery strategy is offered by a combination of Citrix and NVidia. That being said, there is rapid innovation in this field and we expect that the gap will be closed by the other vendors in this space in a short amount of time. It's a very exciting time for rich graphics in virtual desktops.

Please submit any errors or omissions to team Remote Graphics Experts team@teamRGE.com

Thank you for reading this whitepaper, we hope that it was helpful to you.

- TeamRGE (Benny, Ruben and Shawn)