



Citrix 虚拟桌面解决方案

--- 设计用户篇

钱凯

Kai.qian@citrix.com

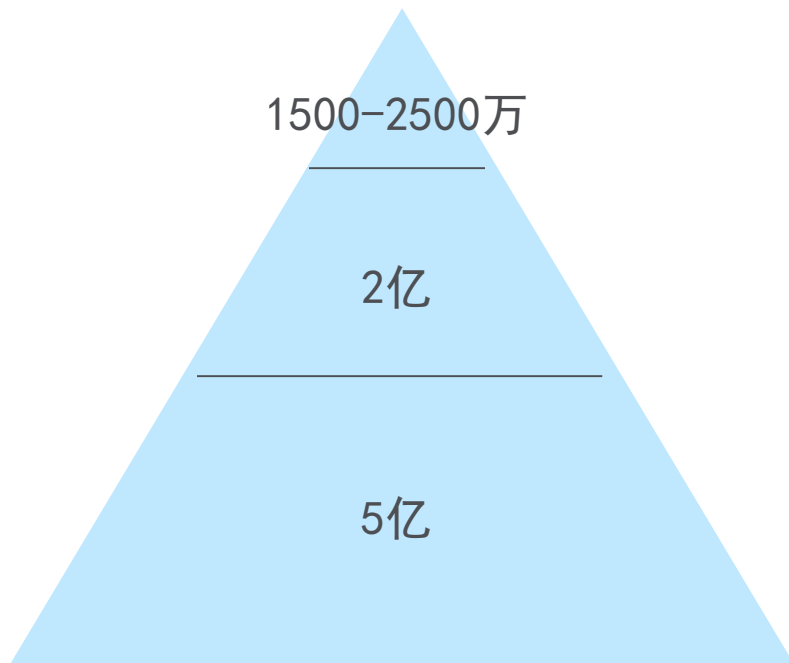
18621276655

针对有3D需求的虚拟桌面客户，我们建议尽可能的避免POC，因为此类客户的POC往往会涉及到服务器和K1/K2卡的准备，以及大量的性能调优工作，会影响到客户的购买周期。而目前我们北京和上海的CBC环境可以提供演示多种3D设计软件的vGPU虚拟桌面，所以建议尽量邀请客户到公司CBC参观。

- 常见设计用户分析
- Citrix 3D解决方案介绍
- Citrix Dedicated GPU 方案
- Citrix Shared GPU 方案
- Citrix vGPU 方案
- 为用户选择合适的方案
- vGPU POC最佳实践
- 案例分析

常见设计用户分析

全球专业图形用户约2.2亿



一级： 高端专业用户
设计，工程

二级： 中端设计用户
用户有编辑查看3D模型的需求

三级： 知识型用户
CPU辅助

常见使用专业显卡的领域



航天航空



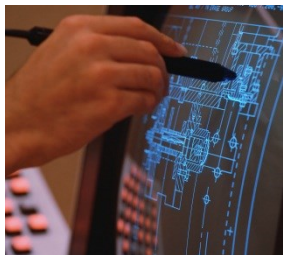
汽车行业



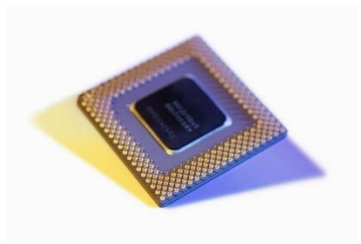
建筑



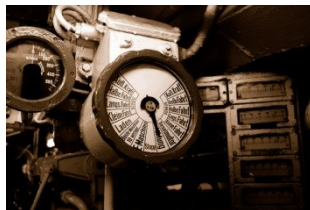
能源



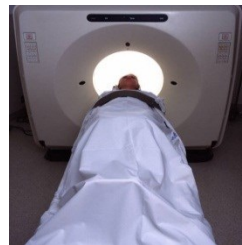
工程项目



高科技、
电子芯片



工业设备



医疗设备

设计用户的一般驱动力



- 全球一体化
- 数据安全、知识产权保护
- 居家办公
- 快速灾难恢复
- 移动化
- 快速交付
- 成本节约



桌面虚拟化核心优势

数据安全

节省成本

如何和3D用户沟通？

- 首先了解主流的3D设计软件

高端产品

- PRO/E、Creo 美国 PTC
- UG 美国 Siemens
- CATIA 法国 DASSAULT



SIEMENS

中端产品

- SOLIDEDGE 美国 Siemens
- SOLIDWORKS 法国 DASSAULT
- INVENTOR 美国 AUTODESK



- 了解客户当前的工作站的情况

一般来说，多数的设计用户都会选择品牌的图形工作站来作为日常设计的平台，最常见的一般为HP的Z系列，比如Z400/600/800/820等，也有部分用户会选择DELL的工作站。还有部分用户出于成本的考虑会选择高端配置的PC机（硬件配置是按照客户自己的需求自行攒出来的）

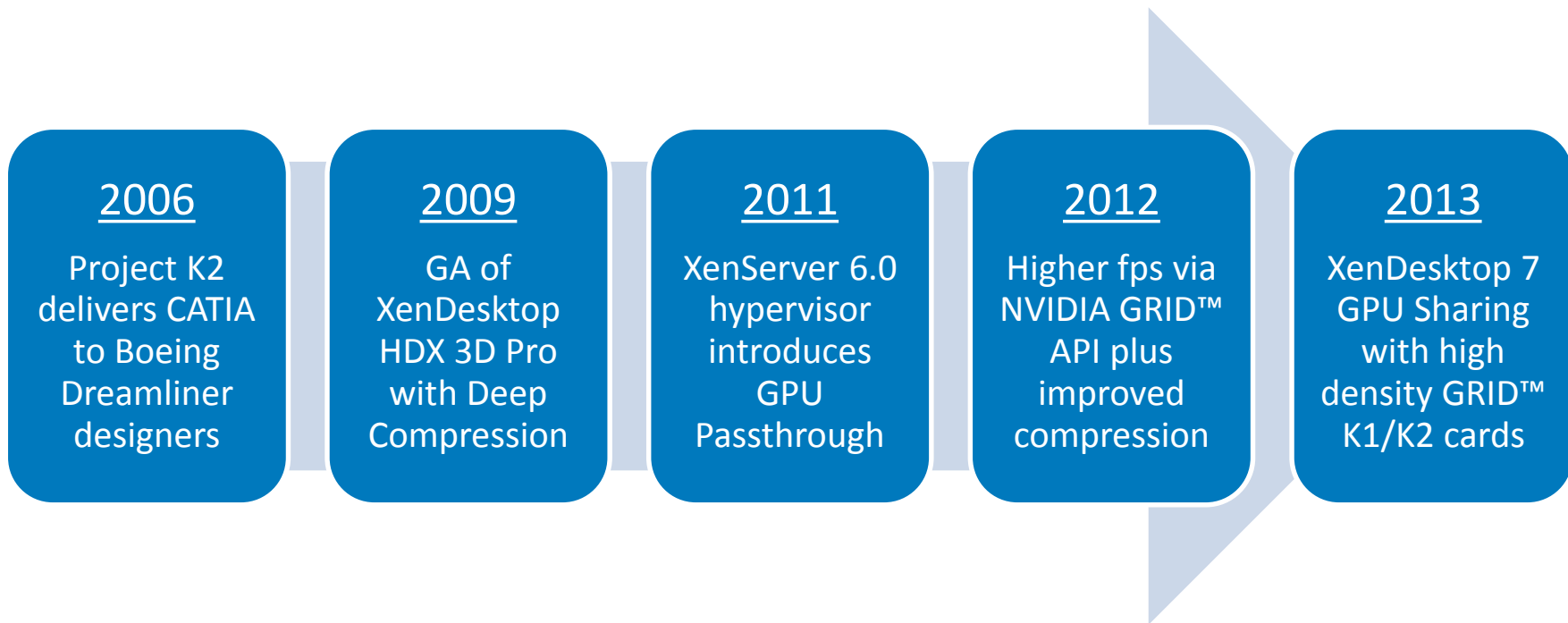
- 显卡型号（方便日后做POC以及硬件的规划）
- CPU的主频（方便日后POC硬件选型和规划）

- 了解客户当前使用哪些手段来保证数据安全
- 外设控制（技术手段、行政手段）
- 数据加密软件
- DLP系统（数据防泄漏系统）

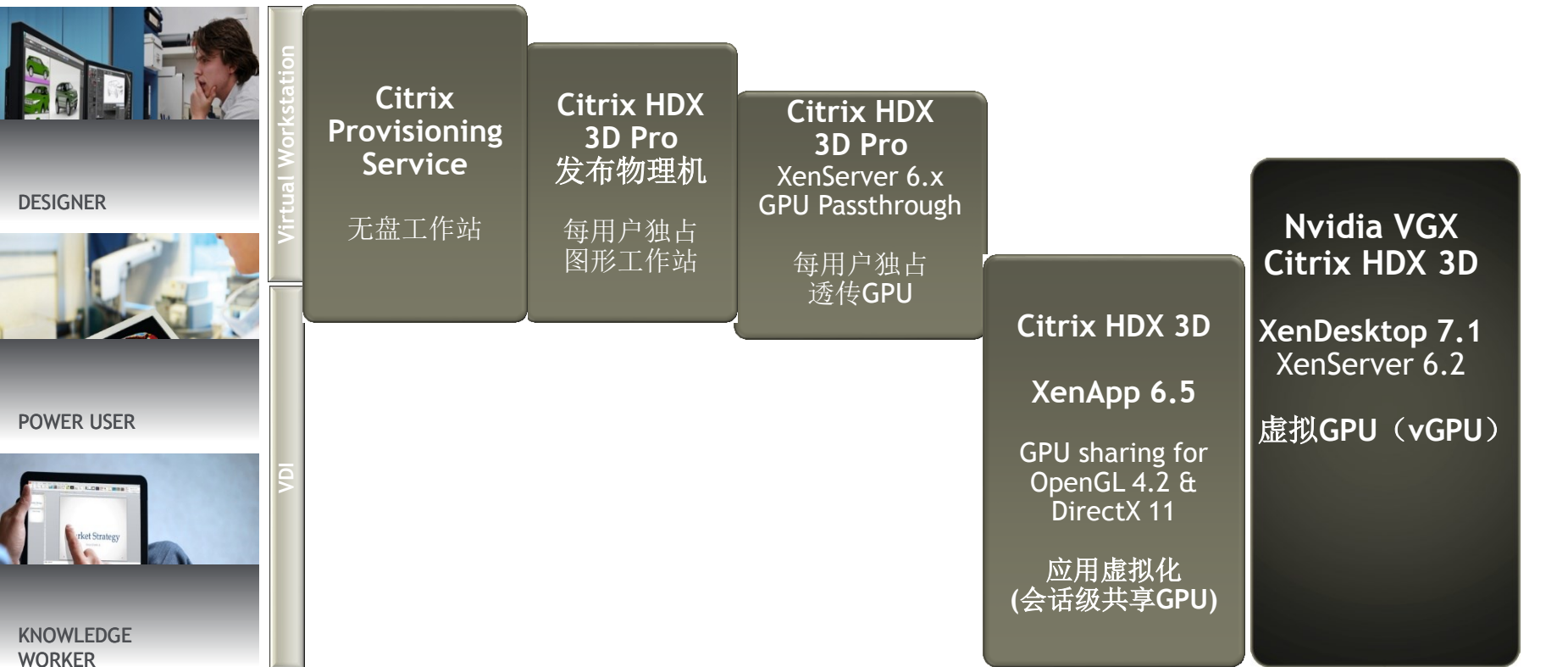
- 了解传统安全方案给客户带来的问题
- 以客户端管理为重点，用户可以物理接触到数据
- 桌面维护困难
- 无法实现快速交付
- 数据的交换困难

Citrix 3D解决方案介绍

Citrix 在 3D 图形交付技术的里程碑

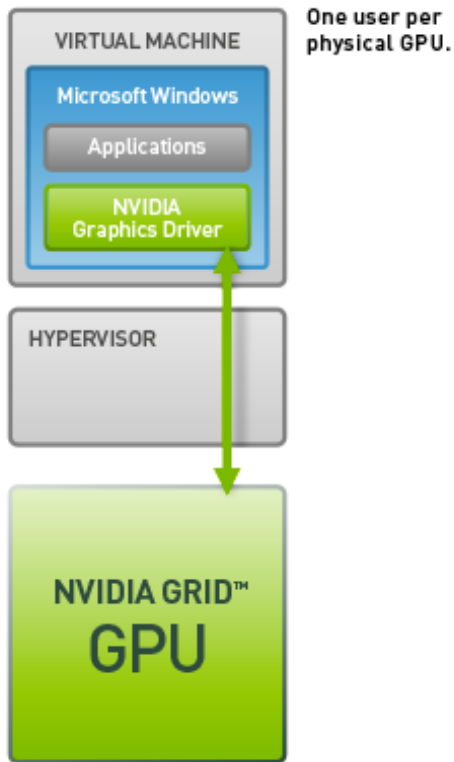


Citrix 3D 解决方案一览



Citrix Dedicated GPU 方案

Xendesktop VDI独享GPU桌面



利用显卡透传技术实现虚拟桌面对GPU资源的独占使用

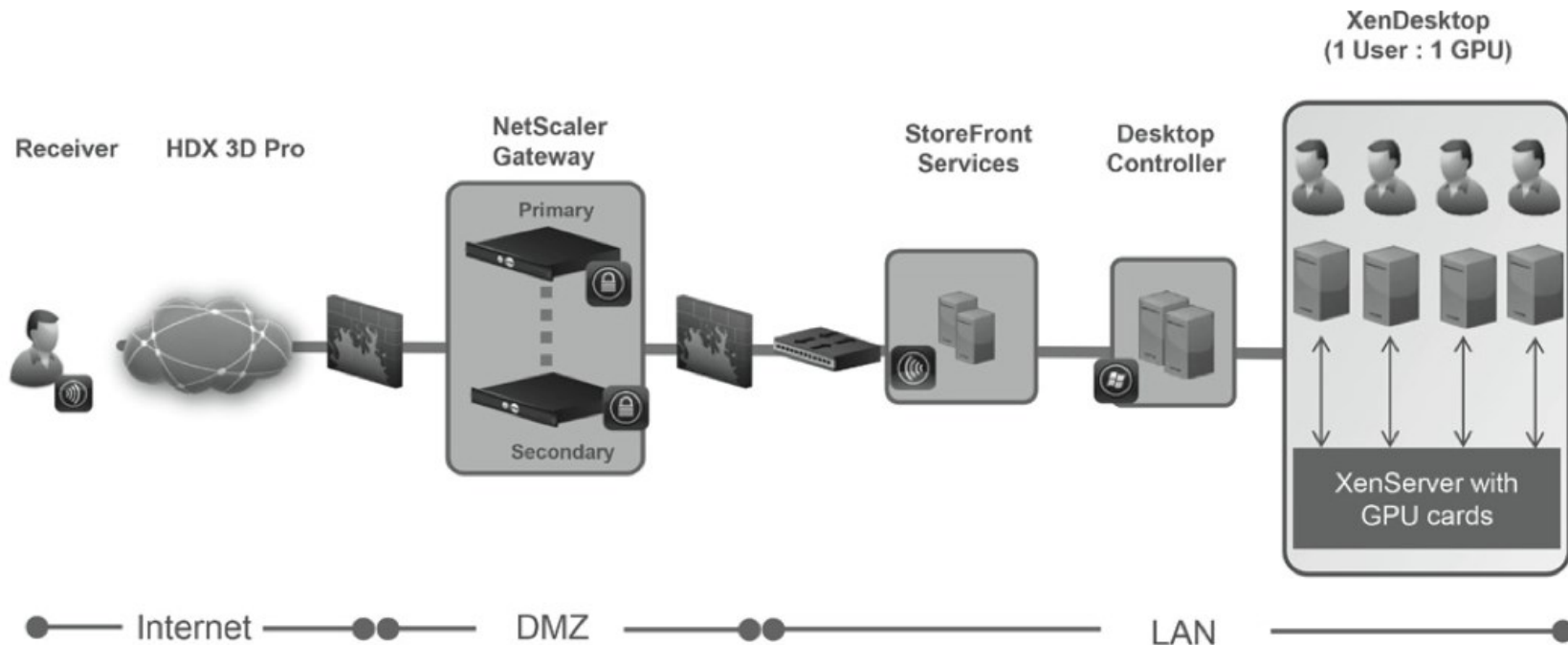
特点:

