k8s使用GPU

一、安装NVIDIA驱动

```
# NVIDIA 官网参考文档，在这检查相关依赖是否满足
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.html

# 官方安装文档
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.html#pre-installation-actions

# 官方下载链接
https://www.nvidia.com/Download/index.aspx

# MD5校验码
https://developer.download.nvidia.com/compute/cuda/12.3.1/docs/sidebar/md5sum.txt

# 参考文档
https://blog.csdn.net/holyvslin/article/details/132299184
```

Native Linux Distribution Support in CUDA 12.3 Update 1

- 后面2列应该是最低满足版本
- X86跟ARM又有不少的区别

Distribution	Kernel1	Default GCC	GLIBC
x86_64			
RHEL 9.y (y <= 2)	5.14.0-284	11.3.1	2.34
RHEL 8.y (y <= 8)	4.18.0-477	8.5.0	2.28
RHEL 7.y (y <= 9)	3.10.0-1160	6.x	2.17
CentOS 7.y (y <= 9)	3.10.0-1160	6.x	2.17
OpenSUSE Leap 15.y (y <= 5)	5.14.21-150500	7.5.0	2.31
Rocky Linux 8.y (y<=8)	4.18.0-477	8.5.0	2.28
Rocky Linux 9.y (y<=2)	5.14.0-284	11.3.1	2.34
SUSE SLES 15.y (y <= 5)	5.14.21-150500	7.5.0	2.31
Ubuntu 22.04.z (z <= 3) LTS	6.2.0-26	11.4.0	2.35
Ubuntu 20.04.z (z <= 6) LTS	5.15.0-67	9.4.0	2.31
Debian 12.2	6.1.0-13	12.2.0	2.36
Debian 11.8	5.10.197-1	10.2.1	2.31
Debian 10.13	4.19.0-21	8.3.0	2.28

Distribution	Kernel1	Default GCC	GLIBC
Fedora 37	6.0.7-301	12.2.1	2.36
KylinOS V10 SP2	4.19.90-25.14.v2101.ky10	7.3.0	2.28
Arm64 sbsa			
RHEL 9.y (y <= 2)	5.14.0-284	11.3.1	2.34
RHEL 8.y (y <= 8)	4.18.0-477	8.5.0	2.28
SUSE SLES 15.y (y <= 5)	5.14.21-150500	7.5.0	2.31
Ubuntu 22.04 LTS (z <= 3) LTS	5.15.0-79	11.4.0	2.35
Ubuntu 20.04.z (z <= 6) LTS	5.4.0-125	9.4.0	2.31
POWER 9			
RHEL 8.y (y <= 8)	4.18.0-477	8.5.0	2.28

二、依赖检查

- 以CentOS7.9为例，检查基础环境

```
# 1、查看OS版本：
[root@k8s-node02 ~]# cat /etc/redhat-release
CentOS Linux release 7.9.2009 (Core)

# 2、查看内核版本
[root@k8s-node02 ~]# uname -r
4.19.12-1.el7.elrepo.x86_64

# 3、检查gcc版本、glibc版本
[root@k8s-node02 ~]# gcc --version
gcc (GCC) 4.8.5 20150623 (Red Hat 4.8.5-44)
 # 版本太低要升级
Copyright (C) 2015 Free Software Foundation, Inc.
This is free software; see the source for copying conditions.  There is NO
warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

[root@k8s-node02 ~]# ldd --version
ldd (GNU libc) 2.17
Copyright (C) 2012 Free Software Foundation, Inc.
This is free software; see the source for copying conditions.  There is NO
warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
Written by Roland McGrath and Ulrich Drepper.

##### 升级 gcc
yum install -y centos-release-scl
yum install -y devtoolset-9
scl enable devtoolset-9 bash

##### 检查gcc版本
[root@k8s-node02 ~]# gcc --version
gcc (GCC) 9.3.1 20200408 (Red Hat 9.3.1-2)
Copyright (C) 2019 Free Software Foundation, Inc.
This is free software; see the source for copying conditions.  There is NO
warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```

三、RHEL 7 / CentOS 7

1. 查看系统显卡情况

centos7.9来实验；

1、使用GPU得有nvidia显卡，所以得先看一下本机的显卡，分别输入a命令和b命令：

```
a、lspci | grep -i nvidia

b、lspci | grep -i vga
```

如果结果显示有nvidia字样还带有型号啥的，说明你机器上有nvidia显卡，可能能够使用GPU（这里用可能二字是因为计算是nvidia显卡也不一定是能用GPU加速计算，对于TensorFlow而言，应该是算力达到一定程度才可以），若果要是输入刚才命令啥也没有显示，而显示：

00:02.0 VGA compatible controller: Cirrus Logic GD 5446

说明机器上没有nvidia显卡，只有一个普通显卡，即便你安装cuda也没用，不能加速计算。

