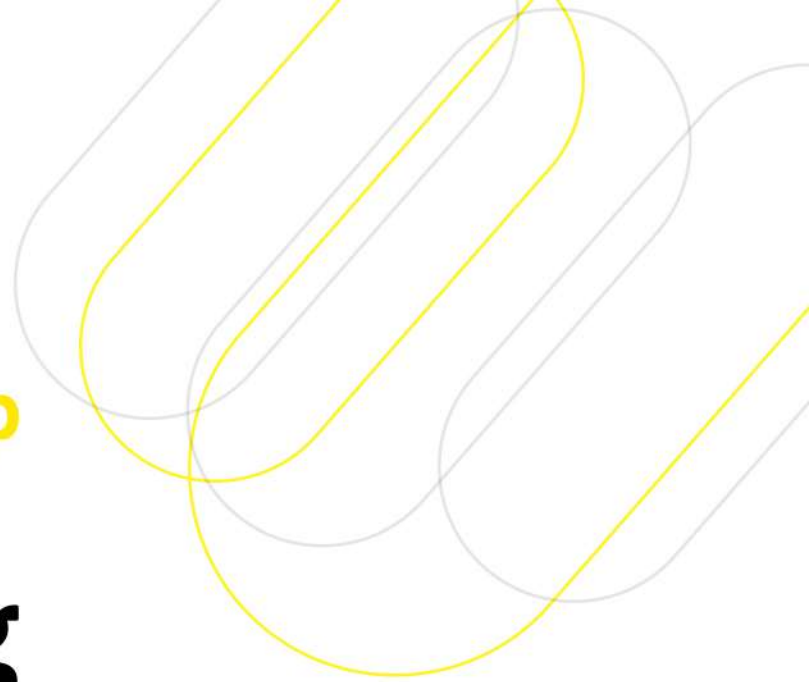




Big Data & Hadoop

e-Learning Course

Zetlan Technologies

Help Desk: +91 8680961847

www.zetlantech.com



COURSE MODULES

Module 1: Introduction to Big Data

- What is Big Data?
- Characteristics of Big Data (Volume, Velocity, Variety, Veracity, Value)
- Traditional Data Processing vs. Big Data Processing
- Challenges in Handling Big Data
- Big Data Applications Across Industries

Module 2: Introduction to Hadoop

- What is Hadoop?
- Evolution of Hadoop & History
- Components of Hadoop Ecosystem
- Hadoop 1.0 vs. Hadoop 2.0 vs. Hadoop 3.0
- Overview of HDFS (Hadoop Distributed File System)
- Hadoop Architecture & Working

Module 3: Hadoop Distributed File System (HDFS)

- HDFS Features & Architecture
- HDFS Blocks, NameNode, DataNode, Secondary NameNode
- File Read/Write Operations in HDFS
- Data Replication & Fault Tolerance
- Hands-on: HDFS Commands & File Operations

Module 4: Hadoop MapReduce Framework

- What is MapReduce?
 - MapReduce Workflow (Mapper, Reducer, Combiner, Partitioning)
 - Input Splits & Output Formats
 - Data Processing Flow in MapReduce
 - Writing and Executing MapReduce Jobs
 - Hands-on: Word Count Program & Other Use Cases
- 

Module 5: Apache YARN (Yet Another Resource Negotiator)

- Introduction to YARN
- YARN Architecture (ResourceManager, NodeManager, ApplicationMaster)
- How YARN Manages Resources
- YARN vs. Traditional MapReduce
- Hands-on: Running Jobs on YARN

Module 6: Hadoop Ecosystem Components

6.1 Apache Hive (Data Warehousing in Hadoop)

- Introduction to Hive & Data Warehousing
- Hive Architecture & Components
- Hive Query Language (HQL) vs SQL
- Working with Hive Tables (Managed vs External)
- Partitioning & Bucketing in Hive
- Hands-on: Running Queries in Hive

6.2 Apache Pig (Data Flow Programming)

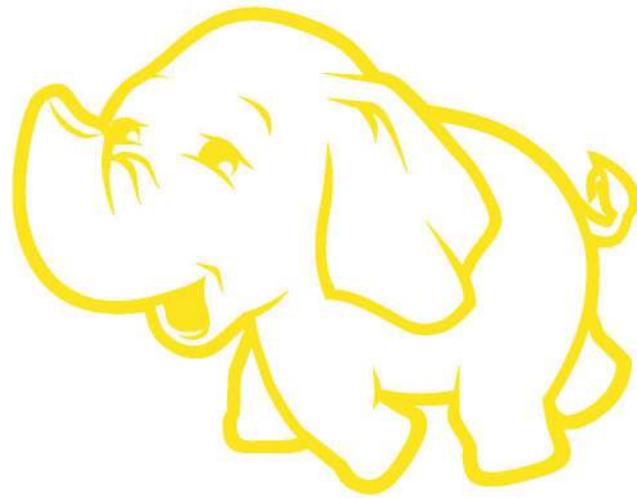
- Introduction to Apache Pig
- Pig Latin Scripting Language
- Data Loading & Transformation using Pig
- Comparing Pig with MapReduce & Hive
- Hands-on: Writing Pig Scripts

6.3 Apache HBase (NoSQL Database in Hadoop)

- Introduction to HBase & NoSQL
- HBase Architecture (Region Servers, Zookeeper, HMaster)
- HBase vs RDBMS
- CRUD Operations in HBase
- Hands-on: HBase Shell Commands

6.4 Apache Sqoop (Data Transfer Tool)

- Sqoop Overview
- Importing Data from RDBMS to Hadoop
- Exporting Data from Hadoop to RDBMS
- Hands-on: Sqoop Import/Export Examples



6.5 Apache Flume (Log & Streaming Data Processing)

- Introduction to Flume
- Flume Agents, Sources, Channels & Sinks
- Real-time Data Ingestion using Flume
- Hands-on: Collecting Data from Web Logs

Module 7: Apache Spark & Real-time Big Data Processing

- Introduction to Apache Spark
- Spark vs MapReduce
- Spark Components (RDDs, DataFrames, Spark SQL, Streaming, MLlib)
- Hands-on: Running Spark Jobs on Hadoop

Module 8: Hadoop Cluster Setup & Administration

- Setting up a Single-Node & Multi-Node Cluster
- Hadoop Configuration Files (core-site.xml, hdfs-site.xml, yarn-site.xml)
- Managing Hadoop Services & Daemons
- Cluster Monitoring & Troubleshooting

Module 9: Big Data Use Cases & Projects

- Real-time Use Cases in Banking, Retail, Healthcare, etc.
- End-to-End Big Data Project Implementation
- Best Practices for Hadoop Development & Deployment