Lesson 1 Review

1.

Which of these kinds of data motivated the Map/Reduce framework?

Large number of patient records that are updated immediately after each patient visit.

Large number of customer internet transactions that are often retrieved by a billing id.

Large number of internet documents that need to be indexed for searching by words.

2.

What is the organizing data structure for map/reduce programs?

A dictionary of words and their semantic value

A list of identification keys and some value associated with that identifier

A set of indices for a table of data values

3.

In map/reduce framework, which of these logistics does Map/Reduce do with the map function?

Gather data distributed across a cluster to the user's computer and run map

Distribute map to cluster nodes, run map on the data partitions at the same time

Distribute map to cluster nodes, run map at one node, wait for it to finish, then run map at the next node, etc,

4.

Map/Reduce performs a ‘shuffle’ and grouping, does that mean it that:

Shuffles <key,value> pairs into random bins and then within a bin it groups keys.

Shuffles <key,value> pairs into different partitions according to the key value, and then aggregates all pairs in 1 partition into 1 group.

Shuffles <key,value> pairs into different partitions according to the key value, and sorts within the partitions by key.

5.

In the word count example, what is the key?

The line number that contains the word.

The document id that contains the word.

The word itself.

6.

Streaming map/reduce allows mappers and reducers to be written in what languages:

java

python

Unix shell commands

R

All of the above

7.

The assignment asked you to run with 2 reducers. When you use 2 reducers instead of 1 reducer, what is the different in global sort order?

- With 1 reducer, but not 2 reducers, the word counts are in global sort order by word.
- With 2 reducers, but not 1 reducer, the word counts are in global sort order by word.
- With 1 reducer or 2 reducers, the word counts are in global sort order by word.
- With 1 reducer or 2 reducers, the word counts are NOT in global sort order by word.