

## 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.leads4pass.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**

From historical data, you know that 50% of students who take Cloudera's Introduction to Data Science: Building Recommenders Systems training course pass this exam, while only 25% of students who did not take the training course pass this exam. You also know that 50% of this exam's candidates also take Cloudera's Introduction to Data Science: Building Recommendations Systems training course.

What is the probability that any individual exam candidate will pass the data science exam?

- A. 3/8
- B. 1/4
- C. 1/8
- D. 1/2

Correct Answer: C

**QUESTION 2**

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?

You choose to perform agglomerative hierarchical clustering on the 10,000 features. How much RAM do you need to hold the distance Matrix, assuming each distance value is 64-bit double?

A. ~ 800 MB

B. ~ 400 MB

C. ~ 160 KB

D. ~ 4 MB

Correct Answer: B

---

### 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

You have just run a MapReduce job to filter user messages to only those of a selected geographical region. The output for this job in a directory named westUsers, located just below your home directory in HDFS. Which command gathers these records into a single file on your local file system?

- A. Hadoop fs getmerge westUsers WestUsers.txt
- B. Hadoop fs get westUsers WestUsers.txt
- C. Hadoop fs cp westUsers/\* westUsers.txt
- D. Hadoop fs getmerge R westUsers westUsers.txt

Correct Answer: B

---

#### 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\` print0 | sargs -0 cat | tr `,\` `t\` > all.tsv`
- B. `Find . name `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 PDF Dumps](#)

[DS-200 Exam Questions](#)

[DS-200 Braindumps](#)