

 **HADOOP Content****Course Content**

What is Hadoop and its usecases (Demo)

HDFS and its architecture

- * NameNode and its functionality
- * DataNode and its functionality

HDFS user commands and admin commands.**MapReduce architecture**

- * JobTracker and its functionality
- * TaskTrack and its functionality
- * Job execution flow

MapReduce Programming Model

- * How to write basic MR job and running in Local
- * Input formatters and its associated Record Readers
 - * Text Input Formatter
 - * KeyValue Input Formatter
 - * Sequence File Input Format
 - * How to write custom Input Formatters and its Record Readers
- * Output formatters and its associated Record Writers
 - * Text Output Formatter
 - * Sequence File Output Formatter
 - * How to write custom Output Formatters and its Record Writers
- * Combiner
- * Partitioner
- * Secondary Sorting
- * Writable and WritableComparables
- * Compression techniques
 - * Snappy
 - * LZO
 - * Zip
- * Schedulers
 - * FIFO, Capacity and Fair
- * Distributed Cache

How to debug MapReduce Jobs in Local and Pseudo cluster Mode.

Unit Testing MR Jobs.

How to Identify Performance Bottlenecks in MR jobs and tuning MR jobs.

Introduction to MapReduce Streaming and Pipes.

Introduction to YARN (Next Generation MapReduce).

 **HADOOP Content****Hbase**

Hbase introduction

Hbase usecases

Hbase basics

- * Column families
- * Scans

Hbase standalone and distributed mode installations

Hbase Architecture

- * Storage
- * Write-Ahead Log
- * Log -Structured Merge-Trees

Mapreduce integration

Mapreduce over Hbase

Usage

- * Key design
- * Bloom Filters
- * Versioning
- * Coprocessors
- * Filters

Clients

- * REST
- * Thrift
- * Hive
- * Web Based UI

Hbase Admin

- * Schema definition
- * Basic CRUD operations

Hive

Hive Introduction

Hive architecture

- * Driver
- * Compiler
- * Semantic Analyzer

HQL

Integration with Hadoop

Hive installation

Starting CLI and Thrift mode

Usage.

PIG

Apache PIG introduction1

PIG set up

PIG hands on

PIG UDF's

Excercises

FLUME**SQOOP**