



# 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.passapply.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.

You want to use the data from the 52 patients in the scenario to improve the ability of doctors being able to distinguish between ALL and AML. What type of data science problem is this?

- A. Classification
- B. Regression
- C. Clustering
- D. Filtering

Correct Answer: D

### QUESTION 2

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 3

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 4

What is one limitation encountered by all systems that employ collaborative filtering and use preferences as input. In order to output product recommendations to consumers?

A. Consumers do not have stable ratings for the same product over time

B. There are too many consumers and too few products

C. Not every product has been rated by every consumer

D. There are too few consumers and too many products

Correct Answer: A

---

### QUESTION 5

What are two defining features of RMSE (root-mean square error or root-mean-square deviation)?

A. It is sensitive to outliers

B. It is the mean value of recommendations of the K-equal partitions in the input data

---



C. It is the square of the median value of the error where error is the difference between predicted rating and actual ratings

D. It is appropriate for numeric data

E. It considers the order of recommendations

Correct Answer: BD

[DS-200 Practice Test](#)

[DS-200 Study Guide](#)

[DS-200 Exam Questions](#)