Hadoop Training

Course Content

1. **Hadoop Setup**
    a. Hadoop Prerequisite
    b. Different Modes for Hadoop
    c. Pseudo Distributed Mode (Setup)
    d. Understanding Important configuration
2. **Quick look on Unix Commands**
    a. File & Process Commands
    b. File Permissions and Search Commands
    c. Compression Commands
    d. Some useful Shortcuts
    e. Setting up Environment Variables
3. **Introduction to Hadoop**
    a. What is Big Data?
    b. Limitation and Solutions of existing Data Analytics Architecture
    c. Introduction to Hadoop
    d. Hadoop Features
    e. Hadoop Ecosystem
    f. Introduction to HDFS and MapReduce
4. **Deep Dive in HDFS**
    a. What is HDFS?
    b. HDFS Key Features
    c. HDFS Architecture
    d. Block Placement Policy
    e. Read Write operation
    f. Network Topology
    g. NameNode, Data Node, HeartBeats, Block Size,Replication Factor
    h. Balancer
    i. HDFS Commands (Hands-On-Experience)
5. **MapReduce**
    a. Introduction to MapReduce
    b. How Mapper and Reducer Works
    c. Combiner, Partitioner and Shuffle and Sort Phase
    d. Advance MapReduce Concept
    e. Job Scheduler
    f. Joining of files
    g. Reduce Joins

Real Time Signals Technologies,
#102, Krishna Grand, Over Marathahalli Bridge,
Bangalore – 37, Phone - +91-80-42008777, 9538566675, mail – info@realtimesig.com

h. Input format
i. Hands-On-Exercise

6. Pig

a. Introduction to Pig
    i. What is Pig
    ii. MapReduce Vs Pig
    iii. Pig Features
    iv. Pig Use Cases
b. Installing and Running Pig
    i. Downloading & Installing Pig
    ii. Pig Running Modes
c. Data Model in Pig
    i. Simple Data Types
    ii. Complex Data Types
    iii. Viewing the Schema
d. Analysis with Pig (Hands-On-Exercise)
    i. Pig Latin Syntax
    ii. Input and Output
    iii. Field Definitions
    iv. Filtering and Sorting Data
    v. Grouping
    vi. Iterating Grouped Data
    vii. Commonly – Used Functions
e. Multi-Data set Operations with Pig (Hands-On-Exercise)
    i. Joining Data Sets in Pig
    ii. Set Operations
    iii. Splitting Data Sets
f. Advanced Pig (Hands-On-Exercise)
    i. Macro & Imports
    ii. UDFs

7. Hive

a. Detailed Installation
    i. Installing Hive
    ii. Hive Command
    iii. Command-Line Interface
b. Introduction to Hive
    i. What is Hive?
    ii. Hive Use Case
    iii. Hive vs Pig
    iv. Hive Architecture and Components
    v. Comparing Hive to Traditional Databases
    vi. Hive Data Models
c. Deep Dive on Hive

i. Interacting With Hive
ii. Hive Databases and Tables
iii. Basic HiveQL Syntax
iv. Data Types
v. Joining Data Sets
vi. Common Built-in Functions

d. Hive Data Management
i. Hive Data Formats
ii. Creating Databases and Hive-Managed Tables
iii. Loading Data into Hive
iv. Altering Databases and Tables
v. Self-Managed Tables
vi. Queries with Views
vii. Storing Query Results
viii. Controlling Access to Data

e. Hive Optimization
i. Understanding Query Performance
ii. Partitioning
iii. Bucketing
iv. Indexing Data

f. Advanced Topics on Hive
i. User-Defined Functions

8. HBase

a. Introduction to HBase
b. Architecture of HBase
c. Where does it fits
d. What is NOSQL

9. Oozie

a. Why Oozie?
b. Installing Oozie
c. Running an example
d. Oozie- workflow engine
e. Example M/R action
f. Word count example
g. Workflow application and Submission
h. Workflow state transitions
i. Oozie job processing
j. Job submission to Hadoop

10. Apache Flume

a. Core concepts
b. Event
c. Clients
d. Agents

e. Source
f. Channels
g. Sinks
h. Interceptors
i. Channel selector
j. Sink processor

11. YARN

a. Introduction to Hadoop YARN
b. Difference between Hadoop 1.0 and 2.0