More on Spark

9/9 points (100%)

Quiz, 9 questions

✔ **Congratulations! You passed!**

Next Item

✅ **1.** Which part of SPARK is in charge of creating RDDs?

1 / 1 points

○ Worker Node

◉ Driver Program

▲

Correct

○ Storage

○ Spark Executor

○ Local CPU

✅ **2.** How does lazy evaluation work in Spark?

1 / 1 points

○ Transformations are queued and executed at a certain threshold.

○ Actions are not executed until the transformation stage.

◉ Transformations are not executed until the action stage.

▲

Correct

○ Actions are queued and executed at a certain threshold.

✔ **3.** What are the consequences of lazy evaluation as mentioned in lecture?

1 / 1
points

○ There are no consequences.

○ Hiccups within the system during queue execution.

◉ Errors sometimes do not show up until the action stage.

Correct

✔ **4.** What is a wide transformation?

1 / 1
points

○ A longer time-taking transformation compared to narrow transformations.

◉ A transformation that requires data shuffling across node partitions.

Correct

○ The name for the most used transformations.

○ Transformations that take a lot of nodes to complete.

✔ **5.** Where does the data for each worker node get sent to after a collect function is called?

1 / 1 points

○ None; Stays in the Same Node

○ Other Worker Nodes

⦿ Spark Context

▲

Correct

○ Spark Streaming

○ Spark SQL

✔ **6.** What are DataFrames?

1 / 1 points

⦿ A column like data format that can be read by Spark SQL.

▲

Correct

○ A type of narrow transformation.

○ A special type of data node that contains framework to manipulate SQL.

✔ **7.** Can RDD's be converted into DataFrames directly without manipulation?

1 / 1 points

○ Yes

⦿ No: lines have to be converted into row.

▲

Correct

○ No: RDD's cannot be converted into DataFrames.

○ No: RDD's needed to be made relational first.

✔ 8. What is the function of Spark SQL as mentioned in lecture? (Choose 3)

1 / 1 points

☐ Better worker node interpolation.

Un-selected is correct

☐ Deploy business intelligence tools over Spark.

Correct

☐ Connect to variety of databases.

Correct

☐ Better ability to manipulate big data.

Un-selected is correct

☐ Enables relational queries on Spark.

Correct

☐ Efficient data manipulation using SQL like structure.

Un-selected is correct

✔

9. What is a triplet in GraphX?

1 / 1
points

   ◯ A type of data to contain vertex info.

   ◯ A type of data to contain both edge and vertex info.

   ◯ A type of data to contain edge info.

   ◉ A type of data to contain the information on connections between vertices and edges.

Correct

👍 👎 🏳