适合于重度设计用户

原生态支持显卡的功能（DirectX9,10,11和OpenGL4.4）

独占GPU资源，无法实现GPU资源的共享

整体架构

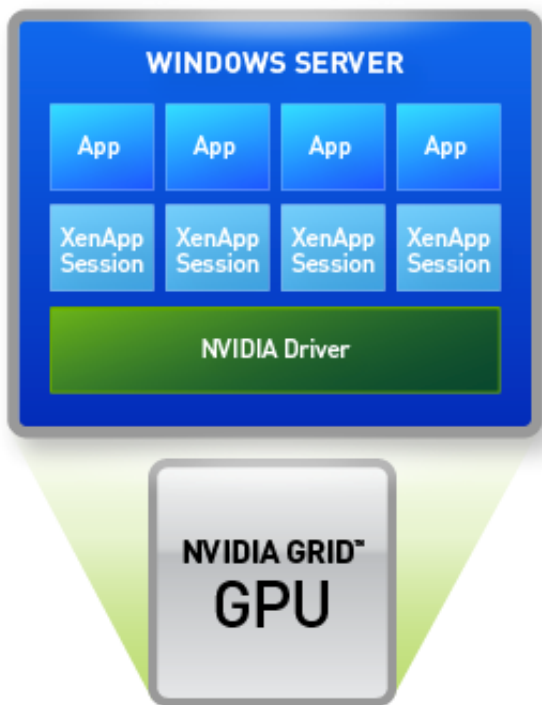


组件要求

- XenDesktop 5.x/7.x
- XenServer 6.x
- NetScaler Gateway 10.x
- Citrix StoreFront Services
- Citrix licensing server
- Citrix Receiver

Citrix Shared GPU 方案

XenApp 共享GPU桌面



利用共享桌面技术实现对GPU资源共享

特点：

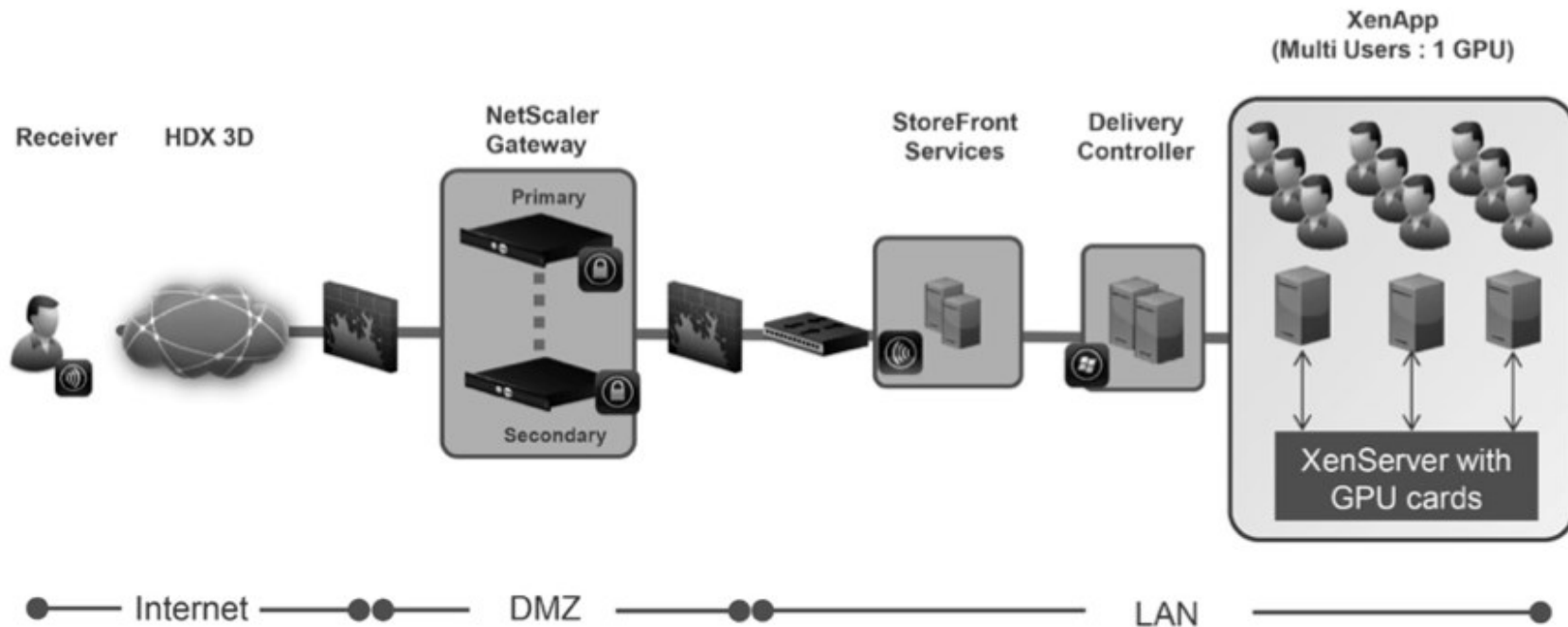
适合于非重度设计和审图用户

原生态支持显卡的功能（DirectX9,10,11和OpenGL4.4）

最大化的利用显卡资源，用户密度高

对应用程序的兼容性有要求

整体架构

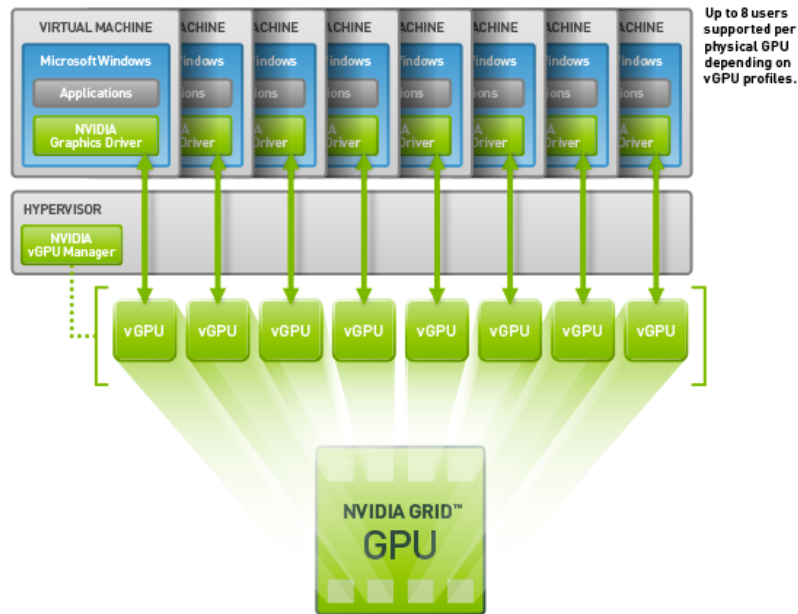


组件要求

- XenApp 6.5 Feature Pack 1 with OpenGL GPU Sharing Add-On/XenDesktop 7.1/XenApp 7.5
- XenServer 6.x
- NetScaler Gateway 10.x
- Citrix StoreFront Services
- Citrix licensing server
- Citrix Receiver

Citrix vGPU 方案

XenDesktop VDI vGPU桌面



利用Nvidia的vGPU技术实现对GPU资源共享

特点：

业内唯一真正的虚拟GPU方案

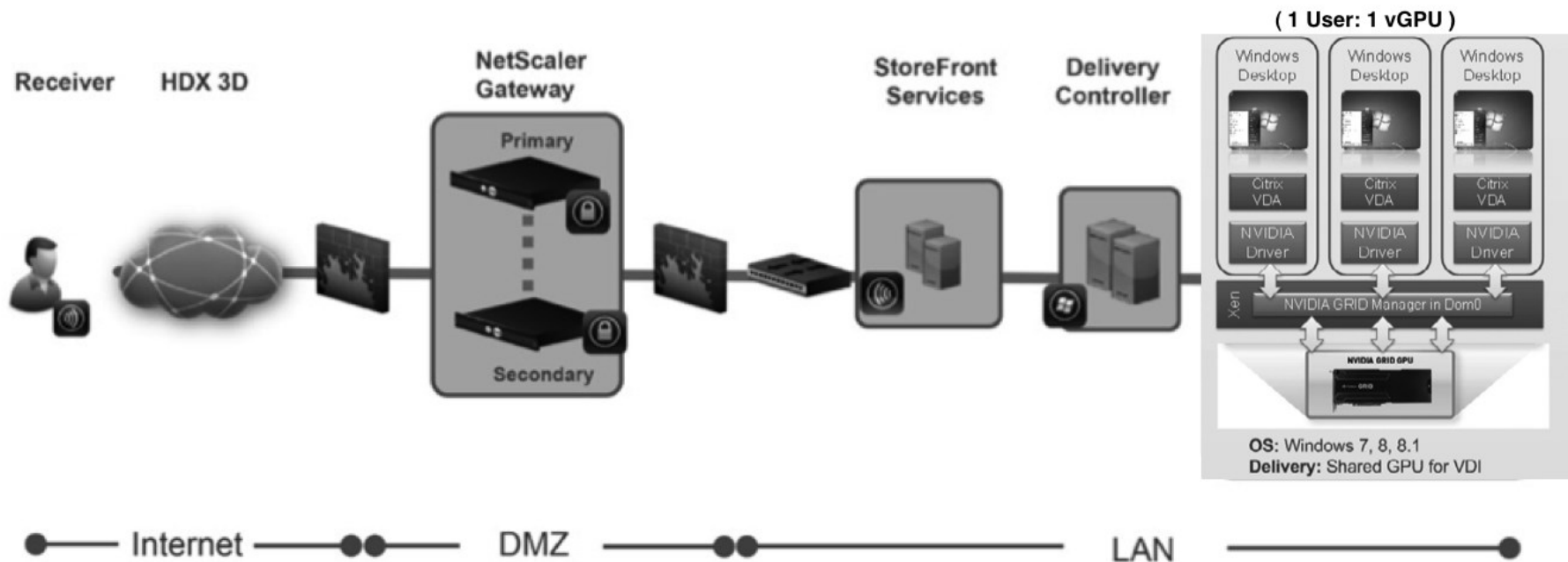
适合于重度设计用户

原生态支持显卡的功能（DirectX9,10,11和OpenGL4.4）

每GPU最大支持切割为8个vGPU

目前支持Windows7/8系统

整体架构



组件要求

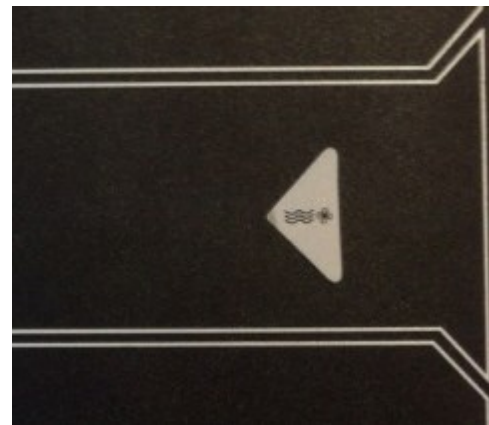
- Citrix XenDesktop 7.1 /7.5
- Citrix XenServer 6.2 SP1
- NetScaler Gateway 10.1
- Citrix StoreFront Services 2.1
- Citrix Licensing Server 11.11
- Citrix Receiver 3.4 or higher

认识K1/K2 Grid卡 - 面向虚拟化专用









Performance Graphs

Move Up

Move Down

Actions ▼

CPU Performance

New Graph...

Edit Graph...

Delete Graph...

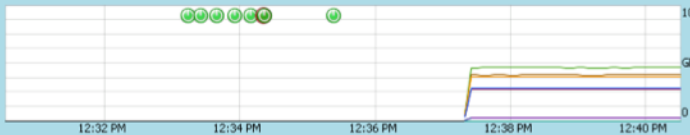
Restore Default Graphs

Performance Graphs

Move Up

Move Down

Actions ▼



GPU

Placement policy: Maximum density: put as many VMs as possible on the same GPU [Edit...](#)

GK104GL [GRID K2]

☒ w7664-K240q-00
GRID K240Q

☒ W508R2-K260q-00
GRID K260Q

[Select All](#)

[Clear All](#)

Allowed vGPU types:

- ✓ Pass-through whole GPU
- ✓ GRID K260Q vGPU (2 per GPU)
- ✓ GRID K240Q vGPU (4 per GPU)
- ✓ GRID K200 vGPU (8 per GPU)

[Edit Selected GPUs...](#)

GK107GL [GRID K1]

☒ MarcelC-K140q
GRID K140Q

☒

☒

☒

[Select All](#)

[Clear All](#)

Allowed vGPU types:

- ✓ Pass-through whole GPU
- ✓ GRID K140Q vGPU (4 per GPU)
- ✓ GRID K100 vGPU (8 per GPU)

[Edit Selected GPUs...](#)

Add Connection and Resources

Studio

- ✓ Connection
- Resources**
- Storage
- Summary

Resources

Select one or more networks for the virtual machines to use:

☒ Network 0
☐ Network 1
☐ Network 2
☐ Network 3

Do you want to use graphics virtualization? [Learn more](#)

☐ No
☒ Yes

Select a GPU type and group:

GRID K200 (Group of NVIDIA Corporation GK104GL [GRID K2] GPUs).

passthrough (Group of NVIDIA Corporation GK104GL [GRID K2] GPUs).

