

BIGDATA HADOOP – ANALYTIC

Module 1: Introduction Big data and Hadoop

- What is Big Data?
- Characteristics of big data
- Big Data challenges
- Popular tools used with big data For storing, processing, analysing & visualization
- Where Hadoop fits in?
- Traditional data analytic's architecture versus Hadoop
- What is Hadoop?
- History of Hadoop
- Hadoop's key characteristics
- Hadoop usage

Module 2: Hadoop Eco-system & Architecture

- Hadoop eco-system & core components
- HDFS architecture & overview of MRv1
- HDFS daemons
- Files and blocks
- Anatomy of a file write & read
- Replication & rack awareness

Module 3: Introduction to YARN

- What is YARN?
- MR1 v MR2
- YARN architecture
- HDFS Federation
- YARN Deamons
- YARN Job execution workflow
- Authentication and high availability in Hadoop

Module 4: Hadoop Cluster Configuration

- Hortonworks sandbox installation & configuration
- Hadoop Configuration files
- Working with Hadoop services using Ambari
- Hadoop deamons
- Browsing Hadoop UI consoles
- Basic Hadoop Shell commands
- Eclipse & winscp installation & configurations on VM

Module 5: Basics of MapReduce on YARN

- Running a MapReduce application in MR2
- MapReduce Framework on YARN
- Fault tolerance in YARN
- Map, Reduce & Shuffle phases
- Understanding Mapper, Reducer & Driver classes

Module 6: MapReduce Programming

- Writing Map Reduce WordCount program
- Executing & monitoring a Map Reduce job
- Use case - Sales calculation using M/R

Module 7: Analysis using Pig

- Background of Pig
- Pig architecture
- Pig Latin basics
- Pig execution modes
- Pig processing – loading and transforming data
- Pig built-in functions
- Filtering, grouping, sorting data
- Relational join operators
- Pig Scripting
- Pig UDF's

Module 8: Analysis using Hive Data Warehousing Infrastructure

- Background of Hive
- Hive architecture
- Hive Query Language
- Derby to MySQL database
- Managed & external tables
- Data processing – loading data into tables
- Hive Query Language
- Using Hive built-in functions
- Partitioning data using Hive
- Bucketing data
- Hive Scripting
- Using Hive UDF's

Module 9: Working with HBase

- HBase overview
- Data model
- HBase architecture
- HBase shell
- Zookeeper & its role in HBase environment
- HBase Shell environment
- Creating table
- Creating column families
- CLI commands – get, put, delete & scan
- Scan Filter operations

Module 10: Importing & Exporting Data using Sqoop

- Importing data from RDBMS to HDFS
- Exporting data from HDFS to RDBMS
- Importing & exporting data between RDBMS & Hive tables

Module 11: Oozie Workflow Management

- Overview of Oozie
- Oozie Workflow Architecture
- Creating workflows with Oozie

Module 12: Using Flume for Analysing Streaming Data

- Introduction to Flume
- Flume Architecture
- Flume Demo

203/RATNMANI BLDG, DADA PATIL WADI, OPP ICICI ATM, THANE WEST
Web: www.nettechindia.com Phone : 9870803004/ 9870803005