

Data Analytics with Hadoop & Spark

Duration: - 7 Days (8-10 hours)

Prerequisites

- Any Debian OS is preferred. Ubuntu is recommended.
- Stable internet connection.

Big Data

Program Duration: 2.5 - 3 Hour

- Big Data Overview
 - What's Big Data
 - Big Data 6V's
 - Use cases and Benefits
- Hadoop (HDFS)
 - Distributed File System
 - What is Hadoop
 - RDBMS vs Hadoop
 - Components and Hadoop
 - Understanding HDFS on multi node cluster
 - Loading data and understanding block and replication factor
 - Basic and advance HDFS operation
 - Mapreduce | Yarn
 - Exploring the architecture of Yarn
 - Solving the sudoku based problem & wordcount
 - Running distributed processing with HDFS and Yarn

Offices : Jaipur || Udaipur || Pune ||

Website: www.regexsoftware.com

Mob : 96028 80219

SPARK

Program Duration : 3 Hours

- Spark Basics
 - Understanding Spark Architecture
 - Feature & benefits of using Spark
 - Spark Vs Map-reduce
- Working with RDD in spark
 - RDD basics
 - RDD operations
 - Transformation and Actions
 - Working with Key-value pair
- Working with Data Frame
 - Basics commands
 - Manipulating dataframe
 - Renaming column.

HIVE

Program Duration : 2 hours

- Understanding Hive
- External & Managed table in Hive
- Understanding the concepts of Partition & Bucketing
- Inserting values from other table
- Changing the dimensions of Table
- Several function available in Hive

NOSQL

Program Duration : 1 Hour

- Understanding the difference between NoSQL & SQL
- How to install it

- Basic understanding of working with HBASE