# 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.leads4pass.com/e20-065.html**

100% Passing Guarantee
100% Money Back Assurance

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

**QUESTION 1**

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

Correct Answer: D

**QUESTION 2**

What is a key beneficial characteristic of the Random Forest algorithm?

A. Provides and explanatory model

B. Distinguishes categorical from continuous variables

C. Support for unstructured data

D. Resiliency to complex, non-linear variable interactions

Correct Answer: D

**QUESTION 3**

How is the relative value of a node visualized in a sunburst?

A. Color

B. Area

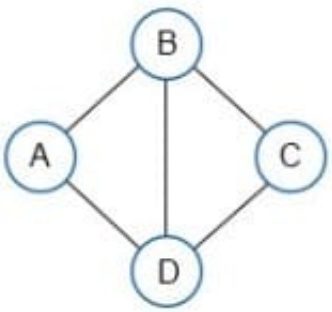
C. Gradient

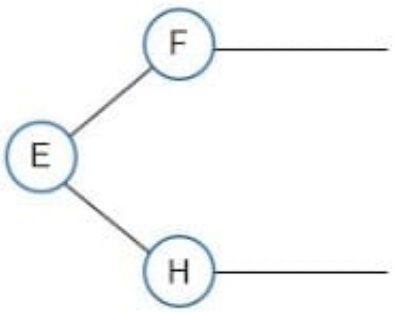
D. Position

Correct Answer: A

**QUESTION 4**

In the graph, which edge would be considered a weak lie? Refer to the exhibit.




A. C-E

B. E-F

C. B-C

D. G-I

Correct Answer: D

**QUESTION 5**

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 6**

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 7**

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 8**

How does Latent Dinchlet Allocation (LDA) interpret a document?

A. As a single-predefined topic

B. As a mixture of pre-defined topics

C. As having a mixture of sentiments

D. As having a single pre-defined sentiment

Correct Answer: B

**QUESTION 9**

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 10**

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 11**

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 12**

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 13**

What is a property of a good color model for ordinal data?

A. Uses a rainbow-like color map for distinction of categories

B. Uses a rainbow-like color map for ease of display and printing

C. Uses perceptually ordinal colors with just-noticeable increments

D. Uses perceptually ordinal colors with linear, perceptual increments

Correct Answer: D

**QUESTION 14**

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 15**

The naive Bayer classifier is trained over 1600 movie reviews and then tested over 400 reviews.

Here is the resulting confusion matrix:

190 (TP) 10(FN)

80 (FP) 120(TN)

What are the precision, recall, and the F1-score values?

A. Precision0.95; Recall: 0704; F1-score: 0.809

B. Precision 0.613, Recall: 0.95, F1-score: 0.745

C. Precision 0.704, Recall: 0.95; F1-score: 0.809

D. Precision 0.95; Recall: 0.613; F1-score: 0.745

Correct Answer: C