

Course Content:

1. Understanding Big Data and Hadoop

Intro to Big Data

Limitations and Solutions of existing Data Analytics Architecture

Hadoop Features

Hadoop Ecosystem

Hadoop 2.x core components

Hadoop Storage: HDFS

Hadoop Processing: MapReduce Framework

Anatomy of File Write and Read

Rack Awareness.

2. Hadoop Architecture and HDFS

Hadoop 2.x Cluster Architecture - Federation and High Availability

Hadoop Cluster Modes

Common Hadoop Shell Commands

Hadoop 2.x Configuration Files

Password-Less SSH

MapReduce Job Execution

Data Loading Techniques: Hadoop Copy Commands FLUME SQOOP.

3. Hadoop MapReduce Framework - I

MapReduce Use Cases

Hadoop 2.x MapReduce Architecture

Hadoop 2.x MapReduce Components

YARN MR Application Execution Flow

YARN Workflow

4. Hadoop MapReduce Framework - II

Input Splits and HDFS Blocks

MapReduce Job Submission Flow

MapReduce: Combiner & Partitioner

5. Advance MapReduce

Counters

- Distributed Cache
- MRunit
- Reduce Join
- Custom Input Format
- Sequence Input Format.

6. Pig

- About Pig
- MapReduce Vs Pig
- Pig Use Cases
- Programming Structure in Pig
- Pig Latin Program
- Data Models in Pig
- Pig Latin commands : Relational Operators, File Loaders, Group Operator, COGROUP Operator, Joins and COGROUP, Union, Diagnostic Operators ,Pig UDF,Pig Data Types.

7. Hive

- Hive Background
- Hive Use Case
- Hive Vs Pig
- Hive Architecture and Components
- Metastore in Hive
- Limitations of Hive
- Comparison with Traditional Database
- Hive Data Types and Data Models, Partitions and Buckets,
- Hive Tables(Managed Tables and External Tables)
- Importing Data
- Querying Data
- Managing Outputs
- Hive Script
- Hive UDF

8. HBase

- Introduction to NoSQL Databases and HBase
- HBase v/s RDBMS
- HBase Components
- HBase Architecture
- HBase Cluster Deployment
- HBase Data Model
- HBase Shell
- HBase Client API

Data Loading Techniques
Filters in HBase.

10. Oozie

Flume and Sqoop

Oozie

Oozie Components

Oozie Workflow

Scheduling with Oozie

Oozie Commands

Oozie Web Console