# www.uplatz.com
**Leading Marketplace for IT and Certification Courses**

## SPARK DEVELOPMENT TRAINING CURRICULUM – 40 HRS

## Prerequisite: SCALA for Spark

## Scala Basics

- What is Scala?
- Why Scala for Spark?
- Intro to Scala REPL : Journey from Java to Scala
- Installing Scala IDE
- Basic Operations
- Defining Functions

## Scala Essentials

- Control Structures in Scala
- loops – ForEach, While, Do-While
- Collections – Array, ArrayBuffer, Map, Tuples, Lists
- If Statements
- Conditional Operators
- Enumerations

## OOP's and FP

- Class and Object Basics
- Scala Constructors
- Nested Classes
- Visibility Rules
- Overriding Methods
- Functional Programming
- Higher Order Functions
- Traits
  - Interfaces
  - Layered Traits

## Prerequisite: BigData and Hadoop Framework

- Introduction to BigData
- Challenges with Bigdata
- Batch Vs. Realtime processing
- Overview- Hadoop Ecosystem
  - HDFS
  - Review of MapReduce
  - Hive
  - Sqoop
  - Flume

# www.uplatz.com
## Leading Marketplace for IT and Certification Courses

## APACHE SPARK

## Introduction to Spark
- ➢ What is Spark?
- ➢ Spark Overview
- ➢ Setting up environment
- ➢ Using Spark Shell
- ➢ Spark Web UI

## Spark Basics

- ➢ RDD's
- ➢ Spark Context
- ➢ Spark Ecosystem
- ➢ In-Memory data – Spark

## Working with RDD's

- ➢ Creating, Loading and Saving RDD
- ➢ Transformations in RDD
- ➢ Actions in RDD
- ➢ Key-Value Pair RDD
- ➢ MapReduce and Pair RDD operations
- ➢ RDD Partitions

## Writing and Deploying Spark Applications

- ➢ Spark Applications vs. Spark Shell
- ➢ Creating Spark Context
- ➢ Building a Spark Application
- ➢ Running a Spark Application
- ➢ Spark and Hadoop Integration-HDFS
- ➢ Handling Sequence Files

## Spark RDD

- ➢ RDD Lineage
- ➢ RDD Persistence Overview
- ➢ Distributed Persistence

## Spark Streaming

- Spark Streaming Architecture
- First Spark Streaming Programming
- Transformations in Spark Streaming

## Spark MLlib

1. What is Machine Learning?
2. ML library for Spark
3. Algorithms
   - Statistics
   - Classification
   - Regression
   - Clustering
   - Collaborative Filtering

## Spark SQL

- Overview on Hive
- Spark SQL Architecture
- SQLContext in Spark SQL
- Working with DataFrames
- Example for Spark SQL
- Integrating Hive and Spark SQL
- DataFrames and RDD's
- Knowing JSON and Parquet File Formats
- Loading of data
- Comparing Spark SQL,Impala and Hive-on-Spark

## GraphX

- Overview of GraphX
- Data Visualisation in Spark
- Common Spark use-cases

## Performance Tuning

- Shared Variables: Broadcast Variables
- Shared Variables: Accumulators
- Common Performance Issues
- Performance tuning tips

## Course Deliverables

- Workshop style coaching
- Interactive approach
- Course material

# www.uplatz.com
## Leading Marketplace for IT and Certification Courses

- POC Implementation
- Hands on practice exercises for each topic
- Quiz at the end of each major topic
- Tips and techniques on Cloudera Certification Examination
- Linux concepts and basic commands
- On Demand Services
- Mock interviews for each individual will be conducted on need basis
- SQL basics on need basis
- Resume preparation and guidance
- Interview questions

- ❖ **Interested to learn? Get in touch with us**
- ❖ **Phone: +44 7836 212635**
- ❖ **Whats App: + 44 7836 212635**
- ❖ **Email: info@uplatz.com**
- ❖ **Website: www.uplatz.com**
- ❖ **Browse wwww.uplatz.com to see more courses.**
- ❖ **Uplatz.com is a Learning Marketplace for IT and Certification courses.**