

GPU 云服务器

操作指南

产品文档



腾讯云

【版权声明】

©2013-2018 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或模式的承诺或保证。

文档目录

操作指南

最佳实践

登录实例

重启实例

安装 NVIDIA 驱动指引

安装 AMD 驱动指引

安装 CUDA 驱动指引

操作指南

最佳实践

最近更新时间：2017-08-29 18:52:51

安全组与网络

- 安全组是一种有状态的提供包过滤功能的虚拟防火墙，用户可通过设置安全组允许受信任的地址访问实例，达到限制访问的目的。创建不同的安全组规则应用于不同安全级别的实例组上，确保运行重要业务的实例无法轻易从外部触达。有关更多信息，请参阅 [安全组](#)。
- 定期修补，更新和保护实例上的操作系统和应用程序。
- 借助弹性公网 IP 地址，快速将地址重新映射到账户中的另一个实例（或 NAT 网关实例），从而屏蔽实例故障。有关更多信息，请参阅 [弹性IP地址](#)。
- 尽量使用 [SSH 密钥](#) 方式登录用户的 Linux 类型实例。使用 [密码登录](#) 的实例需要不定期修改密码。
- 选择使用 [私有网络](#) 进行逻辑区的划分。

存储

- 对于可靠性要求极高的数据，请使用 [腾讯云云硬盘](#) 以保证数据持久可靠存储。
- 对于访问频繁、容量不稳定的数据库，可使用 [腾讯云云数据库](#)。
- 利用 [对象存储 COS](#)，存储静态网页和海量图片、视频等重要数据。

备份与恢复

- 通过 [云主机控制台](#) 回滚备份好的 [自定义镜像](#) 恢复等方式。
- 跨多个可用区部署应用程序的关键组件，并适当地复制数据。
- 定期查看监控数据并设置好适当的告警。有关更多信息，请参阅 [云监控产品文档](#)。

登录实例

最近更新时间：2017-11-17 10:23:28

在购买并启动了 GPU 实例后，您可以连接并登录它。根据您的本地的操作系统、GPU 实例操作系统和 GPU 实例是否可被 Internet 访问，不同情况下可以使用不同的登录方式。

先决条件

- 使用密码登录到 GPU 云服务器时，需要使用管理员帐号和对应的密码；
- 使用密钥登录到 GPU 云服务器时需要创建并下载私钥。

登录指引

若 GPU 实例为 Linux 实例，具体登录指引可参考[登录 Linux 实例](#)

若 GPU 实例为 Ubuntu 实例，具体登录指引可参考[登录 Windows 实例](#)

重启实例

最近更新时间：2017-08-30 11:21:22

重启操作是维护 GPU 云服务器的一种常用方式，重启实例相当于本地计算机的重启操作系统操作。

概述

- **重启准备**：重启期间实例将无法提供正常服务，因此在重启之前，请确保 GPU 云服务器已暂停业务请求。
- **重启操作方式**：建议使用腾讯云提供的重启操作进行实例重启，而非在实例中运行重启命令（如 Windows 下的重新启动命令及 Linux 下的 Reboot 命令）。
- **重启时间**：一般来说重启操作后只需要几分钟时间。
- **实例物理特性**：重启实例不改变实例的物理特性。实例的公网 IP、内网 IP、存储的任何数据都不会改变。
- **计费相关**：重启实例不启动新的实例计费时间。

使用控制台重启实例

1. 登录 [云主机控制台](#)。
2. 重启单个实例：勾选需要重启的实例，在列表顶部，单击【重启】按键。或在右侧操作栏中，单击【更多】 - 【云主机状态】 - 【重启】。
3. 重启多个实例：勾选所有需要重启的实例，在列表顶部，单击【重启】按键。即可批量重启实例。不能重启的实例会显示原因。

