**Q:1** **During the lecture videos, we covered Distributed Computing including Map Reduce.  Briefly describe what happens during the Map phase and the Reduce phase of a standard word count Map Reduce program.**

**Answer: In the map reduce process phase first of all we will put all the documents in the fil distributed system cloud so the file should be available to all the users. Now each word of the file will be assigning a mapping node. After assigning the mapping node will apply mapping function to produce key node pairs. Here the mapping function won’t count anything or any words instead it would produce key value pairs. In the key value pairs the key is the word of the file and value is the number can be ‘1’. After the mapping phase mapper will output with the key value pairs from the file. There is no counting in the map phase yet.**

**After this whole process of map phase, we have reduced phase. In reduce phase we will apply the shuffling function to the data of the file or to the words of the file to reduce it. The task of the reducers would be to take these key value pairs and combine or reduce the same keys. By this the value of reduces keys will increase as per count of the keys. At this place the counting task is done, at the reduce phase. After these all calculations this are the result from the program of MapReduce.**

**For each of the below scenarios, discuss why distributing computing will be appropriate or not appropriate.   
  
a) A credit monitoring company that keeps track of the average number of transactions a user makes per month by aggregating transactions associated with their ID.  The company uses an algorithm to detect and flag anomalies in the average number of transactions over a rolling time window.**

**Answer:** The scenario given above is appropriate for the distributing computing. It is because the credit monitoring company is on big scale, the scalability of this company is so big that it would require the distributing computing. It deals with a large amount of data of the users. For this different task the performance should be also good and fast. There should be no fault in the results, it should provide fault tolerance mechanism. **b) A data analytics firm that tracks trending topics on Facebook and Twitter by measuring the most commonly used words.**

**Answer:** The scenario given above is not appropriate for the distributing computing. It is because the computational cost of the firm is not too big that it can be handle by individual computers. **c) An artificial intelligence research lab that creates image libraries to train Deep Learning algorithms by collecting images and tagging them with metadata.**

**Answer:** The scenario given above is appropriate for the distributing computing. It is because the artificial intelligence research lab contains image processing which need a lot of computational cost. There are vast of amount of data of images. Which also need performance in the task which be only provided through computational processing.

Parallel processing is also involved in this process by which a lot of images can be processed in one go simultaneously.