Quiz 3 - Ch. 6

Please answer all questions in complete sentences. Any responses not given in the form of a complete sentence will be marked incorrect.

1. Define the terms incremental algorithm and recomputation algorithm in the context of a Big Data System.
    a. An incremental algorithm is an algorithm that updates a value as new data arrives. In the context of the batch layer, incremental algorithms update batch views directly when new data arrives. A recomputation algorithm discards previous batch views creates new batch views by recomputing functions over the entire master dataset.

2. Why is it necessary to have a recomputation version of an algorithm in a Big Data System?
    a. It is necessary to have a recomputation version of algorithms in big data systems because this style of algorithm is the only way of ensuring human-fault tolerance in your system and human-fault tolerance is a necessary requirement.

3. What two functions must be implemented to perform a batch computation job with MapReduce?
    a. In order to perform a batch computation job using MapReduce, you must implement a *Map* function and a *Reduce function.*

4. Does MapReduce move code to data or does it transfer data to code? Why?
    a. In order to avoid the need to transfer all data across the network, MapReduce assigns map tasks to servers that host the data to be processed. This reduces the time to complete a computation as there is typically much less code to transfer than data. It also allows for parallelization as the map tasks can each be run on different servers.