Project A5: Parallel Programming

1. What are the basic steps (show all steps) in building a parallel program? Show

at least one example.

1. Breaking down code into tasks
2. Assigning tasks to threads
3. Thread tasks are given instructions to coordinate with each other (master and worker)
4. Steps are given to the processor or processors to run.

An example would be a for loop that uses OpenMP pragmas to run each iteration using 4 threads.

1. What is MapReduce?

A programming model that is geared towards handling big data using parallel function of the threads. It does this by dividing big data into clusters and sorting them by a map functions and then merges similar values that is correlated to each other.

1. What is map and what is reduce?

Map- a construct used to sort large sets of data into smaller sets of organized data.

Reduce- a construct used to gather input from Map and combine similar data into smaller, organized sets.

1. Why MapReduce?

Its is a efficient way of conducting parallel in terms of large quantities of data and doing computations and storing with it.

1. Show an example for MapReduce.

It can keep track of URL access frequency by mapping the logs and then reduce to the values together in order to create a URL count.

1. Explain in your own words how MapReduce model is executed?

It fetches the large sums of data and organizes it into clusters by key associations between the data. Then the map’s work is used by reduce to have the data sorted into even more detailed groups and are performed the instructed functions upon said groups.

1. List and describe three examples that are expressed as MapReduce

computations.

1. A long mathematical computations problem can be divided into smaller and less complex problems.
2. A file with multiple data sets such as numbers for stocks can be divided into smaller sets of coherent data using MapReduce.
3. A large data set(such as a schedule for every class in a school) now needs to have a specific function to be performed on each value(such as converting each class number grade to letter grade) can be done using MapReduce by sorting each case and getting the values and performing said function to each values.
4. When do we use OpenMP, MPI and, MapReduce (Hadoop), and why?

OpenMP-

Best for shared memory parallelism. OpenMP is a directive focused library that can be effective for splitting up work for multiple threads to handle for designated parts of the program. However, it is very sensitive to human error. OpenMP must be carefully implemented in order to achieve program effectiveness.

MPI-

Best used for “scientific applications” that runs synchronous code that flows well, straggling lines of code will make MPI inefficient. It is also best for hybrid programming (using MPI and/or OpenMP for threads within the machine).

MapReduce-

This is best for large quantities of data that needs the data to be ETL(extract, transform, and loaded). It sorts the data to 2 constructs called map and reduce. One for data to be distributed and the other for the function to be applied to each element. One of the main benefits is that it has a higher Fault tolerance than either OpenMP and MPI. However, it does not perform too well when it comes to coupled scientific applications.

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

150 words.

Drug Design in the computational sense is the computer’s ability to detect, generated, and compute ligands in proteins that can be used for medicinal purposes. DNA problem is the errors that computers are susceptible to when computing Drug Design with DNA.