

**100%** Money Back  
**Guarantee**

**Vendor:**EMC

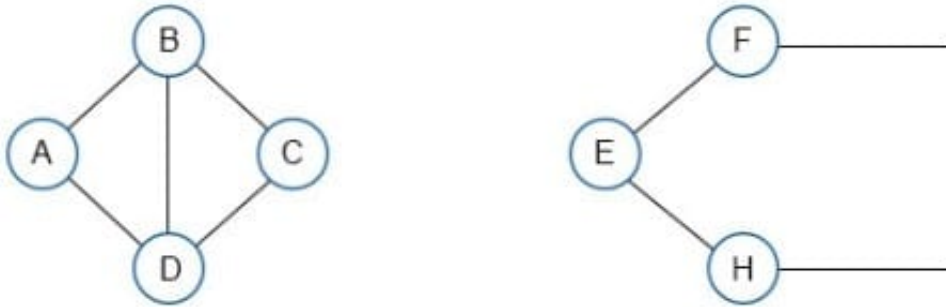
**Exam Code:**E20-065

**Exam Name:**Advanced Analytics Specialist Exam for  
Data Scientists

**Version:**Demo

**QUESTION 1**

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



- A. C-E
- B. E-F
- C. B-C
- D. G-I

Correct Answer: D

---

**QUESTION 2**

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 3**

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 4

A simulation 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 congruential generator (LCG) was used for pseudo-random number generation.

D. Different seeds for the random number generator were used

Correct Answer: C

---

#### QUESTION 5

A data engineer is asked to process several large datasets using MapReduce. Upon initial inspection the engineer realizes that there are complex interdependencies between the datasets.

Why is this a problem?

A. MapReduce works best on unstructured data

B. There is no problem; MapReduce accommodates all the data

C. MapReduce can only parse one file at a time.

D. MapReduce is not ideal when the processing of one dataset depends on another.

Correct Answer: D

---

#### QUESTION 6

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 betweenness

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 7

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 8

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 9

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

- A. Provides an 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 10

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 11**

Why would a company decide to use HBase to replace an existing relational database?

- A. It is required for performing ad-hoc queries.
- B. Varying formats of input data requires columns to be added in real time.
- C. The company's employees are already fluent in SQL.
- D. Existing SQL code will run unchanged on HBase.

Correct Answer: A

---

#### **QUESTION 12**

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