Spark Programming with Scala

Course 3579 – 24 Hours

Overview

Apache Spark™ is a fast and general engine for large-scale data processing, and arguably the first open source software that makes distributed programming truly accessible to data scientists.
Using Apache Spark™ you can write applications quickly using Java, Scala, Python, and R.
This course covers all the fundamentals about Apache Spark with Scala and guides you through everything you need to know about developing Spark applications with Scala.
At the end of this course, you will gain deep understanding about the Spark architecture and general big data analysis and manipulations skills.

Who Should Attend

- Java or Scala developers who would like to start using Spark

Prerequisites

- Java/Scala Experience

Course Contents

Just Enough Scala
- Basic Syntax
- Case Classes
- High Order Functions and Collections
- Pattern Matching

Spark Core
- What is Spark?
- RDD Basics.
- Transformations and Actions.
- Chaining Transformations
- Using Anonymous Functions
- Map Reduce and Shuffles.
- Caching.
- Web Monitoring.
- Usecases.
- Serialization.
- Troubleshooting.

Spark Streaming
- Introduction to Spark Streaming.
- DStream.
- Usecases.

Spark SQL
- Introduction to Spark SQL & DataFrames.
- Integration with Different Data Sources.
- Integration with Hive.

Spark MLlib
- Introduction to Machine Learning.
- The MLlib API.
- Basic Usecases.

Spark GraphX
- Introduction to GraphX.
- The GraphX API.
- Usecases.