使用 API 重启实例

请参考 [RebootInstances 接口](#)。

安装 NVIDIA 驱动指引

最近更新时间：2018-09-08 15:58:28

GPU 云服务器正常工作需安装正确的基础设施软件，对 NVIDIA 系列 GPU 而言，有两个层次的软件包需要安装：

1. 驱动 GPU 工作的硬件驱动程序。
2. 上层应用程序所需要的库。

若把 NVIDIA GPU 用作通用计算，需要安装 Tesla Driver + CUDA，本文仅介绍如何安装 Tesla Driver。

为方便用户，用户可以再创建 GPU 云服务器时，在镜像市场里选择预装特定版本驱动和 CUDA 的镜像。

Linux 驱动安装

Linux 驱动安装有 2 种方式：

1. Shell 脚本安装，适用于任何 Linux 发行版，包括 CentOS，Ubuntu 等；
2. 包安装，适用于不同 Linux 发行版，例如 DEB 包安装，RPM 包安装等。

不管哪种安装方式，NVIDIA Tesla GPU 的 Linux 驱动在安装过程中需要编译 kernel module，所以要求系统安装好了 gcc 和编译 Linux Kernel Module 所依赖的包，例如 kernel-devel-\$(uname -r) 等。

Shell 脚本安装

1. 登录 [NVIDIA 驱动下载](#) 或打开链接 <http://www.nvidia.com/Download/Find.aspx>。

2. 选择操作系统和安装包。以 P40 为例，搜寻驱动，然后选择要下载的驱动版本。

NVIDIA Driver Downloads

Advanced Driver Search

<p>Product Type: <input type="text" value="Tesla"/></p> <p>Product Series: <input type="text" value="P-Series"/></p> <p>Product: <input type="text" value="Tesla P40"/></p>	<p>Operating System: <input type="text" value="Linux 64-bit"/></p> <p>CUDA Toolkit: <input type="text" value="9.2"/></p> <p>Language: <input type="text" value="English (US)"/></p> <p>Recommended/Beta: <input type="text" value="Recommended/Certified"/></p> <p style="text-align: center; border: 1px dashed black; padding: 5px; display: inline-block; background-color: #4CAF50; color: white; margin: 10px 0;">SEARCH</p>
---	---

Name	Version	Release Date	CUDA Toolkit
Tesla Driver for Linux x64	396.44	August 6, 2018	9.2
Tesla Driver for Linux x64	396.37	July 9, 2018	9.2
Tesla Driver for Linux x64	396.26	May 17, 2018	9.2

注意：

操作系统选择 Linux 64-bit 代表下载的是 shell 安装文件，如果选择具体的发行版下载的文件则是对应的包安装文件。

3. 选择特定的版本跳转后，单击【DOWNLOAD】。

TESLA DRIVER FOR LINUX X64

Version: 396.44
Release Date: 2018.8.6
Operating System: Linux 64-bit
Language: English (US)
File Size: 82.54 MB

DOWNLOAD

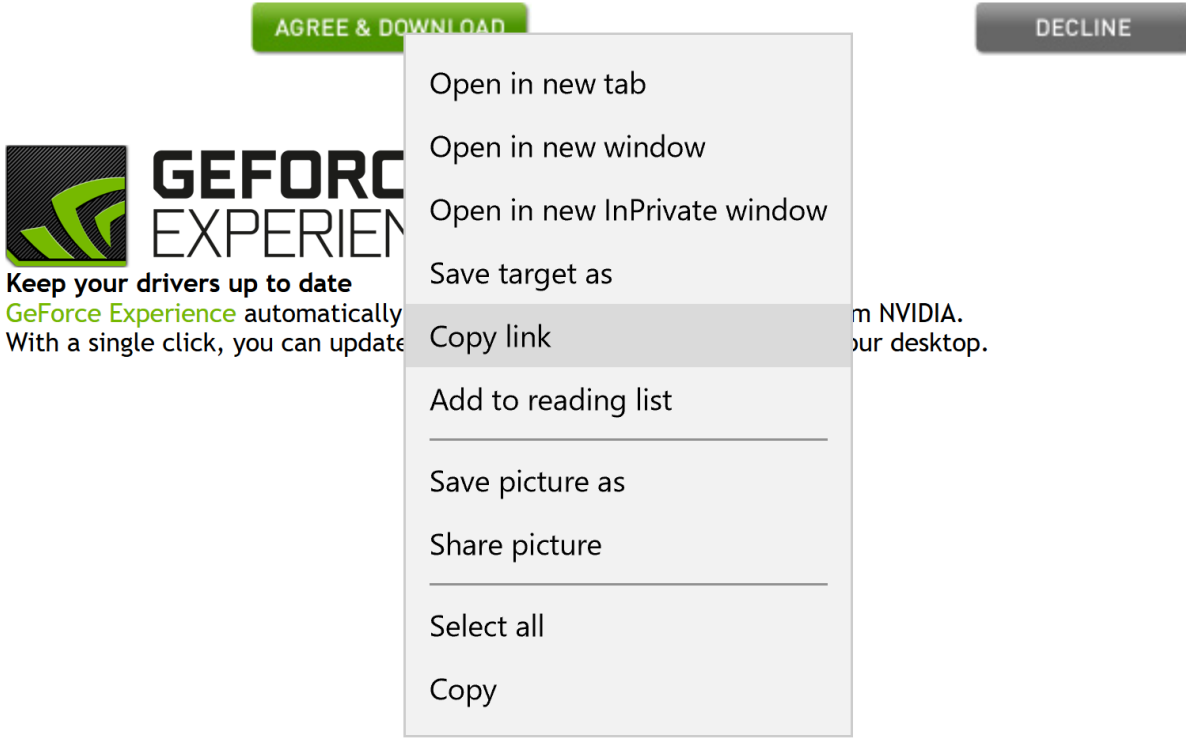
Release Highlights
Supported products
Additional information
What's New

