

# 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

You've built a model that has ten different variables with complicated independence relationships between them, and both continuous and discrete variables that have complicated, multi-parameter distributions. Computing the joint probability distribution is complex, but it turns out that computing the conditional probabilities for the variables is easy. What is the most computationally efficient for computing the expected value?

- A. Method of moments
- B. Markov Chain Monte Carlo
- C. Gibbs sampling
- D. Numerical quadrature

Correct Answer: B

---

### QUESTION 2

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 3

You have a data file that contains two trillion records, one record per line (comma separated). Each record lists two friends and unique message sent between them. Their names will not have commas.

Michael, John, Pabst, Blue Ribbon Tiffany, James, BMX Racing John, Michael, Natural Lemon Flavor

Analyze the pseudo code examples below and determine which set of mappers and reducers in the below pseudo code snippets will solve for the mean number of messages each user sends to all of the friends?

For example pseudo code may have three friends to whom he sends 6, 10, and 200 messages, respectively, so Michael's mean would be  $(6+10+200)/3$ . The solution may require a pipeline of two MapReduce jobs.

- A. `def mapper1 (line): key1, key2, message = line.split( ',' , '\\')` `emit ( (key1, key2) , 1)` `def reducer1(key, values): emit (key, sum(values))` `def mapper2(key, value): key1, key2 = key // unpack both friends name into separate keys` `emit (key1, value)`

def reducer2(key, values):

emit (key, mean (values) )

B. def mapper1 (line): key1, key2, message = line.split ( ` , \\` ) emit ( (key1, key2) , 1) emit ( (key1, key2) , 1) def reducer1(key, values): emit (key, sum(values)) def mapper2(key, value): key1, key2 = key // unpack both friends name into separate keys emit (key1, value) def reducer2(key, values): emit (key, mean (values) )

C. def mapper1 (line): key1, key2, message = line.split ( ` , \\` ) emit ( (key1, key2) , 1) emit ( (key1, key2) , 1) def reducer1(key, values): emit (key, sum(values))

D. def mapper (line) : Key1, key2, message = line.split ( ` , \\` ) Sort (key1, key2) // a fiven pair will always be sorted the same Emit ( ( key 1, key2), 1) Def reducer1(key, values) : Emit (key, sum (values) ) Def Mapper2 (key, value) Key1, key2 = key // unpack both friends names into separate keys Emit (key1, value) Emit (key2, value) Def reducer2(key, values); Emit (key, mean (values) )

Correct Answer: B

#### QUESTION 4

You have a large file of N records (one per line), and want to randomly sample 10% them. You have two functions that are perfect random number generators (through they are a bit slow):

Random\_uniform () generates a uniformly distributed number in the interval [0, 1] random\_permotation (M) generates a random permutation of the number O through M -1.

Below are three different functions that implement the sampling.

Method A

For line in file: If random\_uniform ()

Method B

i = 0

for line in file:

if i % 10 == 0;

print line

i += 1

Method C

idxs = random\_permotation (N) [: (N/10)]

i = 0

for line in file:

if i in idxs:

print line

i +=1

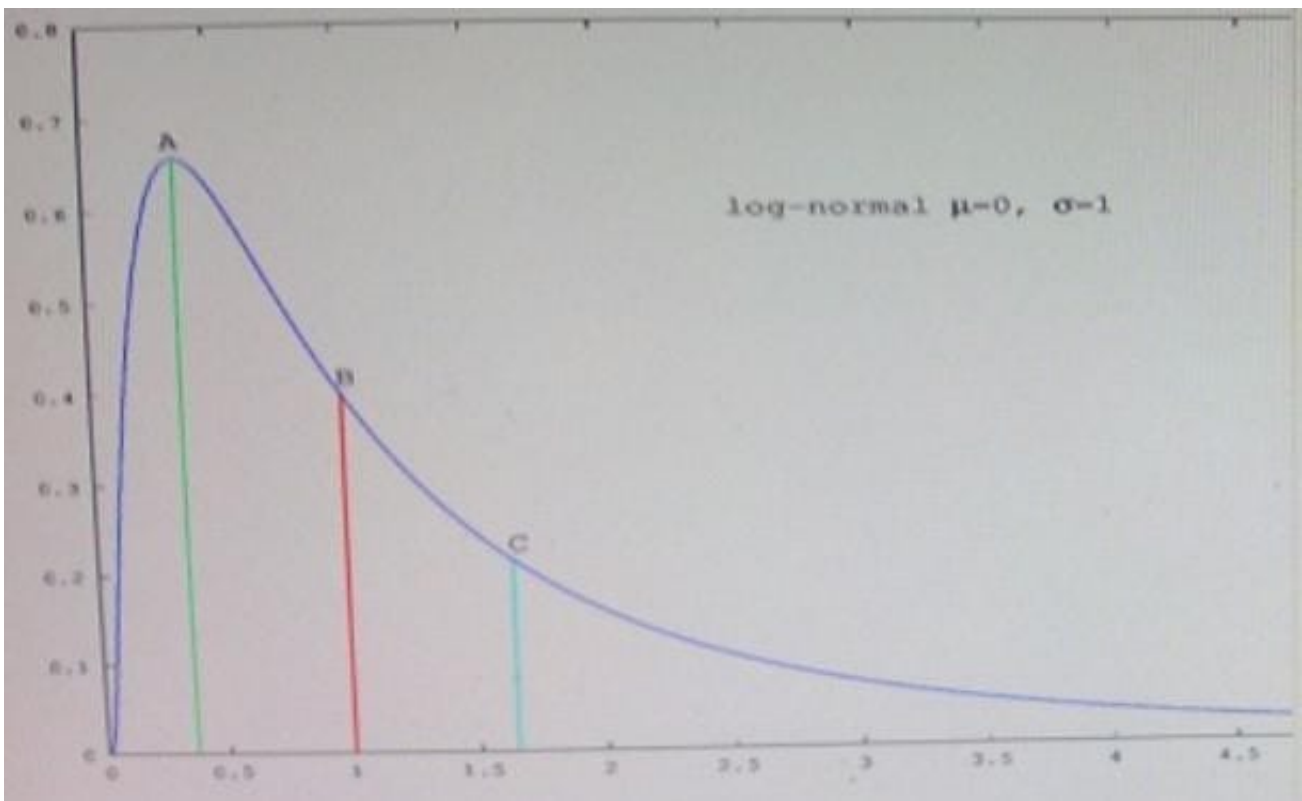
Which method will have the best runtime performance?

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

Correct Answer: A

**QUESTION 5**

Refer to the exhibit.



Which point in the figure is the mean?

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

Correct Answer: B

[DS-200 PDF Dumps](#)

[DS-200 Study Guide](#)

[DS-200 Exam Questions](#)