








## Review Test Submission: Spark Quiz 1

User	Venkata Kowsik Temididapathi
Course	CS 6350.002 - Big Data Management and Analytics - S22
Test	Spark Quiz 1
Started	2/26/22 5:06 PM
Submitted	2/27/22 5:53 PM
Due Date	2/27/22 11:59 PM
Status	Completed
Attempt Score	50 out of 50 points
Time Elapsed	24 hours, 46 minutes
Results Displayed	All Answers, Submitted Answers, Correct Answers

### Question 1

10 out of 10 points






Which of the following are properties of Apache Spark?

- Selected Answer:  Spark achieves higher efficiency than Hadoop MapReduce
- Answers:  Spark uses an abstraction for distributed memory called RDD and performs computations in-memory as far as possible.
- Answers:  Spark is suitable for interactive, real-time processing of multi-stage data
- Answers:  Spark achieves higher efficiency than Hadoop MapReduce
- Answers:  Spark uses an abstraction for distributed memory called RDD and performs computations in-memory as far as possible.
- Answers:  Spark is suitable for interactive, real-time processing of multi-stage data
- Answers:  Spark uses lots of disk reads and writes to persist data

### Question 2

10 out of 10 points

What is RDD?

- Selected Answer:  distributed memory abstraction that lets programmers perform in-memory computations in a fault tolerant manner.
- Answers:  combination of large blocks of disk that are spread across multiple nodes
- Answers:  A form of shared memory that is available within a single machine
- Answers:  distributed memory abstraction that lets programmers perform in-memory computations in a fault tolerant manner.
- Answers:  A form of shared memory that can be written and read simultaneously by multiple cores

### Question 3

10 out of 10 points

Why is Spark better than traditional Hadoop MapReduce for complex algorithms that require multi-stage processing?

Selected



Answers:

Hadoop MapReduce involves tedious programming and Spark provides an easier abstraction for iterative algorithms.



Hadoop MapReduce involves disk reads and writes between the various iterations and stages and this is inefficient.



Spark uses distributed memory known as RDD, which provides efficient and fault-tolerant abstraction for performing fast computation

Answers:



Hadoop MapReduce involves tedious programming and Spark provides an easier abstraction for iterative algorithms.



Hadoop MapReduce involves disk reads and writes between the various iterations and stages and this is inefficient.



Spark uses distributed memory known as RDD, which provides efficient and fault-tolerant abstraction for performing fast computation

Hadoop MapReduce is very fast and interactive.

### Question 4

10 out of 10 points

How does RDD achieve fault tolerance?

Selected



Answers:

By ensuring that only coarse grained transformations are allowed and by recovery through lineage tracking.

Answers:



By ensuring that only coarse grained transformations are allowed and by recovery through lineage tracking.

By replicating each RDD 3 times in memory

By allowing fine grained transformations to RDD and logging every transformation

By writing RDDs to disk and recovery by reading from disk.

### Question 5

10 out of 10 points

What are the ways in which RDD can be created?

Selected Answers:



By applying bulk transformations to files in stable storage or existing RDDs



By parallelizing existing collections

Answers:

By starting with a blank RDD and adding data to it



By applying bulk transformations to files in stable storage or existing RDDs



By parallelizing existing collections

RDDs are available in memory, we just need to fill them with data

Sunday, February 27, 2022 5:53:46 PM CST