```
[root@localhost ~]# lspci | grep -i vga
00:02.0 VGA compatible controller: Red Hat, Inc. QXL paravirtual graphic card
(rev 04)
[root@localhost ~]# lspci | grep -i nvidia
00:0c.0 3D controller: NVIDIA Corporation Device 20b5 (rev a1)
[root@localhost ~]# nvidia-smi
bash: nvidia-smi: 未找到命令...
```

问题：

- **有显卡（NVIDIA Corporation Device 20b5 (rev a1)），无驱动。**

解决方法：

- 00:0c.0 3D controller: NVIDIA Corporation Device 20b5 (rev a1)
  我们要查找的就是：**20b5**
  然后进入网址：PCI Devices

1.1 定位显卡型号

Main -> PCI Devices -> Vendor 10de -> **Device 10de:2208**

Name: GA102 [GeForce RTX 3080 Ti]

Discussion

Name: GA102 [GeForce RTX 3080 Ti]

T4CFantasy

2021-03-01 03:53:08

Discuss

Subsystems

Id	Name

Add item

20b5 Jump

输入并查找，得到

Main -> PCI Devices -> Vendor 10de -> **Device 10de:20b5**

Name: GA100 [A100 PCIe 80GB]

Discussion

Name: GA100 [A100 PCIe 80GB]

T4CFantasy

2021-06-28 14:45:29

Discuss

1.2 下载驱动

下载显卡的官方网址

下载命令：

```
wget https://cn.download.nvidia.cn/tesla/470.141.03/NVIDIA-Linux-x86_64-470.141.03.run
```

2. 安装NVIDIA GPU驱动

安装gcc等依赖包

当安装GPU驱动时，提示缺少相关的依赖包，在此，我们需要提前安装相关的依赖包，目前需要用到的是gcc , g++ , make：

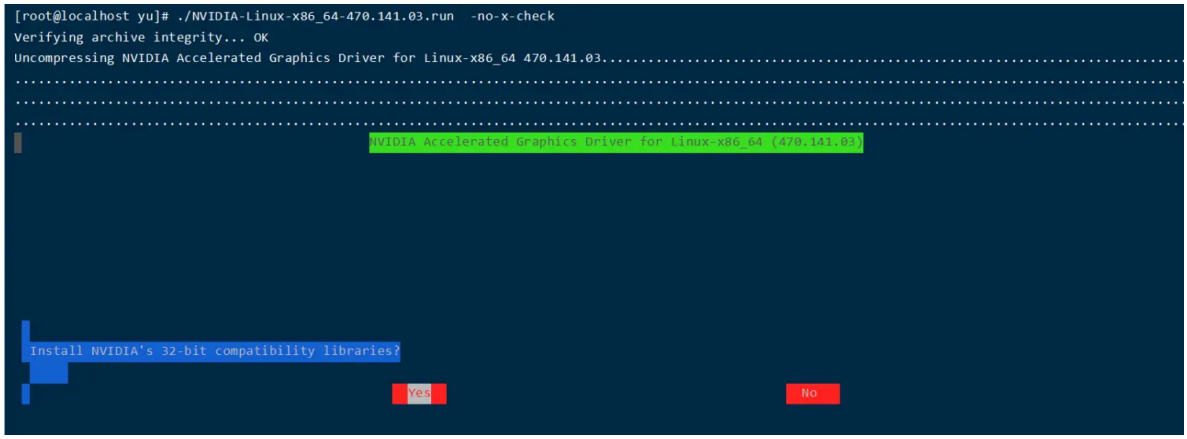
```
[root@localhost ~]# yum -y install gcc gcc-c++ kernel-devel make
```

安装下载的GPU驱动：NVIDIA-Linux-x86_64-470.141.03.run ，目前驱动版本为：470.141.03，如下执行该驱动文件，即可安装。

注：

- 提示权限不足，直接chmod 777 权限
- 在后面加上不对Xserver进行检查的命令（-no-x-check）就可以安装成功！原因 --> 主要是由于安装远程控制lightgm 导致X-server启动。

```
[root@localhost yu]# chmod 777 NVIDIA-Linux-x86_64-470.141.03.run
[root@localhost yu]# ./NVIDIA-Linux-x86_64-470.141.03.run  -no-x-check
```



最后执行`nvidia-smi`验证是否安装成功

```
[root@localhost yu]# nvidia-smi
```

成功安装 大功告成

```
[root@localhost yu]# nvidia-smi
Fri Sep 16 10:36:10 2022
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 470.141.03   Driver Version: 470.141.03   CUDA Version: 11.4     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  NVIDIA A100 80G...  Off  | 00000000:00:0C.0 Off |                    0 |
| N/A   45C    P0    70W / 300W |      0MiB / 80994MiB |      3%      Default |
|                               |                      |             Disabled |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

