BIG DATA- HADOOP (Development & Basic Administration)

What You Will Get From This Course?

- In-depth understanding of Entire Big Data Hadoop and Hadoop Eco System
- Real time idea of Hadoop Development
- Basic Hadoop Administration knowledge
- Detailed Course Materials
- Free Core Java and UNIX Fundamentals
- Real time projects implementation
- Interview Oriented Discussions
- Help in Resume Preparation

Overall Course Structure:

- UNIX/LINUX Basic Commands
- Basic UNIX Shell Scripting
- Basic Java Programming – Core JAVA OOPS Concepts
- Introduction to Big Data and Hadoop
- Working With HDFS
- Hadoop Map Reduce Concepts & Features
- Developing Map Reduce Applications
- Hadoop Eco System Components:
  - HIVE
  - PIG
  - HBASE
  - FLUME
  - SQOOP
  - OOZIE
- Introduction to SPARK & SCALA
- Real Time Tools like Putty, WinSCP, Eclipse, Hue, Cloudera Manager
- Hadoop Installation & Configuration
- Real Time Projects

Pre-Requisite:

- Basic SQL Knowledge
- Computer with Minimum 4GB RAM (8GM RAM Preferred)
- Basic UNIX & Java Programming knowledge is added advantage

Detailed Course Structure:

Introduction to Big Data & Hadoop

- ✓ The Big Data Problem
- ✓ What is Big Data?
- ✓ Challenges in processing Big Data
- ✓ What is Hadoop?
- ✓ Why Hadoop?
- ✓ History of Hadoop
- ✓ Hadoop Components Overview
  - HDFS
  - Map Reduce
- ✓ Hadoop Eco System Introduction
- ✓ NoSQL Database Introduction

Understanding Hadoop Architecture

- ✓ Hadoop 2.x Architecture
- ✓ Introduction to YARN
- ✓ Hadoop Daemons
- ✓ YARN Architecture
  - Resource Manager
  - Application Master
  - Node Manager

Introduction to HDFS (Hadoop Distributed File System)

- ✓ Rack Awareness
- ✓ HDFS Daemons
- ✓ Writing Files to HDFS
  - Blocks & Splits
  - Input Splits
  - Data Replication
- ✓ Reading Files from HDFS
- ✓ Introduction to HDFS Configuration Files

Working with HDFS

- ✓ HDFS Commands
- ✓ Accessing HDFS
  - CLI Approach
  - JAVA Approach **[Introducing HDFS JAVA API]**

Introduction to Map Reduce Paradigm

- ✓ What is Map Reduce?
- ✓ Detailed Map Reduce Flow
  - Introduction to Key/Value Approach
  - Detailed Mapper Functionality
  - Detailed Reducer Functionality
  - Details of Partitioner
  - Shuffle & Sort Process
- ✓ Understanding Map Reduce Flow with Word Count Example

Basic Map Reduce Programming

- ✓ Introduction to Map Reduce API [**New Map Reduce API**]
- ✓ Map Reduce Data Types
- ✓ File Formats
- ✓ Input Formats – Input Splits & Records, text input, binary input
- ✓ Output Formats – Text Output, Binary Output
- ✓ Configuring Development Environment – Eclipse
- ✓ Developing a Map Reduce Application using Default Functionality
  - Identity Mapper
  - Identity Reducer
  - **ToolRunner API** Introduction
- ✓ Developing Word Count Application
  - Writing Mapper, Reducer & Driver Code
  - Building Application
  - Deploying Application
- ✓ Running the Map Reduce Application
  - Local Mode of Execution
  - Cluster Mode of Execution
- ✓ Monitoring Map Reduce Application

Advance Map Reduce Programming

- ✓ Map Reduce Combiner
- ✓ Map Reduce Counters
- ✓ Map Reduce Partitioner
- ✓ Map Reduce Distributed Cache
- ✓ Writing Custom Partitioner
- ✓ Writing Custom Record Reader & Record Writer [Custom Input & Output Formats]
  - Sample Program with PDF Input File
- ✓ Custom Writables & Writable Comparables
- ✓ Map Reduce Compression
- ✓ File Merge Utility

