Reviewer’s Guide for Remote 3D Graphics Apps

Part 1: XenServer GPU Pass-through

with XenDesktop 7 Apps,
Nvidia GRID K1/K2 cards,
Dell R720 Server
**Audience**

Administrators wishing to evaluate the HDX 3D Pro feature in XenDesktop, using the Nvidia GRID K1 or K2 cards. This guide walks you through the physical set up and configuration of NVIDIA GRID K1 GPU card on a Dell R720 server using XenServer hypervisor. The basic steps apply to any Nvidia GRID GPU card and certified server hardware for HDX 3D Pro as per the hardware compatibility list (HCL).

This part focuses on the hardware installation and the process of enabling GPU pass-through in a XenServer environment. The GPU can then be shared by multiple users when delivered from a Windows Server OS using XenDesktop 7 Apps. In subsequent parts of this guide, we learn how to enable GPU sharing on different platforms such as VMware vSphere and XenServer, both for desktop and application workloads in XenDesktop 7.

It is assumed that the reader has good knowledge of networking, virtualization, server hardware, and Windows administration. Familiarity with Citrix and Nvidia products is recommended but not essential to complete these steps. Please see the APPENDIX section for more information.

**Related Documents in this Series**

- **Part 1**: XenServer GPU pass-through for Citrix XenDesktop 7 (includes, physical installation of GPU cards)
- **Part 2**: vSphere GPU pass-through (a.k.a vDGA) for Citrix XenDesktop 7
- **Part 3**: XenServer GPU virtualization (a.k.a vGPU) for Citrix XenDesktop 7

**About the Authors**

Pushpal Ray and Mayunk Jain in the Technical Marketing team of Citrix XenDesktop produced this guide.

Pushpal (@pushpalray) is a Technical Marketing Engineer with over 10 years experience in 3D graphics, infrastructure management, and virtualization. Mayunk (@mayunk) is responsible for competitive marketing, technical demos, and sales enablement for the desktop and cloud solutions at Citrix.
Lab Environment

<table>
<thead>
<tr>
<th>Hardware</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphical Processing Unit (GPU)</td>
<td>NVIDIA GRID K1 (<a href="#">Download K1 and K2 Specs</a>)</td>
</tr>
<tr>
<td>Server hardware</td>
<td>Dell R720 <a href="#">PowerEdge R720 Technical Guide</a></td>
</tr>
</tbody>
</table>
| GPU Installation Kit      | • Power Cables (2 – Internal for GPU)  
                           | • Heat Sink |
| Storage                   | Local/NFS |

<table>
<thead>
<tr>
<th>Software</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypervisor</td>
<td>XenServer 6.2.0-rc4 build 69934c</td>
</tr>
<tr>
<td>NVIDIA GPU driver</td>
<td>320.00</td>
</tr>
</tbody>
</table>
| Guest OS                  | Windows Server 2008 R2 Standard Service Pack 1  
                           | Windows 7 Service Pack 1 |

**Note:**

1. Prior to XenServer 6.2, Citrix XenServer Platinum Licensing is required to enable GPU-Pass through.
2. Snapshot does not work with GPU Cards. It is highly recommended to create/delete/revert snapshot of VM when GPU card is not assigned to VM or else you may have garbage entries in XenServer and VM will not start properly.
3. You should not take default console using XenCenter after assigning GPU Card and Installing NVIDIA Driver. Either take VNC console or take remote console.

**Physical installation of GPU card in Server hardware**

**Checklist for initial setup of server hardware**

Below are the essential components that you need to ensure are in-place to get started with the GPU setup and configuration.

- LAN Network
- Dell Remote Access Card (DRAC), Integrated Lights-Out (iLO), KVM etc.
- Power Supply for both server and GPU card [GPU installation kit]

Please see the APPENDIX for a note about ordering and setting up the Dell R720 Server for GRID GPU.