- Various security issues were addressed, for additional details on the med-high severity issues please review [NVIDIA Product Security](#) for more information

4. 再次跳转后，如有填写个人信息的页面可选择直接跳过，出现下面页面时，右击【AGREE&DOWNLOAD】，右键菜单里复制链接地址。

Download

By clicking the "Agree & Download" button below, you are confirming that you have read and agree to be bound by the [License For Customer Use of NVIDIA Software](#) for use of the driver. The driver will begin downloading immediately after clicking on the "Agree & Download" button below. NVIDIA recommends users update to the latest driver version. Please review [NVIDIA Product Security](#) for more information.



5. 登录 GPU 实例，使用 `wget` 命令，粘贴上述步骤复制的链接地址下载安装包；或通过在本本地系统下载 NVIDIA 安装包，上传到 GPU 实例的服务器。

```
[root@VM_0_4_centos ~]# wget http://us.download.nvidia.com/tesla/396.44/NVIDIA-Linux-x86_64-396.44.run
la/396.44/NVIDIA-Linux-x86_64-3
```

6. 对安装包加执行权限。例如，对文件名为 `NVIDIA-Linux-x86_64-396.44.run` 加执行权限：

```
chmod +x NVIDIA-Linux-x86_64-396.44.run
```

7. 安装当前系统对应的 `gcc` 和 `kernel-devel` 包

```
sudo yum install -y gcc kernel-devel-xxx
```

xxx 是内核版本号，可以通过 `uname -r` 查看。

8. 运行驱动安装程序后按提示进行后续操作。

```
sudo /bin/bash ./NVIDIA-Linux-x86_64-396.44.run
```

9. 安装完成后，运行 `nvidia-smi`。如果有类似如下的 GPU 信息显示出来，说明驱动安装成功。

```
[root@VM_0_4_centos ~]# nvidia-smi
Fri Sep  7 12:09:31 2018
+-----+-----+
| NVIDIA-SMI 396.44                Driver Version: 396.44      |
+-----+-----+
| GPU   Name           Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+
|    0   Tesla P40                Off      | 00000000:00:05.0 Off  |            0         |
| N/A   26C    P0              45W / 250W |      0MiB / 22919MiB |           0%      Default |
+-----+-----+
+-----+-----+
| Processes:                       GPU Memory |
|  GPU       PID    Type    Process name                       Usage |
+-----+-----+
| No running processes found |
+-----+-----+
[root@VM_0_4_centos ~]# █
```

DEB/RPM 包安装

DEB 包安装方式

1. 登录 [NVIDIA 驱动下载](http://www.nvidia.com/Download/Find.aspx) 或打开链接 <http://www.nvidia.com/Download/Find.aspx>。

2. 选择对应的支持 DEB 包的操作系统，例如：Ubuntu 16.04，得到下载链接：`wget`

```
http://us.download.nvidia.com/tesla/396.44/nvidia-diag-driver-local-repo-ubuntu1604-396.44_1.0-1_amd64.deb
```

3. 运行安装软件包命令。

```
dpkg -i nvidia-diag-driver-local-repo-ubuntu1604-396.44_1.0-1_amd64.deb
```

