Your grade: 71.42%

Your latest: **71.42%** • Your highest: **71.42%** • To pass you need at least 70%. We keep your highest score.

Next item →

---

1. What are the three main components of Apache Spark architecture?   **1 / 1 point**

   ○ Mesos; YARN; Kubernetes

   ○ Scala; Java; Python

   ◉ Data; compute interface; resource management

   ○ Storage; HDFS; Python

   ✓ **Correct**
   Correct! The three main components of Apache Spark architecture are data, compute interface, and resource management.

---

2. What are DataFrames in Apache Spark?   **0 / 1 point**

   ○ DataFrames are a distributed collection of data organized into named columns.

   ◉ DataFrames are Spark's built-in machine learning models for predictive analytics.

   ○ DataFrames is a data format for storing graph data structures in Spark.

   ○ DataFrames is a distributed file system in Spark used for storing large data sets efficiently.

   ⊗ **Incorrect**
   Incorrect. Review the Spark SQL and DataFrames video.

---

3. What is Apache Spark?   **1 / 1 point**

   ◉ In-memory framework for distributed data processing

   ○ Hardware manufacturer

   ○ Closed-source data analysis tool

   ○ Cloud storage service

   ✓ **Correct**
   Correct! Spark is an open-source, in-memory application framework for distributed data processing and iterative analysis on massive data volumes.

---

4. What is functional programming?   **1 / 1 point**

   ○ A programming method that prioritizes procedural programming over the use of mathematical functions

   ○ A programming approach that focuses solely on graphical functions and visual designs

   ○ A programming approach that emphasizes the *how to* of the solution as opposed to the *what* of the solution

   ◉ A style of programming that follows the mathematical function format

   ✓ **Correct**
   Correct! In functional programming, the code or program emphasizes the *what* of the solution instead of the *how to* of the solution.

---

5. Which of the following statements defines Resilient Distributed Datasets (RDDs)? Select all that apply.   **1 / 1 point**

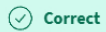   ☐ RDD is a distributed database management system.

   ☑ RDDs are immutable.

☑ RDD is a collection of fault-tolerant elements.

☑ RDD is capable of receiving parallel operations.

6. What is the primary purpose of parallel programming? **1 / 1 point**

○ To break a problem into discrete parts that can be solved sequentially

○ To employ specific control and coordination mechanism

○ To run noncontemporary instructions

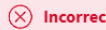◉ To use multiple compute resources to solve a computational problem

7. Which of the following is a benefit of DataFrames? **0 / 1 point**

○ To scale from kilobytes of data on a single laptop to petabytes on a large cluster

◉ To scale from kilobytes of data on multiple laptops to petabytes on a large cluster

○ Supports specific data formats and storage systems

○ To scale small-scale data on a laptop