

## Apache Spark with Scala Course Content

### **SCALA (Object Oriented and Functional Programming)**

- Getting started With Scala.
- Scala Background, Scala Vs Java and Basics.
- Interactive Scala – REPL, data types, variables, expressions, simple functions.
- Running the program with Scala Compiler.
- Explore the type lattice and use type inference
- Define Methods and Pattern Matching.

### **Functional Programming.**

- What is Functional Programming?
- Differences between OOPS and FPP.

### **Collections (Very Important for Spark)**

- Iterating, mapping, filtering and counting
- Regular expressions and matching with them.
- Maps, Sets, group By, Options, flatten, flat Map
- Word count, IO operations, file access, flat Map

### **Introduction**

- Why second generation frameworks?
- Introduction to Spark

### **SPARK CORE.**

- Batch versus real-time data processing
- Introduction to Spark, Spark versus Hadoop
- Architecture of Spark.
- Coding Spark jobs in Scala

- Exploring the Spark shell -> Creating Spark Context.
- RDD Programming
- Operations on RDD.
- Transformations
- Actions
- Loading Data and Saving Data.
- Key Value Pair RDD.
- Broad cast variables.

#### **SPARK STREAMING**

- Introduction of Spark Streaming.
- Architecture of Spark Streaming
- Processing Distributed Log Files in Real Time
- Discretized streams RDD.
- Applying Transformations and Actions on Streaming Data
- Integration with Flume and Kafka.
- Integration with Hbase
- Monitoring streaming jobs.

#### **SPARK SQL**

- Introduction to Apache Spark SQL
- The SQL context
- Importing and saving data
- Processing the Text files,JSON and Parquet Files
- Data Frames
- user-defined functions
- Using Hive



- Local Hive Metastore server

#### **Persistence**

- Configuring and running the Spark cluster.
- Exploring to Multi Node Spark Cluster.
- Cluster management
- Submitting Spark jobs and running in the cluster mode.
- Developing Spark applications in Eclipse
- Tuning and Debugging Spark.

#### **Spark MLib(Intro)**

#### **Spark giraph(Intro)**

eBotix Technologies