

**SLOG SOLUTIONS PRIVATE LIMITED**  
**TECHNOLOGY: BIG DATA**  
**DURATION: MODULE 1 (4 WEEKS)**  
**MODULE 1+2 (6 WEEKS)**

**MODULE 1:**

**BIG DATA Essentials**

**Introduction**

- Introduction to Big Data
- Role of data analyst
- What is data science
- Why data science
- Role of data scientist

**Use of Big Data**

- Finance
- Health
- Retail
- Advertising
- Defense and Intelligence

**Process of Data Analytics**

- Preparation of data
- Preprocessing of data
- Analysis of data
- Post processing of data

**Data presentation**

- Planning
- Data collection
- Data selection

**Data preparation**

- Sqoop
- Flume
- Tools for data preparation
  - Sql database
  - mongo DB

**Data preprocessing**

- Data cleaning
- Data filtering
- Data completion
- Data correction
- Data Transformation
- Pig for data pre processing
- Hive

**Data Analysis**

- Sqoop
- Recommendation
- Classification
- Clustering
- Mahout

**Recommendation**

- Introduction to recommendation
- Making recommendation
- Techniques for making recommendation

**Classification**

- Overview of classification system
- Classification process
- Naïve bayes classifier
- Decision tree
- Clustering

- Clustering basics
- K-means clustering

**Data visualization using R**

- Basics of R
- Data frames
- Vectorized operations on data frames
- Selection
- Projection
- Transformation

**MODULE 2: HADOOP**

**Introduction**

- Introduction to HADOOP
- Components of HADOOP ecosystem
- Commercial HADOOP distributions
- HADOOP architecture distributed storage(HDFS)and YARN

**HDFS**

- What is HDFS
- Need of HDFS
- HDFS architecture and component
- Regular file system vs. HDFS
- Data block split
- Data replication topology
- HDFS command line

## **YARN**

- What is YARN
- YARN use case
- YARN architecture
- Resource manager
- Application master
- How YARN run application
- Tools for YARN developers

## **Data ingestion into Big Data System and ETL**

- Overview of data ingestion
- Apache sqoop
- Sqoop and its uses
- Sqoop preprocessing
- Sqoop connectors
- Apache flume
- Flume model
- Scalability in flume
- Components in flume architecture



SLOG SOLUTIONS PVT.LTD.  
HELPLINE 7456000240/7456000241  
[www.slogsolutions.com](http://www.slogsolutions.com)