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.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

You have acquired a new data source of millions of customer records, and you\\'ve this data into HDFS. Prior to analysis, you want to change all customer registration to the same date format, make all addresses uppercase, and remove all customer names (for anonymization). Which process will accomplish all three objectives?

A. Adapt the data cleansing module in Mahout to your data, and invoke the Mahout library when you run your analysis

B. Pull this data into an RDBMS using sqoop and scrub records using stored procedures

C. Write a script that receives records on stdin, corrects them, and then writes them to stdout. Then, invoke this script in a map-only Hadoop Streaming Job

D. Write a MapReduce job with a mapper to change words to uppercase and to reduce different forms of dates to a single form

Correct Answer: C

QUESTION 2

Certain individuals are more susceptible to autism if they have particular combinations of genes expressed in their DNA. Given a sample of DNA from persons who have autism and a sample of DNA from persons who do not have autism, determine the best technique for predicting whether or not a given individual is susceptible to developing autism?

A. Native Bayes

B. Linear Regression

C. Survival analysis

D. Sequence alignment

Correct Answer: B

QUESTION 3

Under what two conditions does stochastic gradient descent outperform 2nd-order optimization techniques such as iteratively reweighted least squares?

A. When the volume of input data is so large and diverse that a 2nd-order optimization technique can be fit to a sample of the data

B. When the model\\'s estimates must be updated in real-time in order to account for new observations.

C. When the input data can easily fit into memory on a single machine, but we want to calculate confidence intervals for all of the parameters in the model.

D. When we are required to find the parameters that return the optimal value of the objective function.

Correct Answer: AB

QUESTION 4

What is default delimiter for Hive tables?
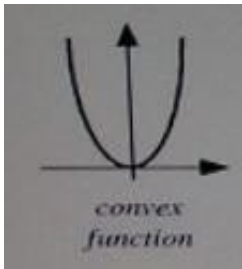
A. ^A (Control-A)

B. , (comma)

C. \t (tab)

D. : (colon)

Correct Answer: A

Reference: http://blog.spryinc.com/2013/10/four-useful-tricks-for-working-with-hive.html (change the delimiter when exporting hive table)

QUESTION 5

Function is convex if the line segment between two points, a and b is greater than equal to the value of the a x b Which two functions are convex?

A. X1/2

B. Ex

C. 2x-1

D. 1-x2

Correct Answer: A