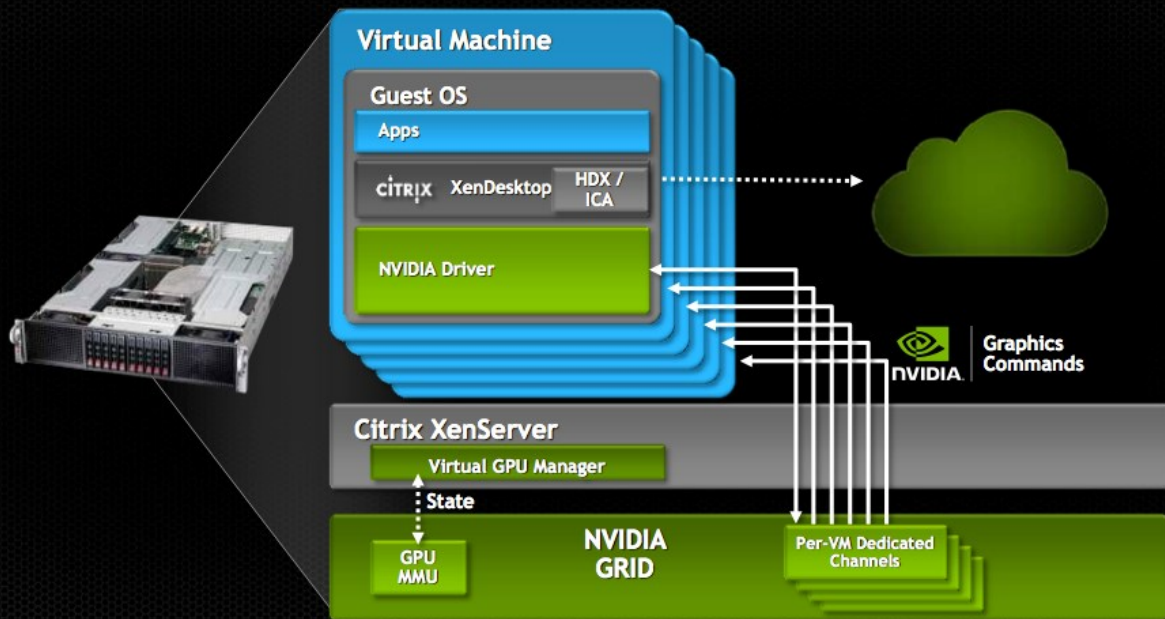
GRID K260Q (Group of NVIDIA Corporation GK104GL [GRID K2] GPUs).

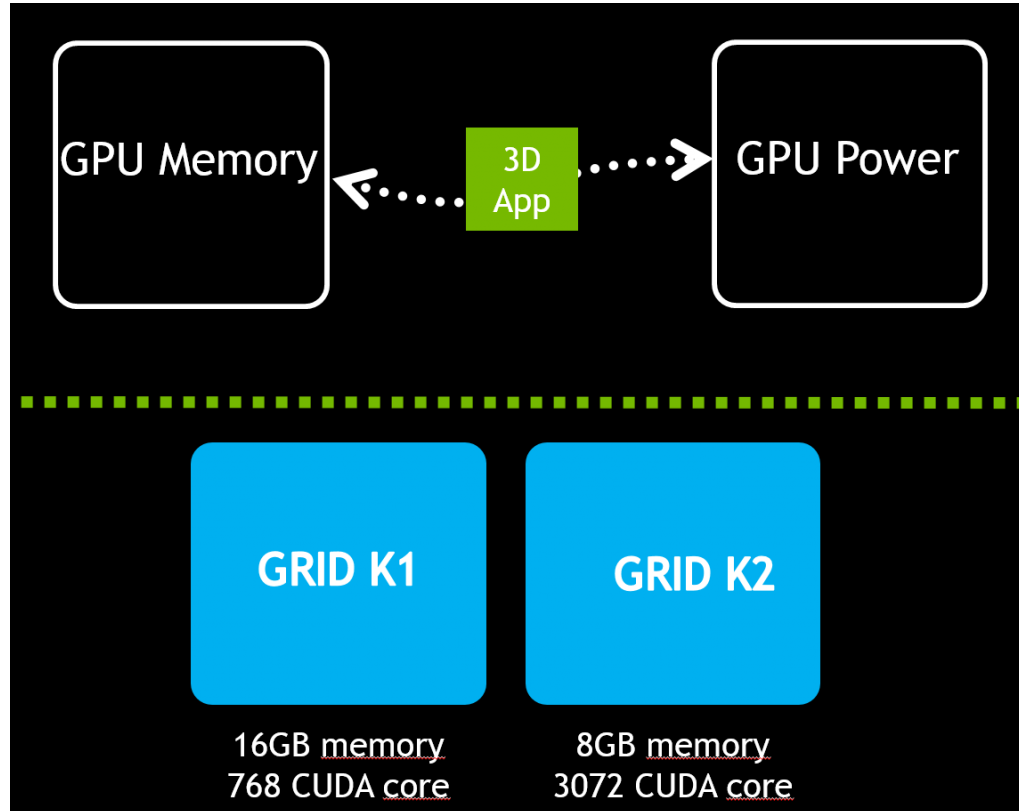
GRID K240Q (Group of NVIDIA Corporation GK104GL [GRID K2] GPUs).

Cancel

vGPU技术原理

NVIDIA GRID Virtual GPU





支持vGPU技术的显卡类型

NVIDIA GRID K1



NVIDIA GRID K2



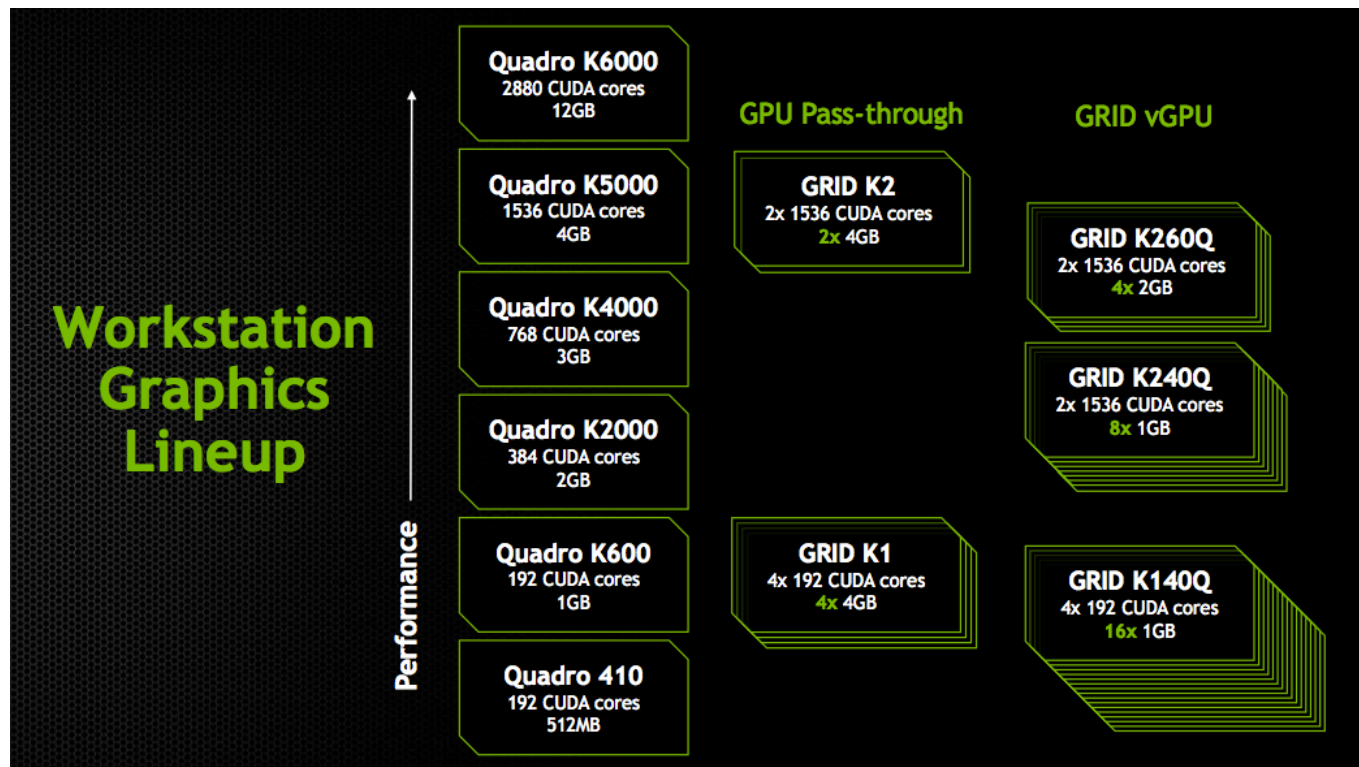
GPU	4 Kepler GPUs	2 High End Kepler GPUs
CUDA cores	768 (192 / GPU)	3072 (1536 / GPU)
Memory Size	16GB DDR3 (4GB / GPU)	8GB GDDR5 (4GB / GPU)
Max Power	130 W	225 W
Form Factor	Dual Slot ATX, 10.5"	Dual Slot ATX, 10.5"
Display IO	None	None
Aux power requirement	6-pin connector	8-pin connector
PCIe	x16	x16
PCIe Generation	Gen3 (Gen2 compatible)	Gen3 (Gen2 compatible)
Cooling solution	Passive	Passive
# users	4 - 100 ¹	2 - 64 ¹
OpenGL	4.3	4.3
Microsoft DirectX	11	11
VGX Hypervisor support	Yes	Yes

¹ Number of users depends on software solution, workload, and screen resolution

如何选择vGPU类型？

NVIDIA GRID CARD	Physically GPUs pr board	vGPU profiles	Surggested use-case	vGPUs per pGPU	vGPUs per board	GPU memory per VM	GPU CUDA per VM	Max Display Per VM	Max Resolution Per User	Max Users Per Graphics Board
NVIDIA GRID K2	2	(pass-through)		1	2	4096 MB	1536	4	2560x1600	2 >*
		K260Q		2	4	1920 MB	768	4	2560x1600	4
		K240Q		4	8	960 MB	384	2	2560x1600	8
		K220Q		8	16	512 MB	96	2	2560x1600	16
		K200		8	16	256 MB	96	2	1920x1200	16
NVIDIA GRID K1	4	(pass-through)		1	4	4096 MB	192	2	2560x1600	4 >*
		K140Q		4	16	960 MB	48	2	2560x1600	16
		K120Q		8	32	512 MB	no Quadro function Support			32
		K100		8	32	256 MB				32

vGPU和物理GPU性能参数对比



vGPU切割原则

GRID K2

Physical GPU 0

Physical GPU 1

Valid configuration with K240Q vGPUs on GPU 0, K260Q vGPUs on GPU 1:

K240Q K240Q K240Q K240Q

K260Q

K260Q

Valid configuration with K200 vGPUs on GPU 0, K240Q vGPUs on GPU 1:

K200 K200 K200 K200 K200 K200 K200

K240Q

K240Q

K240Q

K240Q

Invalid configuration with mixed vGPU types on GPU 0:

K240Q

K240Q

K260Q

K200

K200

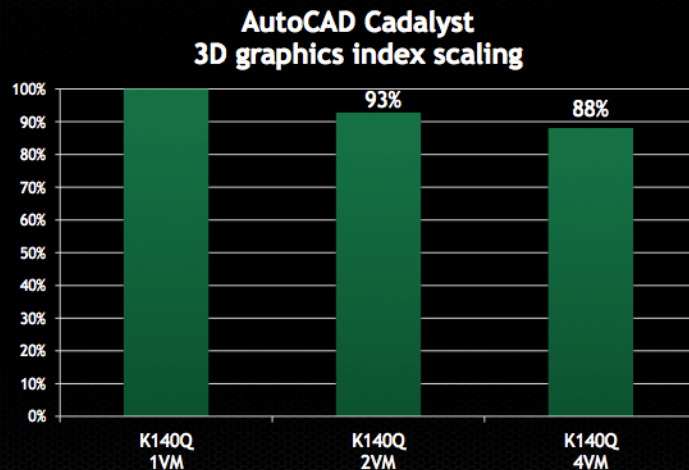
K200

K200

K200

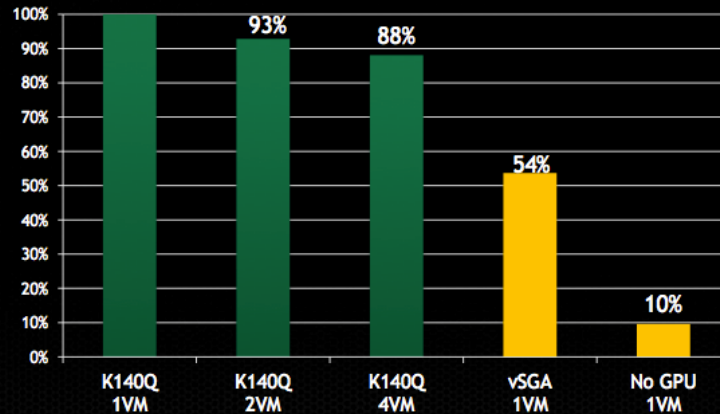
vGPU的性能损耗

Performance Scaling



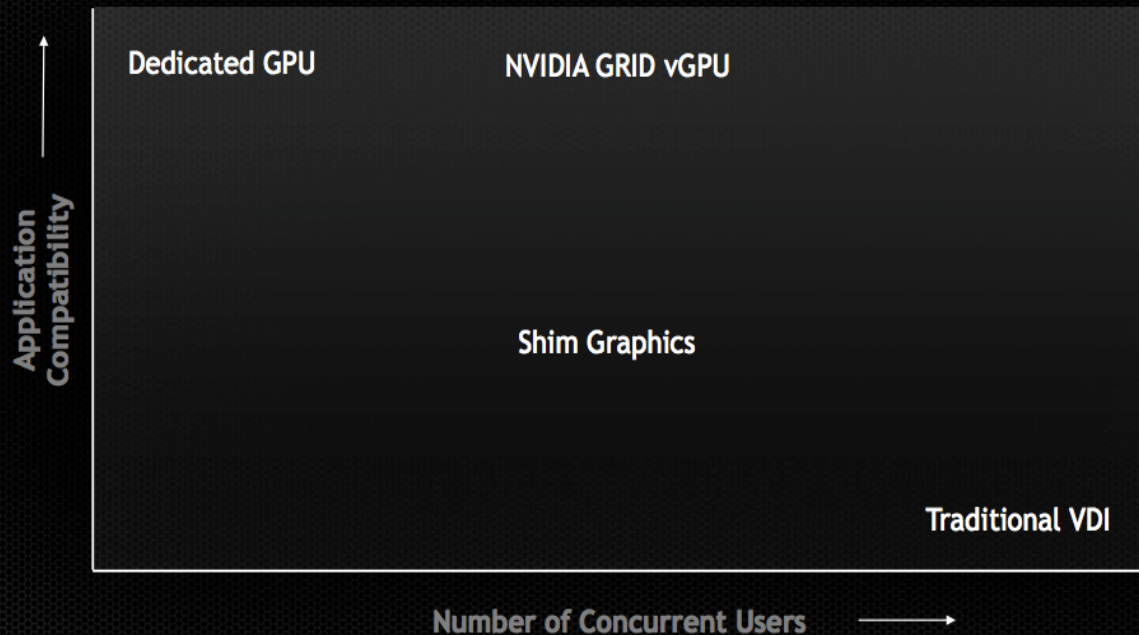
Performance Scaling




AutoCAD Cadalyst
3D graphics index scaling



Citrix vGPU vs VMware vSGA

GRAPHICS ACCELERATED VDI



	Software 3D	Shim Graphics	GRID vGPU	Pass-through
DESIGNER 	✗	✗	✓	✓
POWER USER 	✗	✗ ✓	✓	✓
KNOWLEDGE WORKER 	✓	✓	✓	

Gartner咨询公司针对于Citrix vGPU和VMware vSGA两种显卡虚拟化技术提供的性能分析视频

<http://blogs.gartner.com/gunnar-berger/understanding-virtual-desktop-vdi-gpu-technologies/>

