16.1 GPU 加速求和

搭建GPU计算环境，计算 s=1＋2＋…＋1000000
输出计算结果，并与CPU计算相比较，验证结果正确性；
分析计算效率 （与CPU的加速比）；

源文件 SUM_GPU.cu 与 SUM_GPU.cuh 编译命令 nvcc SUM_GPU.cu -o SUM_GPU.exe

结果展示如下,由于开启优化，计算 s=1~1000000 对于 CPU 负荷不大，因此在计算时候采用 S=1+2+……+2000000000,计算结果两者一致，算法为两步规约。

```
Time used in CPU: 1.518585, result:2000000001000000000
NBlock  NThread         Result            Time
1024    1024    2000000001000000000     0.004213
```

CPU 用时 1.5185s，GPU 加速用时 0.004213s，加速比 360.43，具体计算值随给定的块数和线程数变化,当块数和线程数较小时,计算速度还是 CPU 较快。规约内核代码如下所示：

```
//kernal funcation used in GPU
__global__ void
GPU_SUM(const long long until,long long* device_ptr){
    extern __shared__ long long part[];
    long long sum = 0;
    const int curThread = threadIdx.x;
    for( int index = blockIdx.x * blockDim.x + curThread;
            index <= until;
            index += blockDim.x * gridDim.x)
    {
        sum += (index);
    }
    part[curThread] = sum;
    __syncthreads();
    for(int activeThread = blockDim.x >> 1;
            activeThread;
            activeThread >>= 1)
    {
        if(curThread<activeThread){
            part[curThread] += part[curThread+activeThread];
        }
        __syncthreads();
    }
    if(curThread == 0){
        device_ptr[blockIdx.x] = part[0];
    }
}
```

采用显卡设备规格

显卡型号	核心代号	制造工艺(nm)	流处理器/RT 核心/Tensor 核心	核心频率(MHz)
RTX3060	GA106-300/302	8nm	3584/28/112	1320
加速频率 (MHz)	显存位宽(-bit)	显存容量	显存频率(GHz)	整卡功耗(W)
1777	192	12GB GDDR6	15	170