**Understanding Server hardware and its components**

For your ease and to better understand the components before hand prior to visiting your lab or datacenter, it is useful to know how each of the components look and their location within the server hardware.
Embedded NICs, in this lab, only 1 NIC port is in use.

Dell Remote Access Card port shown in the picture is used to manage your server remotely at a BIOS/firmware level. It requires a separate static IP address and accessed via web browser.

Hot-plug, redundant power supply units: 1100 W each.
Shutdown the server (remotely or pressing power button located on the front side of server).
Unplug the power cables, iDRAC and network cables.

Front view of the Dell R720, 3 servers racked on top of each other. On right-hand bottom corner of each server you can see a blue color label with Intel written on it.
As shown in the image, you need to pull the blue Intel label outward to unlock and slide the complete server out of the rack very slowly, holding on to both the edges of the server.

Once the server is out of the rack, you will see a black screw shaped button which needs to be turned clockwise (just like you unscrew) and then slide back-upward to open the top cover.

Knowing the GPU Installation Kit

A GPU enablement kit is required to support the higher power and heat requirements of GRID GPU cards in the Dell R720 chassis. GPU enablement kit includes:

- Low-profile heat sinks (mandatory)
- Power cables for the GPU cards (mandatory)
- System board support brackets
- Filler brackets with closeout EMI shield for unoccupied PCI-e slots
- Plugs from card to PCI-e riser 2/3
- Low-profile Memory shroud (optional)

One cable supports two or three card connectors. Provides support for cards that require more than 75W power

Source: Dell PowerEdge R720 Owner’s Manual and Dell Support Forum
Low Profile Heat Sink

The figure shows the top and bottom view of the heat sink.

Power cables for GPUs

One cable supports two/three card connectors.

Provides support for cards that require more than 75W power.

Knowing Low-profile Heat Sink

[Note: In this example, server hardware is Dell R720]

What is a heat sink?
A heat sink is an essential component for cooling high-power integrated circuits. In computers, heat sinks are used to cool central processing units or graphics processors. Heat sinks are used with high-power semiconductor devices such as power transistors and optoelectronic devices such as lasers and light emitting diodes (LEDs), wherever the heat dissipation ability of the basic device package is insufficient to control its temperature.
Source: [wiki](https://en.wikipedia.org/wiki/Heat_sink)

Replace the Default Heat Sink with the Low Profile Heat Sink. This is a specific heat sink required for NVIDIA GRID-enabled servers. This is included along with the Power cable(s) in the GPU installation Kit and is a pre-requisite. This must be obtained from the NVIDIA-certified server vendors.
Please see the APPENDIX for more information on other parts of the GPU kit such as Riser card.

**Insert NVIDIA GRID K1 GPU card into the PCIe slot 6 (RISER 3)**

In this example, the NVIDIA GRID K1 insert into the RISER 3 of the PCIe slot.

The figure shows which side of the GPU card goes into the Riser PCIe slot.
<table>
<thead>
<tr>
<th>Image of GPU card being inserted into the slot. You need to tilt slightly to get the card into the slot. Aim for the slot and press the card into the slot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>This is how it looks when the GPU card is plugged into the server.</td>
</tr>
</tbody>
</table>
Internal power supply to GPU card - Plug in the power cables
Once the GPU card is inserted into the expansion-card (slot 6) riser 3 (in otherwords, PCIe slot), follow these steps to insert the power cable into the GPU card to provide internal power supply to the GRID.

Ensure the power cable for the GPU card is handy.

The circle in red shows two ends of the power cable:
1. Power Connector (RISER end)
2. Power Connector (GPU end)

RISER 3 side view

Connect Power connector (white color RISER written on it) to the RISER 3

Plug in the one end of the Power cable to the power connector on the RISER end.
Connect Power connector (black color with six-pin, GPU written on it) to the GPU card.

Plug in the another end of the Power cable to the power connector on the GPU end.

NVIDIA GRID K1 now gets internal power supply. The other end of the power cable (GPU connector) plugs into the GPU card.

