

# DS-200<sup>Q&As</sup>

Data Science Essentials

## Pass Cloudera DS-200 Exam with 100% Guarantee

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

<https://www.pass2lead.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

Which best describes the primary function of Flume?

- A. Flume is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with an infrastructure consisting of sources and sinks for importing and evaluating large data sets
- B. Flume acts as a Hadoop filesystem for log files
- C. Flume Imports data from SQL/relational database into your Hadoop cluster
- D. Flume provides a query languages for Hadoop similar to SQL
- E. Flume is a distributed server for collecting and moving large amount of data into HDFS as it's produced from streaming data flows

Correct Answer: D

---

### QUESTION 2

Which two techniques should you use to avoid overfitting a classification model to a data set?

- A. Include a small number "noise" features that are not through to be correlated with the dependent variable.
- B. Replicate features that are through to be significant predictors of the dependent variable multiple time for each observation.
- C. Separate your input data into a training set that is used for fitting and a test set that is used for evaluating the model's performance
- D. Include a regularization term in the model's objective function to control how precisely the model fits the data
- E. Preprocess the data to exclude a typical observation from the model input

Correct Answer: AE

---

### 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 decide to use the singular value decomposition (SVD; also called principal components analysis PCA)

Refer to the passage above.

What represents the SVD of the Matrix standard  $M$  given the following information:

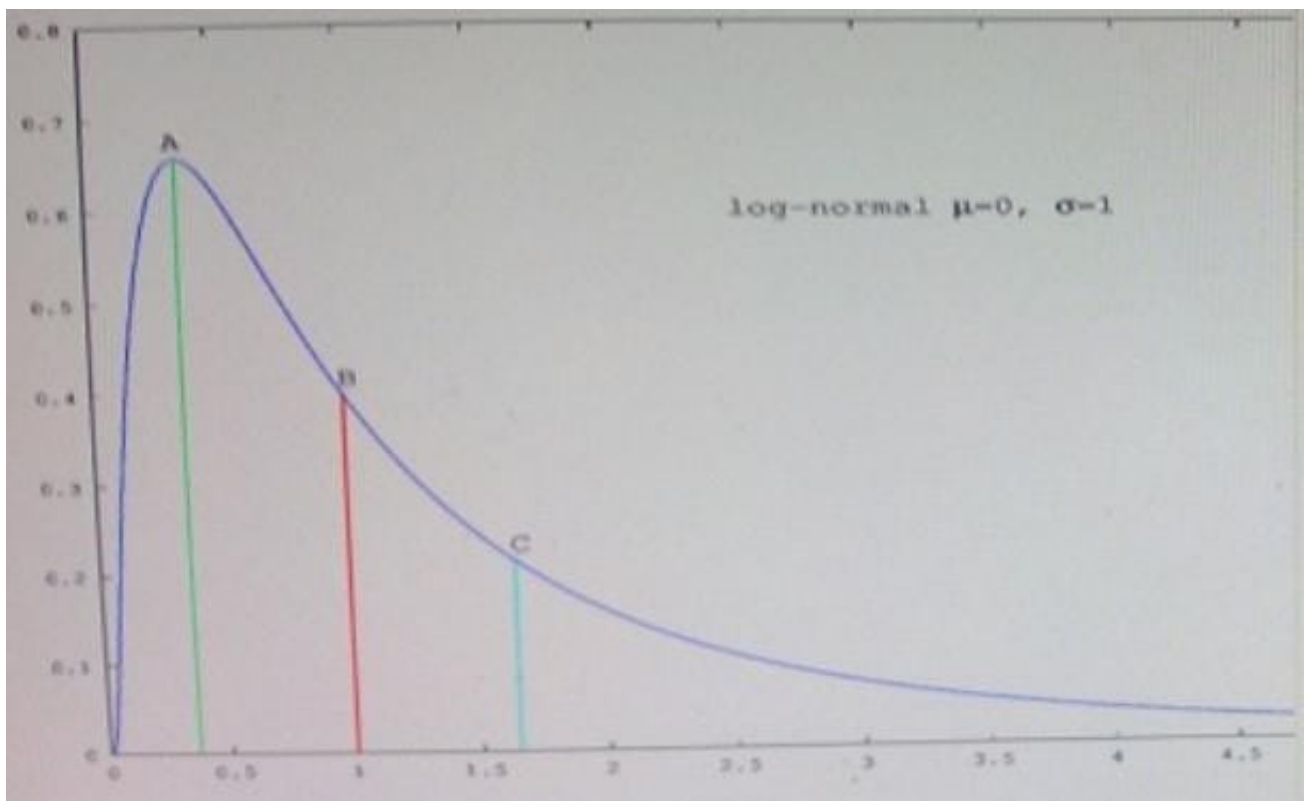
$U$  is  $m \times m$  unitary  $V$  is  $n \times n$  unitary  $S$  is  $m \times n$  diagonal  $Q$  is  $n \times n$  invertible  $D$  is  $n \times n$  diagonal  $L$  is  $m \times m$  lower triangular  $U$  is  $m \times m$  upper triangular

- A.  $M = U S V$
- B.  $M = U P$
- C.  $M = Q D Q^{-1}$
- D.  $M = L U$

Correct Answer: A

**QUESTION 4**

Refer to the exhibit.



Which point in the figure is the mode?

- A. A
- B. B
- C. C

Correct Answer: C

**QUESTION 5**

You are about to sample a 100-dimensional unit-cube. To adequately sample any single given dimension, you need only capture 10 points. How many points do you need to order to sample the complete 100dimensional unit cube adequately?

- A. 10010
- B. 1010
- C.  $\text{Log}_2(100)$
- D. 100
- E. 1000
- F. 1010

Correct Answer: E

[Latest DS-200 Dumps](#)

[DS-200 VCE Dumps](#)

[DS-200 Practice Test](#)