

# K80

```
pacman -S nvidia-470xx-dkms nvidia-470xx-settings nvidia-470xx-util
pacman -U https://archive.archlinux.org/packages/c/cuda/cuda-11.4.2-1-x86_64.pkg.tar.zst
```

nvidia-smi

```
+-----+
--+
| NVIDIA-SMI 470.256.02   Driver Version: 470.256.02   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. |
|                    |                    |                    |
M. |
|=====+=====+=====+
==|
|  0  Tesla K80                 Off | 00000000:03:00.0 Off |
0 |
| N/A   53C    P0     61W / 149W |      0MiB / 11441MiB |      0%
Default |
|                    |                    |
N/A |
+-----+-----+-----+
--+
|  1  Tesla K80                 Off | 00000000:04:00.0 Off |
0 |
| N/A   41C    P0     72W / 149W |      0MiB / 11441MiB |     86%
Default |
|                    |                    |
N/A |
+-----+-----+-----+
--+
+-----+
--+
| Processes:
|
| GPU  GI  CI           PID  Type  Process name          GPU
Memory |
|      ID  ID              |          |          |          |
|=====+=====+=====+
==|
| No running processes found
```

```
|
+-----+
--+
```

```
# cat /proc/driver/nvidia/version
NVRM version: NVIDIA UNIX x86_64 Kernel Module 470.256.02 Thu May 2
14:37:44 UTC 2024
GCC version: gcc version 14.1.1 20240522 (GCC)

# nvcc -V
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Sun_Aug_15_21:14:11_PDT_2021
Cuda compilation tools, release 11.4, V11.4.120
Build cuda_11.4.r11.4/compiler.30300941_0
```

samples

```
git clone --depth 1 --branch v11.4.1 https://github.com/NVIDIA/cuda-
samples.git /opt/cuda-samples
cd /opt/cuda-samples
cd Samples/deviceQuery/
make
```

deviceQuery

```
cd /opt/cuda-samples/bin/x86_64/linux/release
./deviceQuery

./deviceQuery Starting...

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

Detected 2 CUDA Capable device(s)

Device 0: "Tesla K80"
  CUDA Driver Version / Runtime Version          11.4 / 11.4
  CUDA Capability Major/Minor version number:    3.7
  Total amount of global memory:                 11441 MBytes (11997020160
bytes)
  (013) Multiprocessors, (192) CUDA Cores/MP:   2496 CUDA Cores
  GPU Max Clock rate:                            824 MHz (0.82 GHz)
  Memory Clock rate:                             2505 Mhz
  Memory Bus Width:                              384-bit
  L2 Cache Size:                                 1572864 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 shared memory per multiprocessor:   114688 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:                    Enabled
Device supports Unified Addressing (UVA):  Yes
Device supports Managed Memory:           Yes
Device supports Compute Preemption:       No
Supports Cooperative Kernel Launch:       No
Supports MultiDevice Co-op Kernel Launch: No
Device PCI Domain ID / Bus ID / location ID: 0 / 3 / 0
Compute Mode:
  < Default (multiple host threads can use ::cudaSetDevice() with device
simultaneously) >

```

Device 1: "Tesla K80"

```

CUDA Driver Version / Runtime Version     11.4 / 11.4
CUDA Capability Major/Minor version number: 3.7
Total amount of global memory:            11441 MBytes (11997020160
bytes)
(013) Multiprocessors, (192) CUDA Cores/MP: 2496 CUDA Cores
GPU Max Clock rate:                       824 MHz (0.82 GHz)
Memory Clock rate:                         2505 Mhz
Memory Bus Width:                          384-bit
L2 Cache Size:                             1572864 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 shared memory per multiprocessor:   114688 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:                Enabled
Device supports Unified Addressing (UVA): Yes
Device supports Managed Memory:       Yes
Device supports Compute Preemption:   No
Supports Cooperative Kernel Launch:   No
Supports MultiDevice Co-op Kernel Launch: No
Device PCI Domain ID / Bus ID / location ID: 0 / 4 / 0
Compute Mode:
  < Default (multiple host threads can use ::cudaSetDevice() with device
simultaneously) >
> Peer access from Tesla K80 (GPU0) -> Tesla K80 (GPU1) : Yes
> Peer access from Tesla K80 (GPU1) -> Tesla K80 (GPU0) : Yes

deviceQuery, CUDA Driver = CUDART, CUDA Driver Version = 11.4, CUDA Runtime
Version = 11.4, NumDevs = 2
Result = PASS
```

## pytorch

```
pip3 install torch torchvision --index-url
https://download.pytorch.org/whl/cu118

python -c "import torch; print(torch.__version__);
print(torch.cuda.is_available());"
# 2.3.1+cu118
# True
```

## ultralytics

```
pip install ultralytics

python -c "import ultralytics;
print(ultralytics.utils.checks.cuda_is_available());"
# True
```

From:  
<https://wiki.csgalileo.org/> - **Galileo Labs**

Permanent link:  
<https://wiki.csgalileo.org/tips/k80>

Last update: **2024/07/19 07:36**