1. Power Connector (RISER end)
2. Power Connector (GPU end)
Plug in the GPU card into the PCIe slot 6

Adjust the power cables and fit the black plastic sheet to cover the DIMMs and the heat sink for processors 1 and 2.
Post-installation Checklist

Physical verification and configuration of GPU card

Once the GPU installed on the server, mount and rack the server back, insert the power cables and network & DRAC cables. Power-On the server. Follow the steps below to go through the post-install verification checklist.

Method 1: Using XenCenter GUI

Login to the XenServer using XenCenter. Provision a test Windows 7 Enterprise 64-bit virtual machine. Right-click the test VM and select Properties

- Go to the GPU tab, on right panel you should see the NVIDIA GPU listed in the drop-down.
- Select/Assign the NVIDIA GRID K1 to the test VM.

Example:

NVIDIA Corporation GK 107GL (GRID K1) GPUs (4GPUs)

Which means, the GRID K1 has 4 GPUs installed on single board/plane(GK107GL)

This screen gives information and next steps. Read them for your understanding.

Next Step:

- Power-On the test VM
- Install NVIDIA GPU driver on the guest OS
Power-On the test VM

The test VM console in XenCenter also states:

‘This VM has a dedicated GPU assigned. You must connect to it using Remote Desktop’

After the GPU is assigned to the VM, the VM console resolution looks somewhat like this (resolution distorted)

This behavior is seen when a GPU is assigned to a XenServer VM.

Just to show the difference:
XenServer VM console without a GPU card assigned.

Therefore, If you face the console resolution distortion behavior,
Before GPU is attached to the server hardware, Device manager shows only one Standard VGA Graphics Adapter.

Login to the test VM with local Administrative credentials.

Go to Control Panel > Device Manager.

**Tip:** Start > Run > type `devmgmt.msc` and hit Enter.

Under Display adapters, you will see a second VGA Graphics Adapter with an exclamation (!) mark in yellow triangle. **This means the NVIDIA GRID K1 driver is not yet installed.**
### Method 2: Using XenServer command line

**How it looks from XenServer console**

SSH to your XenServer using root credentials.

Type this command to show if the GPU card is identified by the hypervisor.

```
lspci | grep VGA
```

Type this command to show the GPU group ID

```
xe gpu-group-list
```
Type this command to list ALL the GPU cards attached to the XenServer

```
xe pgpu-list
```

**Method 3: Using NVIDIA System Management Interface (NVIDIA-SMI) command line**

This method is applicable when testing GPU on a physical machine and/or on a virtual machine with GPU pass-through or vDGA.

Go to Run → type `cmd` and hit enter.

Using cd command, go to the following directory: `C:\Program Files\NVIDIA Corporation\NVSMI`

- Type `nvidia-smi -h` for all available commands
- Type `nvidia-smi -L`
  - `-L, --list-gpus` Display a list of GPUs connected to the system.
  
  **Example:**
  
  `C:\Program Files\NVIDIA Corporation\NVSMI>nvidia-smi -L`
  
  GPU 0: Quadro 2000 (UUID: GPU-eb75620c-af16-0a19-9c38-027ddeb5e2b2d)

- Type `nvidia-smi` and hit enter.

  This command will display real-time GPU utilization with other information such as Memory, GPU Utilization, GPU temperature, Process ID (PID), Clock, and so on.

  **Example:**

  We saw the following output after rendering the SeaScooter Dolphin for 3-5 minutes using eDrawing viewer and nvidia-smi refresh interval @ 1 second. The values in RED (emphasis mine) shows the GPU Utilization of the 3D program.

