

Office Hours Monday-Friday, 09:00am-07:00pm.

Email conceptlearningcentre@gmail.com/ conceptlearningcentrewagholi@gmail.com

Phone Number +91-9764500015/+91-9860855556.

Webpage www.conceptlearningcentre.com

Course Description:

Hadoop Training for students and Corporate Candidates.

Syllabus Details:

Module 1: Understanding Big Data and Hadoop

Big Data

Limitations and Solutions of existing Data Analytics Architecture

Hadoop

Hadoop Features

Hadoop Ecosystem

Hadoop 2.x core components

Hadoop Storage: HDFS

Hadoop Processing: MapReduce Framework

Hadoop Different Distributions

Module 2: Hadoop Architecture and HDFS

Hadoop 2.x Cluster Architecture - Federation and High Availability

A Typical Production Hadoop Cluster

Hadoop Cluster Modes

Common Hadoop Shell Commands

Hadoop 2.x Configuration Files

Single node cluster and Multi node cluster set up Hadoop Administration

Module 3: Hadoop MapReduce Framework

MapReduce Use Cases

Traditional way Vs MapReduce way

Why MapReduce

Hadoop 2.x MapReduce Architecture

Hadoop 2.x MapReduce Components

YARN MR Application Execution Flow

YARN Workflow

Anatomy of MapReduce Program

Demo on MapReduce

Input Splits

Relation between Input Splits and HDFS Blocks

MapReduce: Combiner & Partitioner

Module 4: Advanced MapReduce

Counters

Distributed Cache

MRunit

Reduce Join

Custom Input Format

Sequence Input Format

Xml file Parsing using MapReduce

Module 5: Pig Scripting

About Pig

MapReduce Vs Pig

Pig Use Cases

Programming Structure in Pig

Pig Running Modes

Pig components

Pig Execution

Pig Latin Program

Data Models in Pig

Pig Data Types

Shell and Utility Commands

Pig Latin : Relational Operators

File Loaders, Group Operator

COGROUP Operator

Joins and COGROUP

Union

Diagnostic Operators

Specialized joins in Pig

Built In Functions (Eval Function, Load and Store Functions, Math function, String Function, Date Function, Pig UDF, Piggybank, Parameter Substitution (PIG macros and Pig Parameter substitution)

Pig Streaming

Testing Pig scripts with Punit

Aviation use case in PIG

Pig Demo on Healthcare Data set

Module 6:Hive

Hive Background

Hive Use Case

About Hive

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 Output

Hive Script

Hive UDF

Retail use case in Hive

Hive Demo on Healthcare Data set

Module 7:Advanced Hive and HBase

Hive QL: Joining Tables

Dynamic Partitioning

Custom Map/Reduce Scripts

Hive Indexes and views Hive query optimizers

Hive : Thrift Server, User Defined Functions

HBase: Introduction to NoSQL Databases and HBase

HBase v/s RDBMS

HBase Components

HBase Architecture

Run Modes & Configuration

HBase Cluster Deployment

Module 8:Advanced HBase

HBase Data Model

HBase Shell

HBase Client API

Data Loading Techniques

ZooKeeper Data Model

Zookeeper Service

Zookeeper

Demos on Bulk Loading

Getting and Inserting Data

Filters in HBase

Module 9: Processing Distributed Data with Apache Spark

What is Apache Spark

Spark Ecosystem

Spark Components

History of Spark and Spark Versions/Releases

Spark a Polyglot

What is Scala?

Why Scala?

SparkContext

RDD

Module 10: Oozie and Hadoop Project

Flume and Sqoop Demo

Oozie

Oozie Components

Oozie Workflow

Scheduling with Oozie

Demo on Oozie Workflow

Oozie Co-ordinator

Oozie Commands

Oozie Web Console

Oozie for MapReduce