3. 安装CUDA 11.1

在nvidia官网下载cuda版本11.1.1，并安装

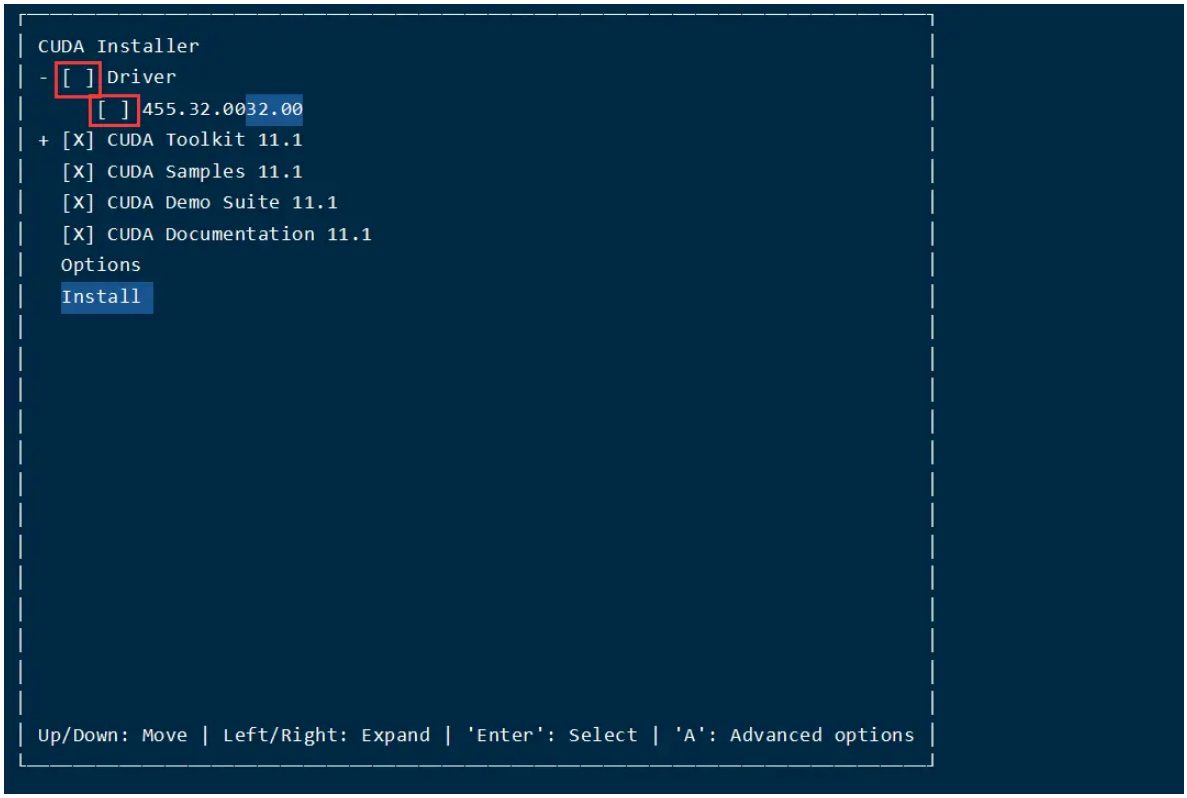
地址为https://developer.nvidia.com/cuda-toolkit-archive。建议采用runfile（**local**）方式下载安装。

Download Installer for Linux CentOS 7 x86_64

The base installer is available for download below.

```
# Installation Instructions:
wget
https://developer.download.nvidia.com/compute/cuda/11.1.1/local_installers/cuda_
11.1.1_455.32.00_linux.run
sudo sh cuda_11.1.1_455.32.00_linux.run
```

如下图，Driver选项不要勾选了，前面已经安装GPU驱动了。




```
[root@localhost yu]# sudo sh cuda_11.1.1_455.32.00_linux.run
===========
= Summary =
===========

Driver:   Not Selected
Toolkit:  Installed in /usr/local/cuda-11.1/
Samples:  Installed in /root/, but missing recommended libraries

Please make sure that
 -   PATH includes /usr/local/cuda-11.1/bin
 -   LD_LIBRARY_PATH includes /usr/local/cuda-11.1/lib64, or, add
/usr/local/cuda-11.1/lib64 to /etc/ld.so.conf and run ldconfig as root

To uninstall the CUDA Toolkit, run cuda-uninstaller in /usr/local/cuda-11.1/bin
***WARNING: Incomplete installation! This installation did not install the CUDA
Driver. A driver of version at least 455.00 is required for CUDA 11.1
functionality to work.
To install the driver using this installer, run the following command, replacing
<CudaInstaller> with the name of this run file:
    sudo <CudaInstaller>.run --silent --driver

Logfile is /var/log/cuda-installer.log
```

