

E20-007^{Q&As}

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.pass2lead.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

Which of the following is an example of quasi-structured data?

- A. OLAP
- B. OLTP
- C. Customer record table
- D. Clickstream data

Correct Answer: D

QUESTION 2

The web analytics team uses Hadoop to process access logs. They now want to correlate this data with structured user data residing in a production single-instance JDBC database. They collaborate with the production team to import the data into Hadoop. Which tool should they use?

- A. Sqoop
- B. Pig
- C. Chukwa
- D. Scribe

Correct Answer: A

QUESTION 3

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?

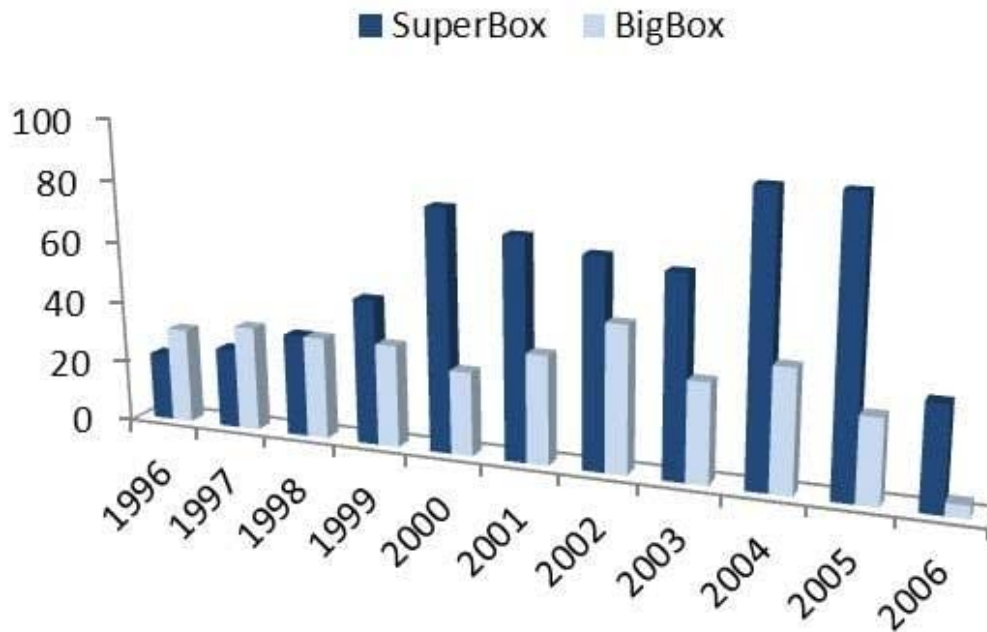
- A. Run MapReduce to transform the data, and find relevant key value pairs.
- B. Divide the data into sets of 1, 000 user profiles, and run K-means clustering in RHadoop iteratively.
- C. Run a Naive Bayes classification as a pre-processing step in HDFS.
- D. Partition the data by XML file size, and run K-means clustering in each partition.

Correct Answer: A

QUESTION 4

Refer to the Exhibit.

Growth of SuperBox Stores (Count of Stores)



In the Exhibit. For effective visualization, what is the chart's primary flaw?

- A. The use of 3 dimensions.
- B. The slanting of axis labels.
- C. The location of the legend.
- D. The order of the columns.

Correct Answer: A

QUESTION 5

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

[Latest E20-007 Dumps](#)

[E20-007 Practice Test](#)

[E20-007 Exam Questions](#)