**Parallel Programming Skills Foundations**

* What are the basic steps (show all steps) in building a parallel program? Show at least

one example.

**The first step is identifying the tasks that could potential run simultaneously then the next step is about checking whether the tasks are independent or not and determine if the process required for each of them. The final step is about implementing the program using the master/worker system and static load balancing if it is possible. Inscribing a square in a circle could be an example.**

* What is Map Reduce?

**Map Reduce is a parallel computation mechanism for processing and distributing large amount of raw data across multiple machines.**

* What is map and what is reduce?

**These are two different functions in the MapReduce library developed by users of this library. Map takes pair as an input and gives a set of key and value pairs as an output. Reduce accepts a key and value from the output of the map function and generate smaller set of values by merging the values together for that key.**

* Why MapReduce?

**MapReduce significantly reduces our program’s run time by allowing parallel computation as a result we will have a faster run time than sequential executing programs**

* Show an example for MapReduce.

**For example, let’ say we have a program that counts the number of occurrences of each word in a document. In this program we can use the map function to map each word with its number of occurrence and then we can use the reduce function to ignore the key/ value pairs that have already been counted.**

* Explain in your own words how MapReduce model is executed?

**First the map reduce library distribute multiple copies of the tasks with the input files in a cluster of a machine. From those copies, there is master one that assigns map or reduce tasks to the workers, which are the rest of the copies of the program. Based on their assigned task these workers read the contents of their input shard then they parse the key/value pairs and send it to the map function which is stored in the local disk. Then these pairs from the local disk will be passed back to the master then the master notifies these locations to the reduce workers. Then the reduce workers read the buffered data from the local disks to the map workers and it sort it by the intermediate keys. The reduce workers iterate over the sorted data and pass these data to the reduce function, which adds the result to the final output file**. **Finally, the master runs the user programs ad the MapReduce call will end and go back to the user code.**

* List and describe three examples that are expressed as MapReduce computations.

**The distributed Grep: produces a line if there is a match between the document and the given pattern. The reduce function copies the given intermediate data to the output.**

**Term-vector per Host: organize the most important word in a document with its frequency of occurrence as a pair.**

**Inverted index: The map function pairs the word with its id. The reduce function accept those pairs and sort them ad produce pairs of word and list then they form an inverted index.**

* When do we use OpenMP, MPI and, MapReduce (Hadoop), and why?

**OpenMP** is a free efficient directive-based library so may developers use it widely to support their programs.

**MPI used for parallel model implementation interface by passing a message and distribute that to the memory.**

**Hadoop: is a programming model that can be used for processing big data in parallel by decomposing the work into independent set of tasks.**

* In your own words, explain what a Drug Design and DNA problem is in no more than 150 words

**The goal of the Drug design and DNA problem is finding drugs or ligands that are good nominees based on previous identification of a protein that are related with the disease through an experiment or molecular modeling computation. In the experiment or in the computation, these selected ligands will be tested against the identified protein of the disease to know if they bind with the protein in meaningful way or not. A score will be set based on the binding properties and the best scores will be marked and become a good nominee to make the drug for that disease.**

**Part 2**

**5) Measure Run-Time**

|  |  |
| --- | --- |
| Implementation | Time 1 thread |
| dd\_serial | 291.35 |
| dd\_omp | 0.02 |
| dd\_threads | 0.02 |

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

**Run time of dd\_omp ad dd\_threads with 2,3 and 4 threads.**

|  |  |  |  |
| --- | --- | --- | --- |
| Implemetation | Time(s) 2 Threads | Time (s) 3 Threads | Time(s) 4 threads |
| dd**\_**omp | 0.02 | 0.04 | 0.34 |
| dd\_threads | 0.02 | 0.04 | 0.23 |

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

**2.3 Discussion Questions**

1. Which approach is the fastest?

**The dd\_threads approach is the fastest. As you can see from the table ad the screenshots, the omp and threads solution have the same run time. However, with 4 threads the omp took more time than the threads approach.**

2. Determine the number of lines in each file (use wc -l). How does the C++11 implementation compare to the OpenMP implementations?

The threads file has **208** lines. The omp file has **194** lines. The sequential file has **171** lines.

The threads program has more lines than the omp program.

A screenshot of a cell phone

Description automatically generated

3. Increase the number of threads to 5 threads. What is the run time for each?

Serial = 5.58

Omp = 2.07

Threads = 1.43

A screenshot of a cell phone

Description automatically generated

Increase the maximum ligand length to 7 and rerun each program. What is the run time for each?

By changing the first input, which is the maximum ligand length of the programs to 7 as you can see in the bottom screenshots, I got those runtimes for omp and the threads program.

A screenshot of a cell phone

Description automatically generated