

Hadoop for Developers

Course Summary

Description

Apache Hadoop is the most popular framework for processing Big Data on clusters of servers. This course will introduce a developer to the Hadoop ecosystem.

Objectives

At the end of this course, students will be able to:

- Hadoop & Big Data
- HDFS
- YARN
- Hive
- HBase

Topics

- Introduction to Hadoop
- HDFS
- YARN
- Hive
- HBase

Audience

This course is designed for developers.

Prerequisites

- Developer background
- Comfortable with SQL and Java programming language (HBase labs are in Java)
- Comfortable in a Linux environment

Duration

Three days

Hadoop for Developers

Course Outline

I. *Introduction to Hadoop*

- A. Hadoop Ecosystem
- B. Hadoop Distributions
- C. High-level architecture
- D. Hardware and software
- E. Lab: first look at Hadoop

II. *HDFS*

- A. Design and architecture
- B. Concepts (horizontal scaling, replication, data locality, rack awareness)
- C. Daemons: Namenode, Secondary Namenode, Datanode
- D. Communications and heart-beats
- E. Data integrity
- F. Read and write path
- G. Namenode High Availability (HA), Federation
- H. Labs: Interacting with HDFS

III. *YARN*

- A. YARN Concepts and architecture
- B. Resource Manager, Node Manager
- C. Writing YARN applications
- D. Labs: Running a sample YARN program

IV. *Hive*

- A. Architecture and design
- B. Hive Data types
- C. HQL
- D. Creating Hive tables and querying
- E. Partitions
- F. Joins
- G. Text processing
- H. Labs: various labs on processing data with Hive

V. *HBase*

- A. Concepts and architecture
- B. HBase vs RDBMS vs Cassandra
- C. HBase Java API
- D. Time series data on HBase
- E. Schema design
- F. Labs: Interacting with HBase using shell; programming in HBase Java API ; Schema design exercise
- G. Browse our courses
- H. We offer instructor-led courses (onsite and online)
- I. Looking for team training?
- J. Up-skill your team with a customized, private training
- K. Public Classes
- L. Suitable for small teams and individuals