

Knowledge Check 11: Data Management in Apache Spark

Due No due date	Points 0	Questions 7
Time Limit None	Allowed Attempts Unlimited	

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	3 minutes	0 out of 0

Submitted Feb 13 at 8:16pm

Question 1

0 / 0 pts

Which of the following was written on top of the Apache Spark software?

☐ Python

☐ Scala

☐ Java

☒ GraphX

GraphX software was written on top of the Apache Spark.

Correct!

Question 2

0 / 0 pts

Which of the following big data software is implemented by Google to rank websites using their popular PageRank algorithm?

Correct!☒ GraphX

Correct! Google uses GraphX to rank websites.

☐ Spark SQL☐ MySQL☐ Oracle**Question 3****0 / 0 pts**

What is the method implemented by Apache Spark to process live streaming data?

☐ binary processing☒ On-demand processing

This is not correct. To process live streaming of data, Apache Spark does not use on-demand processing.

You Answered**Correct Answer**☐ Batch processing☐ Real time processing**Question 4****0 / 0 pts**

Which of the following is an example of live streaming data?

☒ A Twitter hashtag containing a company name.

Correct! Twitter hashtags are live streaming data because numerous users may tweet the same hashtag at a given time.

Correct!

- ☐ Student grades submitted by an instructor.
- ☐ An online banking statement for an individual.
- ☐ A Wikipedia article about a historical figure.

Question 5**0 / 0 pts**

During the processing of live streaming data by Apache Spark, what does each batch correspond to?

- ☐ Second
- ☐ Query
- ☐ Node
- ☒ RDD (Resilient Distributed Dataset)

Correct!

Correct! Apache Spark treats each and every batch as an RDD.

Question 6**0 / 0 pts**

How is spatial data different from traditional data?

- ☐ Spatial data represents simple spaces.
- ☒ Spatial data is tied to physical space.

Correct!

That is correct. For example, location of a house is determined by latitude and longitudinal locations.

- ☐ Spatial data has one dimension.

Question 7**0 / 0 pts**

What is the best definition of a KNN query?



A KNN query is a query that is nested inside a SQL statement and is embedded in the where clause.



A KNN query retrieves all records where a value is between an upper and lower boundary.



KNN query is the nearest neighbor of a given query point q to find k closest objects from q based on it's spatial distance.

That is correct. KNN query can be used to find the best restaurants near a particular airport.

Correct!