## Why Scala

## Scala Installation

## Get deep insights into the functioning of Scala

## Execute Pattern Matching in Scala

## Functional Programming in Scala – Closures, Currying, Expressions, Anonymous Functions

## Know the concepts of classes in Scala

## Object Orientation in Scala – Primary, Auxiliary Constructors, Singleton & Companion Objects

## Traits and Abstract classes in Scala

## Scala Simple Build Tool – SBT

## Building with Maven

## Spark Basics

## What is Apache Spark?

## Spark Installation

## Spark Configuration

## Spark Context

## Using Spark Shell

## Resilient Distributed Datasets (RDDs) – Features, Partitions, Tuning Parallelism

## Functional Programming with Spark

## Working with RDDs

## RDD Operations - Transformations and Actions

## Types of RDDs

## Key-Value Pair RDDs – Transformations and Actions

## MapReduce and Pair RDD Operations

## Serialization

## Spark on a cluster

## Overview

## A Spark Standalone Cluster

## The Spark Standalone Web UI

## Executors & Cluster Manager

## Spark on YARN Framework

## Writing Spark Applications

## Spark Applications vs. Spark Shell

## Creating the SparkContext

## Configuring Spark Properties

## Building and Running a Spark Application

## Logging

## Spark Job Anatomy

## Caching and Persistence

## RDD Lineage

## Caching Overview

## Distributed Persistence

## Improving Spark Performance

## Shared Variables: Broadcast Variables

## Shared Variables: Accumulators

## Per Partition Processing

## Common Performance Issues

## Spark API for different File Formats & Compression Codecs

## Text

## CSV

## Sequence

## Parquet

## ORC

## Compression Techniques – Snappy, Zlib, Gzip

## Spark SQL

## Spark SQL Overview

## HiveContext

## SQL Datatypes

## Dataframes vs RDDs

## Operations on DFs

## Parquet Files with Spark Sql – Read, Write, Partitioning, Merging Schema

## ORC Files

## JSON Files

## Inferring Schema programmatically

## Custom Case Classes

## Temp Tables vs Persistent Tables

## Writing UDFs

## Hive Support

## JDBC Support - Examples

## HBase Support - Examples

## Spark Streaming

## Spark Streaming Overview

## Example: Streaming Word Count

## Other Streaming Operations

## Sliding Window Operations

## Developing Spark Streaming Applications – Integration with Kafka and Hbase