

Hadoop Fundamentals for Developers

This fast-paced, one-day instructor-led workshop takes you on a deep dive into scripting, programming, and coding in Hadoop. MetaScale provides a vendor-agnostic view when training on Pig, Hive, HBase, MapReduce, and Flume. The course has been designed by MetaScale's practicing professionals who have deployed Hadoop technologies at data-intensive enterprises. All students will receive a copy of the instructor's deck, lab deck, a Hadoop cheat sheet, and a copy of the VM software on a USB to take home. If you have a passion for programming and are looking to learn Hadoop, this is the course for you. This course allows you to develop the fundamental skills needed to program in Hadoop.

Audience

- IT Professionals: Managers, System Engineers, Developers, Statisticians, Data Analysts, Data Scientists, Computer Programmers

Prerequisites

- Programming experience required
- Basic knowledge of Linux/UNIX
- Completion of Hadoop 101 course, or equivalent experience

Course Length

- One day

By the end of the workshop, you will:

- Have a good understanding of MapReduce, HBase, Hive, Pig, and Flume in a Hadoop Ecosystem
- Complete ½ day of hands-on training in a Hadoop Ecosystem
- Build a working Hadoop script, start to finish, using raw files in a VM Cluster
- Write a Query using Hive DDL
- Write and convert Pig scripts into MapReduce jobs to aggregate data and get outputs
- View and analyze new outputs in both Hive and HBase through a UI analytics tool

Teaching Methods

- Lecture
 - Exercises
 - Demonstrations
 - Hands-on training in a Hadoop environment
-

Course Outline

Hadoop and Ecosystem Overview

- Introduction to Big Data
- Why use Hadoop?
- Hadoop Architecture Overview
- Hadoop Core Concepts
 - HDFS
 - MapReduce
 - Pig
 - Hive
 - Flume

HDFS

- Introduction
- HDFS Concepts
- HDFS Commands

Map/Reduce

- M / R Paradigm
- M / R Phases Overview & Architecture
- M / R Coding

Hive

Pig

- Introduction
- Pig Architecture
- Pig Latin examples
- Commands to Manipulate Data
- Examples

Flume

Hands-On Labs (4 hours)

- Building a functioning Hadoop script, start to finish, using raw files in a VM Cluster