  `C:\Program Files\NVIDIA Corporation\NVSMI>nvidia-smi`
  
  Tue Jul 30 16:07:53 2013
  +------------------------------------------------------------------+
<p>| 0 | Quadro 2000 [GRID K1] 0.00% utilization 0.00% efficiency 65.3°C |</p>
<table>
<thead>
<tr>
<th>NVIDIA-SMI 4.311.66 Driver Version: 311.66</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GPU Name</strong></td>
</tr>
<tr>
<td>**Fan Temp Perf Pwr:Usage/Cap</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>0 Quadro 2000</td>
</tr>
<tr>
<td>30% 44C P12 N/A / N/A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Compute processes:</strong></td>
</tr>
<tr>
<td><strong>GPU</strong></td>
</tr>
<tr>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>
GPU Software (driver) installation and configuration

The following steps will install GPU driver on guest Operating system and deliver a GPU-powered XenServer VM (Virtual Machine).

From the VM where GPU is installed, go to [http://www.nvidia.com](http://www.nvidia.com) and select **Drivers** from the navigation bar

OR

Simply, go to [www.nvidia.com/drivers](http://www.nvidia.com/drivers)

Select Option 2: Automatically find drivers for my NVIDIA products. *(Recommended)*

This option is used if you wish to download the NVIDIA drivers on the machine where GPU is installed.

Click **GRAPHICS DRIVERS**

Option 1: Manually find drivers for my NVIDIA products.

This option is used if you must download driver from another machine, and then copy it over to the GPU VM. In this case, select the appropriate GPU card from the drop-down.

It may prompt you to install Java if your workstation or VM does not have Java pre-installed.

Click on the ‘Java’ icon to install Java.

The page will redirect to the java website [http://java.com/en/download](http://java.com/en/download) and follow the instructions to install the latest version of the Java plugin.

Click **Agree and Start Free Download**
Open a fresh web browser and go to NVIDIA drivers download page.

Click on the browser alert **Java SE Runtime Environment** and select **Run for this website** (allow)

Check the ‘Do not show this again for apps from the publisher and location above’ box.

Click Run

The webpage will return with the appropriate GPU driver for your GPU attached to the hardware and the Operating System (Guest OS/workstation)

Click Download

On the download page, you will find the following information around the GPU you’re using:

- Release Highlights
- Supported Products
- Additional Information:
  - Quadro/Tesla/GRID Release Notes (v320.00)
  - Control Panel User’s Guide

Click Download
Click **AGREE & DOWNLOAD**

You may get a prompt of Security Warning for Adobe Flash Player.

Click **Install**

Click **Save or Run** to download/install the GPU driver on the operating system of the guest OS or Workstation.

Click **OK** to extract the GPU driver installer.

GPU driver software extraction and installation in progress.

Click **AGREE AND CONTINUE**
Express (Recommended) Installation

Select this option and Click NEXT

Tip:
If required, the Custom Installation lets you select the following components:
- Graphics Driver
- 3D Vision Controller Driver
- 3D Vision Driver
- NVIDIA WMI
- nView

Installation in progress

Click on RESTART NOW to complete the installation.
After the VM reboots, login to the guest OS and ensure the Nvidia drivers are installed. They must show up under:

Start > Control Panel > Programs > Uninstall a Program

Open Device Manager ➔ expand Display adapters.

Where previously we saw the exclamation triangle, it should show NVIDIA GRID K1 and no error icons.

**NOTE:**
If you continue to see an exclamation mark, most likely reasons are:

- GPU driver service is not running
- Server hardware is not supplying sufficient power to the GPU cards
- Incompatible GPU driver
How to verify the applications are using GPU acceleration

Apart from the sheer performance boost easily visible when rendering 3D applications, here are some tools and commands to verify GPU acceleration benefits are available to the VM.

Using XenServer VM GPU properties
The first check is to ensure that your VM is attached to the GPU. In XenCenter, right-click the VM and select Properties. Under the GPU tab, you should see the expected GPU type listed on the right with a comment

*The GPU can only be changed when the VM is shut down*

![GPU properties](image)

Figure 1 The GPU type will display the Nvidia model-name once it is properly installed

Using XenServer Command Line Interface (CLI)
SSH to the XenServer using *putty* and run the following commands to verify GPU is being detected by the hypervisor:

