



# E20-007<sup>Q&As</sup>

Data Science and Big Data Analytics

## Pass EMC E20-007 Exam with 100% Guarantee

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

<https://www.pass4itsure.com/e20-007.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 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 2**

A call center for a large electronics company handles an average of 35, 000 support calls a day. The head of the call center would like to optimize the staffing of the call center during the rollout of a new product due to recent customer complaints of long wait times. You have been asked to create a model to optimize call center costs and customer wait times.

The goals for this project include:

1.  
Relative to the release of a product, how does the call volume change over time?
2.  
How to best optimize staffing based on the call volume for the newly released product, relative to old products.
3.  
Historically, what time of day does the call center need to be most heavily staffed?
- 4.

Determine the frequency of calls by both product type and customer language. Which goals are suitable to be completed with MapReduce?

- A. Goal 2 and 4
- B. Goal 1 and 3
- C. Goals 1, 2, 3, 4
- D. Goals 2, 3, 4

Correct Answer: A

**QUESTION 3**

Refer to the exhibit.

Independent Variable	Coefficient	P-Value
A	1.23	0.001
B	-6.72	0.035
C	4.57	0.020
D	10.23	0.123

After analyzing a dataset, you report findings to your team:

1.

Variables A and C are significantly and positively impacting the dependent variable.

2.

Variable B is significantly and negatively impacting the dependent variable.

3.

Variable D is not significantly impacting the dependent variable.

After seeing your findings, the majority of your team agreed that variable B should be positively impacting the dependent variable.

What is a possible reason the coefficient for variable B was negative and not positive?

- A. Variable B is interacting with another variable due to correlated inputs
- B. Variable B needs a quadratic transformation due to its relationship to the dependent variable
- C. The information gain from variable B is already provided by another variable
- D. Variable B needs a logarithmic transformation due to its relationship to the dependent variable

Correct Answer: A

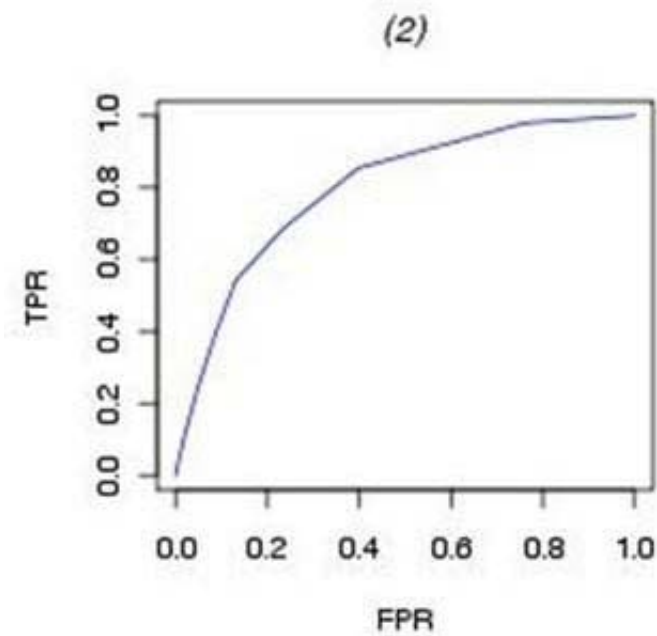
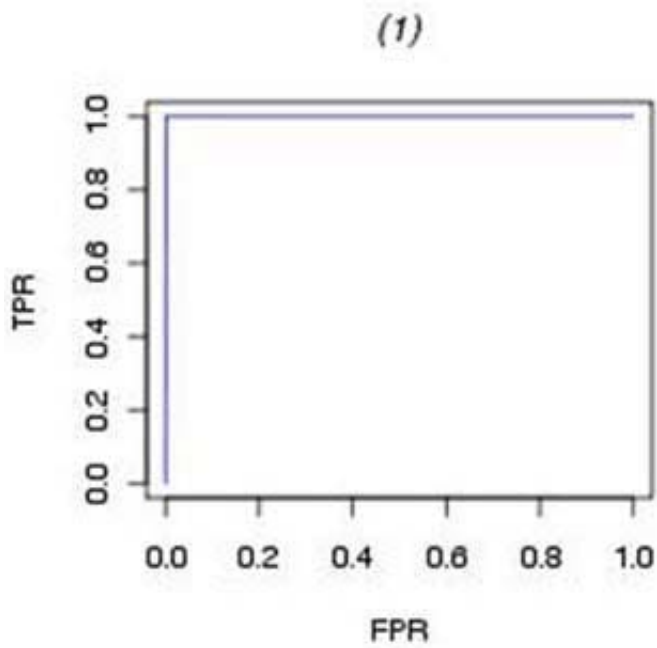
---

