

# E20-026<sup>Q&As</sup>

Enterprise Storage Networking Specialist Exam

**Pass EMC E20-026 Exam with 100% Guarantee**

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

<https://www.leads4pass.com/e20-026.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**

You are studying the behavior of a population, and you are provided with multidimensional data at the individual level. You have identified four specific individuals who are valuable to your study, and would like to find all users who are most similar to each individual. Which algorithm is the most appropriate for this study?

- A. K-means clustering
- B. Linear regression
- C. Association rules
- D. Decision trees

Correct Answer: A

**QUESTION 2**

Refer to the exhibit Consider the training data set shown in the exhibit. What are the classification ( $Y = 0$  or  $1$ ) and the probability of the classification for the tuple  $X(1, 0, 0)$  using Naive Bayesian classifier?

X1	X2	X3	Y
1	1	1	0
1	1	0	0
0	0	0	0
0	1	0	1
1	0	1	1
0	1	1	1

- A. Classification  $Y = 0$ , Probability =  $4/54$

B. Classification  $Y = 1$ , Probability =  $4/54$

C. Classification  $Y = 0$ , Probability =  $1/54$

D. Classification  $Y = 1$ , Probability =  $1/54$

Correct Answer: A

---

**QUESTION 3**

What is Hadoop?

A. Java classes for HDFS types and MapReduce job management and HDFS

B. Java classes for HDFS types and MapReduce job management and the MapReduce paradigm

C. MapReduce paradigm and HDFS

D. MapReduce paradigm and massive unstructured data storage on commodity hardware

Correct Answer: A

---

**QUESTION 4**

When would you prefer a Naive Bayes model to a logistic regression model for classification?

A. When you are using several categorical input variables with over 1000 possible values each.

B. When you need to estimate the probability of an outcome, not just which class it is in.

C. When all the input variables are numerical.

D. When some of the input variables might be correlated.

Correct Answer: A

---

**QUESTION 5**

You have been assigned to run a linear regression model for each of 5,000 distinct districts, and all the data is currently stored in a PostgreSQL database. Which tool/library would you use to produce these models with the least effort?

A. MADlib

B. Mahout

C. R

D. HBase

Correct Answer: A

---

[E20-026 VCE Dumps](#)

[E20-026 Study Guide](#)

[E20-026 Exam Questions](#)