

Hadoop and Spark with Scala



DELTA**FROG**
Technology

1) Introduction to Big Data Hadoop and Spark

Class Objectives: Understand Big Data and its components such as HDFS. You will learn about the Hadoop Cluster Architecture and you will also get an introduction to Spark and you will get to know about the difference between batch processing and real-time processing.

Topics:

- .What is Big Data?
- .Big Data Customer Scenarios
- .How Hadoop Solves the Big Data Problem?
- .What is Hadoop?
- .Hadoop's Key Characteristics
- .Hadoop Ecosystem and HDFS
- .Hadoop Core Components
- .Rack Awareness and Block Replication
- .YARN and its Advantage
- .Hadoop Cluster and its Architecture
- .Hadoop: Different Cluster Modes
- .Big Data Analytics with Batch & Real-time Processing
- .Why Spark is needed?
- .What is Spark?
- .How Spark differs from other frameworks?

2)Introduction to Scala for Apache Spark

Class Objectives: Learn the basics of Scala that are required for programming Spark applications.

Topics:

- .What is Scala?
- .Why Scala for Spark?
- .Scala in other Frameworks
- .Introduction to Scala REPL
- .Basic Scala Operations
- .Variable Types in Scala
- .Control Structures in Scala
- .Foreach loop, Functions and Procedures
- .Collections in Scala- Array
- .ArrayBuffer, Map, Tuples, Lists, and more
- .Scala REPL with example.

3)Functional Programming and OOPs Concepts in Scala

Class Objectives: In this module, you will learn about object-oriented programming and functional programming techniques in Scala.

Topics:

- .Functional Programming
- .Higher Order Functions
- .Anonymous Functions
- .Class in Scala
- .Getters and Setters
- .Custom Getters and Setters
- .Properties with only Getters
- .Auxiliary Constructor and Primary Constructor
- .Singletons
- .Extending a Class
- .Overriding Methods
- .Traits as Interfaces and Layered Traits

4)Introduction to Apache Spark Framework

Class Objectives: Understand Apache Spark and learn how to develop Spark applications. At the end, you will learn how to perform data ingestion using Sqoop.

Topics:

Spark Components & its Architecture

Spark Deployment Modes

Introduction to Spark Shell

Writing your first Spark Job Using SBT

Submitting Spark Job

Spark Web UI

Data Ingestion using Sqoop

5)Understanding Spark RDDs

Class Objectives: Deep understanding of Spark - RDDs and Various operation of RDD.
(Transformations, Actions and Functions performed on RDD).

Topics:

.What is RDD, It's Operations, Transformations & Actions

.Data Loading and Saving Through RDDs

.Key-Value Pair RDDs

.Other Pair RDDs, Two Pair RDDs

.RDD Lineage

.RDD Persistence

.WordCount Program Using RDD Concepts

.RDD Partitioning & How It Helps Achieve Parallelization

.Passing Functions to Spark

6)DataFrames and Spark SQL

Class Objectives: In this Class, you will learn about SparkSQL which is used to process structured data with SQL queries, data-frames and datasets in Spark SQL along with different kind of SQL operations performed on the data-frames. You will also learn about the Spark and Hive integration.

Topics:

- .Need for Spark SQL
- .What is Spark SQL?
- .Spark SQL Architecture
- .SQL Context in Spark SQL
- .User Defined Functions
- .Data Frames & Datasets
- .Interoperating with RDDs
- .JSON and Parquet File Formats
- .Loading Data through Different Sources
- .Spark — Hive Integration

7)Understanding Apache Kafka.

Class Objectives: Understand Kafka and its Architecture. Also, learn about Kafka Cluster, how to configure different types of Kafka Cluster.

Topics:

- .Need for Kafka
- .What is Kafka?
- .Core Concepts of Kafka
- .Kafka Architecture
- .Understanding the Components of Kafka Cluster
- .Configuring Kafka Cluster
- .Kafka Producer and Consumer Java API

8)Apache Spark Streaming - Processing Multiple Batches

Class Objectives: Work on Spark streaming which is used to build scalable fault-tolerant streaming applications. Also, learn about DStreams and various Transformations performed on the streaming data.

Topics:

- .Drawbacks in Existing Computing Methods
- .Why Streaming is Necessary?
- .What is Spark Streaming?
- .Spark Streaming Features
- .Spark Streaming Workflow

.Streaming Context & DStreams

.Transformations on DStreams

9)Integration of Kafka and Spark Streaming.

Class Objectives: -In this class, you will learn Integration of Kafka and Spark Streaming

Hands on demo:-

We will consume data from kafka topic with the help of spark streaming.

We will store data into HDFS after that will process it with the help of spark-sql.

10)RealTime end to end project Demo