List of GPU cards attached to the XenServer: `lspci | grep VGA`

```
[root@XS -6 ~]# lspci | grep VGA
07:00.0 VGA compatible controller: NVIDIA Corporation GK107GL [GRID K1] (rev a1)
08:00.0 VGA compatible controller: NVIDIA Corporation GK107GL [GRID K1] (rev a1)
09:00.0 VGA compatible controller: NVIDIA Corporation GK107GL [GRID K1] (rev a1)
0a:00.0 VGA compatible controller: NVIDIA Corporation GK107GL [GRID K1] (rev a1)
10:00.0 VGA compatible controller: Matrox Electronics Systems Ltd. G200eR2
```

List of GPU Groups created by XenServer and their corresponding UUIDs: `xe gpu-group-list`

```
[root@XS -6 ~]# xe gpu-group-list
uuid ( RO) : 80a4d59a-1cfa-aa2b-6be1-a82885a61394
    name-label ( RW): Group of NVIDIA Corporation GK107GL [GRID K1] GPUs
    name-description ( RW):

uuid ( RO) : c004e835-5142-6dae-bba2-cb74a23c662c
    name-label ( RW): Group of Matrox Electronics Systems Ltd. G200eR2 GPUs
    name-description ( RW):
```
List of all GPU cards attached to the XenServer: `xe pgpu-list`

```
[root@XS-6 ~]# xe pgpu-list
uuid (RO)              : f2bf4eee-6316-329f-dc62-672103ec30e5
    vendor-name (RO): Matrox Electronics Systems Ltd.
    device-name (RO): G200eR2
    gpu-group-uuid (RO): c004e835-5142-6dae-bba2-cb74a23c662c

uuid (RO)              : a8b8c795-6eca-a1d7-b249-29a3be2b2770
    vendor-name (RO): NVIDIA Corporation
    device-name (RO): GK107GL [GRID K1]
    gpu-group-uuid (RO): 80a4d59a-1cfa-aa2b-6bel-a82885a61394

uuid (RO)              : b876f6af-11c6-00ed-3a34-6c9b5b358560
    vendor-name (RO): NVIDIA Corporation
    device-name (RO): GK107GL [GRID K1]
    gpu-group-uuid (RO): 80a4d59a-1cfa-aa2b-6bel-a82885a61394

uuid (RO)              : 0e750526-e214-a855-5af0-84bb7f1a8e1c
    vendor-name (RO): NVIDIA Corporation
    device-name (RO): GK107GL [GRID K1]
    gpu-group-uuid (RO): 80a4d59a-1cfa-aa2b-6bel-a82885a61394

uuid (RO)              : e46c2f2b-ddf5-494e-b3dc-7f7bla26e1ff
    vendor-name (RO): NVIDIA Corporation
    device-name (RO): GK107GL [GRID K1]
    gpu-group-uuid (RO): 80a4d59a-1cfa-aa2b-6bel-a82885a61394
```

List VMs hosted on the XenServer: `xe vm-list`

```
[root@XS-6 ~]# xe vm-list
uuid (RO)              : 49adfc6d-d4dc-2c2c-1ae0-a966ca8a0ba3
    name-label (RW): xd1
    power-state (RO): running

uuid (RO)              : 76eda349-3da1-4575-9114-133e6031d80b
    name-label (RW): Control domain on host: XS-6
    power-state (RO): running
```

Using the UUID listed above, verify the test VM has a GPU assigned: `xe vgpu-list vm-uuid=<uuid of VM>`

In the example output below, `gpu-group-uuid` shows that GPU from that group is assigned to the VM.

```
[root@XS-6 ~]# xe vgpu-list vm-uuid=49adfc6d-d4dc-2c2c-1ae0-a966ca8a0ba3
uuid (RO)              : 131b63d1-909d-4c5e-1b50-cf9058f8c05b
    vm-uuid (RO): 49adfc6d-d4dc-2c2c-1ae0-a966ca8a0ba3
    gpu-group-uuid (RO): 80a4d59a-1cfa-aa2b-6bel-a82885a61394
```

**Tip:** Press tab to complete UUID's after typing first 4-6 characters.
Manually assign and Unassign GPU to VM, using CLI

These commands can be used if manual mapping of GPU to VM is required.