XenDesktop 7.1 NVIDIA GRID vGPU vs vSphere 5.5 VSGA

http://v.youku.com/v_show/id_XNjI4ODU4OTE2.html

NVIDIA GRID vGPU on XenServer vs VMware vSphere vSGA

http://v.youku.com/v_show/id_XNjI4ODU5Nzgw.html



为用户选择合适的方案

设计用户场景定义

场景	建议安敢选型
查图、看图，中、轻量级图形编辑（零部件设计）	Shared GPU
中、重度，大规模的图形编辑、影像处理	vGPU
极重度图形编辑、影像处理	Dedicated GPU

场景、应用选型 – 仅供参考

	应用	备注
Shared GPU（K5000）	AutoCAD, Catia	以Catia v5为例，1G以内的图，6并发。
vGPU – K2	Catia, Solidwork, Maya, Revit, 3D MAX	以Catia v5为例，1G的图纸K240Q无压力，2G以内K260Q无压力。
vGPU – K1	AutoCAD, SolidWork	
Dedicated GPU – K2	Maya, Cero,	以Maya为例，电影、动画级别只能用透传
Dedicated GPU– K1	AutoCAD-3D, SolidWork, 3D MAX	

vGPU软硬件配置 – 参考

服务器配置	中、轻量vGPU交付	中、重度vGPU交付
CPU	Intel E5-2650 v2 * 2 (2.6G 8核)	Intel E5-2690 v2* 2 (3.0G 10核)
RAM	160G	160G
GPU类型	2* K1 (32 K140Q)	2 * K2 (16 K240Q)
电源	1100W *2	1100W *2

虚拟机配置		
vCPU	2	4
RAM	4G	8G
vGPU	K140Q	K240Q

vGPU POC最佳实践

基本步骤

1. 全新安装XenServer6.2，然后下载安装SP1补丁，SP1补丁实现vGPU功能的支持
2. 下载Nvidia VGX安装包，并把此组件安装到XenServer Dom0中
下载地址：<http://www.nvidia.com/download/driverResults.aspx/71474/en-us>
3. 安装XenCenter6.2.2版本，否则在界面上没有控制GPU的菜单和选项
下载地址：<http://support.citrix.com/article/CTX139790>
4. 安装XenDesktop 7.5相关组件，并在虚拟桌面中安装HDX 3D Pro Agent
5. 通过Citrix Studio中通过MCS批量交付带有vGPU的虚拟桌面

详细的图文部署手册：<https://citrix.sharefile.com/d/sbc89d004d3340df9>

选择正确XenCenter版本

- 管理XenServer 6.2SP1的环境，请先升级XenCenter至最新版本，否则无法正常管理此环境。 <http://support.citrix.com/article/CTX139790>



环境建议

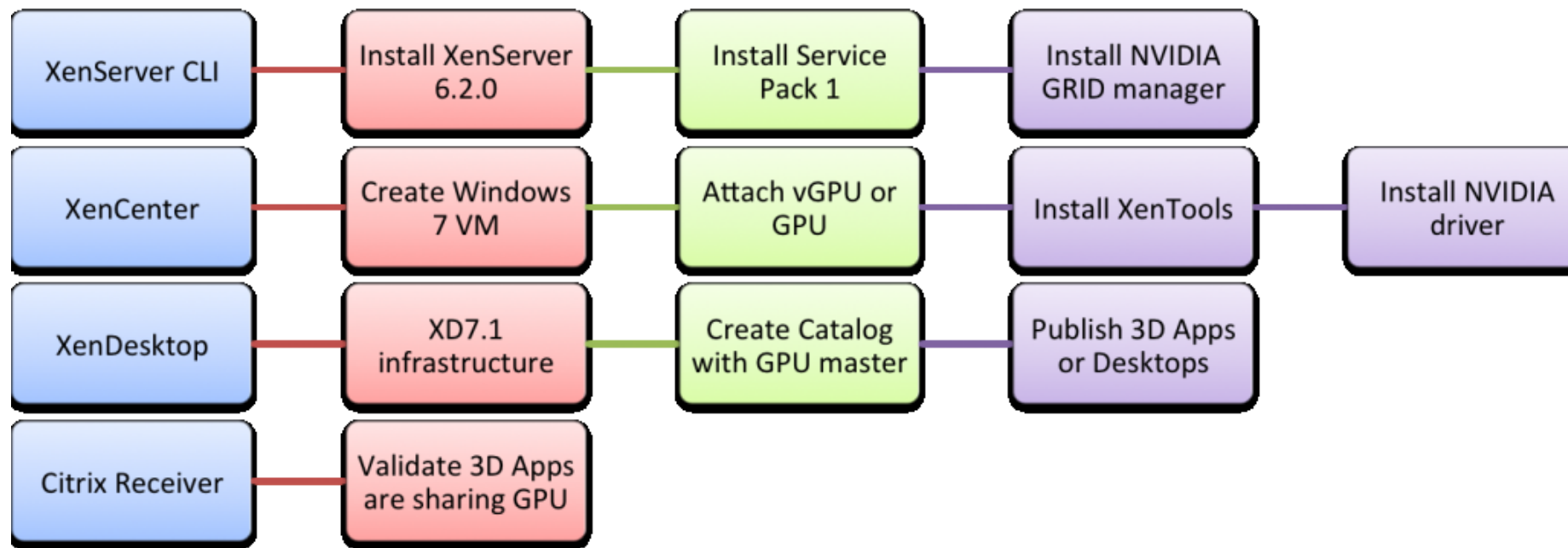
- vGPU交付可以通过MCS完成，现已支持。
- vGPU环境搭配建议：
 - XenServer 6.2 SP1（建议更新到最新补丁） + XenDesktop 7.1/7.5 + Windows 7 SP1 x64 + Receiver 4.1
- 请先安装NV 驱动，再安装HDX 3D VDA。如果安装顺序颠倒，性能最高损耗40%。
- 现阶段官方只支持Windows 7/8/8.1 X86/X64系统。
 - PoC阶段，建议配置为不少于4vCPU，8G内存。
 - vGPU不支持Windows XP系统。
- 带宽需求：
 - LAN: 100M
 - WAN: 1.5Mbps起，5Mbps能保证较好效果。同时网络延时不能超过150ms，50ms以内为好。

XenServer优化 – 参考

- 确保物理服务器上，确保前文提到的Bios中的“4G开关”关闭，
- 确保物理服务器上，性能已经设定为Max Performance（每家服务器设定不同，请联系对应厂商）
- 在高负载高密度的场景下，请将Dom0分配8个vCPU确保XenServer的稳定性。
- 在高负载场景下，请将虚拟机的vCPU与物理CPU进行绑定，减少虚拟机中应用程序运算到CPU与GPU的调度损耗的时间。（如高清视频编辑）
- 确保物理服务器上，C-States 和Turbo Mode选项设定为开启。（**请注意此选项暂时只适用于vGPU环境，勿在普通XenDesktop场景中启用**）
- 在XenServer上运行如下命令并重启：**（此命令需配合C-States and Turbo Mode开关使用，否则将无效）**

```
/opt/xensource/libexec/xen-cmdline --set-xen cpufreq=xen:performance
```

推荐安装顺序



选择经过兼容性认证的硬件平台

GRID CERTIFIED SERVERS

GRID-ENABLED SERVERS			
 <p>Cisco UCS C240 M3 up to 2 GRID K1 or 2 GRID K2</p>		 <p>Dell PowerEdge R720 up to 2 GRID K1 or 2 GRID K2</p>	
 <p>HP ProLiant WS460c Gen8 up to 1 GRID K1 or 1 GRID K2 HP ProLiant SL250s Gen8 up to 3 GRID K2</p>		 <p>IBM iDataPlex dx360 M4 up to 2 GRID K1 or 2 GRID K2 IBM Flex System up to 1 GRID K1 or 1 GRID K2</p>	

Vendor	Device Name/Model	Supported Servers	Supported XS Version
+ NVIDIA	Quadro 2000	Dell Precision R5500, BIOS revision A01	6.0
+ NVIDIA	Quadro 2000	IBM System x3650 M3 [with I/O Riser Card(s)]	6.0
+ NVIDIA	Quadro 4000	HP ProLiant WS460c G6 [with PCI Expansion Blade(s)], BIOS revision I24	6.0
+ NVIDIA	Quadro 4000	SGI Asterism ID112	6.0
+ NVIDIA	Quadro 5000	HP ProLiant WS460c G6 [with PCI Expansion Blade(s)], BIOS revision I24	6.0.2
+ NVIDIA	Quadro 6000	HP ProLiant WS460c G6 [with PCI Expansion Blade(s)], BIOS revision I24	6.0.2
+ NVIDIA	Tesla M2070Q	HP ProLiant WS460c G6 [with PCI Expansion Blade(s)], BIOS revision I24	6.0.2
+ NVIDIA	Grid K1	Dell PowerEdge R720, minimal BIOS revision 1.6.0	6.1, 6.2
+ NVIDIA	Grid K1	Cisco UCS C240 M3, BIOS revision C240M3.1.5.3b.0	6.1, 6.2
+ NVIDIA	Grid K2	Dell PowerEdge R720, minimal BIOS revision 1.6.0	6.1, 6.2
+ NVIDIA	Grid K2	Cisco UCS C240 M3, BIOS revision C240M3.1.5.3b.0	6.0.2, 6.1, 6.2
+ NVIDIA	Quadro 6000	HP WS460c Gen8, BIOS revision 9/20/13	6.0.2, 6.1, 6.2

Vendor	Device Name/Model	Supported Servers	Supported XS Version
+ NVIDIA	GRID K2	Dell PowerEdge R720 BIOS Version 2.0.19	6.2
+ NVIDIA	GRID K1	Cisco C240 M3 BIOS Version 1.5.3b	6.2
+ NVIDIA	GRID K1	Dell PowerEdge R720 BIOS Version 2.0.19 08/29/2013	6.2
+ NVIDIA	GRID K2	IBM iDataPlex DX360 M4 BIOS Version 1.31	6.2

vGPU驱动程序注意事项

NVIDIA GRID vGPU SOFTWARE RELEASE 331.59.01/332.83

版本:	332.83 WHQL
发布日期:	2014.8.22
操作系统:	Citrix XenServer 6.2
语言:	Chinese (Simplified)
文件大小:	

下载

发布重点

产品支持列表

其他信息

The release package includes both Windows Display Driver (332.83) and GRID vGPU Manager (331.59.01)

This software package enables NVIDIA GRID vGPU providing shared hardware based graphics acceleration for Windows VMs running on Citrix XenServer 6.2 SP1 or higher.

Note: If you do not run NVIDIA GRID vGPU, NVIDIA recommends using driver version 332.76 or higher.

New in this Release:

- Support for Citrix XenServer hotfix XS62ESP1009, see <http://support.citrix.com/article/CTX141191>
- Various bug fixes

Existing Support:

- Support for OpenGL 4.4
- Support for DirectX 11.1

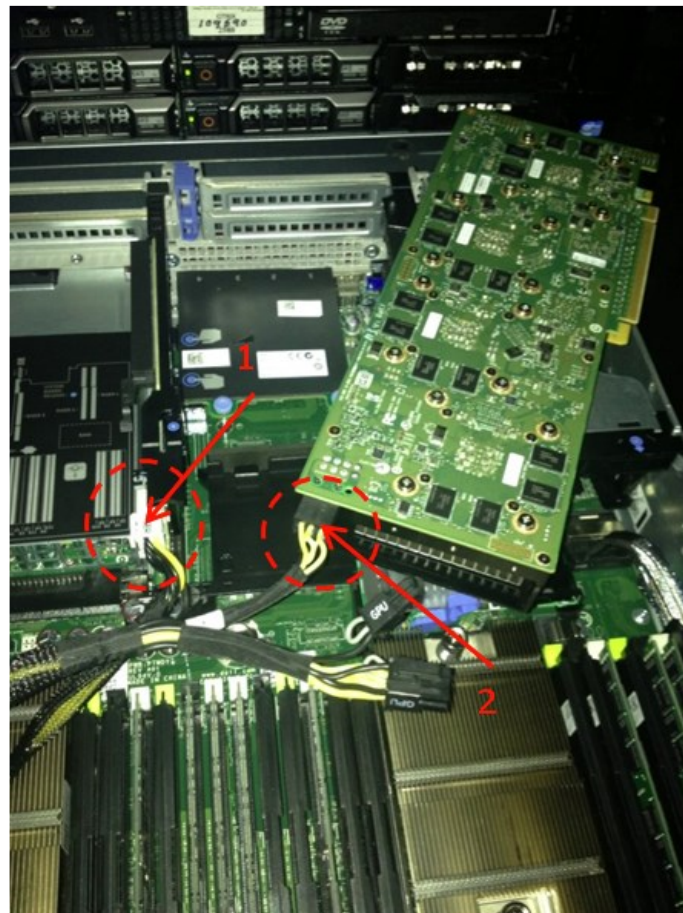
服务器的BIOS设置

- 部分早期生产的服务器，由于BIOS相对较旧，故可能出现兼容等问题，建议参考HCL列表对照，以确保服务器的BIOS要新于HCL列表所列。
 - <http://hcl.xensource.com/GPUPass-throughDeviceList.aspx>
- 部分服务器在需要BIOS进行如下设置，vGPU功能才可以正常工作
如何通过BIOS设置解决GPU内存映射的问题
<http://support.citrix.com/article/CTX139834>

K1/K2卡混插的问题

- 目前GIRD 认证的4家6款服务器都不支持K1/K2卡的混插。
- 目前确认可以支持K1/K2混插的为SuperMicro的服务器，其他渠道品牌暂时无法确认是否支持。

8Pin电源线 – 显卡感叹号



协议支持

Dedicated GPU	vGPU	Shared GPU
驱动版本： 331.82 <ul style="list-style-type: none">• OpenGL 4.4• OpenCL 1.1• CUDA• Direct X 9/10/11	驱动版本： 331.82 <ul style="list-style-type: none">• OpenGL 4.4• Direct X 9,10,11	驱动版本： 331.65 <ul style="list-style-type: none">• OpenGL 4.4• OpenCL 1.1• CUDA• Direct X 9/10/11

终端设备的选择

- 测试阶段，尽量使用普通PC/笔记本来做测试，机器老点不怕。
- 设备主频最好在2.0GHZ以上，不要少于1.6GHZ。
- 瘦客户机，个人还未严格测试过。但是建议使用Citrix Ready中已经通过HDX Premie认证的机器。
 - 对于 x86 设备，处理器速度至少为 1.6 GHz，才能在典型分辨率（例如 1280 x 1024）下很好地显示单监视器会话。如果使用 HDX 3D Pro 功能，需要安装本机硬件加速的图形驱动程序，且处理器的最低速度为 2 GHz。
 - 对于 ARM 设备，需要使用硬件 H.264 解码器才能同时实现常规 H.264 支持和 HDX 3D Pro 功能。此外，处理器时钟速度越快，性能越优异。
 - Intel Atom芯片貌似测试效果较差，慎重选用。

