Apache Spark and Scala Certification Course Agenda

Lesson 1: Course Preview

- Course overview
- Objectives

Lesson 2: Introduction to Spark

- Limitations of MapReduce in Hadoop Objectives
- Batch vs. Real-time analytics
- Application of stream processing
- How to install Spark
- Spark vs. Hadoop Eco-system

Lesson 3: Introduction to Programming in Scala

- Features of Scala
- Basic data types and literals used
- List the operators and methods used in Scala
- Concepts of Scala

Lesson 4: Using RDD for Creating Applications in Spark

- Features of RDDs
- How to create RDDs

- RDD operations and methods
- How to run a Spark project with SBT
- Explain RDD functions and describe how to write different codes in Scala

Lesson 5: Running SQL queries Using SparkSQL

- Explain the importance and features of SparkSQL
- Describe methods to convert RDDs to DataFrames
- Explain concepts of SparkSQL
- Describe the concept of hive integration

Lesson 6: Spark Streaming

- Explain a concepts of Spark Streaming
- Describe basic and advanced sources
- Explain how stateful operations work
- Explain window and join operations

Lesson 7: Spark ML Programming

- Explain the use cases and techniques of Machine Learning (ML)
- Describe the key concepts of Spark ML
- Explain the concept of an ML Dataset, and ML algorithm, model selection via cross validation

Lesson 8: Spark GraphX Programming

- Explain the key concepts of Spark GraphX programming
- Limitations of the Graph Parallel system

- Describe the operations with a graph
- Graph system optimizations

For information on the course, visit: http://www.simplilearn.com/big-data-and-analytics/apache-spark-scala-certification-training