



MindsMapped Consulting – Online Big Data and Hadoop Training

Big Data & Hadoop Training

Pre-requisites: Core Java/Oracle, Basics of Unix

Duration: 50 Hours(15 Session)

Call us at (+1) (435) 610-1777 / (801) 901-3010 info@mindsmapped.com

Session 1: Introduction and Overview of Hadoop

- ✚ What is Hadoop?
- ✚ History of Hadoop.
- ✚ Building Blocks - Hadoop Eco-System.
- ✚ Who is behind Hadoop?
- ✚ What Hadoop is good for and what it is not?

Session 2: Hadoop Distributed File System (HDFS)

- ✚ HDFS Overview and Architecture
- ✚ HDFS Installation
- ✚ HDFS Use Cases
- ✚ Hadoop File System Shell
- ✚ File System Java API
- ✚ Hadoop Configuration

Session 3: HBase – The Hadoop Database

- ✚ HBase Overview and Architecture
- ✚ HBase Installation
- ✚ HBase Shell
- ✚ Java Client API
- ✚ Java Administrative API
- ✚ Filters
- ✚ Scan Caching and Batching
- ✚ Key Design
- ✚ Table Design

Session 4: Map/Reduce 2.0/YARN

- ✚ Decomposing Problems into MapReduce Workflow
- ✚ Using JobControl
- ✚ Oozie Introduction and Architecture
- ✚ Oozie Installation
- ✚ Developing, deploying, and Executing Oozie Workflows

Session 5: Pig

- ✚ Pig Overview
- ✚ Installation
- ✚ Pig Latin
- ✚ Developing Pig Scripts
- ✚ Processing Big Data with Pig
- ✚ Joining data-sets with Pig

Session 6: Hive

- ✚ Hive Overview
- ✚ Installation
- ✚ Hive QL

Session 7: Sqoop

- ✚ Introduction
- ✚ Sqoop Tools
- ✚ Sqoop Import
- ✚ Sqoop Import all tables
- ✚ Sqoop Export
- ✚ Sqoop Job
- ✚ Sqoop metastore
- ✚ Sqoop Eval
- ✚ Sqoop Codegen
- ✚ Sqoop List Databases and List Tables
- ✚ Sqoop Create Hive Table

Session 8: Mockup Interview Session

Session 9: Integrating Hadoop Into The Workflow

- ✚ Relational Database Management Systems
- ✚ Storage Systems
- ✚ Importing Data from RDBMSs With Sqoop
- ✚ Hands-on exercise
- ✚ Importing Real-Time Data with Flume
- ✚ Accessing HDFS Using FuseDFS and Hoop

Session 10: Delving Deeper Into The Hadoop API

- ✚ More about ToolRunner
- ✚ Testing with MRUnit
- ✚ Reducing Intermediate Data With Combiners
- ✚ The configure and close methods for Map/Reduce Setup and Teardown
- ✚ Writing Partitioners for Better Load Balancing
- ✚ Hands-On Exercise
- ✚ Directly Accessing HDFS
- ✚ Using the Distributed Cache

Session 11: Common Map Reduce Algorithms

- ✚ Sorting and Searching
- ✚ Indexing
- ✚ Machine Learning With Mahout
- ✚ Term Frequency – Inverse Document Frequency
- ✚ Word Co-Occurrence

Session 12: Using Hive and Pig

- ✚ Hive Basics
- ✚ Pig Basics



MindsMapped Consulting – Online Big Data and Hadoop Training

Session 13: Practical Development Tips and Techniques

- ✚ Debugging MapReduce Code
- ✚ Using LocalJobRunner Mode For Easier Debugging
- ✚ Retrieving Job Information with Counters
- ✚ Logging
- ✚ Splittable File Formats
- ✚ Determining the Optimal Number of Reducers
- ✚ Map-Only MapReduce

Session 14: Resume Preparation

Session 15: More Advanced Map Reduce Programming

- ✚ Custom Writables and WritableComparables
- ✚ Saving Binary Data using SequenceFiles and Avro Files
- ✚ Creating InputFormats and OutputFormats

Course Highlights:

- ✚ Hands on Assignments from each session
- ✚ Instructor led learning sessions
- ✚ Interactive sessions & Hands on Practice
- ✚ Lifetime access to Knowledge Base
- ✚ Interview and Job perspectives

For more information, call us on (+1) (435) 610-1777 / (801) 901-3010 Or email us at info@mindsmapped.com

Session 16: Joining Data Sets in Map

- ✚ Map-Side Joins
- ✚ The Secondary Sort
- ✚ Reduce-Side Joins

Session 17: Graph Manipulation in Hadoop

- ✚ Introduction to graph techniques
- ✚ Representing graphs in Hadoop
- ✚ Implementing a sample algorithm: Single Source Shortest Path

Session 18: Creating Workflows With Oozie

- ✚ The Motivation for Oozie
- ✚ Oozie's Workflow Definition Format

Session 19: Project Works

- ✚ Java to MapReduce Conversion
- ✚ MapReduce Project
- ✚ Hive Project
- ✚ Pig Project

Session 20: Mockup Interview Session

