COMPUTAÇÃO DE ALTO DESEMPENHO

2017/2018

Test 2

14/06/2018

1. Consider the following CUDA kernel, where array big (with size n) is much bigger than array small (of size m, typically less than 32). Optimize this kernel so that it makes a better use of the GPU's memory hierarchy.

```
__global__ void my_kernel(int* out, int* big, int n, int* small, int m) {
    unsigned int i = blockIdx.x*blockDim.x + threadIdx.x;
    if (i < n)
       out[i] = big[i] + small[(i*4)%m];
}
```

2. Consider that you want to develop a computational kernel to be executed simultaneously on a multi-core CPU and on one, or more, GPUs.
   a) Is it possible to program a single kernel to be compiled for both types of processors? If so, what technology would use (CUDA, OpenCL, OpenMP, some other)? Justify. If not, how can this multi-platform execution be achieved?
   b) Following your answer of a), how could your solution allow for optimizations tailored for one of the two processor types? You are allowed to introduce modifications to your answer of a), if needed.

3. In the context of the Spark distributed parallel computing framework:
   a) Implement a function that, given two RDDs of key-value pairs (Integer, Integer), returns the key for which the sum of the values of both RDDs is maximum.
   b) What is the result of your solution when applied to the following RDDs?
      { (1,1), (2,20), (3,3) } and { (1,1), (3,3) }

4. Consider the following Spark pseudo-code:

```
JavaPairRDD<Integer, Integer> rdd = …
JavaPairRDD<Integer, Integer> rdd2 = …
while (someCondition) {
  int value = rdd.groupByKey().reduce((x,y) -> …);
  rdd = rdd.join(rdd2).mapToPair(p -> (… value …));
}
```

   a) Assuming an execution on a cluster of multiple nodes, identify the (Spark) stages that the execution spawns.
   b) Identify which parts of the code may be optimized, justifying your choices.
   c) Propose optimizations for the code parts identified in b)

5. Comment the following statement, clearly sustaining your arguments: "Spark and MPI both follow a Single Program Multiple Data execution model".

6. Suppose you want to accelerate the execution of multiple chained Spark map operations by resorting to GPUs or the Xeon Phi.
   a) Which accelerator would you choose? Justify your answer.
   b) What if we add filter operations to maps, your choice remains the same? Justify your answer.