NVIDIA SMI

`nvidia-smi` (System Management Interface) is a CLI tool that comes with NVIDIA drivers and allows you to monitor, manage, and control NVIDIA GPU devices on your system.

It talks directly with the NVIDIA kernel driver and uses the NVML (NVIDIA Management Library) under the hood.

Key Features:
- View GPU and process usage
- Control persistence & compute mode
- Manage MIG (Multi-Instance GPU)
- Power & thermal management
- Logging & diagnostics
- ECC error tracking
- Kill GPU processes
- Query detailed properties

How It works:
- Installed as part of the NVIDIA driver package (`nvidia-utils`, `nvidia-smi`).
- Communicates via NVML API with the kernel module.
- Works for Tesla , A-series ,RTX and data center GPUs (A100,L40 ,etc)

Basic Commands:
- nvidia -smi (get gpu info)

```
+-----------------------------------------------------------------------------------------+
| NVIDIA-SMI 550.54.14              Driver Version: 550.54.14      CUDA Version: 12.2      |
|-----------------------------------------+------------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id          Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |           Memory-Usage | GPU-Util  Compute M. |
|=========================================+========================+======================|
|   0  NVIDIA A100-SXM4           On      | 00000000:82:00.0 Off   |                    0 |
| 31%   38C    P0             65W / 300W  |      0MiB / 40960MiB   |     0%       Default |
+-----------------------------------------+------------------------+----------------------+
```

Shows:
- GPU name
- Driver version
- Power usage
- Memory usage
- Utilization %
- Compute Mode (Default /Exclusive)

- watch -n 1 nvidia-smi (Watch GPU in Real-Time)
- nvidia-smi pmon -c 1(list of running processes on GPU)
        Or
- nvidia-smi --query-compute-apps=pid,process_name,used_memory --format=csv

Power ,Performance , and Persistence:

- sudo nvidia-smi -pm 1 (Enable Persistence Mode-Keeps GPU driver and memory active even when no process is using it (avoid startup delay))
- sudo nvidia-smi -pm 0 (Disable Persistence)

Change GPU Compute Mode:
    Affects how multiple processes can share GPU resources.
- sudo nvidia-smi -c 0   # Default
- sudo nvidia-smi -c 1   # Exclusive Thread
- sudo nvidia-smi -c 2   # Prohibited
- sudo nvidia-smi -c 3   # Exclusive Process

Set Power limit :
- nvidia-smi -pl 250 (Sets max power draw to 250W)
- nvidia-smi -q -d POWER (To see supported range)

Query Fan Speed, Clock, Temperature:
- nvidia-smi -q -d TEMPERATURE
- nvidia-smi -q -d FAN
- nvidia-smi -q -d CLOCK

MIG Management (Multi-Instance GPU):
- sudo nvidia-smi -i 0 -mig 1 (Enable MIG Mode)

- sudo nvidia-smi mig -cgi 19,19,19 -C (Create MIG Profiles)
  - `-cgi`: List of profile IDs (e.g. 19 = 1g.10gb)
  - `-C`: Commit the profile
- sudo nvidia-smi mig -dci (Delete MIG Instances)
- nvidia-smi -L (List Profiles and Instances)

Advanced Query / Logging :

- nvidia-smi -q (Detailed Info Per GPU)
  Shows :
    - ECC status
    - BIOS info
    - Clock speeds
    - Power limits
    - Active processes
- nvidia-smi -q -x > report.xml (Save Logs to File - Gives an XML diagnostic dump.)
- sudo nvidia-smi --gpu-reset -i 0 # Reset GPU
- sudo nvidia-smi --pid=12345 --kill   (Kill GPU Process)

Full Capability List:
- nvidia-smi --help (To see all commands & options)
- nvidia-smi --help-query-gpu (nvidia-smi --help-query-gpu)