Week 1

Introduction to Big Data Quiz

1. Which of the following does not belong to traditional database technology?
   a. RDBMS
   b. DBMS
   c. Flat Files
   d. NOSQL

Answer: d

2. If a big data analyst were to analyze data from a database of call logs provided by a telecom service provider, which element of big data would he be dealing with?
   a. Volume
   b. Variety
   c. Velocity
   d. All of the above

Answer: d

3. Some people call this data as "structured but not relational" . Which data are we talking now?
   a. Structured Data
   b. Unstructured Data
   c. Semi structured Data
   d. Mixed Data

Answer: c

4. You have a database of size 500 GB and the size of the database did not drastically increase over a 6 month period and even the forecast does not indicate a dramatic increase in size in the future. This is a Big Data problem and it is best to convert this database (and its dependencies/jobs etc.) in to a Hadoop based solution.
   a. True
   b. False

Answer: False

5. Scenario: Due to increase in the data volume you are asked to re-engineer an online shopping website which is currently using MySQL database server as it backend. The website fetches the data from the database and also insert/delete records in the database. Users should not see a performance degradation after the re-engineering effort. Your Approach: Since Hadoop is a batch and not an interactive solution it is best to choose a NoSQL database solution.
   A. True
   B. False

Answer : True

6. Scenario: You have an ETL job in Oracle that takes 2 hours to complete. Business users are requesting you to bring down the execution time of the job to 1 hour. Best Approach: It is recommended to add additional servers to solve this problem?
   a. True
   b. False

Answer: False

7. What is spurious correlation?
   a. Correlation that occurs because of poor data
   b. Correlation that occurs because of poor analysis
   c. Correlation that is not really present but seems to appear so because of hidden variables
   d. Correlation that is not really present but seems to appear because of fault in the system

Answer: c

8. In Big Data, What is a Model?
   a. A human construct that helps understand data
   b. A human construct that helps understand real world problem
   c. A human construct that helps understand big data generated by humans
   d. A machine construct that helps understand data.

Answer: b

9. In the Peter Norvig video, he says that machine translation is an example where domain knowledge is not needed because
   a. Machine translations are based upon a probabilistic model of language which does not include any linguistic language
   b. Domain knowledge is not really needed to understand a passage
   c. Good translations are possible without use of domain knowledge
   d. Semantics of words need to be input to create a good machine translation

Answer: a

10. You have to analyze data of players using input from commentary, but you see that the same player's name are often expressed in different ways. For example, Virat Kohli, V Kohli, Virat are often found in the commentary and refer to the same person. In terms of Big Data, this is an example of?

    A. Veracity, since not all the data was in the same format.
    B. Veracity, since the data was unstructured
    C. Variety, since sometimes the data had a comma and sometimes it did not
    D. Volume, since the number of stadiums is very large.

Answer: c

11. The difference between old-style data and new style data is that
    a. New style data has a schema while old style data has no schema
    b. New style data is generated by humans, old style by machine

c. New style data has no schema while old style data has schema
d. New style data may be structures, unstructured or semi structured while old style data is only structured.

Answer: d