Knowledge Check 12: Spark and HDFS

Due No due date **Points** 0 **Questions** 10
Time Limit None **Allowed Attempts** Unlimited

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	4 minutes	0 out of 0

Submitted Feb 13 at 8:21pm

Question 1 — 0 / 0 pts

Which company has created the MapReduce framework as a concept?

- Oracle
- Google (Correct!)

  Google engineers created the MapReduce framework as a concept.
- Microsoft
- Amazon

Question 2 — 0 / 0 pts

Which company has implemented Hadoop an an open-source version of MapReduce?

Google

Amazon

Microsoft

Correct!

Yahoo

Yahoo has implemented Hadoop as an open-source version of MapReduce.

Question 3

0 / 0 pts

Which of the following is true about the Hadoop file system?

ou Answered

Each node stores distinct data blocks

This is not correct. Data blocks are replicated across several nodes.

rrect Answer

Files are append-only

Meta node stores metadata

Files split in to 1 GB blocks

Question 4

0 / 0 pts

What does HDFS stand for?

Correct!

Hadoop File System

HDFS stands for Hadoop File System.

Highly Distributed File System

- Highly Disturbed File System
- High Definition File System

Question 5

0 / 0 pts

What is the data type used by Hadoop for a MapReduce process?

- Document-based
- Column-based
- Graph-based

Correct!

- Key-value

Hadoop uses key-value data type for a MapReduce process.

Question 6

0 / 0 pts

What is the output of the Map function in a MapReduce process?

- List of table columns
- List of network nodes

Correct!

- List of key-value pairs

The Map function generates a list of key-value pairs.

- List of graph nodes

Question 7

0 / 0 pts

Where do mapper nodes save their outputs before serving to reducer nodes?

- Central node
- Master node

ou Answered

- Another node

  This is not correct. Mapper nodes save their outputs to local disk first.

rrect Answer

- Local Disk

Question 8

0 / 0 pts

What does Hadoop do with a task that crashes in a node?

- The task is failed.

rrect Answer

- The task is retried on another node.

ou Answered

- The node is rebooted.

  This is not correct. Hadoop does not reboot a node just because there is a single crash of a task.

- The node is shut down.

Question 9

0 / 0 pts

Apache Spark sorts its data processing operations, such as collect, filter, and sort, by building a graph called DAG. What does DAG stand for?

Distributed Asymmetric Graph

Derived Apache Graph

Correct!

Directed Acyclic Graph

DAG stands for Directed Acyclic Graph.

Distributed Apache Graph

Question 10

0 / 0 pts

Which of the following statements about the difference between Hadoop and Spark is true?

Hadoop provides multiple built-in data processing operations such as filter and join.

Hadoop supports in-memory cluster computing.

Hadoop is faster than Spark.

Correct!

Both Hadoop and Spark can load data from Hadoop File System (HDFS)

Hadoop and Spark are designed to load data from HDFS.