

## Big Data and Hadoop

### **Fundamentals of Big Data**

- What is Big Data?
- Application of Big Data
- Tools for Big Data
- How the financial service sector can employ analytics
- Role of Big data in the strategic decision making in financial services industry.
- Big data Cyber Security Analytics
- Scalable and distributed data mining

### **Big Data Management**

- Distributed file systems for Big data Storage, access and analytics
- Frameworks and tools for big data cyber security analytics
- Performance modeling , simulation and analysis
- Big Data applications in cyber security
- Parallel and distributed algorithms for Big Data analytics
- Big Data case studies and applications

### **Hadoop**

- What is hadoop?
- Why Hadoop and its uses
- Different Components of Hadoop
- Introduction to Apache Pig
- Map Reduce Vs Apache Pig
- SQL vs Apache Pig
- Different Data Types in Pig
- Modes of Execution in Pig
  - Local Mode
- Execution Mechanism
  - Grunt Shell
  - Script
  - Embedded
- Transformations in Pig
- How to write a simple pig script
- UDFs in Pig

### **HDFS(Hadoop Distributed File System)**

- Significance of HDFS in Hadoop
- Features of HDFS
- Daemons of Hadoop

## HIVE

- HIVE Introduction
- HIVE Architecture
- HIVE Meta Store
- HIVE Integration with Hadoop

## Map Reduce (Map Reduce Architecture)

- Map Reduce Programming Mode
- Different phases of MapReduce Algorithm
- Different Data Types in MapReduce
- How to write a basic program
- Joining Datasheets in MapReduce Jobs- Map Joins and Reduce Joins
- Creating Input and Output Formats in MapReduce Jobs
  - Text Input Format
  - Key Value Input Format
  - Sequence File Input Format
- How to Debug MapReduce
- Jobs in Local and Pseudo Cluster Mode
- Unit Testing in MR Jobs
- How to debug MapReduce
- Streaming and Pipelining
- Data Localization in MapReduce
- Combiner ( Mini Reducer and Partitioner)

## Hbase

- Overview
- Installation
- Hbase Shell
- Hbase General Command

## Sqoop

- What Sqoop Does?
- Data Imports
- Parallel Data Transfer
- Fast Data Copies

## Deliverables to Participant after the conduction of workshop

1. Software toolkit to each participant.
2. Training Material (eBooks) for each participant.