Maxwell.

- 4 GPCs (1 GPC 에 4 SMM 존재.)
- SMM : (32 CUDA + 8 LD/ST + 8 SFU) x 4
  - 4 개 warp scheduler
    - 각 warp scheduler 는 clock 당 2 instruction dispatch.
    - 32 CUDA cores x 4 = 128 CUDA core per SM
    - 디자인 shared memory + L1 cache 구조 = on-chip memory X
      96 KB shared memory 비교도 존재.
      L1 cache 는 texture caching function 과 share.

Volta

- 6 GPCs
  - 각 GPC는 7 TPCs (각각 2SMs 포함), 14 SMs
- 84 SMs
  - 각 SM은 64 FP32 cores, 64 INT32 cores
    32 FP64 cores, 8 Tensor cores
    4 texture units.
- 8 512 bit memory controllers (4096 bits total)

Volta Streaming Multiprocessor.

- GV100 SM은 4 processing blocks
  각 processing block은.
  - 16 FP32, 16 INT32, 8 FP64 2 Tensor cores
    L0 instruction cache. one warp scheduler
    one dispatch unit. 64KB Register file.
  - Shared memory + L1 data cache ~~96~~ 128 KB