ASPIRE TECHSOFT PVT. LTD.

# ASPIRE TECHSOFT PVT. LTD.

**ISO 9001: 2015 Certified Training Institute**

# HADOOP Development

**Batches: Weekdays/ Weekend**

## Big-Data and Hadoop

- Introduction to big data and Hadoop
- **H**adoop **D**istributed **F**ile **S**ystem Architecture
- Installing Ubuntu with Java 1.8 on VM Workstation 11
- Hadoop Versioning and Configuration
- Single Node Hadoop 1.2.1 installation on Ubuntu 14.4.1
- Multi Node Hadoop 1.2.1 installation on Ubuntu 14.4.1
- Linux commands and Hadoop commands
- Cluster architecture and block placement
- Modes in Hadoop
  - Local Mode
  - Pseudo Distributed Mode
  - Fully Distributed Mode
- Hadoop Daemon
  - Master Daemons (Name Node, Secondary Name Node, Job Tracker)
  - Slave Daemons (Job tracker, Task tracker)
- Task Instance
- Hadoop HDFS Commands
- Accessing HDFS
  - CLI Approach
  - Java Approach

## Map-Reduce

- Understanding Map Reduce Framework
- Inspiration to Word-Count Example
- Developing Map-Reduce Program using Eclipse Luna
- HDFS Read-Write Process
- Map-Reduce Life Cycle Method
- Serialization (Java)
- Data types
- Comparator and Comparable (Java)
- Custom Output File
- Analyzing Temperature dataset using Map-Reduce
- Custom Practitioner & Combiner
- Running Map-Reduce in Local and Pseudo Distributed Mode

## Advanced Map-Reduce

- Enum (Java)
- Custom and Dynamic Counters
- Running Map-Reduce in Multi-node Hadoop Cluster
- Custom Writable
- Site Data Distribution
  - Using Configuration
  - Using Distributed Cache
  - Using stringifie




Address: Malhar Building, Plot No. 2, Office No.2, 2nd Floor, H. A Society, Paud Phata, Kothrud, Pune – 38
Mob No. : 7058198728/48, Email-Id:info@aspireit.net

- Input Formatters
  - NLine Input Formatter
  - XML Input Formatter
- Sorting
  - Primary Reverse Sorting
  - Secondary Sorting
- Compression Technique
- Working with Sequence File Format
- Working with AVRO File Format
- Testing Map Reduce with MR Unit
- Working with NYSE Datasets
- Working with Million Song Datasets
- Running Map-Reduce in Cloud era Box

**HIVE**

- Hive Introduction & Installation
- Data Types in Hive
- Commands in Hive
- Exploring Internal and External Table
- Partitions
- Complex data types
- UDF in Hive 4.7.1. Built-in UDF 4.7.2. Custom UDF
- Thrift Server
- Java to Hive Connection
- Joins in Hive
- Working with HWI
- Bucket Map-side Join
- More commands
  - View
  - Sort By
  - Distribute By
  - Lateral View
- Running Hive in Cloud era

**SQOOP**

- Sqoop Installations and Basics
- Importing Data from Oracle to HDFS
- Advance Imports
- Real Time Use Case
- Exporting Data from HDFS to Oracle
- Running Sqoop in Cloud era
- PIG
- Installation and Introduction
- Word Count in Pig
- NYSE in Pig
- Working With Complex Datatypes
- Pig Schema
- Miscellaneous Command
  - Group
  - Filter
  - Order
  - Distinct
  - Join
  - Flatten
  - Co-group
  - Union
  - Illustrate
  - Explain
- UDFs in Pig
- Parameter Substitution and Dry Run
- Pig Macros
- Running Pig in Cloud era

**SPARK**

**OOZIE**

- Installing Oozie
- Running Map-Reduce with Oozie
- Running Pig and Sqoop with Oozie




Address: Malhar Building, Plot No. 2, Office No.2, 2nd Floor, H. A Society, Paud Phata, Kothrud, Pune – 38
Mob No. : 7058198728/48, Email-Id:info@aspireit.net