(15p) What are the basic steps (show all steps) in building a parallel program? Show

at least one example.

When it comes to building a parallel program, one should first understand what can be broken down to parallel. Some functions require previous input to finish while others can work in conjunction with one another. The example give was the Master worker workload where the dataset is split into even sequence amongst the workers, and then processed. Finding PI is an example of how this can be enforced.

o (5p) What is MapReduce?

MapReduce allows us to process large datasets with parallel algorithms by distributing workloads.

o (10p) What is map and what is reduce?

Similar to HashMaps Map is a function the generates key value pairs as well as intermediate value pairs which afterwards Reduce merges these pairs.

o (5p) Why MapReduce?

Operation Management. MapReduce gives the benefit of system management when checking load and has a developed tolerance to manage stress.

o (5p) Show an example for MapReduce.

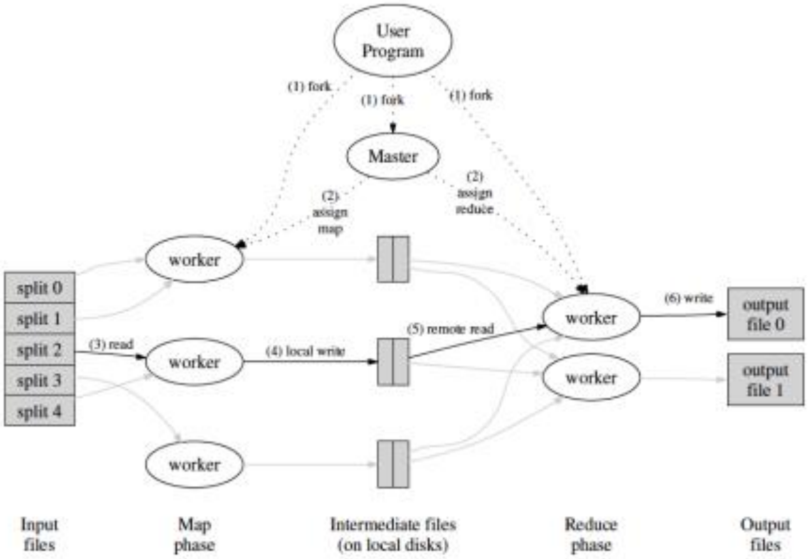

Figure 1: Execution overview

o (10p) Explain in your own words how MapReduce model is executed?

o (6p) List and describe three examples that are expressed as MapReduce

computations.

Distributed Grep : Map function will return a line if a given pattern is matched. Reduce function copies intermediate data to output.

Count of URL Access Frequency : Map function processes logs of web pages request and outputs. The reduce function then adds all together for the same URL and returns a pair. If I'm correct Google actually uses this for data representation.

Term-Vector per Host : Term vector summarize the most important words in a document in pairs of frequency tied to word, and then emits a pair for each input. Reduce function passes all vectors for a host and then adds together throwing away infrequent terms.

- (6p) When do we use OpenMP, MPI and, MapReduce (Hadoop), and why?

OpenMP: We use OpenMP when we want to introduce shared memory parallelism in our code. It is very useful in taking singular loads and distributing them.

MPI: Message Passing Interface is useful for parallel code that runs over multiple machines. It is useful for parallel implementation for scientific applications

MapReduce: We use MapReduce to reduce the amount of data used in operations over large sets of data.

- (14p) In your own words, explain what a Drug Design and DNA problem is in no more than

150 words.

The process for this involves an object called ligands. These are needed to try for proteins. We will give a score to each of these and then determine the highest bond ratio. We will use parallel programming to hash the ligands and distribute the association process