4. 使用 `apt-get` 命令更新软件包。

```
apt-get update
```

5. 运行 `apt-get` 命令安装驱动。

```
apt-get install cuda-drivers
```

6. 运行 `reboot` 命令重启。
7. 运行 `nvidia-smi` 能输出正确信息代表驱动安装成功。

RPM 包安装方式

1. 登录 [NVIDIA 驱动下载](#) 或打开链接 <http://www.nvidia.com/Download/Find.aspx> 。

1.支持 RPM 包的操作系统，例如：rhel 7.x，得到下载链接：`wget`

```
http://us.download.nvidia.com/tesla/396.44/nvidia-diag-driver-local-repo-rhel7-396.44-1.0-1.x86_64.rpm
```

2. 使用 `rpm` 命令安装 rpm 包。

```
rpm -i nvidia-diag-driver-local-repo-rhel7-396.44-1.0-1.x86_64.rpm
```

3. 使用 `yum` 命令清除缓存。

```
yum clean all
```

4. 使用 `yum` 命令安装驱动。

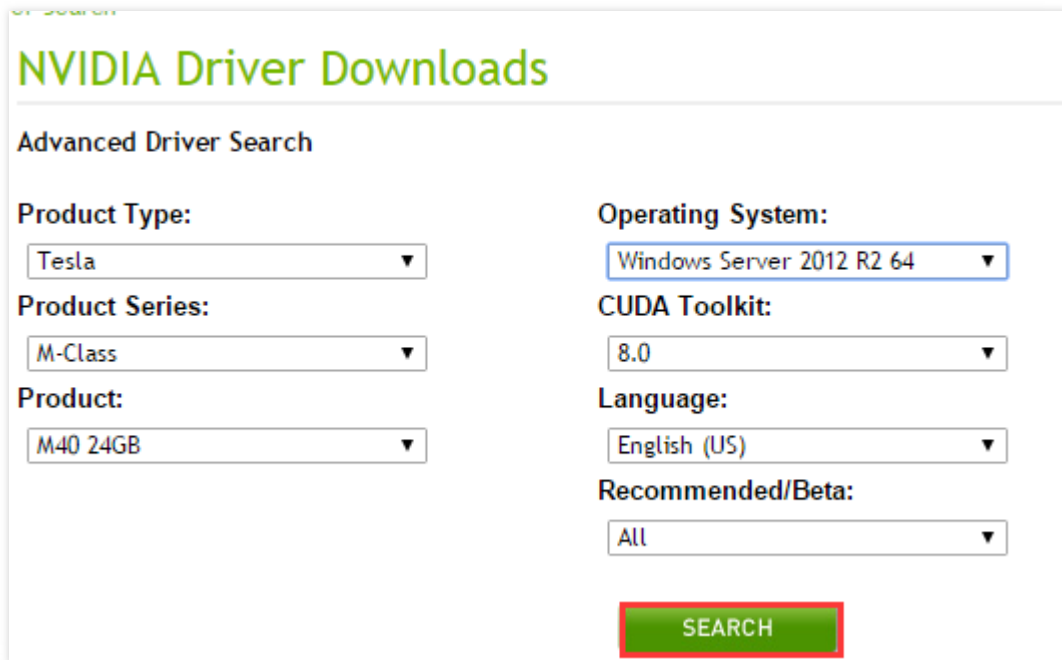
```
yum install cuda-drivers
```

5. 运行 `reboot` 命令重启。
6. 运行 `nvidia-smi` 能输出正确信息代表驱动安装成功。

Windows 驱动安装

1. 登录 [NVIDIA 驱动下载官网](#)。

2. 选择操作系统和安装包。以 M40 为例，选择如下驱动程序：



NVIDIA Driver Downloads

Advanced Driver Search

Product Type:
Tesla ▼

Product Series:
M-Class ▼

Product:
M40 24GB ▼

Operating System:
Windows Server 2012 R2 64 ▼

CUDA Toolkit:
8.0 ▼

Language:
English (US) ▼

Recommended/Beta:
All ▼

SEARCH

3. 打开下载驱动程序的文件夹，然后双击安装文件以启动它。按照说明安装驱动程序并根据需要重启实例。要验证 GPU 是否正常工作，请检查设备管理器。

安装失败原因

Linux 系统驱动安装失败表现为 nvidia-smi 无法工作，一般有下面几个常见原因：

1. 系统缺乏编译 kernel module 所需要的包，如 gcc，kernel-devel-xxx 等，导致无法编译，最终安装失败。
2. 系统里面存在多个版本的 kernel，由于 DKMS 的不正确配置，导致驱动编译为非当前版本 kernel 的 kernel module，导致 kernel module 安装失败。
3. 安装驱动后，升级了 kernel 版本导致原来的安装失效。

