E20-065 Q&As

Advanced Analytics Specialist Exam for Data Scientists

Pass EMC E20-065 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.pass4itsure.com/e20-065.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by EMC Official Exam Center

⚙ **Instant Download** After Purchase

⚙ **100% Money Back** Guarantee

⚙ **365 Days** Free Update

⚙ **800,000+** Satisfied Customers

QUESTION 1

What is an important simu-lation design consideration?

A. Ensure model Inputs align with reality

B. Use different seed values to regenerate results

C. For rare event models, minimize number of trials

D. A complex model is better than a simple model

Correct Answer: A

QUESTION 2

What process must address acoustic ambiguity in NLP?

A. Part-of-speech tagging

B. Word sense disambiguation

C. Speech recognition

D. Discourse

Correct Answer: C

QUESTION 3

What runs more efficiently because of Apache Tez?

A. Pig and Hive

B. Hive and HBase

C. Yarn and Spark

D. All MapReduce jobs

Correct Answer: D

QUESTION 4

What is the most likely reason for an HBase table to contain millions of columns?

A. Data is imported from a relational database table

B. Data is stored in the column qualifier

C. There are thousands of columns families

D. The column names are randomly generated

Correct Answer: B

QUESTION 5

In which step in the visualization lifecycle would you determine how the raw data is stored?

A. Visualization Planning

B. Data Preparation

C. Visualization Building

D. Discovery

Correct Answer: B

QUESTION 6

Which scenario would be ideal for processing Hadoop data with Hive?

A. Structured data, real-time processing

B. Unstructured data; batch processing

C. Unstructured data; real-time processing

D. Structured data; batch processing

Correct Answer: B

QUESTION 7

What is a characteristic of the trigram language model?

A. Based on the second-order Markov process

B. Equivalent to trigram hidden Markov models

C. Uses smoothing to reduce the high dimensionality in text

D. Can be used for part-of -speech tagging

Correct Answer: A

QUESTION 8

Which graph structure would best model the relationship between job seekers and employers?

A. Bipartite

B. Weighted

C. Directed acyclic

D. Ranked

Correct Answer: A

QUESTION 9

What best describes the meaning behind the phrase "Six Degrees of Separation\\"?

A. Ability to use about six hops to reach any other node in an extremely large social network

B. Erdos number of all scholars having written papers with Paul Erdos

C. Maximum number of edges between nodes in a graph with a diameter of six

D. Typical distance between nodes that are connected by triadic closure

Correct Answer: A

QUESTION 10

What is an intended application of the MapReduce framework?

A. Processing can be broken into smaller pieces

B. Processing a large number of small files

C. Processing in real time is required

D. Processing a small subset of data

Correct Answer: A

QUESTION 11

What advantage does replication provide while storing a file in HDFS?

A. Data protection and scheduling flexibility

B. Elimination of requirement for a combiner process

C. Elimination of requirement for Shuffle and Sort process

D. Memory optimization and minimizing tasks to run

Correct Answer: A

QUESTION 12

Which library is NOT part of the Apache Spark distribution?

A. MLib

B. NLTK

C. GraphX

D. Spark SQL

Correct Answer: B

QUESTION 13

What elements are needed to determine the time complexity of finding all the cliques of size k in social network analysis?

A. Eigenvector centrality and betwenness

B. Clique size and total number of nodes in the network

C. Number of edges in the network and centrality measure of the cliques

D. Clique size and betweenness centrality

Correct Answer: B

QUESTION 14

What does YARN provide over and above MapReduce?

A. Separate cluster and resource management

B. Parallelized processing

C. Serialized processing

D. Access to HDFS data

Correct Answer: A

QUESTION 15

A simul-ation to compare two different sales models yields different results for the same set of input

variables in different runs.

What is the likely cause?

A. bit operating system was used

B. The same number of trials was used.

C. A linear congruenlial generator (LCG) was used for pseudo-random number generation.

D. Different seeds forthe random number generator were used

Correct Answer: C