

Spark with Scala Developer



IT EDUCATION CENTRE

Hadoop Overview

1. Need of Hadoop technology
2. Hadoop Cluster and Racks in detail
3. Overview of Map Reduce
4. Big data Concepts and data types
5. Concept of Streaming data and different tools utilization
6. HDFS and Basic hadoop commands

Scala Programming

1. Scala overview and Environment Setup
2. OOps concepts in scala
3. Control Structure and Functions
4. Closures and Collections
5. Exception Handling in scala

Apache Spark 2 x Installation

1. Download release and set
2. Working with eclipse
3. Installing Scala IDE with spark
4. Testing with different OS



IT EDUCATION CENTRE

Working with Apache Spark

1. RDD and its Transformations
2. Working with Eclipse Maven, Spark context and RDD
3. Working with different file formats
4. Introduction to Spark DataFrame
5. Data Frames and RDD's with with Spark 1.x and 2.x style
6. Creating Multiple Spark Context and Spark Sessions
7. Applying Own Schema to the DataFrame and basic operations
8. Creating Datasets and its basic operations
9. Dataset vs DataFrame Performance
10. Running Spark Job in Yarn/cluster Mode From IDE
11. Spark with Mysql, transformations On MySQL Table-DataFrame API
12. Query Push Down to MySQL Database
13. Creating Partitioned Table with Spark
14. Spark built-in functions and UDF
15. Examples with spark Sql and RDD's
16. Spark job submit

Spark Streaming

1. Working with data stream
2. Example of network
3. Twitter data stream
4. Twitter data analysis cases



IT EDUCATION CENTRE

Kafka

1. Fundamentals of kafka, Work Flow and Basic Operations
2. Creating Topics, Partition, Replication, Broker and Kafka cluster
3. Working with Producer and Consumer Examples
4. Creating Consumer Group, Leaders, Followers
5. Starting brokers, Listing and modifying topics
6. Single Node-Multiple Brokers Configuration and Basic Topic Operations
7. Creating Producer, Consumer and Consumer group application
8. Running a jar files from terminal

Real time case studies:

1. Working on different data sets
2. Working on nse live data

SparkMlib

1. Classification algorithm
2. Clustering algorithm
3. Sequence Mining algorithm

Spark Graphx

1. Graph analysis with Spark
2. Graphx for graphs
3. Graph parallel computation



IT EDUCATION CENTRE