**QUESTION 4**

Which ROC curve represents a perfect model fit?

A.

B.

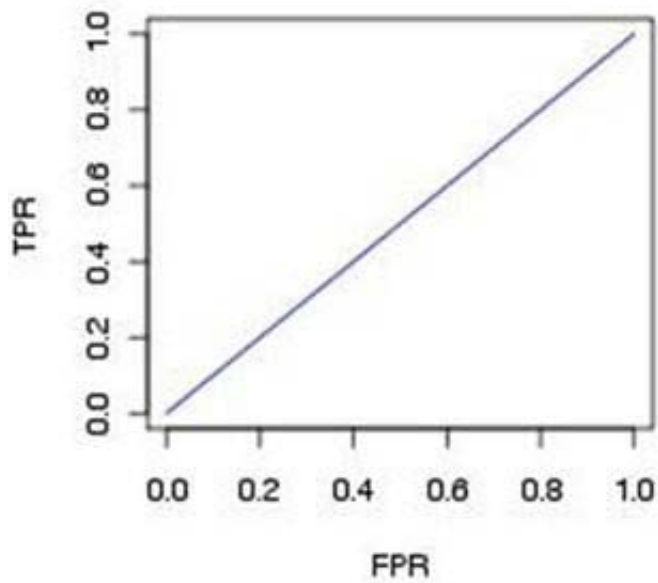


C.

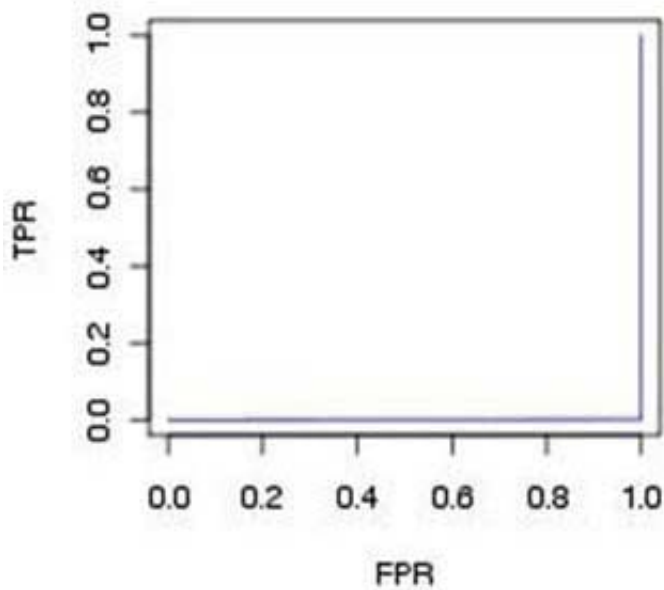
D.



(3)



(4)



A. Exhibit A

B. Exhibit B

C. Exhibit C

D. Exhibit D

Correct Answer: A



### QUESTION 5

You are having a discussion with a business colleague. The colleague mentions that they want to perform K-means clustering on text file data stored in HDFS.

Which tool should be recommended?

- A. Mahout
- B. HBase
- C. Scribe
- D. Sqoop

Correct Answer: A

[E20-007 PDF Dumps](#)

[E20-007 VCE Dumps](#)

[E20-007 Exam Questions](#)