创建环境变量，编辑 `~/.bashrc` 文件：

```
[root@localhost ~]# vim ~/.bashrc
```

将下面命令追加到文件最后面：

```
export CUDA_HOME=/usr/local/cuda
export PATH=$PATH:$CUDA_HOME/bin
export LD_LIBRARY_PATH=/usr/local/cuda-
11.1/lib64${LD_LIBRARY_PATH:+:${LD_LIBRARY_PATH}}
```

为 `nvcc` 命令创建一个软连接到/usr/bin目录：

```
[root@localhost ~]# sudo ln -s /usr/local/cuda/bin/nvcc /usr/bin/nvcc
```

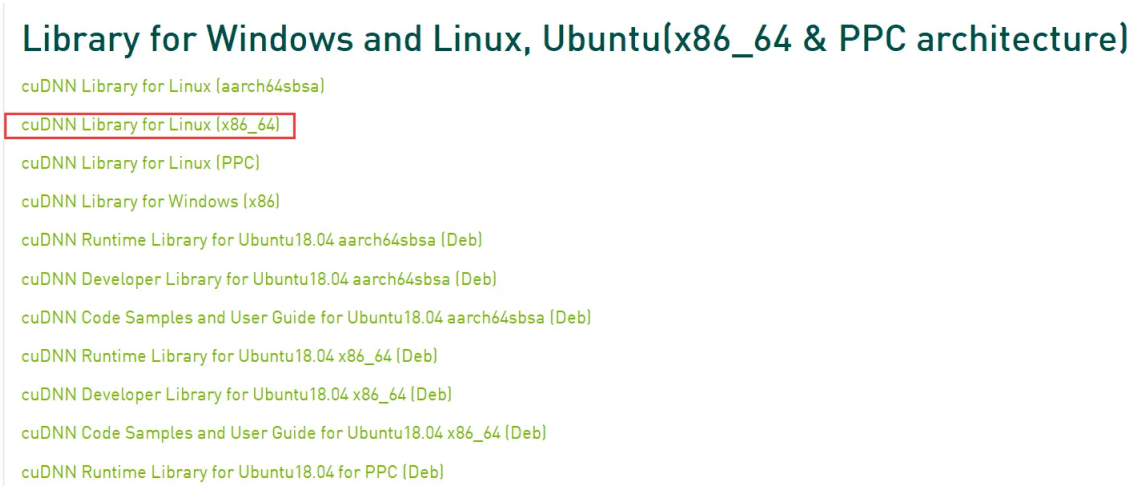
使用nvcc命令查看cuda的版本：

```
[root@localhost ~]# nvcc --version
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2020 NVIDIA Corporation
Built on Mon_Oct_12_20:09:46_PDT_2020
Cuda compilation tools, release 11.1, V11.1.105
Build cuda_11.1.TC455_06.29190527_0
```

4. cudnnan安装

从nvidia官网下载cudnn8.0

https://developer.nvidia.com/rdp/cudnn-archive



安装:

```
tar -xzvf cudnn-11.1-linux-x64-v8.0.4.30.tgz
sudo cp cuda/include/cudnn*.h /usr/local/cuda-11.1/include
sudo cp -P cuda/lib64/libcudnn* /usr/local/cuda-11.1/lib64
sudo chmod a+r /usr/local/cuda-11.1/include/cudnn*.h /usr/local/cuda-
11.1/lib64/libcudnn*
```

安装验证:

```
cat /usr/local/cuda-11.1/include/cudnn_version.h | grep CUDNN_MAJOR -A 2
```

```
[root@localhost local]# cat /usr/local/cuda/include/cudnn_version.h | grep CUDNN_MAJOR -A 2
#define CUDNN_MAJOR 8
#define CUDNN_MINOR 0
#define CUDNN_PATCHLEVEL 4
--
#define CUDNN_VERSION (CUDNN_MAJOR * 1000 + CUDNN_MINOR * 100 + CUDNN_PATCHLEVEL)

#endif /* CUDNN_VERSION_H */
```

```
[root@localhost local]# ll
总用量 4
drwxr-xr-x.  2 root root    6 4月   11 2018 bin
lrwxrwxrwx.  1 root root   21 9月   16 11:36 cuda -> /usr/local/cuda-11.1/
drwxr-xr-x. 16 root root 4096 9月   16 11:37 cuda-11.1
drwxr-xr-x.  2 root root    6 4月   11 2018 etc
drwxr-xr-x.  2 root root    6 4月   11 2018 games
```