✖

Spark Lesson 1

6 questions

1
point

1.

Apache Spark was developed in order to provide solutions to shortcomings of another project, and eventually replace it. What is the name of this project?

○ Hadoop

○ Pig

◉ MapReduce

○ HDFS

1
point

2.

Why is Hadoop MapReduce slow for iterative algorithms?

○ The Java Virtual Machine uses too much memory

○ Communication is a bottleneck

◉ It needs to read off disk for every iteration

○ Iterative algorithms do not scale well

1
point

3.

What is the most important feature of Apache Spark to speedup iterative algorithms?

○   Python interface

○   Resiliency to data loss

○   Caching datasets on disk

◉   Caching datasets in memory

---

1 point

4.

Which other Hadoop project can Spark rely to provision and manage the cluster of nodes?

○   Pig

○   HDFS

◉   YARN

○   MapReduce

---

1 point

5.

When Spark reads data out of HDFS, what is the process that interfaces directly with HDFS?

○   Cluster Manager

○   Driver

◉   Executor

○   YARN

<table>
<tr><td>1
point</td></tr>
</table>

6.

Under which circumstances is preferable to run Spark in Standalone mode instead of relying on YARN?

○ When we want to mix MapReduce and Spark jobs.

◉ When you only plan on running Spark jobs

○ For iterative algorithms

○ Never

Submit Quiz

👍 👎 🏳