Receiver版本的选择

- 在XD7.x HDX 3D场景中，传输协议已经全部使用H.264解码规范，所以支持H.264的Windows Receiver 4.1、MAC Receiver 11.8.2、iOS Receiver 5.8、Android Receiver 3.4、Linux Receiver 13.0都可以用来访问。测试阶段推荐使用Windows版本(目前推荐版本：4.1.200)。
- 未来Receiver for HTML5也将支持H.264。

虚拟桌面资源规划一般原则

虚拟桌面类型	vCPU	Memory	HardDisk
vGPU/Dedicated GPU	4-8vGPU	每虚拟桌面4-16GB 内存	30iops
Shared GPU	每共享桌面平均 占用1CPU资源	每共享桌面保证 1GB内存	4-8iops

物理CPU的主频建议选择主频3.0GHZ以上，CPU主频对用户体验的影响很大，因为目前大多数的设计软件不支持多核。

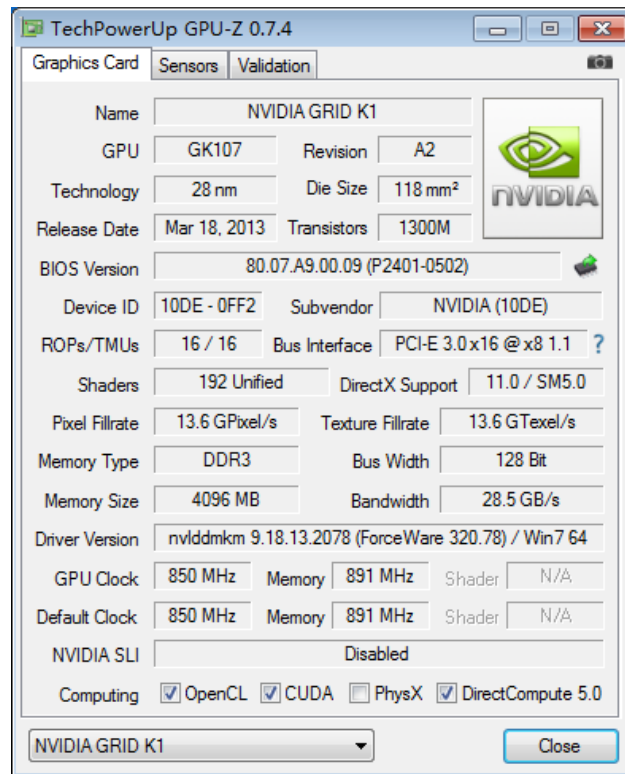
详细的规划设计文档：<https://citrix.sharefile.com/d/s9178c823a6a4101b>

虚拟机资源规划

- 目前大部分3D制图软件，仍然是**靠CPU完成解码**。所以在配置物理GPU之后，CPU性能的高低与GPU一起决定着3D场景的体验效果。
- 目前绝大多数的制图软件仍然是单线程程序，所以服务器端CPU的主频高低也直接影响操作、出图效果。
- 过多的vCPU并不一定都能起到更好的效果，VDI场景个人建议4vCPU~6vCPU，GPU Sharing使用8vCPU。8vCPU以上，从目前测试结果看效果不明显。
- 由于目前支持vGPU的服务器多以2路服务器为主，故最多只能有20个core(E5系列v2的CPU)。在高性能的3D的场景建议1个core对应一个VM。所以在不考虑超线程的情况下，建议在一台服务器就承载16个VM，正好对应的就是2块K2分为K240Q。

POC常用工具推荐

- 为了验证GPU的状态是否正常，可以考虑使用如下工具：
- GPU-Z
- GPU Caps View



GPU Caps View

Geeks3D GPU Caps Viewer - GP...

GPU | OpenGL | CUDA | OpenCL | Tools | About

GPU 1: GRID K2

GPU: unknown Shader: 1536

TDP: BIOS: 80.04.60.00.2f

Mem size: 4096MB Mem type: GDDR5, 256-bit

Device ID: 10DE - 11BF Subdevice ID: NVIDIA (10DE-100D)

GPU temp: 46.0°C / 114.8°F Fan speed:

Current clock: GPU core: 744 MHz, mem: 2497 MHz

Max clock: base: 745 MHz, boost: 745 MHz, mem: 2500 MHz

GPU load: 15.0% VDDC: current: 1.012 V

GPU power: current: 40.0 % TDP, target: 100.0 % TDP

Driver: 9.18.13.2078 (8-9-2013), R320.78 (branch: r319_00-304)

OpenGL: OpenGL 4.3 (Quadro FX GRID K2/PCIe/SSE2 with 306 ext.)

OpenCL: OpenCL 1.1, GRID K2 compute units:8@745MHz

CUDA: GRID K2 CC:3.0, multiprocessors:8@745MHz

PhysX: GPU PhysX (GRID K2)

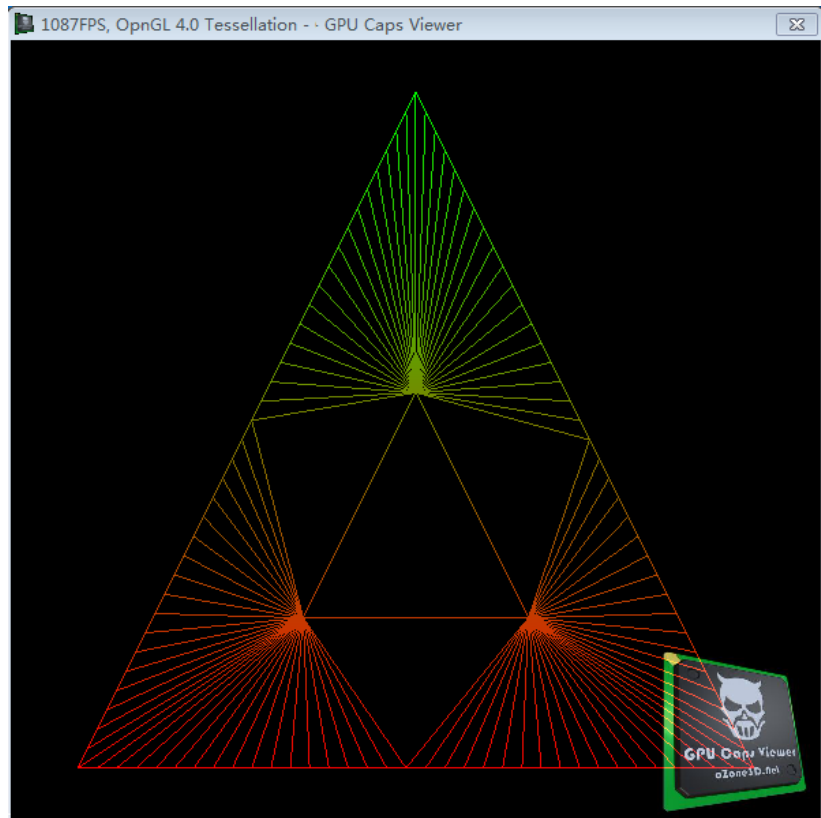
Multi-GPU: no multi-GPU support (1 physical GPUs)

OpenGL and OpenCL demos

GL 4.x - Tessellation CL GPU - 4D Quaternion Julia Se

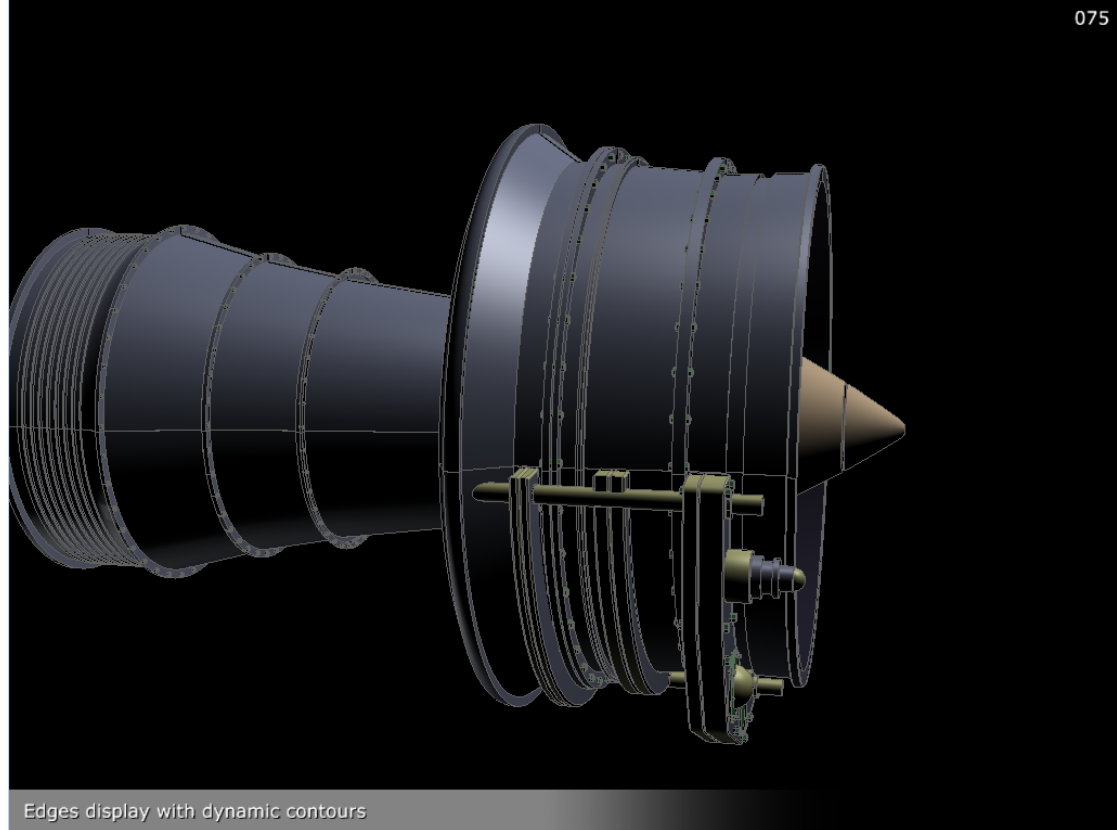
OpenGL demos Start OpenCL demos Start

Build: 1.18.1 [Jun 17 2013 @ 16:55:36] Quit



Benchmark Tools

- Redway
TurbineDemo(OpenGL)
- Heaven Benchmark(DX)
- SPECviewperf11win64



显卡性能调优

选择一项任务...

- 3D 设置
 - 通过预览调整图像设置
 - 管理 3D 设置
- 工作站
 - 查看系统布局
- 显示
 - 更改分辨率
 - 设置多个显示器
- 视频
 - 调整视频颜色设置
 - 调整视频图像设置

全局设置 程序设置

全局预设 (G):

基本配置文件

恢复 (T)

设置 (S):

功能	设置
CUDA - GPUs	全部
OpenGL 渲染 GPU	自动选择
三重缓冲	关
各向异性过滤	应用程序控制的
启用覆盖功能	关
垂直同步	关
多显示器/混合 GPU 加速	多显示器性能模式
导出的像素类型	颜色索引覆盖 (8 bpp)
平滑处理 - FXAA	关
平滑处理 - 模式	应用程序控制的

查看GPU利用率 – NVSMI命令

- 在使用GPU运行时，在Windows系统的cmd下，运行如下命令，可查看GPU的利用率：
 - `C:\Program Files\NVIDIA Corporation\NVSMI\nvsmi.exe -a -l -d utilization`
- 此命令同样适用于在XenServer上查看GPU的利用率，`nvsmi.exe`更改为`nvidia-smi`即可

AutoCAD

- 在AutoCAD中运行3dconfig命令，确保在“手动调节”选项中的“启用硬件加速”都勾选。
- AutoCAD请使用2012以后的版本，过久的版本效果较差。
- AutoCAD 2008之前的版本请不用在vGPU场景中测试，其Windows XP平台不在支持之列。



AutoCAD鼠标指针的问题

- 问题描述:

当在应用程序使用非标准的Windows Cursor时会出现显示不正常。典型的应用如: Autodesk AutoCAD, 它在非工作区时Cursor是箭头, 工作区是十字架, 拖动时是手型。在PC机中, 这些鼠标变化都是正常的, 但是在XenDesktop中会出现鼠标显示不正常的状态。这种情况会严重影响专业人员作图和设计。

- 环境:

XenDesktop 7.5\AutoCAD 201X\Sobey

- VDA补丁建议:

建议VDA安装 004、005 (可以修补vGPU/GPU鼠标黑框问题, Blank Box, 如上图所示)

解决方法详细文档下载: 链接: <http://pan.baidu.com/s/1kTBm6ZP> 密码: o5ij

AutoCAD在Shared desktop环境中的优化

- Solution 1:

On AutoCAD, go to 3D Config and enable h/w accelerator (Tick the box)

- Solution 2:

On both Windows Server 2008 R2 and Windows Server 2012 create EnableWPFHook:

```
[HKEY_LOCAL_MACHINE\SOFTWARE\Citrix\CtxHook\AppInit_Dlls\ Multiple Monitor Hook] "EnableWPFHook"=dword:00000001
```

```
[HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\Citrix\CtxHook\AppInit_Dlls\ Multiple Monitor Hook] "EnableWPFHook"=dword:00000001
```


AutoDesk 相关软件安装问题

- Revit 2014 安装的时候建议断网（否则由于联网下载组件的问题导致安装时间非常长）
- AutoCAD安装的时候不要选择安装补丁，来提高安装速度

Inventor 优化

- 定位到Inventor如下目录的C:\Program Files\AutoDesk\Inventor 2014\Configuration文件夹
- 用文本编辑器编辑HardwareLibraryDM.xml文件
- 在<SETTINGS FeatureLevelLimit="50" />下增加如下条目：（注：OS value=xx有多个，请每个下面都按此方式添加。

```
<SYSVENDOR value="Citrix.*">
```

```
<!--Citrix Virtual OS Systems-->
```

```
<SETTINGS FeatureLevelLimit="50" />
```

```
</SYSVENDOR>
```

```
<OS value="Win7">
```

```
<SETTINGS FeatureLevelLimit="50" />
```

```
<SYSVENDOR value="Citrix.*">
```

