

# 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.certbus.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

In which lifecycle stage are appropriate analytical techniques determined?

- A. Model planning
- B. Model building
- C. Data preparation
- D. Discovery

Correct Answer: A

---

### QUESTION 2

You have used k-means clustering to classify behavior of 100,000 customers for a retail store. You decide to use household income, age, gender and yearly purchase amount as measures. You have chosen to use 8 clusters and notice that 2 clusters only have 3 customers assigned. What should you do?

- A. Decrease the number of clusters
- B. Increase the number of clusters
- C. Decrease the number of measures used
- D. Identify additional measures to add to the analysis

Correct Answer: A

---

### QUESTION 3

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  $X(0, 0, 1)$  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 = 1,Probability = 4/54
- B. Classification Y = 0,Probability = 1/54
- C. Classification Y = 1,Probability = 1/54
- D. Classification Y = 0,Probability = 4/54

Correct Answer: A

#### QUESTION 4

What is required in a presentation for project sponsors?

- A. The "Big Picture" takeaways for executive level stakeholders
- B. Data warehouse design changes
- C. Line by line review of the developed code
- D. Detailed statistical basis for the modeling approach used in the project

Correct Answer: A

#### QUESTION 5

Which word or phrase completes the statement? Emphasis color is to standard color as \_\_\_\_\_ .

- A. Main message is to context
- B. Main message is to key findings
- C. Frequent item set is to item
- D. Pie chart is to proportions

Correct Answer: A

---

#### QUESTION 6

Which word or phrase completes the statement? Structured data is to OLAP data as quasi-structured data is to\_\_\_\_\_

- A. Clickstream data
- B. XML data
- C. Text documents
- D. Image files

Correct Answer: A

---

#### QUESTION 7

Your colleague, who is new to Hadoop, approaches you with a question. They want to know how best to access their data. This colleague has previously worked extensively with SQL and databases. Which query interface would you recommend?

- A. Hive
- B. Pig
- C. Howl
- D. HBase

Correct Answer: A

---

#### QUESTION 8

Review the following code:

```
SELECT pn, vn, sum(prc*qty)
```

```
FROM sale
```

GROUP BY CUBE(pn, vn)

ORDER BY 1, 2, 3;

Which combination of subtotals do you expect to be returned by the query?

- A. (pn,vn)
- B. ( (pn,vn),(pn) )
- C. ( (pn,vn),(pn),(vn) )
- D. ( (pn,vn),(pn),(vn),( ) )

Correct Answer: D

---

#### QUESTION 9

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 10

In linear regression, what indicates that an estimated coefficient is significantly different than zero?

- A. A small p-value
- B. R-squared near 1
- C. R-squared near 0
- D. The estimated coefficient is greater than 3

Correct Answer: A

---

#### QUESTION 11

Which process in text analysis can be used to reduce dimensionality?

- A. Stemming
- B. Parsing

C. Digitizing

D. Sorting

Correct Answer: A

---

#### QUESTION 12

Which data type value is used for the observed response variable in a logistic regression model?

A. Any positive real number

B. Any integer

C. A binary value

D. Any real number

Correct Answer: C

---

#### QUESTION 13

What is the primary bottleneck in text classification?