Assign GPU to VM using CLI

[root@XS-6 ~]# xe vgpu-create vm-uuid=<UUID of the VM> gpu-group-uuid=<group UUID of the GPU>

Unassign GPU from VM using CLI

Make a note of the UUID of the VM. Verify if VM has any GPU already assigned. If yes, then shut down the Virtual Machine and destroy that assignment. Commands to destroy any pre-assigned GPU(s):

[root@XS-6 ~]# xe vgpu-list vm-uuid=<UUID of the VM>
[root@XS-6 ~]# xe shutdown uuid=<UUID of the VM>
[root@XS-6 ~]# xe vgpu-destroy uuid=<UUID of the GPU>

[to verify if GPU unassigned successfully]
[root@XS-6 ~]# xe vgpu-list vm-uuid=<UUID of the VM>

Using third party Utilities

NOTE: Citrix and Nvidia do not officially support tools such as GPU-Z; they are presented here only as free admin utilities

TechPowerUp GPU-Z

From within the VM, use TechPowerUp GPU-Z to easily verify that the Nvidia GRID card is installed and recognized as a known device by your hypervisor and guest OS. GPU-Z is an open source, lightweight system utility designed to provide vital information about your video card and graphics processor.

The Graphics Card tab shows detailed information about the GPU card such as:
- Make and Model/Version
- Memory & Processor

The Sensors tab shows the real-time GPU utilization in terms Memory and Processing power
Figure 2 GPU-Z shows detailed information about the state of running GPU and driver version, etc.

Non-working/Unutilized GPU
(values showing 'unknown' and GPU Core/Memory Clock is 0.0 MHz)

Figure 3 GPU-Z shows "Unknown" value if the GPU is not installed or not detected
GPU Caps Viewer

GPU Caps Viewer is an OpenGL and OpenCL graphics card utility. To download go to http://www.ozone3d.net/gpu_caps_viewer/

OpenGL Extension Viewer

OpenGL Extensions Viewer displays useful information about the current OpenGL 3D accelerator, such as the vendor name, the version implemented, the renderer name and the extensions of the current OpenGL 3D accelerator. OpenGL Extensions Viewer is available for Windows 32bit and 64bit and MacOS X, iOS, and Android.

Summary

In this first part of the HDX 3D Pro Reviewer’s Guide, we learn how to identify the different hardware components of an HDX 3D Pro solution and complete the physical installation. We also configured the graphics drivers and tested the GPU being ready for use inside the virtual machine (VM). Please refer to the XenDesktop 7 Reviewer’s Guide to learn how these VMs act as the base image for HDX 3D delivery using Citrix XenDesktop. It goes through the process of setting up the XenDesktop infrastructure and accessing applications from thin-clients and standard PCs using Citrix Receiver.

In this document, GPU pass-through was enabled in XenServer environment. The next part looks at the same process in a VMware vSphere environment.
Related Documents in this Series

**Part 1**: XenServer GPU pass-through for Citrix XenDesktop 7 (includes, physical installation of GPU cards)
**Part 2**: vSphere GPU pass-through (a.k.a vDGA) for Citrix XenDesktop 7
**Part 3**: XenServer GPU virtualization (a.k.a vGPU) for Citrix XenDesktop 7

Also reference, [Reviewer’s Guide for XenDesktop 7](#)

Documents and Resources

**Server hardware related**

- [PowerEdge R720 Technical Guide](#)
- [PowerEdge R720 Owner's Manual](#)
- NVIDIA GRID-enabled Servers – [Where to Buy](#)

**Citrix HDX 3D Pro related**

- [XenServer 6.0 Multi-GPU Passthrough for XenDesktop HDX 3D Pro Graphics](#)
- [Citrix GPU Hardware Compatibility List](#)
- [Key Project Design Guide – HDX 3D Pro](#)