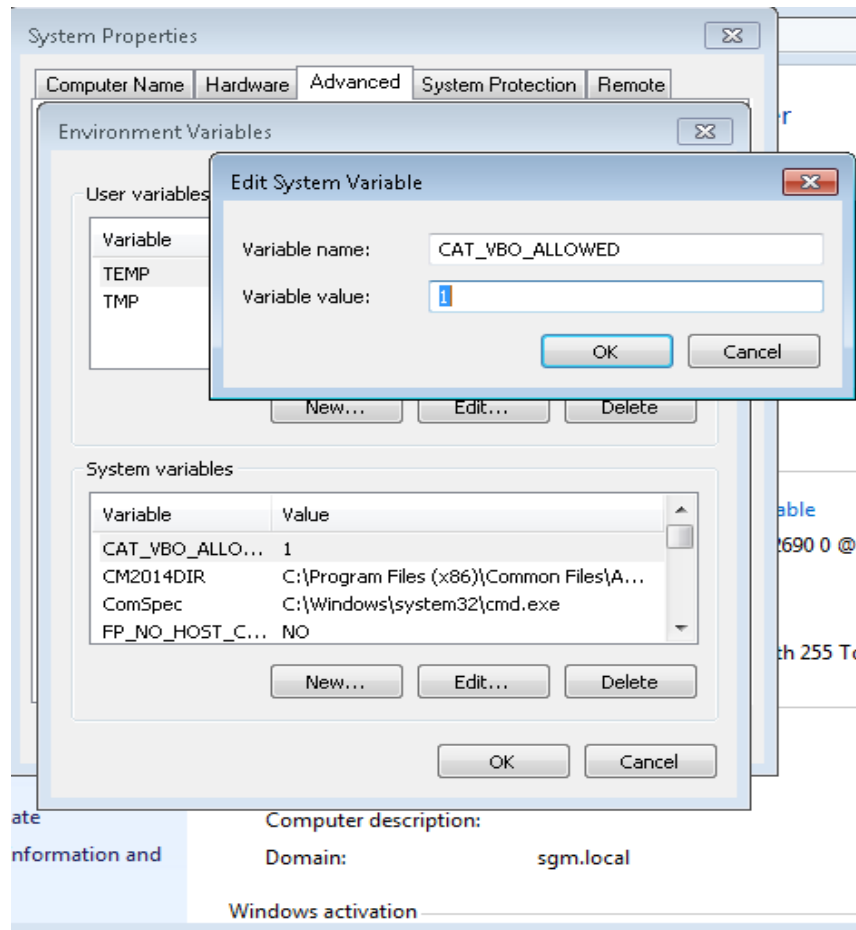
```
<!--Citrix Virtual OS Systems-->
```

```
<SETTINGS FeatureLevelLimit="50" />
```

```
</SYSVENDOR>
```

Catia V5优化

- Catia v5
- 添加系统环境变量，参数：
CAT_VBO_ALLOWED，数值
为：1



Photoshop CS

- Photoshop CS 2在XD75中会有鼠标指针无法正常切换的问题，切换到CS6版本，没有此问题。

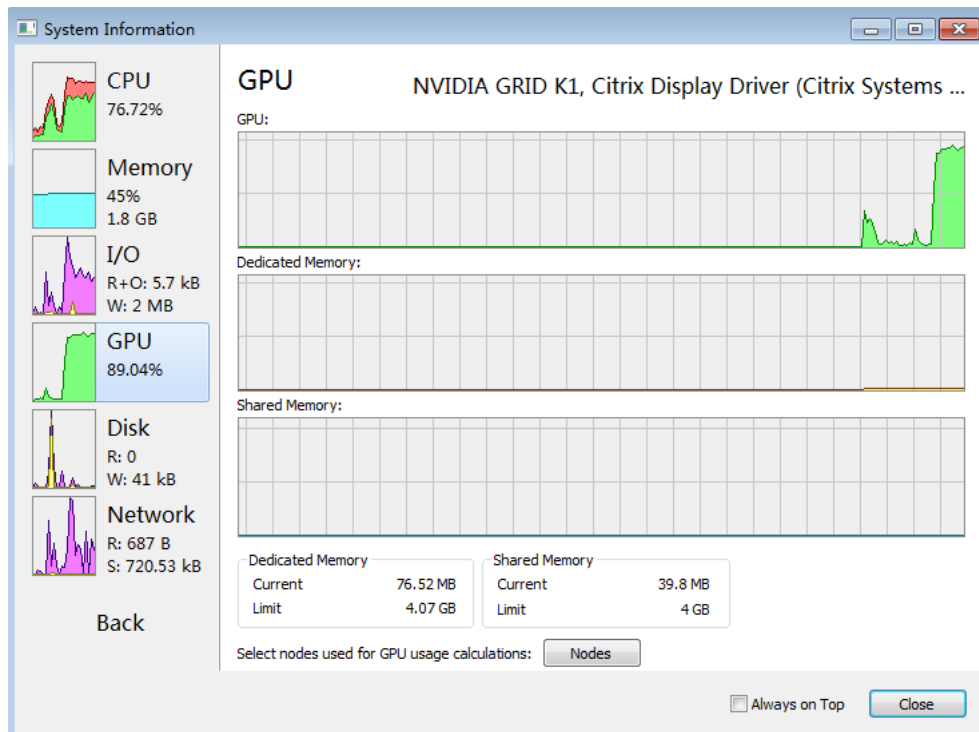
应用已知问题

- Catia v6可能会遇到光标消失的问题。
 - 定位到 C:\Program Files\Citrix\ICAService
 - 将Hd3dProvidersApollo.dll 更名 Hd3dProvidersApollo_backup.dll
 - 从如下地址下载替换的HDX DLL 文件
http://downloadns.citrix.com.edgesuite.net/8193/3DPro_vGPU_cursor_fix.zip
 - 将Hd3dProvidersApollo.dll 放置到C:\Program Files\Citrix\ICAService
 - 打开注册表并定位到
HKLM\Software\Citrix\HDX3D\BitMapRemotingConfig
 - 创建一个 DWORD: HKLM_CatiaV6Cursor
 - 设置数值为： 1
 - 将VM重启

光标显示黑色方块

- 在使用部分3D软件时，鼠标键的光标会显示为黑色的方块。
- 已经验证出现此问题并解决的有：ProE，PTC Creo 2.0，AutoDesk Inventor，方正排版软件。
- 请安装Hotfix XD710ICAWSWX--007
 - <http://support.citrix.com/article/CTX140736>
 - <http://support.citrix.com/article/CTX140737>

查看GPU利用率 - Process Hacker



案例分析

XenApp 共享GPU桌面 – 建筑设备行业

(设计人员使用的软件Catiav5\SolidEdge ST4) 当前已上线用户100人， 共计300用户

	传统PC	共享GPU桌面 - XenApp	对比
硬件投入	高性能PC机 * 120	服务器*5 + NetApp 存储 + Citrix License + 瘦客户机*120	投入成本接近
新员工桌面交付时间	半个工作日	10分钟	更有效
设计文件安全	困难 – 数据加密系统\外设控制\DLP系统	简便 – 通过HDX策略控制	更安全
为供应商提供设计模型	困难 – 很容易导致数据泄露	简便 – 所有数据存储在数据中心	更安全
员工更换工位	2小时/员工	20分钟/员工	更有效
桌面运维支持	<100 PC/人	>1000 虚拟桌面/人	更有效
硬件更新周期	3年	超过五年	硬件投资保护

XenDesktop vGPU 桌面 – 汽车制造行业

(设计人员使用的软件Catiav5/Siemens NX) 一期试点用户为30人，总计200设计人员

客户需求：解决设计图纸安全

POC: Citrix vGPU vs VMware vSGA

POC结果：VMware vSGA虚拟桌面无法满足用户对性能的要求而出局

客户设计用户日常处理的图形文件大小600M-1.5GB左右

设计文件(600MB/1.5GB)	HP Z800	vGPU k240Q
软件启动速度，文件读取速度，操作、编辑速度，出图速度	当前设计人员所使用的图形工作站	整体体验优于HP Z800

Windows 7 Aero & Windows 8 desktops/apps

DirectX将替换GDI/GDI+成为新的Windows的图形标准

DirectX 渲染技术		RDS Workloads	VDI Workloads	VDI-in-a-Box
Server	“Software GPU”	XD7.0 (req. WS12)	XD7.0	ViaB 5.3
	Hardware GPU acceleration	XA6.0, XD7.0	XD7.0 (reg key) [XD4: HDX 3D Pro]	ViaB 5.3 (reg key)
Client	Desktop Composition Redirection	Not applicable	XD5.5	ViaB 5.0

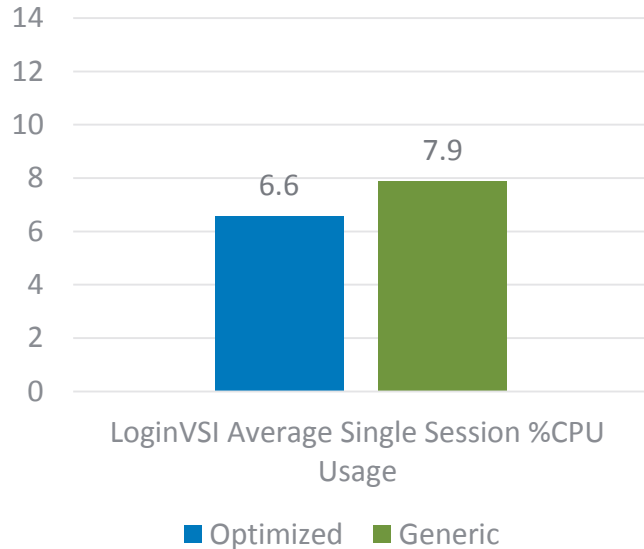
HDX 3D vs HDX 3D Pro

- OpenGL support
- Lossless

There are two types of VDAs and it is important to understand the naming convention and support for each. For traditional VDI, the **XenDesktop Standard** provides virtual machines, access to leveraging a GPU, for certain graphics engines such as Direct X11. This is a new feature and only available in XenDesktop 7.1 and above. More about this can be found [here](#). In the past to leverage a GPU for a desktop you needed HDX3DPro. This is no longer the case for XenDesktop 7.1 as the Standard VDA can use a GPU ONLY with applications that use the Direct X11 graphics engine like Microsoft Office and Google Earth. The second VDA is the **HDX 3D Pro**. **HDX 3DPro** provides direct access to graphics vendor's hardware and supports the latest in graphics engines such as OpenGL and so forth. **The HDX 3D Pro is ONLY SUPPORTED BY CITRIX WITH NVIDIA CARDS AT THIS TIME.**

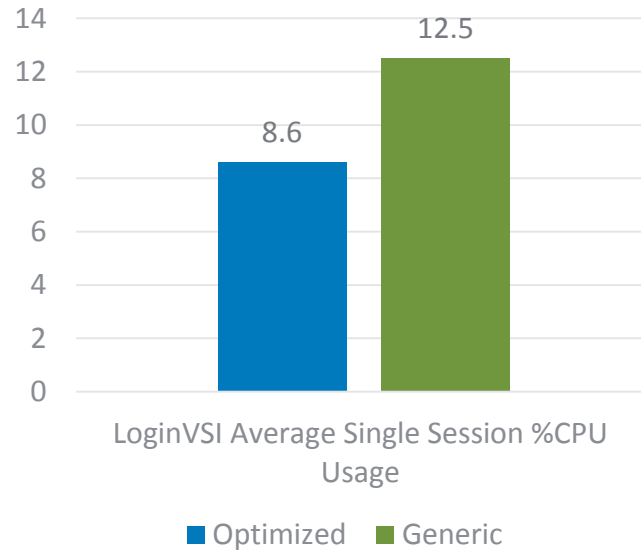
Desktop Composition Redirection optimization benefits

%CPU Usage on
Win8 VDI



15-20% benefit in Single Server Scalability
using Desktop Composition Redirection

%CPU Usage on
Win7 Aero VDI

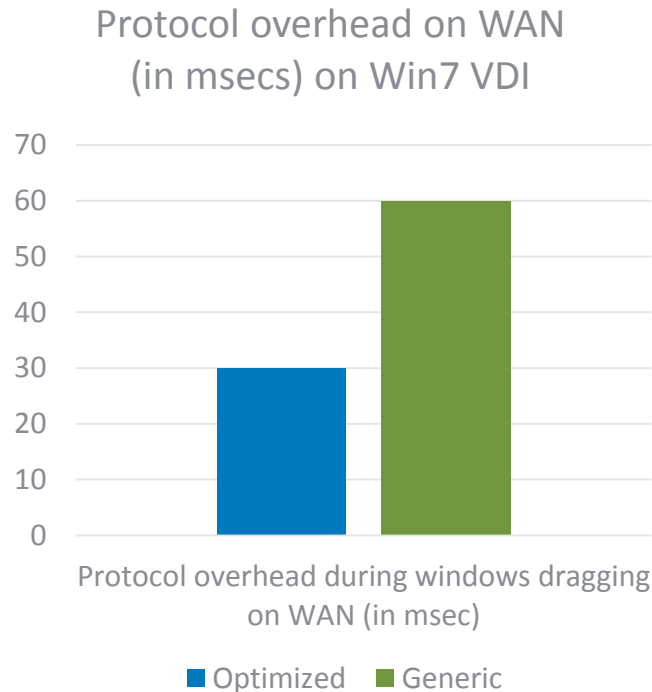


35-45% benefit in Single Server Scalability
using Desktop Composition Redirection

Note:

- Animations & window shadows disabled in generic solution to minimize server scalability impact
- Transparency is enabled on Win7

Desktop Composition Redirection benefits – cont'd



Desktop Composition Redirection provides
highest frame rate, most interactive experience

3D设计场景服务器推荐

Cisco UCS C240 M3	E5-2690v2/E5-2667 v2(130w)	128GB	SSD x4	1000Mbps	K2
IBM iDataPlex DX360 M4	E5-2690v2/E5-2667 v2 (130w)	128GB	SSD x4	1000Mbps	K2
DELL R720	E5-2680 v2 (115w)	128GB	SSD x4	1000Mbps	K2

- 1.型号
- 2.电源
- 3.CPU

Hardware Selection Considerations for Virtualized Workstations

	Dell PowerEdge R720	Dell Precision R7610 – May 2013
Operating System	VM Ware ESX or Citrix XenServer	Citrix XenServer/HDX 3D Pro
Compute (CPU)	Xeon EP up to 115W	Xeon EP up to 150W
Memory	768GB	512GB
Manageability	iDRAC, ASM	OMCI, vPro, ASM
Network	4x 1G up to 2x 10G and 2x 1G NDC	2x 1G Native or 2x 10G on NIC
Dedicated GPU Users	2-4 Users NV GRID K1,K2 , AMD S7000/S9000	2-6* Users NV K2000, K4000, K5000, AMD W5000 *NV GRID K2A (Q4)
Shared GPU Options	NV GRID K1/K2 w OpenGL 4.X - Citrix Q3*	3x NV GRID K2A (Q4)
Scalability / Slots	7 Slots – 4 FH, 3 LP 2x 225W Double Wide	6 Slots – 5 FH/FL 1 LP 3x 225W Double wide
Storage	16x 2.5 SATA/SAS/SSD	6x 2.5 SATA/SAS/SSD
Hardware (OS level) ISV Certification	No	~100 Professional Applications
ISV Certification for Virtualized workstation Environments	Select coming	Coming
Dell Precision WS Support	No	Yes

Best Performing or Feature Rich for WS virtualization - low # of med-high end users

Cisco UCS C240 M3

Product ID (PID)	Intel Number	Clock Freq (GHz)	Power (W)	Cache Size (MB)	Cores	QPI	Highest DDR3 DIMM Clock Support (MHz) ¹
Intel Xeon E5-2600 v2							
UCS-CPU-E52697B	E5-2697 v2	2.70	130	30	12	8 GT/s	1866
UCS-CPU-E52695B	E5-2695 v2	2.40	115	30	12	8 GT/s	1866
UCS-CPU-E52690B	E5-2690 v2	3.00	130	25	10	8 GT/s	1866
UCS-CPU-E52680B	E5-2680 v2	2.80	115	25	10	8 GT/s	1866
UCS-CPU-E52670B	E5-2670 v2	2.50	115	25	10	8 GT/s	1866
UCS-CPU-E52667B	E5-2667 v2	3.30	130	25	8	8 GT/s	1866
UCS-CPU-E52660B	E5-2660 v2	2.20	95	25	10	8 GT/s	1866
UCS-CPU-E52650B	E5-2650 v2	2.60	95	20	8	8 GT/s	1866
UCS-CPU-E52640B	E5-2640 v2	2.00	95	20	8	7.2 GT/s	1600
UCS-CPU-E52637B	E5-2637 v2	3.50	130	15	4	8 GT/s	1866
UCS-CPU-E52630B	E5-2630 v2	2.60	80	15	6	7.2 GT/s	1600
UCS-CPU-E52620B	E5-2620 v2	2.10	80	15	6	7.2 GT/s	1600
UCS-CPU-E52643B	E5-2643 v2	3.50	130	25	6	8 GT/s	1866
UCS-CPU-E52650LB	E5-2650L v2	1.70	70	25	10	8 GT/s	1600
UCS-CPU-E52630LB	E5-2630L v2	2.40	60	15	6	7.2 GT/s	1600
UCS-CPU-E52609B	E5-2609 v2	2.50	80	10	4	6.4 GT/s	1333

