

Hadoop

Spark Scala

Course Content



1. 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

2. 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

3. 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

4. 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

5. Spark Streaming

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

6. 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

7. 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



Visit: www.Webtrackker.com
E-mail: info@webtrackker.com

Noida Office

Webtrackker Technology (P)Ltd
c-67 Sector 63 Noida, India

Other Office

**F -1 Sector 3 (Near Sector
16 metro station) Noida, India.**