



# 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.geekcert.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

There are 20 patients with acute lymphoblastic leukemia (ALL) and 32 patients with acute myeloid leukemia (AML), both variants of a blood cancer.

The makeup of the groups as follows:

ALL GROUP			
	Male	Female	
Caucasian	14	1	15
Asian-American	5	0	5
	19	1	20

AML GROUP			
	Male	Female	
Caucasian	9	4	13
Asian-American	7	12	19
	16	16	32

Each individual has an expression value for each of 10000 different genes. The expression value for each gene is a continuous value between -1 and 1.

With which type of plot can you encode the most amount of the data visually?

Rather than use all 10,000 features to separate AML from ALL, you pick a small subset of features to

separate them optimally. You feature vectors have 10,000 dimensions while you only have 52 data points. You use cross-validation to test your chosen set of features. What three methods will choose the features in an optimal way?

- A. Singular value Decomposition
- B. Bootstrapping
- C. Markov chain Monte Carlo
- D. Hidden Markov
- E. Bayesian Information Criterion
- F. Mutual Information

Correct Answer: CDF

### QUESTION 2



In what format are web server log files usually generated and how must you transform them in order to make them usable for analysis in Hadoop?

- A. XML files that you need to convert to JSON
- B. Text files that require parsing into useful fields
- C. CSV files that require parsing into useful fields
- D. HTML files that you need to convert to plain text or CSV
- E. Binary files that may require decompression and conversion using AVRO

Correct Answer: AB

---

### QUESTION 3

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?

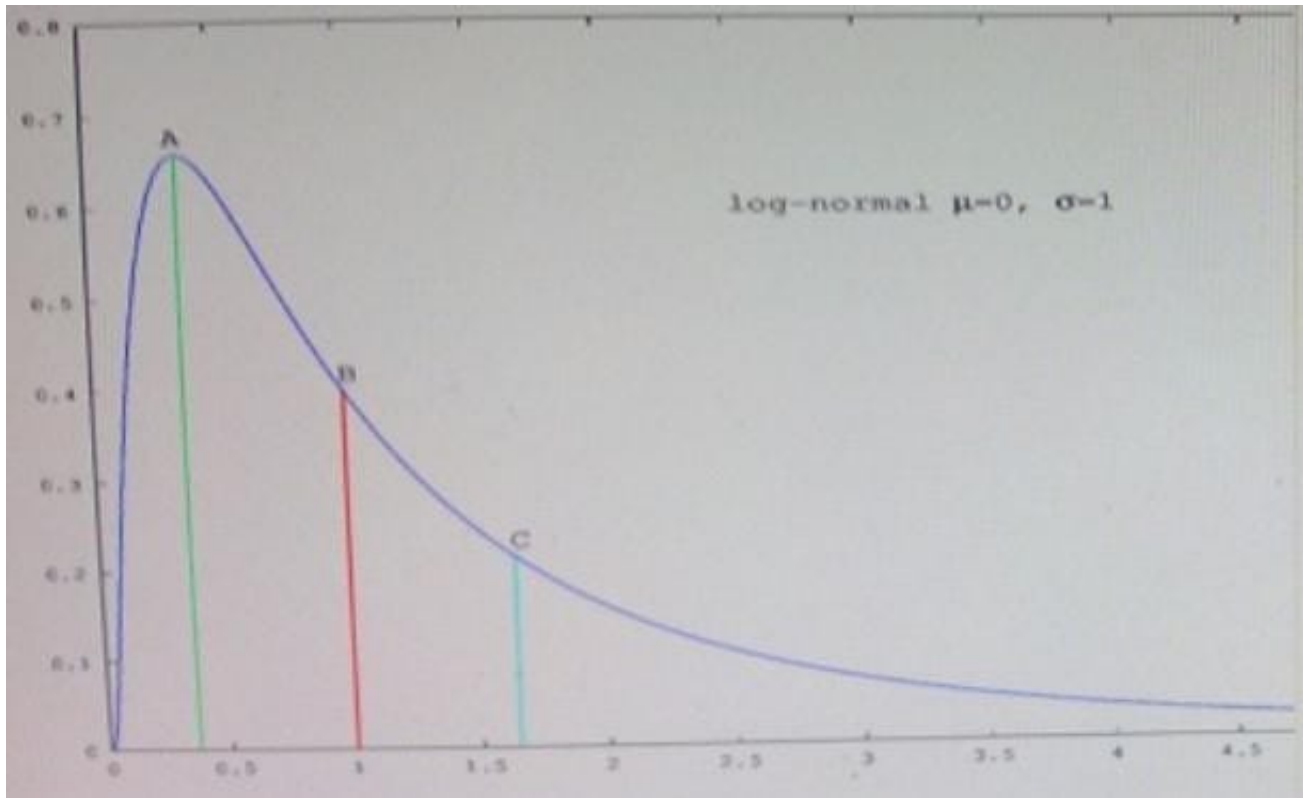
- A. 10010
- B. 1010
- C.  $\text{Log}_2(100)$
- D. 100
- E. 1000
- F. 1010

Correct Answer: E

---

### 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 have a directory containing a number of comma-separated files. Each file has three columns and each filename has a .csv extension. You want to have a single tab-separated file (all .tsv) that contains all the rows from all the files.

Which command is guaranteed to produce the desired output if you have more than 20,000 files to process?

- A. `find . -name '*.CSV' | xargs -0 cat | tr ',' '\t' > all.tsv`
- B. `find . -name '*.CSV' | cat | awk 'BEGIN {FS = "," OFS = "\t"} {print $1, $2, $3}' > all.tsv`
- C. `find . -name '*.CSV' | tr ',' '\t' | cat > all.tsv`
- D. `find . -name '*.CSV' | cat > all.tsv`
- E. `Cat *.CSV > all.tsv`



Correct Answer: B

[DS-200 Practice Test](#)

[DS-200 Exam Questions](#)

[DS-200 Braindumps](#)