

Hadoop Essential Training for Administrators and Developers

Description

Big data is here! In this Hadoop Essential Training for Administrators and Developers class, students will learn the fundamentals of setting up a Hadoop cluster as well as the “soup” of related technologies like Hive and Pig. Come prepared to learn how to access the Hadoop file system, write MapReduce jobs using java, Pig, and Hive, as well as how to use Pig, Hive, and Oozie. Every participant will work with their own installation of a Hadoop 2, single node cluster.

Delegates will learn

- an understanding of the Hadoop File System (HDFS).
- what MapReduce is and why you should care.
- how to write a MapReduce job with java, Pig, and Hive.
- how the different Hadoop technologies interoperate to provide a cohesive big data solution.
- basic management of a Hadoop cluster.
- how to perform basic unit testing of your MapReduce jobs.
- the different modes that Hadoop can be run in to support massive amounts of data as well as your MapReduce jobs during development.

Outline

Hadoop Overview

HDFS

HBASE

Map Reduce on YARN

Hadoop Streaming

MapReduce Workflows

Oozie

Pig

Hive