

Hadoop and MapReduce Fundamentals

Learn the basics of Hadoop and MapReduce in this comprehensive training course.

Category: Big Data · Duration: 2 days

WHO IT IS FOR

- Data analysts and data scientists
- Developers of any background
- Beginners welcome: no prior experience required

This training program provides a deep understanding of Hadoop and MapReduce, exploring key concepts and techniques for effective data processing. Participants will gain hands-on experience in setting up Hadoop clusters, writing MapReduce programs, and optimizing data processing workflows. By the end of this course, participants will be equipped with the necessary skills to harness the power of Hadoop and MapReduce for big data analysis.

Prerequisites

Basic programming knowledge and familiarity with Linux command line.

Syllabus

- Introduction to Big Data and Hadoop
- HDFS – Hadoop Distributed File System
- MapReduce Concepts and Principles
- Writing MapReduce Programs in Java
- Advanced MapReduce Techniques
- Hadoop Cluster Setup and Configuration
- Hadoop Data Ingestion
- Data Processing with Hive
- Data Processing with Pig
- Working with HBase
- Introduction to Apache Spark
- Optimizing MapReduce Jobs

- Monitoring and Debugging Hadoop Applications
- Securing Hadoop Clusters
- Integrating Hadoop with Existing Systems
- Best Practices for Hadoop Development
- Real-world Use Cases and Case Studies
- Performance Tuning and Optimization
- Introduction to YARN – Yet Another Resource Negotiator
- Big Data Analytics and Machine Learning with Hadoop