A. The availability of tagged training data.

B. The ability to parse unstructured text data.

C. The high dimensionality of text data.

D. The fact that text corpora are dynamic.

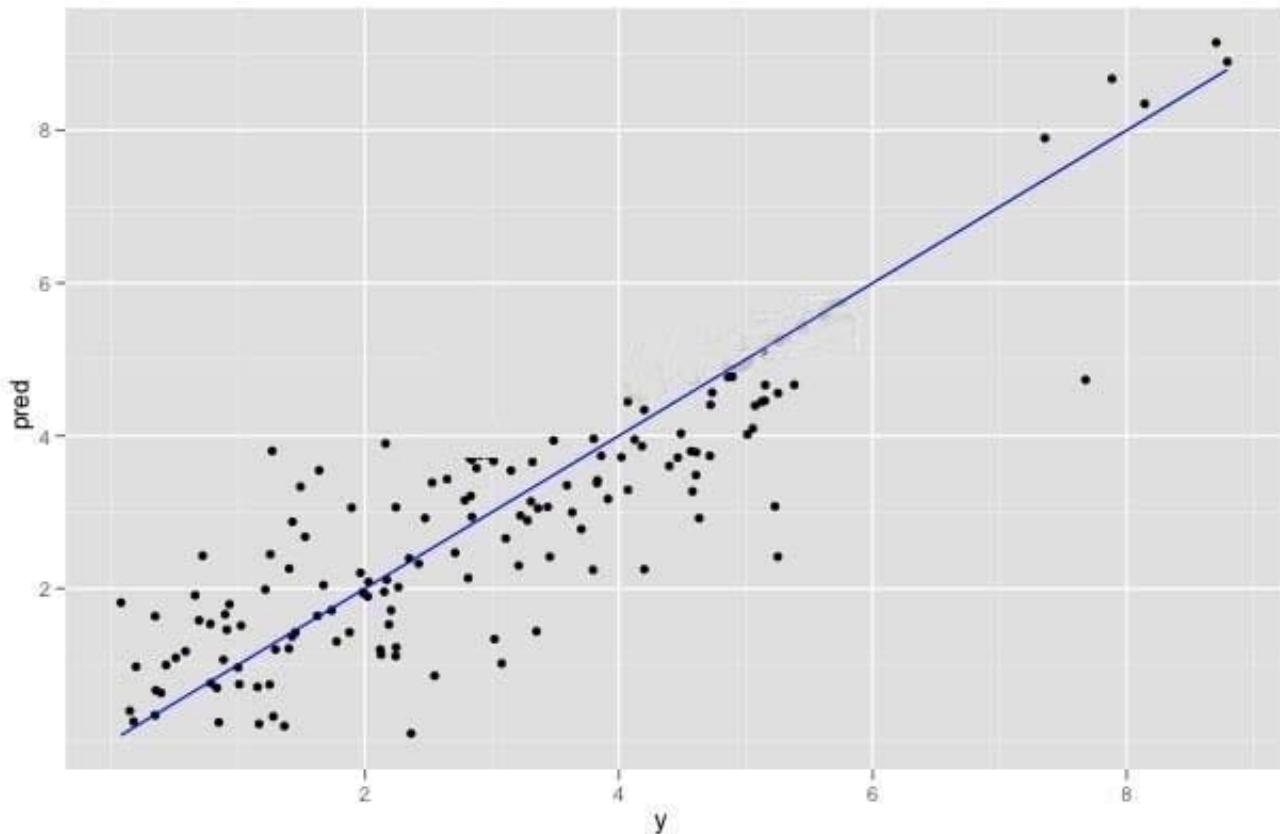
Correct Answer: A

---

#### QUESTION 14

Refer to the exhibit.

You have run a linear regression model against your data, and have plotted true outcome versus predicted outcome. The R-squared of your model is 0.75. What is your assessment of the model?



- A. The R-squared may be biased upwards by the extreme-valued outcomes. Remove them and refit to get a better idea of the model's quality over typical data.
- B. The R-squared is good. The model should perform well.
- C. The extreme-valued outliers may negatively affect the model's performance. Remove them to see if the R-squared improves over typical data.
- D. The observations seem to come from two different populations, but this model fits them both equally well.

Correct Answer: A

### QUESTION 15

What describes the use of UNION clause in a SQL statement?

- A. Operates on queries and potentially increases the number of rows
- B. Operates on queries and potentially decreases the number of rows
- C. Operates on tables and potentially decreases the number of columns
- D. Operates on both tables and queries and potentially increases both the number of rows and columns

Correct Answer: A

[E20-026 PDF Dumps](#)

[E20-026 Study Guide](#)

[E20-026 Braindumps](#)