

# *Dell EMC*

*D-AA-OP-23*

*Dell Technologies Data Scientist Advanced Analytics Optimize 2023*



**For More Information – Visit link below:**

**<https://www.examsboost.com/>**

## **Product Version**

- ✓ Up to Date products, reliable and verified.
- ✓ Questions and Answers in PDF Format.

---

# Latest Version: 6.0

## Question: 1

According to Metcalfe's law, what is true about the value of a network?

Response:

- A. Proportional to the number of edges
- B. Proportional to the logarithm of the number of edges
- C. Unrelated to the number of edges
- D. Proportional to the square of the number of edges

**Answer: C**

## Question: 2

What are key characteristics of Random Graphs?

Response:

- A. Low clustering coefficients; high network diameters
- B. Low clustering coefficients; small network diameters
- C. High clustering coefficients; high network diameters
- D. High clustering coefficients; small network diameters

**Answer: B**

## Question: 3

Which scenario is a proper use case for multinomial logistic regression?

Response:

- A. A marketing firm wants to estimate the personal income of a group of potential customers. Using inputs such as age, education, marital status, and credit card expenditures, a data scientist is building a model that will estimate a person's income
- B. A logistic distribution company wants to minimize the distance traveled by its delivery trucks. A data scientist is building a model to determine the optimal route for each of its trucks
- C. To improve the initial routing of a loan application, a financial institution plans to classify a loan application as Approve, Reject, or Possibly\_Approve. Based on the company's historical loan application data, a data scientist is building a model to assign one of these three outcomes to each submitted application.

D. A manufacturer plans to determine the optimal number of workers to employ in an assembly line process. Utilizing the observed distributions of the task durations of each process step, a data scientist is building a model to mimic the interactions and dependencies between each stage in the manufacturing process.

**Answer: C**

#### Question: 4

You develop a Python script "logisticpy" to evaluate the logistic function denoted as  $f(y)$  for a given value  $y$  that includes the following Pig code:

Register 'logistic.py' using jython as udf; z = FOREACH y GENERATE \$0, udf.logistic (\$0); DUMP z;

What is the expected output when the Pig code is executed?

Response:

- A. 0
- B. Jython is not a supported language
- C. Value of  $f(y)$  for ally
- D. Tuples  $(y, f(y))$

**Answer: D**

#### Question: 5

What is a characteristic of stop words?

Response:

- A. Used in term frequency analysis
- B. Include words such as "a", "an", and "the"
- C. Meaningful words requiring a parser to stop and examine them
- D. Don't occur often in text

**Answer: B**

#### Question: 6

Why would a company decide to use HBase to replace an existing relational database?

Response:

- A. It is required for performing ad-hoc queries.
- B. Varying formats of input data requires columns to be added in real time.
- C. The company's employees are already fluent in SQL.

D. Existing SQL code will run unchanged on HBase.

**Answer: A**

### Question: 7

Which problem type is best suited for simulation?

Response:

- A. One with a few, non-random input variables
- B. One that compares "what-if scenarios"
- C. One with numerous, non-random input-variables
- D. One that has a closed-form solution

**Answer: B**

### Question: 8

Which scenario would be ideal for processing Hadoop data with Hive?

Response:

- A. Structured data, real-time processing
- B. Unstructured data; batch processing
- C. Unstructured data; real-time processing
- D. Structured data; batch processing

**Answer: B**

### Question: 9

You conduct a TFIDF analysis on 3 documents containing raw text and derive TFIDF ("data", document y) = 1.908. You know that the term "data" only appears in document 2. What is the TF of "data" in document 2?

Response:

- A. 2 based on the following reasoning:  $TFIDF = TF \cdot IDF = 1.908$  You then know that  $IDF = \log(3/1) = 0.954$  Therefore,  $TFIDF = TF \cdot 0.954 = 1.908$  TF will then round to 2
- B. 4 based on the following reasoning:  $TFIDF = TF \cdot IDF = 1.908$  You then know that  $IDF = \log(3/1) = 0.477$  Therefore,  $TFIDF = TF \cdot 0.477 = 1.908$  TF will then round to 4
- C. 6 based on the following reasoning:  $TFIDF = TF \cdot IDF = 1.908$  You then know that  $IDF = \log(3/1) = 3$  Therefore,  $TFIDF = TF / 3 = 1.908$  TF will then round to 6

---

D. 11 based on the following reasoning:  $TFIDF = TF \cdot IDF = 1908$  You then know that IDF will equal  $\text{LOG}(3/2) = 0.176$  Therefore,  $TFIDF = TF \cdot 0.176 = 1.908$  TF will then round to 11

**Answer: B**

### Question: 10

You are analyzing written transcripts of focus groups conducted on product X. Your approach is to use TF-IDF for your analysis. What combination of TF-IDF scores should you examine to ensure you only report on the most important terms?

Response:

- A. High TF score and high DF score
- B. High TF score and high IDF score
- C. High TF score and low IDF score
- D. Low TF score and low DF score

**Answer: C**

---

# Thank You for Trying Our Product

For More Information – **Visit link below:**

**<https://www.examsboost.com/>**

15 USD Discount Coupon Code:

**G74JA8UF**

## FEATURES

- ✓ **90 Days Free Updates**
- ✓ **Money Back Pass Guarantee**
- ✓ **Instant Download or Email Attachment**
- ✓ **24/7 Live Chat Support**
- ✓ **PDF file could be used at any Platform**
- ✓ **50,000 Happy Customer**



Visit us at <https://www.examsboost.com/test/d-aa-op-23/>