

# BIG DATA

## Hadoop (Development)



## CURRICULUM

### » UNDERSTANDING BIG DATA AND HADOOP

- Understand What Is Big Data
- Analyze Limitations And Solutions Of Existing Data Analytics Architecture
- Understand What Is Hadoop And Its Features
- Hadoop Ecosystem
- Understand Hadoop 2.x Components
- Perform Read And Write In Hadoop
- Understand Rack Awareness Concept

### » HADOOP ARCHITECTURE AND HDFS

- Run Hadoop In Different Cluster Nodes
- Implement Basic Hadoop Commands On Terminal
- Prepare Hadoop 2 Configuration Files Analyze The Parameters In It.
- Implement Password-less Ssh On Hadoop Cluster
- Analyze Dump Of A Mapreduce Program
- Implement Different Data Loading Techniques

### » HADOOP MAPREDUCE FRAMEWORK – I

- Analyze Different Use-cases Where Mapreduce Is Used
- Differentiate Between Traditional Way And Mapreduce Way
- Learn About Hadoop 2.x Mapreduce Architecture And Components
- Understand Execution Flow Of Yarn Mapreduce Application
- Implement Basic Mapreduce Concepts
- Run A Mapreduce Program

### » HADOOP MAPREDUCE FRAMEWORK – II

- Analyze Mapreduce Job Submission Flow
- Implement Combiner And Partitioner In Mapreduce
- Understand Mapreduce Codes In Details
- Code In Mapreduce For A Given Problem Statement
- Understand Input Splits Concepts In Mapreduce
- Module 5– Advance Mapreduce
- Implement Counter In Mapreduce
- Numerical Summarizations
- Counting With Counters
- Top K Records
- Distinct Records
- Total Order Sorting
- Reduce Side Join
- Replicated Join
- Implement Distributed Cache Concept In Mapreduce
- Customizing Input And Output In Hadoop
- Implement Custom Input Format In Mapreduce
- Implement Sequence Input Format In Mapreduce
- Implement Xml Input Format In Mapreduce

## » INTRODUCTION TO CLOUDERA AND UNDERSTANDING PIG

- Pig And Its Need
- Difference Between Pig And Mapreduce
- Pig Features And Programming Structure
- Pig Running Modes
- Pig Components And Data Models
- Basics Operations In Pig
- Udf In Pig

## » UNDERSTANDING HIVE

- Hive And Its Use Cases
- Hive Vs. Pig
- Hive Architecture And Components
- Primitive And Complex Type In Hive
- Data Models In Hive
- Query Efficiency Measures
- Partitioning
- Bucketing
- Hive Script And Hive Udf

## » UNDERSTANDING SQOOP AND FLUME

- Implement Flume Job To Download Data From Twitter
- Implement Flume Job To Download Data From Other Sources
- Implement Sqoop To Import Table From Rdbms Into Hdfs.
- Implement Sqoop To Import All Tables From Rdbms Into Hdfs.
- Implement Sqoop To Import Table From Rdbms Into Hive.
- Implement Sqoop To Import Schema And Tables Details Rdbms.
- Implement Sqoop To Export Data To Rdbms (insert And Update Mode)
- Implement Sqoop To Generate Java Classes Which Encapsulate And Interpret Imported Records

## » INTRODUCTION TO NOSQL AND WORKING WITH OOZIE

- Understand Oozie
- Schedule Job In Oozie
- Implement Oozie Workflow
- Implement Oozie Coordinator

## CORE SPARK

### » INTRODUCTION TO CORE SPARK

- Introduction to Scala/Python
- Apache Map Reduce vs Apache Spark
- Functional Programming With Spark
- Interactive Shell
- Resilient Distributed Data Sets(RDDs)
- Transformation
- Action
- Using HDFS With Spark

### » SPARK QL

- Running SQL
- Queries Using Spark QL
- Explain Me Importance And Features of Spark QL
- Describe Methods to Convert RDDs to Data Frames
- Explain Concepts of Spark SQL
- Describe Me Concept of Hive Integration

### » PROJECT

- MultiNode Cluster Setup on Linux
- Amazon Ec2 Online Cluster Setup
- Crime Data Analysis
- E-Commerce Data Analysis
- Twitter Data Fetching

Partners :



Java



development | consultancy | training

E-mail: [info@ducatindia.com](mailto:info@ducatindia.com)

Visit us: [www.ducatinidia.com](http://www.ducatinidia.com)

[www.facebook.com/ducateducation](http://www.facebook.com/ducateducation)

#### NOIDA

A-43 & A-52, Sector-16,  
Noida - 201301, (U.P.) INDIA  
Ph. : 0120-4646464  
Mb. : 09871055180

#### GURGAON

1808/2, 2nd floor old DLF ,  
Near Honda Showroom,  
Sec.-14, Gurgaon (Haryana)  
Ph. : 0124-4219095-96-97-98  
Mb. : 09873477222-333

#### GREATER NOIDA

F 205 Neelkanth Plaza Alpha 1  
commercial Belt Opposite to Alpha  
Metro Station Greater Noida  
Ph. : 0120-4345190-91-92 to 97  
Mb. :09899909738, 09899913475

#### JAIPUR

38,Jai Jawan Colony 3rd,  
Near Gaurav Tower,JLN  
Marg, Jaipur (Rajsthan)  
Ph. : 0141-2550077, 2550202  
Mb : 08824246937

#### GHAZIABAD

1, Anand Industrial Estate,  
Near ITS College, Mohan Nagar ,  
Ghaziabad (U.P.)  
Ph.: 0120-4835400..98-99  
Mb. : 09810831363 / 9818106660  
: 08802288258 - 59-60

#### FARIDABAD

SCO-32, 1st Floor , Sec.-16,  
Faridabad (HAR YANA)  
Ph. : 0129-4150605-09  
Mb. : 0981 1612707