

BÁO CÁO BÀI TẬP 02

29/10/2019

Họ và tên: Huỳnh Minh Huấn

MSSV: 1612858

Câu 1

```
idl612858@9aa3d017fc93:~/BT02$ nvcc bt02.cu -o bt02
idl612858@9aa3d017fc93:~/BT02$ ./bt02
*****GPU info*****
Name: GeForce GTX 1080 Ti
Compute capability: 6.1
Num SMs: 28
Max num threads per SM: 2048
Max num warps per SM: 64
GMEM: 11713052672 bytes
*****

Input size: 16777217

Kernel 1
Grid size: 16385, block size: 512
Kernel time = 11.267744 ms, post-kernel time = 0.036832 ms
CORRECT :)

Kernel 2
Grid size: 16385, block size: 512
Kernel time = 0.332256 ms, post-kernel time = 0.036320 ms
INCORRECT :(

Kernel 3
Grid size: 16385, block size: 512
Kernel time = 0.347392 ms, post-kernel time = 0.037056 ms
INCORRECT :(
idl612858@9aa3d017fc93:~/BT02$
```

Câu 2:

Block size	Grid size	Num Blocks / SM	Occupancy (%)	Kernel time (ms)	Post-kernel time (ms)	Total time (ms)
1024	8193	2	100	13.808448	0.022912	13.83136
512	16385	4	100	11.2654	0.037504	11.302904
256	32769	8	100	9.65488	0.063168	9.718048
128	65537	16	100	8.846656	0.116608	8.963264

Tính toán:

- Phiên bản sử dụng CC 6.1
 - o Max number blocks per SM là: 32 (blocks/SM)
 - o Max number threads per SM là: 2048 (threads/SM)

- Max number warps per SM là: 64 (warps/SM)
- Warp size: 32
- Num blocks / SM:
 - Blk size 1024: $2048 / 1024 = 2$ (blocks/SM) (< 32)
 - Blk size 512: $2048 / 512 = 4$ (blocks/SM) (< 32)
 - Blk size 256: $2048 / 256 = 8$ (blocks/SM) (< 32)
 - Blk size 128: $2048 / 128 = 16$ (blocks/SM) (< 32)
- Occupancy: (num warps per SM) / (max num warps per SM)
 Công thức là : $((\text{num_threads} / \text{warp_size}) * (\text{blocks/SM})) / (\text{max_warps_per_SM})$
 Num_threads / warp_size -> số warps trong 1 blocks
 Blocks/SM tính từ trên.
 - Blk size 1024: $(1024 / 32) * 2 / 64 = 100\%$
 - Blk size 512: $(512 / 32) * 4 / 64 = 100\%$
 - Blk size 256: $(256 / 32) * 8 / 64 = 100\%$
 - Blk size 128: $(128 / 32) * 16 / 64 = 100\%$

Nhận xét:

- Khi blockSize tăng thì, số lượng phép tính cần tính trong device tăng lên (dùng nhiều thread hơn)
-> thời gian kernel tăng.
- Còn số phần tử output trả ra sẽ giảm vì số phần tử output chính là gridSize. Được tính ở host, nên khi blockSize tăng thì thời gian post-kernel giảm.

Câu 3:

```
id1612858@9aa3d017fc93: ~/BT02
Kernel time = 5.750560 ms, post-kernel time = 0.037120 ms
CORRECT :)
id1612858@9aa3d017fc93:~/BT02$ ./bt02 512
*****GPU info*****
Name: GeForce GTX 1080 Ti
Compute capability: 6.1
Num SMs: 28
Max num threads per SM: 2048
Max num warps per SM: 64
GMEM: 11713052672 bytes
*****

Input size: 16777217

Kernel 1
Grid size: 16385, block size: 512
Kernel time = 11.261888 ms, post-kernel time = 0.037184 ms
CORRECT :)

Kernel 2
Grid size: 16385, block size: 512
Kernel time = 10.294816 ms, post-kernel time = 0.052000 ms
CORRECT :)

Kernel 3
Grid size: 16385, block size: 512
Kernel time = 5.698816 ms, post-kernel time = 0.036576 ms
CORRECT :)
id1612858@9aa3d017fc93:~/BT02$
```

Câu 4:

Block có kích thước là 128 thread -> 4 warps lần lượt là 1, 2, 3, 4

Đối với kernel 3 thì stride ngược lại.

stride	warp phân kỳ		
	Kernel 1	Kernel 2	Kernel 3
1	không có warp nào	không có warp nào	1
2	tất cả	Không có warp nào	1
4	tất cả	Không có warp nào	1
8	tất cả	Không có warp nào	1
16	tất cả	1	Không có warp nào
32	tất cả	1	Không có warp nào
64	1, 3	1	Không có warp nào
128	1	1	Không có warp nào

Câu 5:

```
id1612858@9aa3d017fc93: ~/BT02
* Documentation:  https://help.ubuntu.com
* Management:    https://landscape.canonical.com
* Support:       https://ubuntu.com/advantage
Last login: Fri Oct 25 09:23:42 2019 from 172.29.65.66
id1612858@9aa3d017fc93:~$ cd BT02
id1612858@9aa3d017fc93:~/BT02$ ./bt02_p2
*****GPU info*****
Name: GeForce GTX 1080 Ti
Compute capability: 6.1
Num SMs: 28
Max num threads per SM: 2048
Max num warps per SM: 64
GMEM: 11713052672 bytes
*****

Input size: 33554433

Processing time (use host): 60.502014 ms

Processing time (use device partly): 42.715839 ms
CORRECT :)

Processing time (use device fully): 43.345570 ms
CORRECT :)
```

HẾT