

# **EMC**

## **Exam E20-065**

### **Advanced Analytics Specialist Exam for Data Scientists**

**Verson: Demo**

**[ Total Questions: 10 ]**

**Question No : 1**

What do first-order and second-order Markov processes have in common concerning next word prediction?

- A. Both use WordNet to model the probability of the next word
- B. Both are unsupervised methods
- C. Both provide the foundation to build a trigram language model
- D. Neither makes assumptions about the probability of the next word

**Answer: C**

**Question No : 2**

What are two visualization tools used for trivariate data?

- A. Scatter plot matrix
- B. Hexbin plot and heatmap
- C. Scatter plot matrix and density plot
- D. Scatter plot matrix and heatmap

**Answer: B**

**Question No : 3**

How can you improve processing performance in HIVE?

- A. Partition tables
- B. Run the SET hive.exec.parallel = false command
- C. Ensure highly normalized tables and use joins
- D. Minimize bucketing

**Answer: A**

**Question No : 4**

A hotel chain runs a simulation on room pricing. They want to estimate revenue, per hotel, within +/- \$10 with 95% confidence ( $Z_{\alpha/2}=1.96$ ). The estimated revenue standard deviation is \$5000 based on previous booking data.

What is the optimal number of simulation trials to run?

- A. A 32-bit operating system was used
- B. The same number of trials was used
- C. A linear congruential generator (LCG) was used (or pseudo-random number generation)
- D. Different seeds for the random number generator were used.

**Answer: C**

**Question No : 5**

What is an effective use of color in visualization?

- A. Use self-explanatory colors so a legend is unnecessary
- B. Maximize use of color to make a more lasting impression
- C. Use high contrast colors such as red and blue
- D. Minimize use of color except for emphasis

**Answer: C**

**Question No : 6**

In multinomial logistic regression, what is used to calculate the probability of outcome occurring?

- A. Logistic function applied to a linear combination of the input and outcome variables
- B. Linear regression applied to a combination of input variables

- C. Linear regression applied to a combination of input and outcome variables
- D. Logistic function applied to a linear combination of the input variables

**Answer: B**

**Question No : 7**

After a client submits a job request to the YARN ResourceManager, what happens next?

- A. The scheduler allocates a container to run an ApplicationMaster
- B. The ResourceManager allocates containers to run map and reduce tasks
- C. The Resource Manager requests load data from the NodeManagers
- D. The ApplicationManager starts an ApplicationMaster

**Answer: D**

**Question No : 8**

A marketing team creates a graph using a square for each data point, where the length of each side is set to the data value. The data values are 10 and 20.

What is the lie factor of the graph?

- A. 1
- B. 2
- C. 3
- D. 6

**Answer: B**

**Question No : 9**

Which HDFS feature protects against user errors causing accidental loss of data?

- A. Encryption
- B. Replication
- C. Namenode federation
- D. Snapshots

**Answer: B**

**Question No : 10**

What is a characteristic of spark?

- A. Unable to run map -> reduce execution plans
- B. Supports applications written in Python, Java, and Scala
- C. Less efficient processing small files than Hadoop MapReduce
- D. Supports workflows that can return to previous work steps

**Answer: B**