安装 AMD 驱动指引

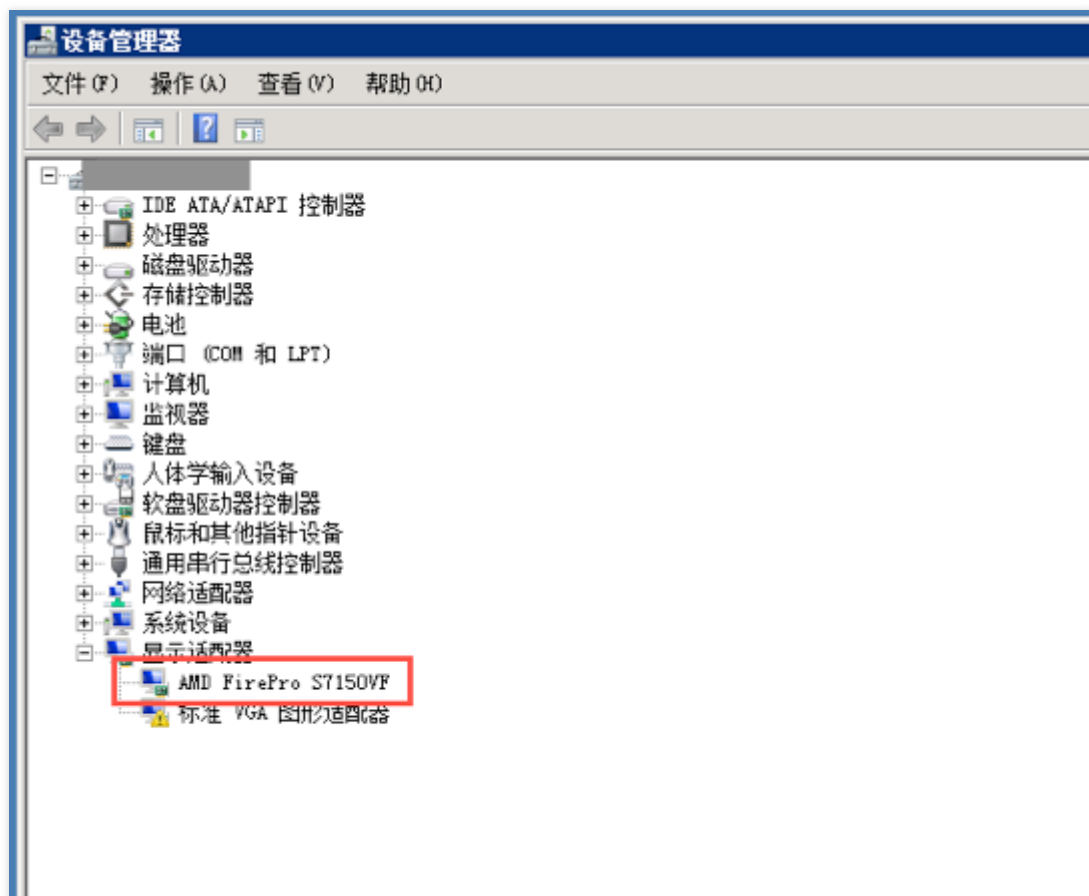
最近更新时间：2017-09-28 00:41:14

GPU 云服务器必须具备相应的 GPU 驱动才可正常运用，对于AMD GPU 云服务器必须针对您的编译环境安装合适的 AMD GPU 驱动程序。

GA2 实例驱动安装

Windows 安装驱动

1. 请从您购买的GPU云服务器内访问该链接下载 AMD GPU 驱动
http://mirrors.tencentyun.com/install/windows/s7150_guest_driver.7z
2. 打开下载驱动程序的文件夹，然后双击安装文件以启动它。要验证 GPU 是否正常工作，请检查设备管理器
3. 安装完成后，到设备管理器内查看，如下显示表明安装成功



