E20-026^{Q&As}

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.passapply.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

What is an appropriate data visualization to use in a presentation for an analyst audience?

A. Pie chart

B. Area chart

C. Stacked bar chart

D. ROC curve

Correct Answer: D

QUESTION 2

You have completed your model and are handing it off to be deployed in production. What should you deliver to the production team, along with your commented code?

A. The production team needs to understand how your model will interact with the processes they already support. Give them documentation on expected model inputs and outputs,and guidance on error-handling.

B. The production team are technical,and they need to understand how the processes that they support work,so give them the same presentation that you prepared for the analysts.

C. The production team supports the processes that run the organization,and they need context to understand how your model interacts with the processes they already support. Give them the same presentation that you prepared for the project sponsor.

D. The production team supports the processes that run the organization,and they need context to understand how your model interacts with the processes they already support. Give them the executive

summary.

Correct Answer: A

QUESTION 3

You are testing two new weight-gain formulas for puppies. The test gives the results:

Control group: 1% weight gain Formula A. 3% weight gain Formula B. 4% weight gain A one-way ANOVA returns a p-value = 0.027

What can you conclude?

A. Either Formula A or Formula B is effective at promoting weight gain.

B. Formula B is more effective at promoting weight gain than Formula A.

C. Formula A and Formula B are both effective at promoting weight gain.

D. Formula A and Formula B are about equally effective at promoting weight gain.

Correct Answer: A

QUESTION 4

Which word or phrase completes the statement? A data warehouse is to a centralized database for reporting as an analytic sandbox is to a _____?

A. Collection of data assets for modeling

B. Collection of low-volume databases

C. Centralized database of KPIs

D. Collection of data assets for ETL

Correct Answer: A

QUESTION 5

What does R code nv = 1000 to nv

D. Selects values of vector v less than 1000,modifies v,and makes a copy to nv

Correct Answer: A

QUESTION 6

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 7

You are given 10, 000, 000 user profile pages of an online dating site in XML files, and they are stored in HDFS. You are assigned to divide the users into groups based on the content of their profiles. You have been instructed to try K-means clustering on this data. How should you proceed?
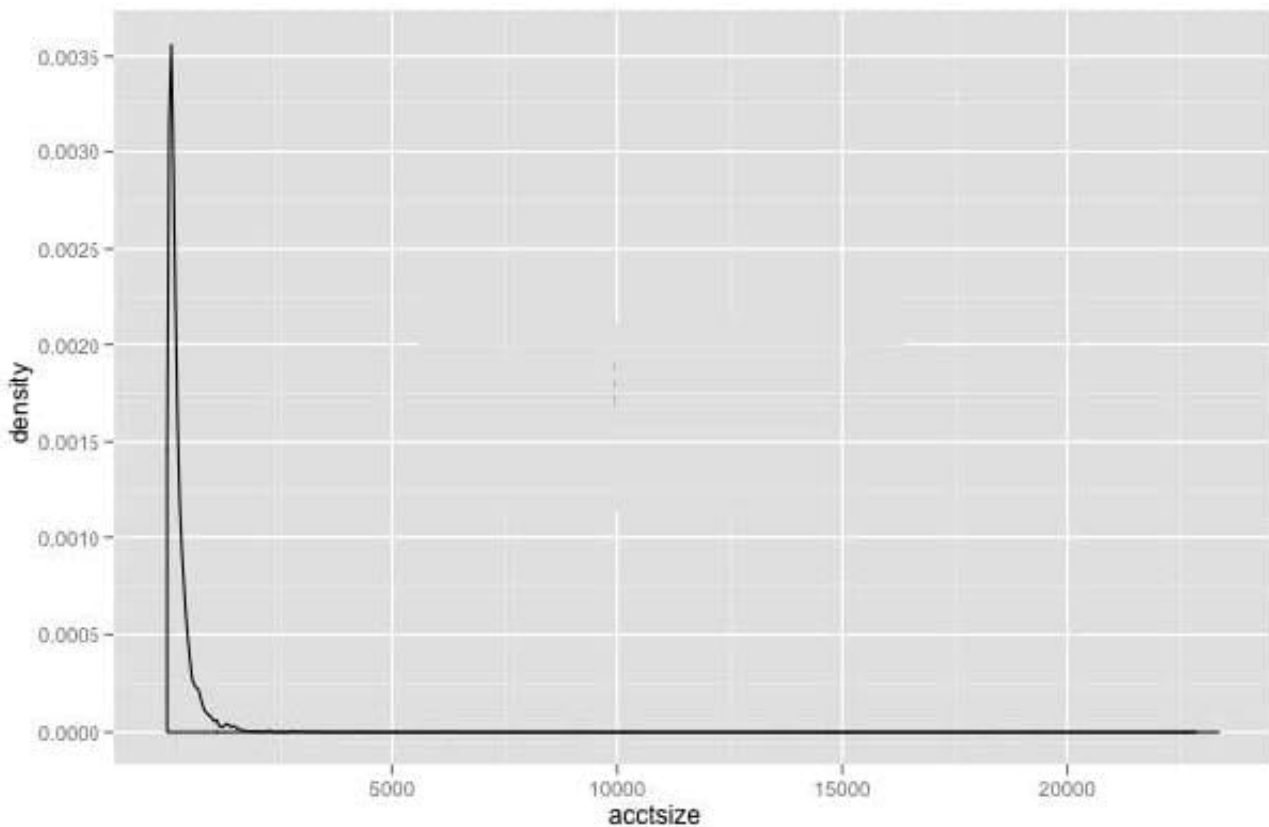
A. Run MapReduce to transform the data,and find relevant key value pairs.

