Mind Q Systems Private Limited

Course Content for Hadoop Developer Introduction to BigData, Hadoop

- Big Data Introduction
- Hadoop Introduction
- What is Hadoop? Why Hadoop?
- Hadoop History?
- Different types of Components in Hadoop?
- HDFS, MapReduce, PIG, Hive, SQOOP, HBASE, OOZIE, Flume, Zookeeper and so on..

- What is the scope of Hadoop?

HDFS (for Storing the Data):

- Introduction of HDFS
- HDFS Design
- HDFS role in Hadoop
- Features of HDFS
- Daemons of Hadoop and its functionality
- Name Node
- Secondary Name Node
- Job Tracker
- Data Node
- Task Tracker
- Anatomy of File Wright
- Network Topology
- Nodes
- Racks
- Data Center
- Parallel Copying using DistCp
- Basic Configuration for HDFS
- Data Organization
- Blocks and
- Replication
- Rack Awareness
- Heartbeat Signal
- How to Store the Data into HDFS
- How to Read the Data from HDFS
- Accessing HDFS (Introduction of Basic UNIX commands)
- CLI commands

MapReduce using Java (Processing the Data):

- The introduction of MapReduce.
- MapReduce Architecture
- Data flow in MapReduce
- Splits
- Mapper
- Portioning
- Sort and shuffle
- Combiner
- Reducer
- Understand Difference Between Block and InputSplit
- Role of RecordReader
- Basic Configuration of MapReduce
- MapReduce life cycle
- Driver Code
- Mapper
- and Reducer
- How MapReduce Works
- Writing and Executing the Basic MapReduce Program using Java
- Submission & Initialization of MapReduce Job.
- File Input/Output Formats in MapReduce Jobs
- Text Input Format
- Key Value Input Format
- Sequence File Input Format
- NLine Input Format
- Joins
- Map-side Joins
- Reducer-side Joins
- Word Count Example
- Partition MapReduce Program
- Side Data Distribution
- Distributed Cache (with Program)
- Counters (with Program)
- Types of Counters
- Task Counters
- Job Counters
- User Defined Counters
- job Scheduling

Mind Q Systems Private Limited

PIG:

- ✓ Introduction to Apache PIG
- ✓ Introduction to PIG Data Flow Engine
- ✓ MapReduce vs. PIG in detail
- ✓ When should PIG use?
- ✓ Data Types in PIG
- ✓ Basic PIG programming
- ✓ Modes of Execution in PIG
- ✓ Local Mode and
- ✓ MapReduce Mode
- ✓ Execution Mechanisms
- ✓ Grunt Shell
- ✓ Script
- ✓ Embedded
- ✓ Operators/Transformations in PIG
- ✓ PIG UDF's with Program
- ✓ Word Count Example in PIG
- ✓ he difference between the MapReduce and PIG

SQOOP:

- ✓ Introduction to SQOOP
- ✓ Use of SQOOP
- ✓ Connect to MySQL database
- ✓ SQOOP commands
- ✓ Import
- ✓ Export
- ✓ Eval
- ✓ Codegen etc…
- ✓ Joins in SQOOP
- ✓ Export to MySQL
- ✓ Export to Hbase

HIVE:

- ✓ Introduction to HIVE
- ✓ HIVE Meta Store
- ✓ HIVE Architecture
- ✓ Tables in HIVE
- ✓ Managed Tables
- ✓ External Tables
- ✓ Hive Data Types
- ✓ Primitive Types
- ✓ Complex Types

- ✓ Partition
- ✓ Joins in HIVE
- ✓ HIVE UDF's and UADF's with Programs
- ✓ POC

HBASE:

- ✓ Introduction to HBASE
- ✓ Basic Configurations of HBASE
- ✓ Fundamentals of Hbase
- ✓ What is NoSQL?
- ✓ HBase Data Model
- ✓ Table and Row
- ✓ Column Family and Column Qualifier
- ✓ Cell and its Versioning
- ✓ Categories of NoSQL Data Bases
- ✓ Key-Value Database
- ✓ Document Database
- ✓ Column Family Database
- ✓ HBASE Architecture
- ✓ Hmaster
- ✓ Region Servers
- ✓ Regions
- ✓ MemStore
- ✓ Store
- ✓ SQL vs. NOSQL
- ✓ How HBASE is differed from RDBMS
- ✓ HDFS vs. Hbase
- ✓ Client-side buffering or bulk uploads
- ✓ HBase Designing Tables
- ✓ HBase Operations
- ✓ Get
- ✓ Scan
- ✓ Put
- ✓ Delete

MongoDB:

- ✓ What is MongoDB?
- ✓ Where to Use?

Setup:

- ✓ Downloading Virtual Box
- ✓ Installing Hadoop & ecosystems
- ✓ Installing Hadoop

- ✓ Creating Cluster
- ✓ Monitoring the Cluster Health
- ✓ Starting and Stopping the Nodes

Zookeeper:

- ✓ Introduction Zookeeper
- ✓ Data Modal
- ✓ Operations

OOZIE:

- ✓ Introduction to OOZIE
- ✓ Use of OOZIE
- ✓ Where to use?

Flume:

- ✓ Introduction to Flume
- ✓ Uses of Flume
- ✓ Flume Architecture
- ✓ Flume Master
- ✓ Flume Collectors
- ✓ Flume Agents

Project Explanation with Architecture