安装 CUDA 驱动指引

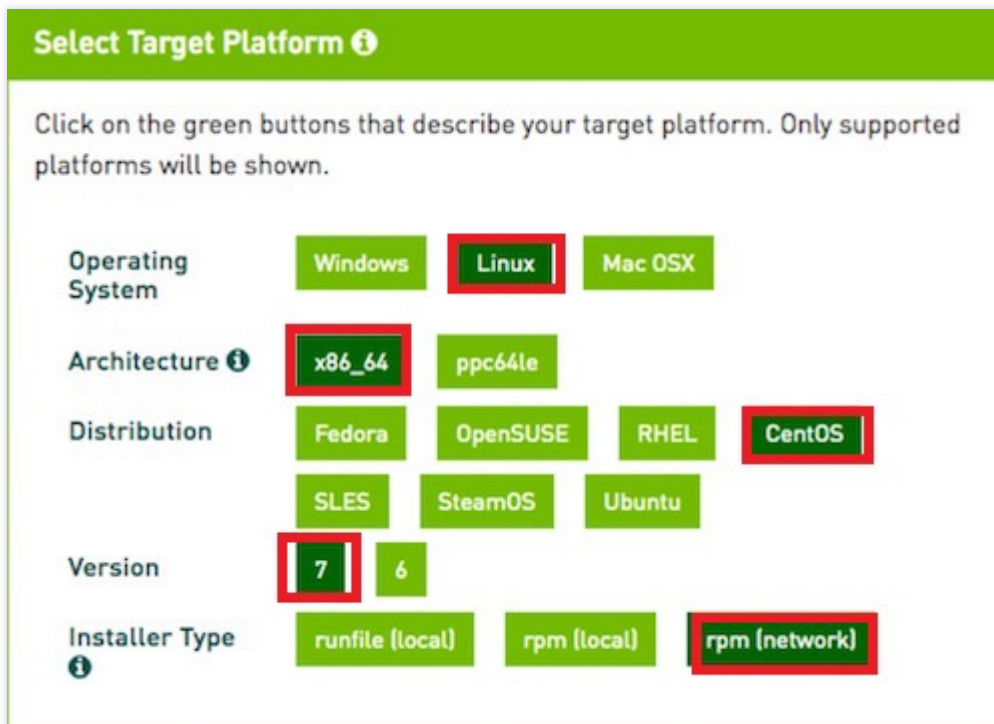
最近更新时间：2017-09-22 16:40:57

CUDA (Compute Unified Device Architecture) 是显卡厂商 NVIDIA 推出的运算平台。CUDA™ 是一种由 NVIDIA 推出的通用并行计算架构，该架构使 GPU 能够解决复杂的计算问题。它包含了 CUDA 指令集架构 (ISA) 以及 GPU 内部的并行计算引擎。开发人员现在可以使用 C 语言, C++, FORTRAN 来为 CUDA™ 架构编写程序，所编写出的程序可以在支持 CUDA™ 的处理器上以超高性能运行。

GPU 云服务器采用 NVIDIA 显卡，需要安装 CUDA 开发运行环境。以目前最常用的 CUDA 7.5 为例，可参照以下步骤进行安装。

Linux 系统指引

1. 登录 [CUDA驱动下载](https://developer.nvidia.com/cuda-75-downloads-archive) 或复制链接 <https://developer.nvidia.com/cuda-75-downloads-archive>。
2. 选择操作系统和安装包。以 CentOS 7.2 64 位为例，可按如下方式进行选择：



注意：

Installer Type 推荐选择 rpm (network)。

network：网络安装包，安装包较小，需要在主机内联网下载实际的安装包。
 local：本地安装包。安装包较大，包含每一个下载安装组件的安装包。

3. 右击【Download】 - 【复制链接地址】。



4. 登录 GPU 实例，使用 `wget` 命令，粘贴上述步骤复制的链接地址下载安装包；或通过在本本地系统下载 CUDA 安装包，上传到 GPU 实例的服务器。

```

# wget http://developer.download.nvidia.com/compute/cuda/7.5/Prod/local_installers/cuda-repo-rhel7-7-5-local-7.5-18.x86_64.rpm
http://developer.download.nvidia.com/compute/cuda/7.5/Prod/local_installers/cuda-repo-rhel7-7-5-local-7.5-18.x86_64.rpm
Resolving developer.download.nvidia.com (developer.download.nvidia.com)... 111.202.43.167, 111.202.43.168, 111.202.43.169, ...
Connecting to developer.download.nvidia.com (developer.download.nvidia.com):111.202.43.167:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1211912148 (1.1G) [application/x-rpm]
Saving to: 'cuda-repo-rhel7-7-5-local-7.5-18.x86_64.rpm'
    
```

