

Your grade: **85.71%**

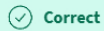
Your latest: **85.71%** • Your highest: **85.71%** • To pass you need at least 70%. We keep your highest score.

Next item →

1. Which of the following statements is true about a directed acyclic graph (DAG)? Select all that apply.

1 / 1 point

- ☐ A tabular data structure with rows and columns
- ☒ In Apache Spark, RDDs are represented by the vertices



Correct

Correct! In Apache Spark, RDDs are indeed represented by the vertices of a DAG, while directed edges represent transformations and actions.

- ☒ A data structure with edges and vertices



Correct

Correct! DAGs are graphical structures with edges and vertices.

- ☒ A new edge is obtained from an older vertex



Correct

Correct! Every new edge in a DAG is derived from an existing vertex, reflecting the directional relationship between elements.

2. Which function is applied to create a data set from a sequence?

1 / 1 point

- ☐ seqDS()
- ☒ toDS()
- ☐ DSRdd()
- ☐ Create()



Correct

Correct! You can utilize the 'toDS()' function to create a data set from a sequence in Apache Spark.

3. Which of the following is a feature of Tungsten?

1 / 1 point

- ☒ Utilizes CPU registers for storing intermediate data
- ☐ Depends on the JVM object model
- ☐ Enhances security by restricting developers from adding data source-specific rules
- ☐ Generates an optimized physical query plan from the logical query plan



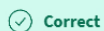
Correct

Correct! Tungsten leverages CPU registers to efficiently store and process intermediate data.

4. While adhering to best practices, in what order does a typical data engineer perform operations on Apache Spark?

1 / 1 point

- ☒ Read, analyze, transform, load, and write
- ☐ Read, analyze, load, transform, and write
- ☐ Analyze, read, transform, load, and write
- ☐ Analyze, read, load, transform, and write



Correct

Correct! Read, analyze, transform, load, and write is the order in which a typical data engineer performs Extract, Transform, and Load (ETL) operations on Apache Spark.

5. Which data sources can be utilized with Apache Spark SQL?

1 / 1 point

- ☐ MongoDB
- ☐ External APIs
- ☒ Parquet files
- ☐ Custom file formats



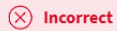
Correct

Correct! Spark SQL supports reading and writing data from Parquet files while preserving the data schema.

6. How can you create a Global Temporary view in Spark SQL?

0 / 1 point

- ☐ Use the createView function with a "Global" prefix
- ☐ Use the createGlobalTempView function
- ☒ Use the createGlobalView function
- ☐ Use the createTempView function with a "Global" prefix



Incorrect

Incorrect. Review the Real-World Usage of SparkSQL video.

- 7.
1. Spark creates a Directed Acyclic Graph (DAG) during the creation of a Resilient Distributed Dataset (RDD).
 2. The DAG is associated with the new RDD.
 3. If there is an action, the driver program, which invokes calls the action, evaluates the DAG after Spark completes the action.
 4. The pointer responsible for transforming the RDD returns to the Spark driver program.
 5. Spark utilizes the DAG Scheduler to perform a transformation and updates the DAG accordingly.

1 / 1 point

Select the sequence that explains RDD transformation and actions.

- ☐ 1 >> 2 >> 3 >> 4 >> 5
- ☒ 1 >> 5 >> 2 >> 4 >> 3
- ☐ 2 >> 3 >> 4 >> 1 >> 5
- ☐ 1 >> 3 >> 4 >> 5 >> 2



Correct

Correct! This process explains RDD transformation and actions.

Spark creates the DAG when creating an RDD. It then enables the DAG Scheduler to perform a transformation and update the DAG. The DAG points to the new RDD. The pointer that transforms RDD is returned to the Spark driver program. If there is an action, the driver program that calls the action evaluates the DAG only after Spark completes the action.