

# Hadoop 101: Begin Your Big Data Journey

This full day instructor-led workshop gives both seasoned technical professionals and beginners an opportunity to understand Big Data and Hadoop. MetaScale offers a vendor-agnostic view of the Hadoop Ecosystem and introduces you to the key technologies that make up a Hadoop environment. All students will perform hands-on activities in a VM cluster, gaining firsthand experience working in a Hadoop environment. 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 Linux/UNIX cheat sheet and a copy of the VM software on a USB drive to take home. If you are ready to take a practical look at Hadoop, we strongly encourage you to attend this comprehensive one-day course and begin on your Big Data Journey.

---

## Audience

- IT Professionals: Managers, System Engineers, Developers, Statisticians, Data Analysts, Data Scientists
- C-levels
- Business Managers
- Educators

## Prerequisites

- None

## Course Length

- One day

## By the end of the workshop, you will:

- Understand "What is Hadoop" and the "History of Hadoop"
- Understand when to use "Hadoop" and when not to use "Hadoop"
- Develop a high-level understanding of the following Hadoop components: Hive with MySQL, Pig, MapReduce, HDFS, Sqoop, Flume, HBase, Oozie, and ZooKeeper
- Perform basic hands-on tasks using Pig, Hive, HBase, and HDFS in a customized VM cluster.

## Teaching Methods

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

## Course Outline

### History of Hadoop & Big Data Overview

- An Internet Minute
- Hadoop history
- What is Big Data?

### What is Hadoop?

- Introduction
- Features
- Hadoop 1.0 vs. Hadoop 2.0
- What Hadoop is not...
- Using the right tool for the right job

### How Hadoop works

- Throughput in Hadoop
- File extraction and distribution

### Hadoop Components (High Level Introduction)

- Overview
- Hadoop Ecosystem
- HDFS
- MapReduce
  - Mapper vs. Reducer
  - How Map Reduce works
- Flume
- Hadoop Streaming
- HBase
- Hive

- Hue
- Pig
- Oozie
- Sqoop
- ZooKeeper

### Non-Hadoop Components

- Overview
- Monitoring Tools
- BI Tools
- Analytic tools
- Scripting Languages

### Hands-On Labs (2 hours)

- Hive
- Pig
- HBase
- HDFS