

## Analysis of Algorithm Quiz

### 1. Multiple Choice [4 Points]

- a. A MapReduce algorithm should be considered when:
  - i. Large volume of data (petabyte) need to be analyzed and there is one really fast computer is available.
  - ii. Moderate (Gigabyte) amount of data need to be analyzed.
  - iii. Large volume of data need to be analyzed and a cluster of computers is available
  - iv. All of the above
- b. Which one of the following is a correct statement?
  - i. The Mapper implementation processes one line at a time via map function.
  - ii. Reduce() function is responsible for combining the results produced by each of the Map() functions.
  - iii. All map tasks must complete before the reduce task can start.
  - iv. All of the above
- c. All of the following accurately describe Hadoop/MapReduce, EXCEPT:
  - i. Open source
  - ii. Real-time
  - iii. Java-based
  - iv. Master-Worker approach
- d. MapReduce is **not** suitable when
  - i. The computation can be partitioned into separate tasks where tasks need to communicate with each other throughout the processing.
  - ii. The computation can be partitioned into separate tasks where tasks can run concurrently on multiple processors.
  - iii. There is a need to handle massive amounts of data.
  - iv. There is a need for handling lengthy computations in a fault-tolerant way.

### 2. Write the map and reduce function for the following scenarios [4 Points]

- a. Find the number of times a particular product has been recommended. Input data is a set of lines in the following format  
<Recommender\_id> <space> <Product\_id>
- b. Find the categorization of the regions according to its real estate sale values. If the average house price of a region is in the range of >100K to <300K, the category should be "Normal" and if it is in the range of >300K to <500K, the category should be "Wealthy" and if it is >500K, the category should be "Super Rich". The input data is a set of lines in the following format  
<property\_id> <space> <zip> <space> <sale\_date> <space> <sale-value>