**Nvidia GPU related**

- NVIDIA GRID Datasheet [Download K1 and K2 Specs](#)
- For more information on [NVIDIA GRID K1 and K2 GPUs](#)

<table>
<thead>
<tr>
<th>Demo Apps</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Earth</td>
<td><a href="http://www.google.com/earth">http://www.google.com/earth</a></td>
</tr>
<tr>
<td>Adobe Photoshop (trial)</td>
<td><a href="http://www.adobe.com/photoshop">http://www.adobe.com/photoshop</a></td>
</tr>
<tr>
<td>Autodesk Inventor</td>
<td><a href="http://www.autodesk.com/inventor">http://www.autodesk.com/inventor</a></td>
</tr>
</tbody>
</table>
A Note about the Dell R720 Server Hardware

- Ensure that you order Dell R720 system with the redundant **1100W** power supplies and **GPU Kit**.
- The PowerEdge R720 must be installed with two processors.
- Ensure the GPU enablement kit is delivered with your Dell R720. For details, please see [knowing the GPU Installation Kit](#).
- Currently, all GPU cards in one server must be of the same type and model. The supported configurations are:
  - up to two double-wide GPU cards. This requires a Riser 3 card.
  - up to four single-wide GPU cards.
- Due to the high power consumption of GPUs, the ambient system inlet temperature is restricted to 30 °C to ensure adequate system cooling when one or more GPU cards are installed in PowerEdge R720. Note that this is less than the standard environmental specification of 35 °C.
- Internal GPU cards are supported on the PowerEdge R720 and not on the PowerEdge R720xd.

*Source:* Page 74 in Dell PowerEdge R720 *Owner’s Manual*
Knowing the PCI Express and Riser, Memory and Processor Location

Top view of the internal server components.

Below picture shows the Memory, Processor and the PCIe slots and Risers for GPU card to plug in.

**Components**
1. expansion-card riser 1  
2. expansion-card riser 2  
3. expansion-card riser 3  
4. heat sink for processor 1  
5. heat sink for processor 2  
6. DIMMs (24)

**Riser Layout**
**Internal GPU support**
- 2 x 300W (double wide)
- 4 x 150W (single wide)

**PCI Express**
Total PCIe Slots: 7  
PCIe version: 3.0  
- One x16 full-length, full-height  
- Three x8 full-length, full-height  
- Three x8 half-length, half-height

This is the layout of the PCIe Slot and the location of the riser planes.
**PCIe expansion**

For information on card installation, requirements, and slot priorities, see the *PowerEdge R720 R720xd Systems Owner’s Manual* on [Support.Dell.com/Manuals](http://Support.Dell.com/Manuals).

**PCIe slots**

The R720 and R720xd provide greatly expanded PCIe slot capability over their predecessor servers. This is made possible by the 40 PCIe lanes available from each processor in the system. The R720 and R720xd have been designed to be PCIe 3.0 compliant in order to take full advantage of the processor capabilities. Table 16 details the R720 and R720xd PCIe slots.

<table>
<thead>
<tr>
<th>System</th>
<th>R720</th>
<th>R720xd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slots*</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slot types</th>
<th>R720</th>
<th>R720xd</th>
</tr>
</thead>
<tbody>
<tr>
<td>One x16</td>
<td>One x16 full-length, full-height</td>
<td>Two x16 full-length, full-height</td>
</tr>
<tr>
<td>Three x8</td>
<td>Three x8 full-length, full-height</td>
<td>One x8 full-length, full-height</td>
</tr>
<tr>
<td>Three x8 half-length, half-height</td>
<td>Three x8 half-length, half-height</td>
<td></td>
</tr>
</tbody>
</table>

*In a single-processor configuration, slots 1–4 are not usable*
In this example, the NVIDIA GRID K1 goes into the expansion-card slot 6 in Riser 3.

The power cable connector goes into the power connector (for GPU cards).

The above two components are highlighted in the figure.

Source: Dell PowerEdge R720 Owner's Manual