

Introduction to Hadoop Development

Description

Hadoop training “Introduction to Hadoop Development” is designed for professionals who want to understand and master the fundamentals of Big Data. In today’s context, where data has become a strategic asset, knowing how to process massive volumes of information is essential. This Hadoop course offers a clear, step-by-step, and practical approach to entering the world of Hadoop and its key components.

A Hadoop training tailored to Big Data needs

The program covers the complete Hadoop ecosystem. Participants will discover the role of HDFS and MapReduce, and then learn how to use essential tools such as Pig, Hive, Flume, Sqoop, and YARN. The goal is to provide a practical and operational perspective, helping learners understand how to set up and manage an efficient Hadoop environment.

Course Content

Module 1: Introduction to Big Data & Hadoop and its Ecosystem, Map Reduce and HDFS Module 2: Hadoop Installation & setup Module 3: Deep Dive in Mapreduce Module 4: Graph Problem Solving Module 5: Detailed understanding of Pig Module 6: Detailed understanding of Hive Module 7: (AVRO) Data Formats Module 8: Introduction to Hbase architecture Module 9: Hadoop Cluster Setup and Running Map Reduce Jobs Module 10: Advance Mapreduce

Lab / Exercises

- Working with HDFS
- Writing WordCount Program
- Writing custom partitioner
- Mapreduce with Combiner
- Map Side Join
- Reduce Side Joins
- Unit Testing Mapreduce
- Running Mapreduce in Local Job Runner Mode

Documentation

- Digital courseware included

Participant profiles

- Software developers, analytics, BI, ETL, and data warehousing professionals
- Big Data Hadoop developers, architects and testing personnel

Prerequisites

- No prerequisites

Objectives

- Learn the Hadoop Architecture and Hadoop basics for beginners
- Learn what is Hadoop, HDFS and MapReduce framework

- Write MapReduce programs and deploy Hadoop clusters
- Develop applications for Big Data using Hadoop Technology
- Develop YARN programs on the Hadoop 2.X version
- Work on Big Data analytics using Hive, Pig and YARN
- Integrate MapReduce and HBase to do advanced usage and Indexing
- Learn fundamentals of Spark framework and its working
- Understand RDD in Apache Spark
- Learn Hadoop development best practices
- Job scheduling using Oozie

Description

Hadoop Training

Niveau

Fondamental

Classroom Registration Price (CHF)

2300

Virtual Classroom Registration Price (CHF)

2150

Duration (in Days)

3

Reference

HDP-01