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.certbus.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 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

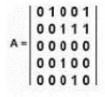D. Minimize use of color except for emphasis

Correct Answer: C

QUESTION 2

What is NOT a category of a NoSQL data store?

A. Columnar

B. Document

C. Key/Value

D. Flat File

Correct Answer: D

QUESTION 3

Assuming the node index starts at 1, what is the out-degree of node 3 in the adjacency matrix shown? Refer to the exhibit.

$$A = \begin{vmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{vmatrix}$$

A. 0

B. 1

C. 2

D. 3

Correct Answer: A

QUESTION 4

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

Correct Answer: B

QUESTION 5

Which scenario is a proper use case for multinomial logistic regression?

A. A marketing firm wants to estimate the personal income of a group of potential customers. Using inputs such as age, education, marital status, and credit card expenditures, a data scientist is building a model that will estimate a person\\'s income

B. A logistic distribution company wants to minimize the distance traveled by its delivery trucks. A data scientist is building a model to determine the optimal route for each of tis trucks

C. To improve the initial routing of a loan application, a financial institution plans to classify a loan application as Approve, Reject, or Possibly_Approve. Based on the company\\'s historical loan application data, a data scientist is building a model to assign one of these three outcomes to each submitted application.

D. A manufacturer plans to determine the optimal number of workers to employ in an assembly line process. Utilizing the observed distributions of the task durations of each process step, a data scientist is building a model to mimic the interactions and dependencies between each stage in the manufacturing process.

Correct Answer: C

QUESTION 6

What do lemmatization and stemming have in common?

A. Use WordNet

B. Remove common words in a natural language

C. Reduce the high dimensionality in text

D. Use a set of heuristics

Correct Answer: C

QUESTION 7

What are key characteristics of regular lattices?

A. Low clustering coefficients, high network diameters

B. High clustering coefficients, small network diameters

C. Low clustering coefficients; small network diameters

D. High clustering coefficients; high network diameters

Correct Answer: D

QUESTION 8

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 9

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

Correct Answer: B

QUESTION 10

You develop a Python script "logisticpy" to evaluate the logistic function denoted as f(y) for a given value y that includes the following Pig code:

Register \\'logistic.py\\' using jython as udf;

z = FOREACH y GENERATE $0, udf.logistic ($0);

DUMP z;

What is the expected output when the Pig code is executed?

A. 0

B. Jython is not a supported language

C. Value of f(y) for ally

D. Tuples (y, f(y))

Correct Answer: D

QUESTION 11

Which representation is most suitable for a small and highly connected network?

A. Edge list

B. Adjacency matrix

C. Eigenvector centrality

D. Adjacency list

Correct Answer: B

QUESTION 12

What are the major components of the YARN architecture?

A. ResourceManager and NodeManager

B. Task Tracker and NameNode

C. HDFS, Tez, and Spark

D. Avro, ZooKeeper, and HDFS

Correct Answer: A

QUESTION 13

Which Hadoop Files System shell command copies data from a local file system into HDFS?

A. rm

B. cp

C. put

D. get

Correct Answer: C

QUESTION 14

In a social network, what does it mean for a node to have a high degree but low betweenness?

A. The node is adjacent to a few nodes, each of each has high Page Ranks.

B. The node has the only edge connecting its community to the rest of the graph.

C. The node can be easily bypassed by communications taking other shorter paths.

D. The node acts as the hub of the graph.

Correct Answer: D

QUESTION 15

What is an ideal use case for HDFS?

A. Storing files that are updated frequently

B. Storing files that are written once and read many times

C. Storing results between Map steps and Reduce steps

D. Storing application files in memory

Correct Answer: B

E20-065 Practice Test   E20-065 Study Guide   E20-065 Exam Questions