



Big Data & Hadoop Training

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

Duration: 50 Hours(18 Session)

Call us at (+1) (801) 901-3035 / (801) 901-3010

E-mail:info@mindsmapped.com

Session 1: Big Data Opportunities & Challenges

- Introduction to 3V
- BigData & Hadoop
- OOPS & Java Fundamentals

Session 2: Understanding Linux Commands

- Linux commands required for Hadoop

Session 3: Introduction to Hadoop

- Concept of Hadoop Distributed file system(HDFS)
- Design of HDFS
- Common challenges
- Best practices for scaling with your data
- Configuring HDFS
- Interacting with HDFS
- HDFS permission and Security
- Additional HDFS Tasks
- Data Flow – Anatomy of a File Read
- Anatomy of a File Write
- Hadoop Archives

Session 4: Getting Started with Hadoop

- Creating & Running your program

Session 5: Pseudo Cluster Environment – Setting up Hadoop Cluster

- Cluster specification
- Hadoop Configuration (Environment Settings, Hadoop Daemon- Properties, Addresses and Ports)
- Basic Linux and HDFS Commands
- Setup a Hadoop Cluster

Session 6: MapReduce

- Hadoop Data Types
- Functional-Concept of Mappers
- Functional-Concept of Reducers
- The Execution Framework
 - Concept of Partioners
- Functional- Concept of Combiners
- Hadoop Cluster Architecture
- MapReduce types
- Input Formats (Input Splits and Records, Text Input, Binary Input, Multiple Inputs)
- OutPut Formats (TextOutput, BinaryOutPut, Multiple Output).
- Writing Programs for MapReduce

Session 7: PIG

- Installing and Running Pig Grunt
- Pig's Data Model
- Pig Latin
- Developing & Testing Pig Latin Scripts
- Writing Evaluation
- Filter
- Loads & Store Functions

Session 8: HIVE

- Hive Architecture
- Running Hive
- Comparison with Traditional Database (Schema on Read versus Write, Updates, Transactionsand Indexes)
- HiveQL (Data Types, Operators and Functions)
- Tables (Managed and External Tables, Partitions and Buckets, Storage Formats, Importing Data)
- Altering Tables, Dropping Tables
- Querying Data (Sorting And Aggregating, Map Reduce Scripts, Joins & Subqueries & Views
- Map and Reduce site Join to optimize Query
- User Defined Functions
- Appending Data into existing Hive Table
- Custom Map/Reduce in Hive
- Perform Data Analytics using Pig and Hive

Session 9: HBASE

- Introduction
- Client API- Basics
- Client API- Advanced Features
- Client API – Administrative Features
- Available Client
- Architecture
- MapReduce Integration
- Advanced Usage
- Advanced Indexing
- Implement HBASE

Session 10: SQOOP

- Database Imports
- Working with Imported data
- Importing Large Objects
- Performing Exports
- Exports- A Deeper look



Session 11: ZooKeeper

- The Zookeeper Service (Data Modal, Operations, Implementation, Consistency, Sessions, States)

Session 12: Oozie

- Workflow
- Coordinator
- Flume
- Concepts and Real time data streaming

Session 13: Spark Introduction

- What is Spark? Why Spark?
- Spark Ecosystem
- Overview of Scala
- Why Scala?
- Mapreduce Vs Spark

Session 14: Hello Spark

- my First Program in Spark
- Overview of RDD

Session 15: Spark Installation

- Installing Spark on Standalone cluster

Session 16: RDD Fundamentals

- Purpose and structure of RDD's
- Transformations
- Actions
- Programming API

Session 17: Spark SQL/DataFrames

- Dataframes / SQL APIs
- Uses of it

Session 18: Spark Streaming

- Sources and Tasks
- Dstream APIs
- Reliability and Fault Recovery

Course Highlights:

- Industry based Project work.
- Life time access to Knowledge base.
- Hands on project execution.
- Resume preparation and Mock Interviews
- Get Hadoop Certified
- Introduction to Kafka
- Projects related to MapReduce, Pig, Hive and Sqoop & Flume
- Projects related to Spark SQL will be assigned

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