5. 在 CUDA 安装包所在目录下运行如下命令：

```

sudo rpm -i cuda-repo-rhel7-7.5-18.x86_64.rpm
sudo yum clean all
sudo yum install cuda
    
```

6. 在 `/usr/local/cuda-7.5/samples/1_Uutilities/deviceQuery` 目录下，执行 `make` 命令，可以编译出 `deviceQuery` 程序。

7. 执行 deviceQuery 正常显示如下设备信息，此刻认为 CUDA 安装正确。

```
./deviceQuery Starting...

CUDA Device Query (Runtime API) version (CUDA static linking)

Detected 1 CUDA Capable device(s)

Device 0: "Tesla M40 24GB"
  CUDA Driver Version / Runtime Version      8.0 / 7.5
  CUDA Capability Major/Minor version number: 5.2
  Total amount of global memory:            24505 MBytes (25695092736 bytes)
  (24) Multiprocessors, (128) CUDA Cores/MP: 3072 CUDA Cores
  GPU Max Clock rate:                       1112 MHz (1.11 GHz)
  Memory Clock rate:                        3004 Mhz
  Memory Bus Width:                          384-bit
  LZ Cache Size:                             3145728 bytes
  Maximum Texture Dimension Size (x,y,z)    1D=(65536), 2D=(65536, 65536), 3D=(4096, 4096, 4096)
  Maximum Layered 1D Texture Size, (num) layers 1D=(16384), 2048 layers
  Maximum Layered 2D Texture Size, (num) layers 2D=(16384, 16384), 2048 layers
  Total amount of constant memory:          65536 bytes
  Total amount of shared memory per block:   49152 bytes
  Total number of registers available per block: 65536
  Warp size:                                 32
  Maximum number of threads per multiprocessor: 2048
  Maximum number of threads per block:      1024
  Max dimension size of a thread block (x,y,z): (1024, 1024, 64)
  Max dimension size of a grid size (x,y,z): (2147483647, 65535, 65535)
  Maximum memory pitch:                     2147483647 bytes
  Texture alignment:                         512 bytes
  Concurrent copy and kernel execution:     Yes with 2 copy engine(s)
  Run time limit on kernels:                 No
  Integrated GPU sharing Host Memory:       No
  Support host page-locked memory mapping:  Yes
  Alignment requirement for Surfaces:       Yes
  Device has ECC support:                   Disabled
  Device supports Unified Addressing (UVA):  Yes
  Device PCI Domain ID / Bus ID / location ID: 0 / 0 / 7
  Compute Mode:
    < Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) >

deviceQuery, CUDA Driver = CUDART, CUDA Driver Version = 8.0, CUDA Runtime Version = 7.5, NumDevs = 1, Device0 = Tesla M40
Result = PASS
```

Windows 系统指引

要在 Windows 实例上安装 CUDA，请使用远程桌面以管理员的身份登录您的 Windows 实例。

1. 在 [CUDA 驱动官网](#) 下载 CUDA 安装包。

2. 选择操作系统和安装包。以 Win Server 2012 R2 64 位为例，可按如下方式进行选择:

Select Target Platform ⓘ

Click on the green buttons that describe your target platform. Only supported platforms will be shown.

Operating System	Windows	Linux	Mac OSX
Architecture ⓘ	x86_64		
Version	10	8.1	Server 2012 R2
	Server 2008 R2		
Installer Type ⓘ	exe (network)	exe (local)	

Download Target Installer for Windows Server 2012 R2 x86_64

cuda_7.5.18_windows_network.exe (md5sum:
bd166ba460e7507e2ddab3e324653263)

Download (8.5 MB)

Installation Instructions:

1. Double click cuda_7.5.18_windows_network.exe
2. Follow on-screen prompts

For further information, see the [Installation Guide for Microsoft Windows](#) and the [CUDA Quick Start Guide](#).

3. 启动安装程序，按提示进行安装，如果最后出现完成对话框，则安装成功。

