Nathan Heckman

A5 Parallel Programming Foundations

**What are the basic steps in building a parallel program?**

1. Identify tasks that have the ability to run concurrently
2. Determine if the task has no dependencies and if the same processing is required for each computational element
3. Implement the program using the master/worker technique and static load balancing if applicable

The example given revolved around approximating the value for pi. Inscribe a circle within a square and then have each worker generates two random numbers that lie within the square. The workers then see if the coordinate lies within the circle, and adds one to countCircle if it does. The master then computes pi from the given countCircle value.

**What is MapReduce?**

MapReduce is a programming model for generating big data sets with a parallel algorithm on a cluster. It was developed by Google as way to process large amounts of raw data by distributing it across thousands of machines in order to complete the process in a reasonable amount of time.

**What is map and what is reduce?**

In Lisp, map takes a function and sequence of values as input and then applies the function to all the values. In MapReduce, map takes input and produces intermediate key pairs and then passes all of them to the reduce function. Reduce combines the intermediate keys given by map and a set of values to produce a smaller set of values that is easier to understand.

**Why MapReduce?**

MapReduce allows engineers to perform simple computations quickly while hiding the underlying details of the program.

**Show an example for MapReduce.**

A good example of MapReduce is the example given in the MapReduce article about counting occurrences of words in a large document collection. The map function emits each word plus an associated count of said word and the reduce functions adds all the counts for a particular word.

**How is the MapReduce model executed?**

Map segments are automatically distributed across multiple machines. Data is split into sets of shards and these are executed in parallel on different machines. Reduce segments are distributed by breaking up the intermediate key into R pieces using a partitioning function which is user-specified.

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

Distributed Grep – map function emits a line if it matches the given pattern. Reduce function copies the supplied intermediate data to the output.

Count of URL Access Frequency – map function processes web page request logs and outputs. Reduce function adds together all values for the logs and emits a <URL, total count> pair.

Reverse Web-Link Graph – map function outputs target-source pairs to a target URL. Reduce function concatenates all source URLs that have something to do with a target URL and emits a pairing

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

OpenMP – OpenMP is a great directive-based library to use shared memory parallelism in code. Using #pragma parallel we can split the loop into threads and each thread then handles a chunk of the loop iterations.

MPI – Message Passing Interface is used when developing parallel scientific applications. The code is tightly synchronous and well load balanced. Can be used to develop most parallel code that runs over multiple machined and gives access to hybrid programming.

MapReduce (Hadoop) – Hadoop gives two constructs that can be applied over large sets of data (map and reduce). Makes it possible to chain maps and reductions to design complex problems.

**Explain what a Drug Design and DNA problem is:**

The strategy for drug design software involves generate ligands, computing a fitness score for each ligand, and identifying the highest ligand score and then testing it. This helps scientists develop medicine to help repair proteins and transform them into a desirable shape. DNA contains the instructions for producing proteins and their different shapes and drug design software can help us learn how to alter the protein structure in a body to produce a certain shape.