Table 3 Available Intel CPUs: Intel Xeon E5-2600 v2 and E5-2600 series processor family CPUs

Product ID (PID)	Intel Number	Clock Freq (GHz)	Power (W)	Cache Size (MB)	Cores	QPI	Highest DDR3 DIMM Clock Support (MHz) ¹
Intel Xeon E5-2600²							
UCS-CPU-E5-2690	E5-2690	2.90	135	20	8	8 GT/s	1600
UCS-CPU-E5-2680	E5-2680	2.70	130	20	8	8 GT/s	1600
UCS-CPU-E5-2670	E5-2670	2.60	115	20	8	8 GT/s	1600
UCS-CPU-E5-2667	E5-2667	2.90	130	15	6	8 GT/s	1600
UCS-CPU-E5-2665	E5-2665	2.40	115	20	8	8 GT/s	1600
UCS-CPU-E5-2660	E5-2660	2.20	95	20	8	8 GT/s	1600
UCS-CPU-E5-2650	E5-2650	2.00	95	20	8	8 GT/s	1600
UCS-CPU-E5-2650L	E5-2650L	1.80	70	20	8	8 GT/s	1600
UCS-CPU-E5-2643	E5-2643	3.30	130	10	4	8 GT/s	1600
UCS-CPU-E5-2640	E5-2640	2.50	95	15	6	7.2 GT/s	1333
UCS-CPU-E5-2630	E5-2630	2.30	95	15	6	7.2 GT/s	1333
UCS-CPU-E5-2630L	E5-2630L	2.00	60	15	6	7.2 GT/s	1333
UCS-CPU-E5-2620	E5-2620	2.00	95	15	6	7.2 GT/s	1333
UCS-CPU-E5-2609	E5-2609	2.40	80	10	4	6.4 GT/s	1066

Notes . . .

1. If higher or lower speed DIMMs are selected than what is shown in the table for a given CPU, the DIMMs will be clocked at the lowest common denominator of CPU clock and DIMM clock.
2. 1866-MHz DIMMs cannot be used with Intel Xeon E5-2600 CPUs

IBM DX360 M4

Processor options

The dx360 M4 supports the processor options listed in the following table.

Table 4. Processor options (Part 1)

Feature code	Intel Xeon processors*	Where used
Intel Xeon Processor E5-2600 family (requires dx360 M4 base feature code A1T3)		
A1TE	Intel Xeon Processor E5-2603 4C 1.8GHz 10MB Cache 1066MHz 80W	-
A1TG	Intel Xeon Processor E5-2609 4C 2.40GHz 10MB Cache 1066MHz 80W	42x
A1TH	Intel Xeon Processor E5-2620 6C 2.0GHz 15MB Cache 1333MHz 95W	-
A1TJ	Intel Xeon Processor E5-2630 6C 2.3GHz 15MB Cache 1333MHz 95W	-
A1TS	Intel Xeon Processor E5-2630L 6C 2.0GHz 15MB Cache 1333MHz 60W	22x
A1TF	Intel Xeon Processor E5-2637 2C 3.0GHz 5MB Cache 1600MHz 80W	-
A1TK	Intel Xeon Processor E5-2640 6C 2.5GHz 15MB Cache 1333MHz 95W	-
A2FV	Intel Xeon Processor E5-2648L 8C 1.8GHz 20MB Cache 1600MHz 70W	-
A1TL	Intel Xeon Processor E5-2650 8C 2.0GHz 20MB Cache 1600MHz 95W	-
A1TT	Intel Xeon Processor E5-2650L 8C 1.8GHz 20MB Cache 1600MHz 70W	-
A1TR	Intel Xeon Processor E5-2658 8C 2.1GHz 20MB Cache 1600MHz 95W	-
A1TM	Intel Xeon Processor E5-2660 8C 2.2GHz 20MB Cache 1600MHz 95W	62x
A2FW	Intel Xeon Processor E5-2665 8C 2.4GHz 20MB Cache 1600MHz 115W	-
A1TQ	Intel Xeon Processor E5-2667 6C 2.9GHz 15MB Cache 1600MHz 130W	-
A1TU	Intel Xeon Processor E5-2670 8C 2.6GHz 20MB Cache 1600MHz 115W	-
A1TN	Intel Xeon Processor E5-2680 8C 2.7GHz 20MB Cache 1600MHz 130W	-
A3MY	Intel Xeon Processor E5-2690 8C 2.9GHz 20MB Cache 1600MHz 135W	-

Table 4. Processor options (Part 2)

Feature code	Intel Xeon processors*	Where used
Intel Xeon Processor E5-2600 v2 family (requires dx360 M4 base feature code A4BV)		
A455	Intel Xeon Processor E5-2603 v2 4C 1.8GHz 10MB Cache 1333MHz 80W	-
A456	Intel Xeon Processor E5-2609 v2 4C 2.5GHz 10MB Cache 1333MHz 80W	-
A45F	Intel Xeon Processor E5-2618L v2 6C 2.0GHz 15MB Cache 1333MHz 50W	-
A3ZA	Intel Xeon Processor E5-2620 v2 6C 2.1GHz 15MB Cache 1600MHz 80W	13x
A45G	Intel Xeon Processor E5-2628L v2 8C 2.2GHz 20MB Cache 1600MHz 70W	-
A457	Intel Xeon Processor E5-2630 v2 6C 2.6GHz 15MB Cache 1600MHz 80W	-
A45D	Intel Xeon Processor E5-2630L v2 6C 2.4GHz 15MB Cache 1600MHz 60W	-
A45A	Intel Xeon Processor E5-2637 v2 4C 3.5GHz 15MB Cache 1866MHz 130W	-
A458	Intel Xeon Processor E5-2640 v2 8C 2.0GHz 20MB Cache 1600MHz 95W	-
A45B	Intel Xeon Processor E5-2643 v2 6C 3.5GHz 25MB Cache 1866MHz 130W	-
A45H	Intel Xeon Processor E5-2648L v2 10C 2.0GHz 25MB Cache 1866MHz 70W	-
A40M	Intel Xeon Processor E5-2650 v2 8C 2.6GHz 20MB Cache 1866MHz 95W	63x
A45E	Intel Xeon Processor E5-2650L v2 10C 1.7GHz 25MB Cache 1600MHz 70W	-
A45J	Intel Xeon Processor E5-2658 v2 10C 2.4GHz 25MB Cache 1866MHz 95W	-
A3ZB	Intel Xeon Processor E5-2660 v2 10C 2.2GHz 25MB Cache 1866MHz 95W	-
A45C	Intel Xeon Processor E5-2667 v2 8C 3.3GHz 25MB Cache 1866MHz 130W	-
A40N	Intel Xeon Processor E5-2670 v2 10C 2.5GHz 25MB Cache 1866MHz 115W	83x
A3ZC	Intel Xeon Processor E5-2680 v2 10C 2.8GHz 25MB Cache 1866MHz 115W	-
A459	Intel Xeon Processor E5-2690 v2 10C 3.0GHz 25MB Cache 1866MHz 130W	-
A3TZ	Intel Xeon Processor E5-2695 v2 12C 2.4GHz 30MB Cache 1866MHz 115W	-
A3TY	Intel Xeon Processor E5-2697 v2 12C 2.7GHz 30MB Cache 1866MHz 130W	-

* Processor detail: model, core speed, cores, L3 cache, memory speed, TDP power

Recommended configuration from Walter:

	2 K2 adapters	2 K1 adapters
<u>Passthru</u>	<u>no</u> need to disable any PCI slots	Need to disable in UEFI one of the following slots: <u>#1</u> ,#2 or mezzanine
<u>vGPU</u>	Need to disable in UEFI one of the following slots: <u>#1</u> ,#2 or mezzanine	Not able to load minimum device drivers in available 32 bit memory. Do not use.

Further testing shows that 1 K1 adapter does not work in an iDataplex using vGPU mode. 1 K1 adapter works in the x240/PEN for vGPU mode.

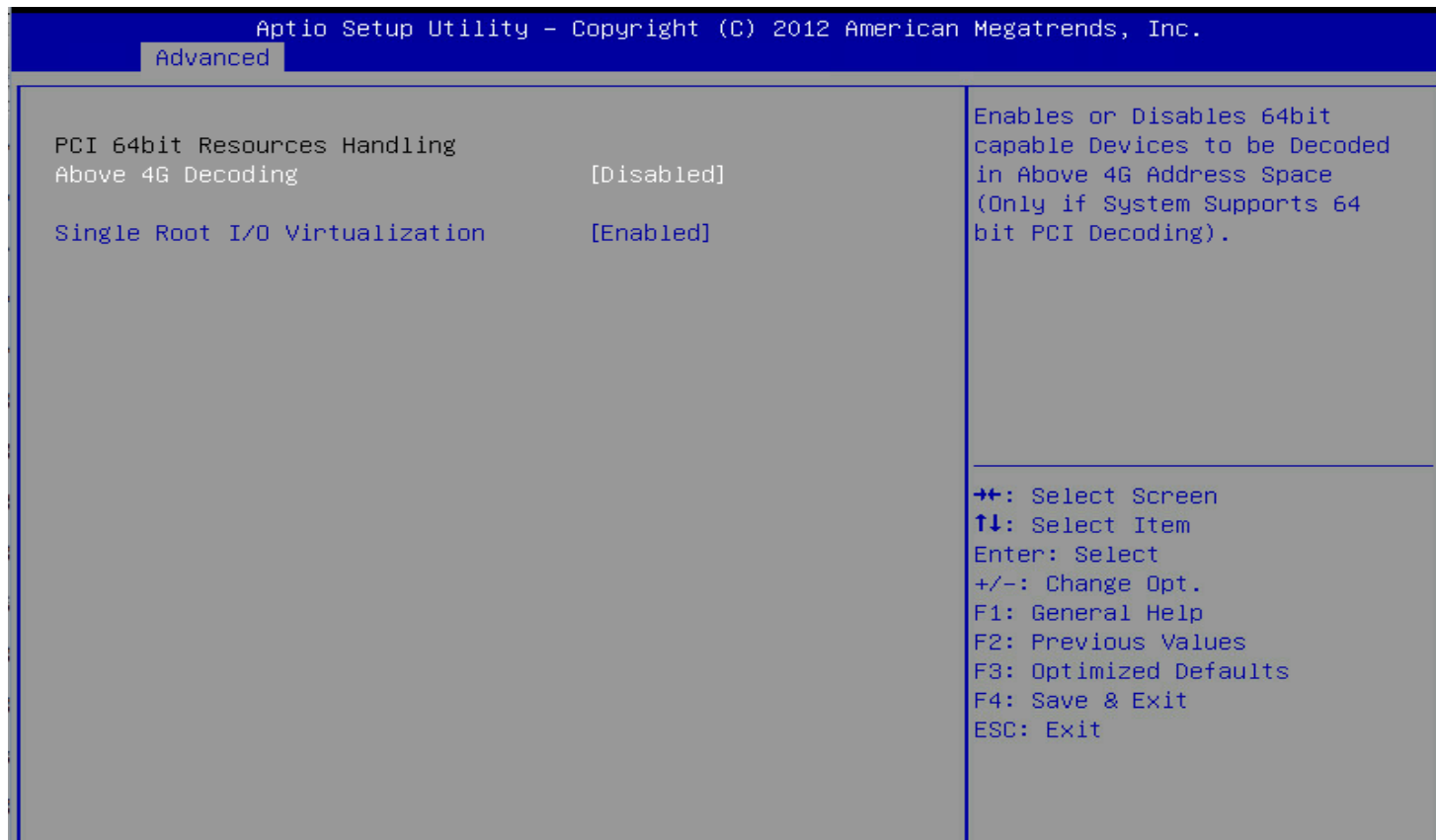
iDataplex UEFI level for vGPU mode

Recommend using UEFI 1.31 and IMM2 3.10 with k2 adapters for vGPU/XenServer. There is a bug in UEFI 1.40 that causes PCI64=Disable to be reset to Enable after a reboot.

Link for UEFI 1.31 and IMM2 3.10: http://www-933.ibm.com/support/fixcentral/systemx/quickorder?parent=System+x+iDataPlex+dx360+M4+server&product=ibm/systemx/7912&&platform=All&function=fixId&fixids=ibm_fw_uefi_tde134e-1.31_anyos_32-64&includeSupersedes=0&source=fc

If you need to fallback from UEFI 1.40, see page 254 in the following manual: <http://www-947.ibm.com/support/entry/portal/docdisplay?Indocid=migr-5089727>

BIOS设置(重点)



前端设备的选择

To access desktops or applications delivered with HDX 3D Pro, users must install Citrix Receiver. For more information on Receiver system requirements, see [Receiver and Plug-ins](#).

HDX 3D Pro supports all monitor resolutions that are supported by the GPU on the host computer. However, for optimum performance with the minimum recommended user device and GPU specifications, Citrix recommends maximum monitor resolutions for users' devices of 1920 x 1200 pixels for LAN connections and 1280 x 1024 pixels for WAN connections.

Citrix recommends that user devices include at least 1 GB of RAM and a CPU with a clock speed of 1.6 GHz or higher. Use of the default deep compression codec, which is required on low-bandwidth connections, requires a more powerful CPU unless the decoding is done in hardware. For optimum performance, Citrix recommends that users' devices are equipped with at least 2 GB of RAM and a dual-core CPU with a clock speed of 3 GHz or higher. For multi-monitor access, Citrix recommends user devices equipped with quad-core CPUs.

