

# Latest Version: 6.0

## Question: 1

What is the difference between the array and list data structures in R?  
Response:

- A. Arrays contain only the same data type;  
Lists can contain different data types
- B. Arrays are only 2-dimensional;  
Lists are N-dimensional
- C. Arrays can contain different data types;  
Lists can contain only the same data type
- D. Arrays are N-dimensional;  
Lists are only 2-dimensional

**Answer: A**

## Question: 2

You have an automotive database containing numeric characteristics such as engine size, horsepower, and top speed. Which technique could you use to group similar cars together?  
Response:

- A. Naïve Bayes classifier
- B. Association rules
- C. K-means clustering
- D. Logistic regression

**Answer: C**

## Question: 3

Which activity might be performed in the Operationalize phase of the Data Analytics Lifecycle?  
Response:

- A. Run a pilot
- B. Try different analytical techniques
- C. Try different variables
- D. Transform existing variables

**Answer: A**

### Question: 4

The web analytics team uses Hadoop to process access logs. They now want to correlate this data with structured user data residing in their massively parallel database. Which tool should they use to export the structured data from Hadoop?

Response:

- A. Sqoop
- B. Pig
- C. Chukwa
- D. Scribe

**Answer: A**

### Question: 5

What is a property of window functions in SQL commands?

Response:

- A. They can be used to calculate moving averages over various intervals.
- B. They group rows into a single output row.
- C. They can be used between the keywords FROM and WHERE in a SELECT command.
- D. They don't require ordering of data within a window.

**Answer: A**

### Question: 6

Your colleague, who is new to Hadoop, approaches you with a question. They want to know how best to access their data

- a. This colleague has a strong background in data flow languages and programming.

Which query interface would you recommend?

Response:

- A. Pig
- B. Hive
- C. Howl

D. HBase

**Answer: A**

**Question: 7**

In the Map Reduce framework, what is the purpose of the Map Function?

Response:

- A. It processes the input and generates key-value pairs
- B. It collects the output of the Reduce function
- C. It sorts the results of the Reduce function
- D. It breaks the input into smaller components and distributes to other nodes in the cluster

**Answer: A**

**Question: 8**

Refer to the exhibit.

**Table A: Term inverse document frequencies**

	Solid	State	Disk
IDF	2	1.5	0.5

**Table B: Term per-document frequencies**

Document	Tf		
	Solid	State	Disk
A	0.1	0.1	0.1
B	0.1	0.1	0.5
C	0.2	0.2	0.1
D	0.1	0.3	0.1

Click on the calculator icon in the upper left corner. An analyst is searching a corpus of documents for the topic "solid state disk".

In the Exhibit, Table A provides the inverse document frequency for each term across the corpus. Table B provides each term's frequency in four documents selected from corpus.

Which of the four documents is most relevant to the analyst's search?

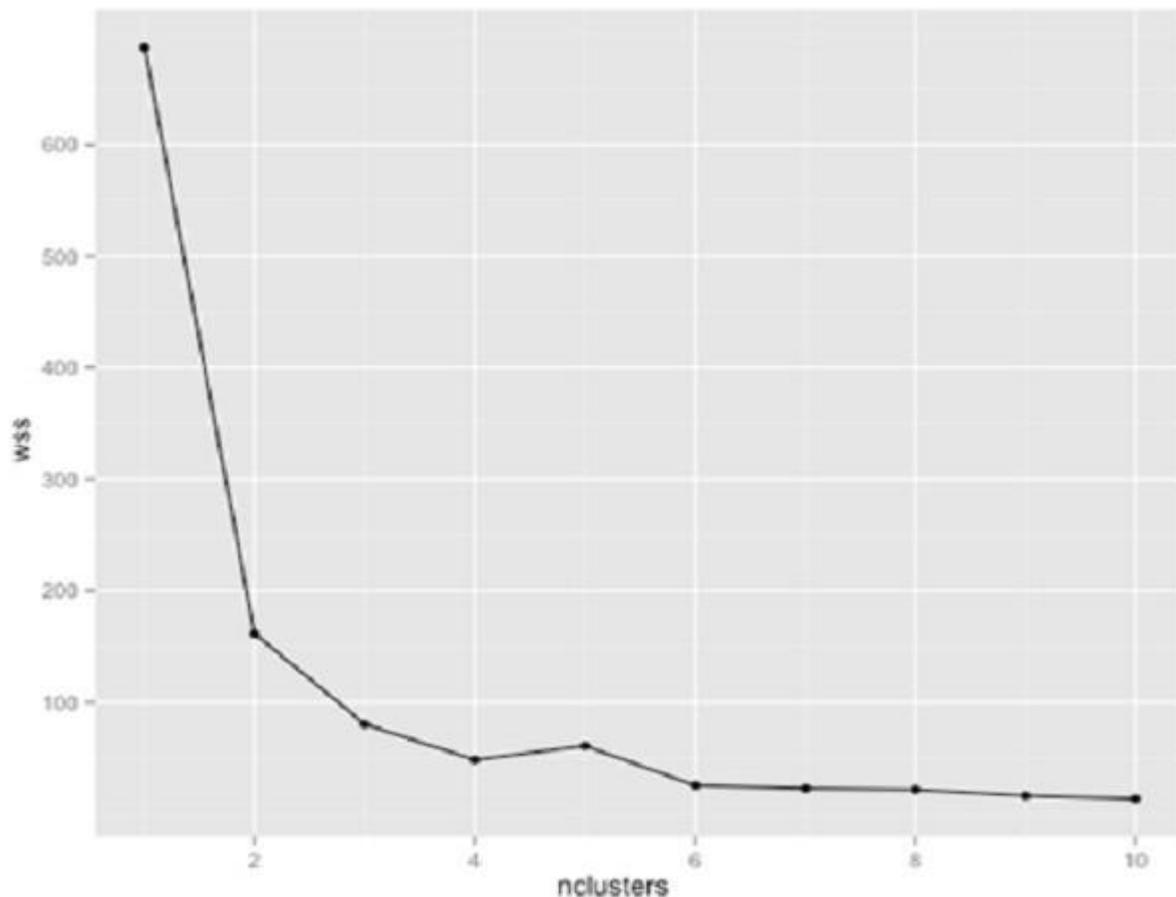
Response:

- A. Document C
- B. Document A
- C. Document B
- D. Document D

**Answer: A**

### Question: 9

Refer to the exhibit.



You are using k-means clustering to discover groupings within a data set. You plot within- sum-of-squares (wss) of multiple cluster sizes. Based on the exhibit, how many clusters should you use in your analysis?

Response:

- A. 4
- B. 2
- C. 8
- D. 10

**Answer: A**

### Question: 10

Which key role for a successful analytic project can provide business domain expertise with a deep understanding of the data and key performance indicators?

Response:

- A. Business Intelligence Analyst
- B. Project Manager
- C. Project Sponsor
- D. Business User

**Answer: A**