

Hadoop, scala, spark Training



Apache Hadoop is a collection of open-source software utilities that facilitates using a network of many computers to solve problems involving massive amounts of data and computation.

Scala is a strong statically typed general-purpose programming language which supports both object-oriented programming and functional programming.

Apache Spark is an open-source unified analytics engine for large-scale data processing.

Table of Content:

Hadoop Introduction

- Big data intro
- Hadoop Intro
- Features

Hadoop Architecture

- Servers
- Daemon And Name node
- Job Tracker
- Load Balance
- Namenode
- Yarn

MapReduce

- Mapper
- Reducer
- Mapper Only Functionality
- Reducer Req

Spark

- Introduction & Advantages
- Spark Components
- Architecture
- Rdd, Caching And Persistence

Spark Program

- Create Rdd And Aggregations
- Aggregation
- Multiple Aggregation
- Merging And Cross Joins

Scala

- Introduction
- Map Collection
- Transformation Over Collections

- Perform Transformations
- Function, Case Classes