现有的缺陷

- vGPU不支持OpenCL CUDA（透传方案）
- Linux VDA目前不支持
- 服务器端硬编码H.264不支持（NV自有方案支持）
- 前端设备无法采用GPU解码H.264

Deep Compression for WAN and 3G

- Deep Compression codec now runs on CPU only (**no GPU option in XD7**)
- Significantly better performance than older CUDA-based implementation
- Doesn't compromise graphics rendering (doesn't compete for CUDA cores)
- H.264 decoding, previously only in Citrix Receiver for Windows and Linux, now also in Mac Receiver 11.8, iOS Receiver 5.8, Android Receiver 3.4 (Jul'13)
- Improved decoding efficiency for thin clients: 1.6 GHz CPU is sufficient now

后续计划:

- **H.264 decode in HTML5 Receiver**
- **Hardware-based H.264 decode for Windows thin clients (as on Linux)**

性能优化

- XenServer dom0 vGPU Pin
- SR-iov
- GPU CPU PIN (10% Performance enhance)
- CPU Turbo(10% Performance enhance)
- XenTools (CPU load decrease 40-50% Solidworks)
- VDA Patch (Mouse issues, vGPU performance enhance)

Pin Domain-0 vCPUs

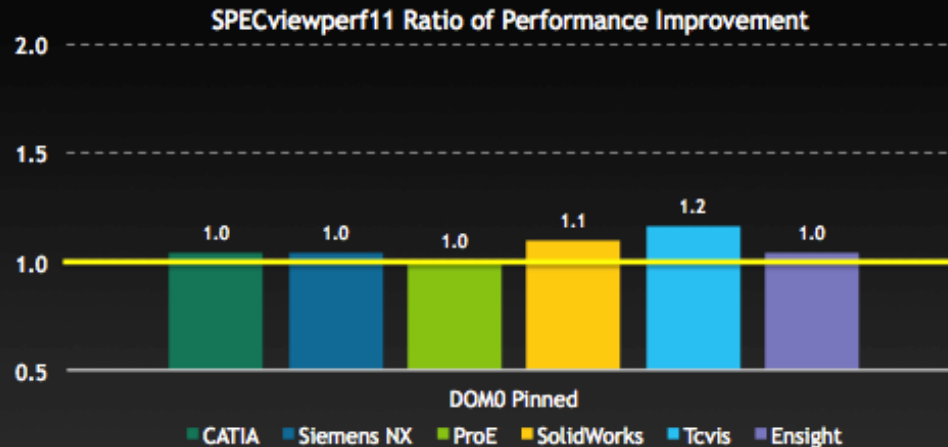
- For a fully populated two(2) socket NUMA system add the following to /etc/rc.local

```
# pin the DOM0 vCPUs to 4 logical cores on Socket 0
xl vcpu-pin 0 0 0
xl vcpu-pin 0 1 1
xl vcpu-pin 0 2 2
xl vcpu-pin 0 3 3

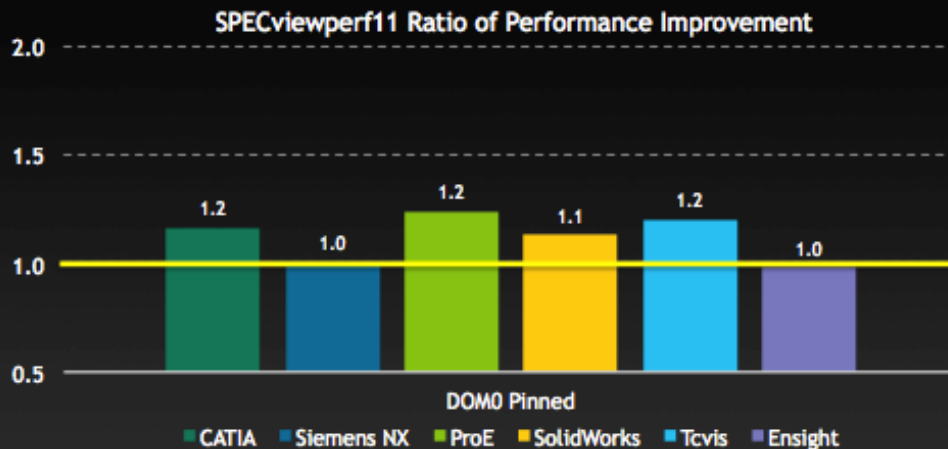
# pin the DOM0 vCPUs to 4 logical cores on Socket 1
xl vcpu-pin 0 4 16
xl vcpu-pin 0 5 17
xl vcpu-pin 0 6 18
xl vcpu-pin 0 7 19
```

- Reboot the XenServer host

Domain-0 Pinning Pass-through: Average 10% increase



Domain-0 Pinning vGPU 1 VM: Average 11% increase

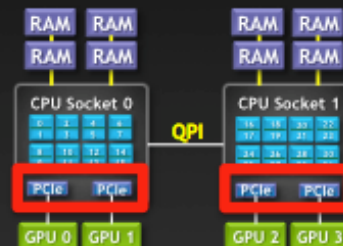


CPU's PCI bus and GPU locality

- **Determine the CPU socket and PCI bridge association**

```
[root@nvrgrid ~]# lspci | grep "PCI bridge"
00:01.0 PCI bridge: Intel Corporation Xeon E5/Core 17 IIO PCI Express Root Port 1a (rev 07)
00:01.1 PCI bridge: Intel Corporation Xeon E5/Core 17 IIO PCI Express Root Port 1b (rev 07)
...
00:01.0 PCI bridge: Intel Corporation Xeon E5/Core 17 IIO PCI Express Root Port 1a (rev 07)
00:02.0 PCI bridge: Intel Corporation Xeon E5/Core 17 IIO PCI Express Root Port 2a (rev 07)
...
```

- '0x:xx.x' roots are associated with CPU Socket 0
- '2x:xx.x' or '8x:xx.xx' roots are associated with CPU Socket 1
 - All non-zero roots should be associated with CPU Socket 1
- In this example
 - Socket 0 is associated with the '0x:xx.x' addresses
 - Socket 1 is associated with the '8x:xx.x' addresses



vCPU Pinning Socket 1

Pinning virtual machines to Socket 1

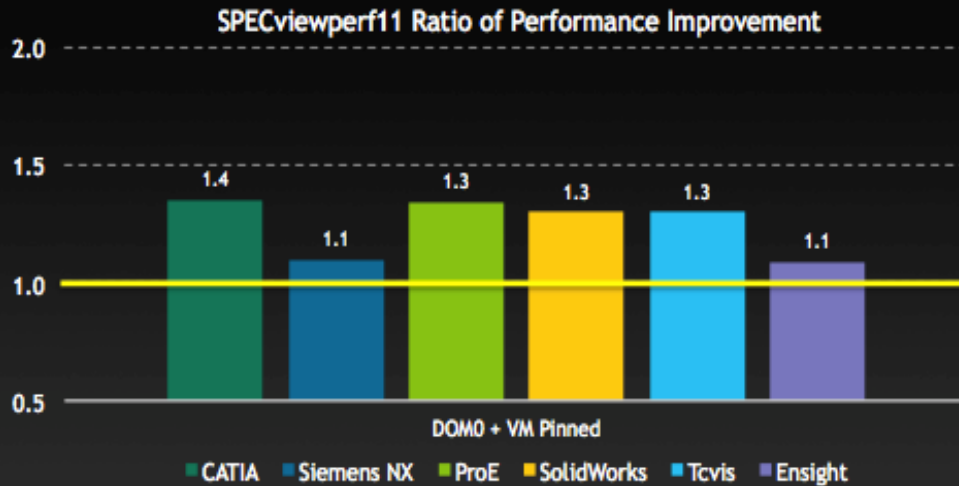
- Virtual machines that are to use the GPU on Socket 1's bus should have their vCPUs pinned to Socket 1
- In this example we will pin the vGPU virtual machines to Socket 1

```
root@nvrtd ~]# xe vm-param-set uuid=SVG1ID VCPUs-params:mask=20,21,22,23,24,25,26,27,28,29,30,31
root@nvrtd ~]# xe vm-param-set uuid=SVG2ID VCPUs-params:mask=20,21,22,23,24,25,26,27,28,29,30,31
root@nvrtd ~]# xe vm-param-set uuid=SVG3ID VCPUs-params:mask=20,21,22,23,24,25,26,27,28,29,30,31
root@nvrtd ~]# xe vm-param-set uuid=SVG4ID VCPUs-params:mask=20,21,22,23,24,25,26,27,28,29,30,31
```

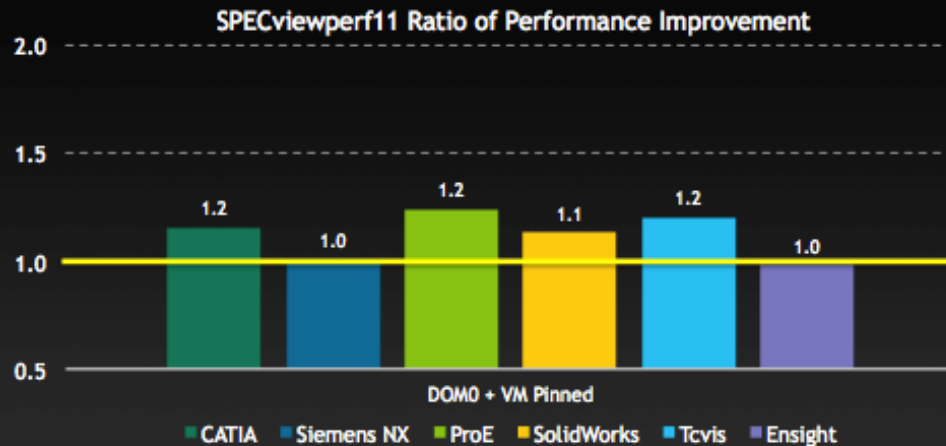
NOTE: Before executing the pinning commands, ensure the virtual machines are halted.

- vCPU cores 16, 17, 18 and 19 have been excluded from the mask as they have been reserved and pinned for Domain-0 use on Socket 1

Domain-0 + VM Pinning Pass-through: Average 26% increase

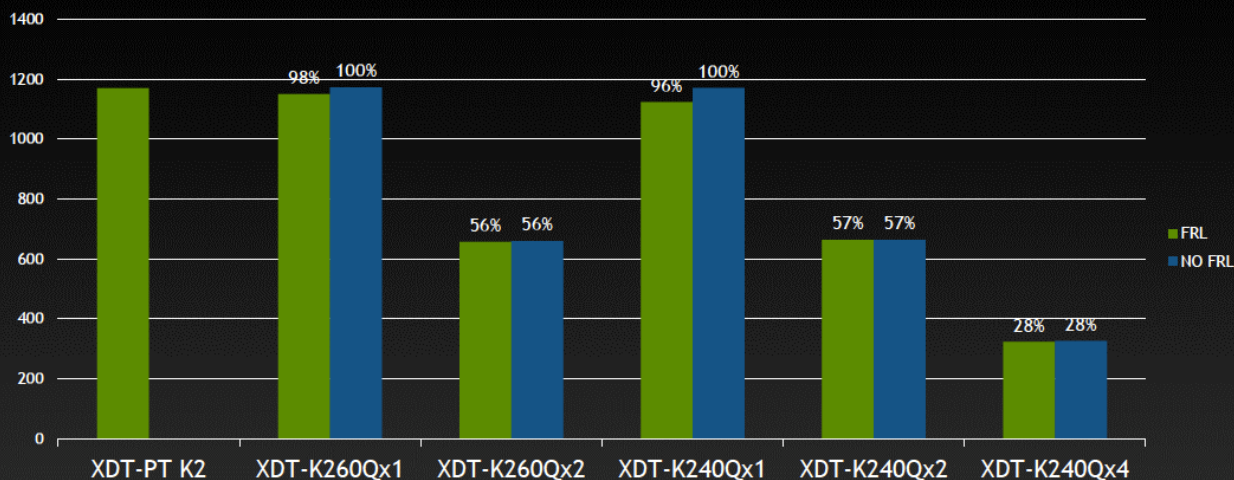


Domain-0 + VM Pinning vGPU 1 VM: Average 11% increase



vGPU Heaven Benchmark

E5-2620 4 vCPU 8GB 4xSSD stripe GRID K2 OpenGL 1280x720 Quality Ultra Tessellation Extreme



benchmark is the most GPU awareness, so FRL and CPU does not directly impact final results, it's been proved perfectly!

多GPU场景-XenApp

Registry settings for GPU scheduling on bare metal deployments

For unvirtualized RDS workloads on bare metal (Either XenApp 6.5 FP2 or XenDesktop 7 with a windows server OS),if you have multiple homogeneous GPUs:

OpenGL apps will be assigned to GPUs on a round-robin basis, determined by the session id.

By default, DirectX apps will all share the same single GPU.

There is an undocumented registry setting to enable experimental support for assigning DirectX apps to GPUs on a round-robin basis as well:

Create the following two registry values

[HKEY_LOCAL_MACHINE\SOFTWARE\Citrix\CtxHook\ApplInit_Dlls\Graphics Helper] "DirectX"=dword:00000001

[HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\Citrix\CtxHook\ApplInit_Dlls\Graphics Helper] "DirectX"=dword:00000001

分配多个GPU到VM

The instructions

You can't use the XenCenter GUI or `xe gpu-group` constructs to do the GPU assignment. No VMs can be active on the machine with GPU assignments made using the GUI or `xe gpu-group`

You need to use the `other-config:pci` parameter to assign the GPUs:

```
xe vm-param-set uuid=<vm-uuid> other-config:pci=0/0000:04:0.0,0/0000:05:0.0
```

Replace 04:0.0, 05:0.0 with the bus/device/func of the GPUs to be assigned.

关于CUDA和OpenCL的支持

For Citrix products, *Vendor Drivers* provide the support for DirectX, OpenGL, OpenCL and CUDA. All of Citrix's GPU technologies rely on GPU vendors own drivers to avoid an additional layer of support and certification. This ensures that the support for DirectX and OpenGL is always up to date and available as for physical servers, this white paper from HP details how [support on other hypervisors significantly lags Citrix. We have deliberately avoided using synthetic drivers to avoid the serious problems encountered by other vendors on benchmarks such as Redway3D.](#)

OpenGL and CUDA support is also provided by the GPU vendors such as NVIDIA. If NVIDIA enable CUDA support for vGPU as they have for GPU passthrough, then Citrix will automatically gain support.

vGPU桌面对PVS的支持情况

发件人: Praveen B Prakash <praveenb.prakash@citrix.com>

日期: 2014年9月10日 GMT+821:13:20

收件人: Rachel Berry <rachel.berry@citrix.com>, Mayunk Jain <Mayunk.Jain@citrix.com>, Jie Zhang <jie.zhang@citrix.com>

抄送: Stuart Moore <stuart.moore@citrix.com>

主题: 回复: Can PVS support vGPU desktop?

You can deploy vGPU based VM's via PVS. But, post creation of VM, you need to manually assign vGPU to the VM.

Thanks,
Praveen

曙光天阔I620-G15服务器Grid K2显卡断电保护问题解决 方法

、故障描述：

K2显卡划分4个K260Q vGPU运行三维设计软件，GPU负载较轻和没有调用GPU运行时状态正常。
当运行高GPU负载的应用程序进行三维设计或调用GPU进行渲染时，出现分配了vGPU的虚拟桌面蓝屏死机。
但是未分配GPU的虚拟机运行正常，在XenServer中查看不到显卡，运行nvidia-smi命令报错无法显示显卡状态。
只能对服务器通过关机冷启动方式重启才能重新加载显卡和vGPU。

3、问题原因：

Nvidia Grid K2显卡一般情况下的工作温度为60-90℃，当物理显卡扇热不畅或者服务器供电不足时显卡会自动断电停

4、解决方法：

a) I620-G15服务器如果配显卡必须配置1200W电源。该服务器出厂时电源有550W、750W、1200W三种规格可选。

b) 物理服务器风扇必须为7个。

c) 在物理服务器BIOS中禁止自动风扇控制：

Advanced菜单-Hardware Health Configuration - Auto Fan Control - Disable

d) 在物理服务器BIOS中关闭CPU省电模式

Advanced菜单 - CPU Power Management Configuration - Power Technology

- Energy Efficient



Work better. Live better.