B. Divide the data into sets of 1,000 user profiles,and run K-means clustering in RHadoop iteratively.

C. Run a Naive Bayes classification as a pre-processing step in HDFS.

D. Partition the data by XML file size,and run K-means clustering in each partition.

Correct Answer: A

QUESTION 8

Refer to the exhibit.

You have plotted the distribution of savings account sizes for your bank. How would you proceed, based

on this distribution?

A. The data is extremely skewed. Replot the data on a logarithmic scale to get a better sense of it.

B. The data is extremely skewed,but looks bimodal; replot the data in the range 2,500-10,000 to be sure.

C. The accounts of size greater than 2500 are rare,and probably outliers. Eliminate them from your future analysis.

D. The data is extremely skewed. Split your analysis into two cohorts: accounts less than 2500,and accounts greater than 2500

Correct Answer: A

QUESTION 9

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 10

You have two tables of customers in your database. Customers in cust_table_1 were sent an e-mail promotion last year, and customers in cust_table_2 received a newsletter last year. Customers can only be entered in once per table. You want to create a table that includes all customers, and any of the

communications they received last year. Which type of join would you use for this table?

A. Full outer join

B. Inner join

C. Left outer join

D. Cross join

Correct Answer: A

QUESTION 11

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 12

A data scientist is asked to implement an article recommendation feature for an on-line magazine. The

magazine does not want to use client tracking technologies such as cookies or reading history. Therefore,

only the style and subject matter of the current article is available for making recommendations. All of the

magazine\\'s articles are stored in a database in a format suitable for analytics.

Which method should the data scientist try first?

A. K Means Clustering

B. Naive Bayesian

C. Logistic Regression

D. Association Rules

Correct Answer: A

QUESTION 13

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 14

When creating a project sponsor presentation, what is the main objective?

A. Show that you met the project goals

B. Show how you met the project goals

C. Show how well the model will meet the SLA (service level agreement)

D. Clearly describe the methods and techniques used

Correct Answer: A

QUESTION 15

In R, functions like plot() and hist() are known as what?

A. generic functions

B. virtual methods

C. virtual functions

D. generic methods

Correct Answer: B