Programming with HIVE

- ✓ Introduction to HIVE
- ✓ Hive Architecture
- ✓ Types of Meta store
- ✓ Introduction to Hive Configuration Files
- ✓ Hive Data Types
  - Simple Data Types
  - Collection Data Types
- ✓ Types of Hive Tables
  - Managed Table
  - External Table
- ✓ Hive Query Language (HQL or HIVE QL)
  - Creating Databases
  - Creating Tables
  - Joins in Hive
  - Group BY and Distinct operations
  - Partitioning
    - Static Partitioning
    - Dynamic Partitioning
  - Bucketing
  - Lateral View & Explode [**Introduction to Hive UDFs → UDF, UDAF & UDTF**]
  - XML Processing in HIVE
  - JSON processing in HIVE
  - URL Processing in HIVE
- ✓ Hive File Formats [**Introduction to Hive SERDE**]
  - Parquet
  - ORC
  - AVRO
- ✓ Introduction to HIVE Query Optimizations
- ✓ Developing Hive UDFs in JAVA
- ✓ Hive JDBC Client

Programming with PIG

- ✓ Introduction to PIG
- ✓ PIG Architecture
- ✓ Introduction to PIG Configuration Files
- ✓ PIG vs. HIVE vs. Map Reduce
- ✓ Introduction to Data Flow Language
- ✓ Pig Data Types
- ✓ Pig Programming Modes
- ✓ Pig Access Modes
- ✓ **Detailed PIG Latin Programming**

- ✓ PIG UDFs & UDF Development in JAVA
- ✓ PIG Macros
- ✓ Hive - PIG Integration → Introduction to HCATALOG
- ✓ Processing XML Data in PIG
- ✓ Introduction to PIG Optimization

NoSQL & HBASE

- ✓ Introduction to NoSQL Databases
- ✓ Types of NoSQL Databases
- ✓ Introduction To HBASE
- ✓ HBASE Architecture
- ✓ HBASE Shell Interface
  - Creating Data Bases and Tables
  - Inserting Data in tables
  - Accessing data from Tables
  - HBase Filters
- ✓ Hive & HBASE Integration
- ✓ PIG & HBASE Integration
- ✓ HBASE JAVA API

Introduction to Streaming & FLUME

- ✓ Introduction to Streaming
- ✓ Introduction to FLUME
- ✓ FLUME Architecture
- ✓ Flume Agent Setup
- ✓ Types of Source, Channel & Sinks
- ✓ Developing Sample Flume Applications
- ✓ Introduction to **KAFKA**

SQOOP

- ✓ Introduction to SQOOP
- ✓ Connecting to RDBMS Using SQOOP
- ✓ SQOOP Import
  - Import to HDFS
  - Import to HIVE
  - Import to HBASE
  - Bulk Import
    - Full Table
    - Subset of a Tables
    - All tables in DB
  - Incremental Import
- ✓ SQOOP Export

- Export from HDFS
- Export from Hive

Oozie

- ✓ Oozie Fundamentals
- ✓ Oozie Architecture
- ✓ Oozie XML File Specifications
- ✓ Workflow Creation
- ✓ Job Submission, Monitoring & Debugging
- ✓ Job Coordinators & Bundles

Introduction to SPARK & SCALA

- ✓ Introduction to Spark
- ✓ Spark vs. Map Reduce
- ✓ Concepts of Transformation & Action
- ✓ Sample Word Count Program in Spark with Scala

Hadoop Installation & Configuration

- ✓ Set up a Single Node Hadoop Cluster
- ✓ Hadoop Configuration Files
- ✓ HIVE Installation (Hands on Installation on Laptops)
- ✓ PIG Installation (Hands on Installation on Laptops)
- ✓ SQOOP Installation (Hands on Installation on Laptops)
- ✓ HBase Installation (Hands on Installation on Laptops)
- ✓ OOZIE Installation (hands on Installation on Laptops)
- ✓ Introduction to Name Node Federation