



DS-200^{Q&As}

Data Science Essentials

Pass Cloudera DS-200 Exam with 100% Guarantee

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

<https://www.pass4itsure.com/ds-200.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Cloudera
Official Exam Center

- ⚙ **Instant Download** After Purchase
- ⚙ **100% Money Back** Guarantee
- ⚙ **365 Days** Free Update
- ⚙ **800,000+** Satisfied Customers





QUESTION 1

What is the most common reason for a k-means clustering algorithm to return a sub-optimal clustering of its input?

- A. Non-negative values for the distance function
- B. Input data set is too large
- C. Non-normal distribution of the input data
- D. Poor selection of the initial controls

Correct Answer: C

QUESTION 2

You have user profile records in an OLTP database that you want to join with web server logs which you have already ingested into HDFS. What is the best way to acquire the user profile for use in HDFS?

- A. Ingest with Hadoop streaming
- B. Ingest with Apache Flume
- C. Ingest using Hive's LOAD DATA command
- D. Ingest using Sqoop
- E. Ingest using Pig's LOAD command

Correct Answer: BD

Reference: https://thinkbiganalytics.com/leading_big_data_technologies/ingestion-and-streaming-withstorm-kafka-flume/

QUESTION 3

You have a large $m \times n$ data matrix M . You decide you want to perform dimension reduction/clustering on your data and have decided to use the singular value decomposition (SVD; also called principal components analysis PCA)

You performed singular value decomposition (SVD; also called principal components analysis or PCA) on your data matrix but you did not center your data first. What does your first singular component describe?

- A. The mean of the data set
- B. The variance of the data set
- C. The standard deviation of the data set



D. The maximum of the data set

E. The median of the data set

Correct Answer: C

QUESTION 4

Many machine learning algorithm involve finding the Global minimum of a convex loss function, primarily because:

A. The additive inverse of a convex function is concave

B. The derivative of convex function is always defined

C. The second derivative of a convex function is a constant

D. Any local minimum of a convex is also a global minimum

Correct Answer: B

QUESTION 5

You are working with a logistic regression model to predict the probability that a user will click on an ad. Your model has hundreds of features, and you're not sure if all of those features are helping your prediction. Which regularization technique should you use to prune features that aren't contributing to the model?

A. Convex

B. Uniform

C. L2

D. L1

Correct Answer: A

[Latest DS-200 Dumps](#)

[DS-200 PDF Dumps](#)

[DS-200 Study Guide](#)