

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.leads4pass.com/ds-200.html>

100% Passing Guarantee
100% Money Back Assurance

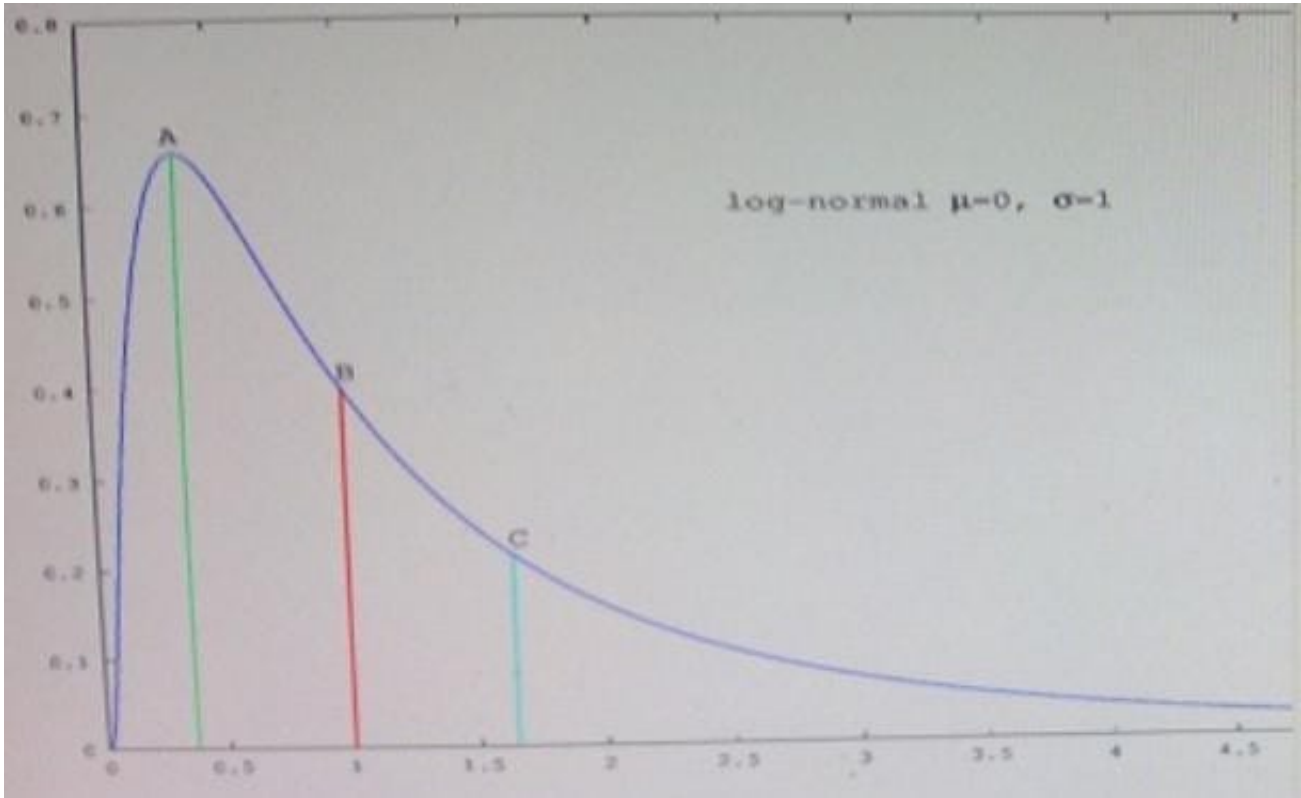
Following Questions and Answers are all new published by Cloudera
Official Exam Center

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



QUESTION 1

Refer to the exhibit.



Which point in the figure is the mode?

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

Correct Answer: C

QUESTION 2

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. $\log_2(100)$

D. 100

E. 1000

F. 1010

Correct Answer: E

QUESTION 3

Given the following sample of numbers from a distribution:

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89

How do high-level languages like Apache Hive and Apache Pig efficiently calculate approximately percentiles for a distribution?

A. They sort all of the input samples and then lookup the samples for each percentile

B. They maintain index of input data as it is loaded into HDFS and load them into memory

C. They use pivots to assign each observations to the reducer that calculate each percentile

D. They assign sample observations to buckets and then aggregate the buckets to compute the approximations

Correct Answer: C

QUESTION 4

Consider the following sample from a distribution that contains a continuous X and label Y that is either A or B:

X	Y
1	A
2	A
3	A
4	B
5	A
6	B
7	A
8	B
9	B
10	B

Which is the best cut point for X if you want to discretize these values into two buckets in a way that minimizes the sum of chi-square values?

- A. $X = 8$
- B. $X = 6$
- C. $X = 5$
- D. $X = 4$
- E. $X = 2$

Correct Answer: D

QUESTION 5

You want to build a classification model to identify spam comments on a blog. You decide to use the words in the comment text as inputs to your model. Which criteria should you use when deciding which words to use as features in order to contribute to making the correct classification decision?

- A. Choose words for your sample that are most correlated with the Spam label
- B. Choose words for your sample that occur most frequently in the text
- C. Choose words, for your sample that have the largest mutual information with the spam label
- D. Choose words for your sample that are least correlated with the spam label

Correct Answer: A

[DS-200 PDF Dumps](#)

[DS-200 Practice Test](#)

[DS-200 Exam Questions](#)