



# 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.pass4itsure.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**

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 2**

What is an appropriate data visualization to use in a presentation for a project sponsor?

- A. Bar chart
- B. Pie chart
- C. Box and Whisker plot
- D. Density plot

Correct Answer: A

---

**QUESTION 3**

Which type of numeric value does a logistic regression model estimate?

- A. Probability
- B. A p-value
- C. Any integer
- D. Any real number

Correct Answer: A

---

**QUESTION 4**

In the MapReduce framework, what is the purpose of the Reduce function?

- A. It aggregates the results of the Map function and generates processed output
- B. It distributes the input to multiple nodes for processing
- C. It writes the output of the Map function to storage
- D. It breaks the input into smaller components and distributes to other nodes in the cluster 26 / 55

Correct Answer: A

---

**QUESTION 5**

You are provided four different datasets. Initial analysis on these datasets show that they have identical mean, variance and correlation values. What should your next step in the analysis be?

- A. Visualize the data to further explore the characteristics of each data set
- B. Select one of the four datasets and begin planning and building a model
- C. Combine the data from all four of the datasets and begin planning and building a model
- D. Recalculate the descriptive statistics since they are unlikely to be identical for each dataset

Correct Answer: A

---

**QUESTION 6**

You have been assigned to do a study of the daily revenue effect of a pricing model of online transactions. When have you completed the analytics lifecycle?

- A. You have written documentation, and the code has been handed off to the Data Base Administrator and business operations.
- B. You have a completely developed model, and the results have shown statistically acceptable results.
- C. You have presented the results of the model to both the internal analytics team and the business owner of the project.
- D. You have a completely developed model based on both a sample of the data and the entire set of data available.

Correct Answer: A

---

**QUESTION 7**

Consider these itemsets:



(hat, scarf, coat) (hat, scarf, coat, gloves) (hat, scarf, gloves) (hat, gloves) (scarf, coat, gloves)

What is the confidence of the rule (gloves -> hat)?

- A. 75%
- B. 60%
- C. 66%
- D. 80%

Correct Answer: A

---

### QUESTION 8

In MADlib what does MAD stand for?

- A. Magnetic, Agile, Deep
- B. Machine Learning, Algorithms for Databases
- C. Mathematical Algorithms for Databases
- D. Modular, Accurate, Dependable

Correct Answer: C

---

### QUESTION 9

Consider a database with 4 transactions:

Transaction 1: {cheese, bread, milk} Transaction 2: {soda, bread, milk} Transaction 3: {cheese, bread} Transaction 4: {cheese, soda, juice}

You decide to run the association rules algorithm where minimum support is 50%. Which rule has a confidence equal to 25%?

- A. {cheese} => {bread}
- B. {juice} => {cheese}
- C. {milk} => {soda}
- D. {soda} => {milk}

Correct Answer: A

---

### QUESTION 10

You have been assigned to do a study of the daily revenue effect of a pricing model of online transactions. All the data



currently available to you has been loaded into your analytics database; revenue data, pricing data, and online transaction data. You find that all the data comes in different levels of granularity. The transaction data has timestamps (day, hour, minutes, seconds), pricing is stored at the daily level, and revenue data is only reported monthly. What is your next step?

- A. Report back to the business owner that the current data model does not support the business question.
- B. Interpolate a daily model for revenue from the monthly revenue data.
- C. Aggregate all data to the monthly level in order to create a monthly revenue model.
- D. Disregard revenue as a driver in the pricing model, and create a daily model based on pricing and transactions only.

Correct Answer: A

---

### QUESTION 11

For which class of problem is MapReduce most suitable?

- A. Embarrassingly parallel
- B. Minimal result data
- C. Simple marginalization tasks
- D. Non-overlapping queries

Correct Answer: A

---

### QUESTION 12

Which word or phrase completes the statement? Mahout is to Hadoop as MADlib is to \_\_\_\_\_ .

- A. PostgreSQL
- B. R
- C. Excel
- D. SAS

Correct Answer: A

---

### QUESTION 13

What is holdout data?

- A. a subset of the provided data set selected at random and used to validate the model
- B. a subset of the provided data set selected at random and used to initially construct the model



- C. a subset of the provided data set that is removed by the data scientist because it contains data errors
- D. a subset of the provided data set that is removed by the data scientist because it contains outliers

Correct Answer: A

---

#### QUESTION 14

In which lifecycle stage are appropriate analytical techniques determined?

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

Correct Answer: A

---

#### QUESTION 15

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

[E20-026 PDF Dumps](#)

[E20-026 Study Guide](#)

[E20-026 Exam Questions](#)