



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**

When optimizing a function using stochastic gradient descent, how frequently should you update your estimate of the gradient?

- A. Once after every pass through the data set
- B. Once per observation
- C. For each observation with a probability that you choose ahead of time
- D. After a random number of observations
- E. Once every N observations, where you decide N ahead of time

Correct Answer: AC

QUESTION 2

You want to understand more about how users browse your public website. For example, you want to know which pages they visit prior to placing an order. You have a server farm of 200 web servers hosting your website. Which is the most efficient process to gather these web servers access logs into your Hadoop cluster for analysis?

- A. Sample the web server logs from web servers and copy them into HDFS using curl
- B. Channel these click streams into Hadoop using Hadoop Streaming
- C. Write a MapReduce job with the web servers for mappers and the Hadoop cluster nodes for reducers
- D. Import all user clicks from your OLTP databases into Hadoop using Sqoop
- E. Ingest the server web logs into HDFS using Flume

Correct Answer: C

QUESTION 3

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 choice of cut points for X if you want to discretize these values into three buckets that minimizes the sum of chi-square values?

- A. X 5 and X 8
- B. X 4 and X 6
- C. X 3 and X 8
- D. X 3 and X 6
- E. X 2 and X 9

Correct Answer: E

QUESTION 4

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 5**

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_permutation (M) generates a random permutation of the number 0 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_permutation (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



VCE & PDF

Pass4itSure.com

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

2024 Latest pass4itsure DS-200 PDF and VCE dumps Download

[Latest DS-200 Dumps](#)

[DS-200 Practice Test](#)

[DS-200 Braindumps](#)