

Topics

Day 1- Introduction to Big Data and prerequisite technologies

Theory

- Introduction to Big Data
- Basic architecture of Hadoop
- Basics of Unix and useful Unix commands
- Restful Calls

Lab Sessions

- Hands-on running restful calls and Unix commands

Day 2- Hadoop - HDFS

Theory

- Prime techniques to handle/process to Big Data
- HDFS
- Basic architecture of HDFS
- Various ways to access HDFS
- Step-by-step example explained in detail

Lab Sessions

- Hands-on example of accessing HDFS

Day 3- MapReduce algorithm and implementation

Theory

- Understanding of Map and Reduce
- MapReduce algorithm
- Hadoop Framework, it's component and related technologies (e.g. HIVE, PIG, HBase etc.)
- Run through of one or two real use-cases of Hadoop in the industry
- Step-by-step simple word count problem explained in detail

Lab Sessions

- Hands-on implementation of word count problem
- **(and if time permits)** Dictionary translation (or log processing)

Day 4- NoSQL DBs and Real Time Analytics

Theory

- Big Data Analytics and NoSQL Storage
- Few associated technologies
- Introduction to HBase
- Interactive Analytics using -
 - Elasticsearch, Logstash and Kibana
- End-2-end data flow with a real use case explained
- Data ingestion and data access mechanism/techniques
- Analytics using Kibana

Lab Sessions

- Few hands-on example of data insertion and data access from Elasticsearch
- Prepare Analytics Dashboard for Apple Stock Trends (E2E)
- **(and if time permits)** New York City Traffic Incidents navigation

Day 5- Fitment of Cloud, Message Broker into Analytics

Theory

- Introduction to Cloud(AWS)
- Message brokers and their fitment into the data flow
- Walk through of Popular Message Brokers
 - ActiveMQ, RabbitMQ, Kafka etc.
- Convergence of Hadoop, Cloud, NoSQL, Message Broker towards Analytics

Lab Sessions

- Carried over assignments from previous day
- ActiveMQ
 - GUI InterfaceHands-on with sending and receiving messages using producers and consumers

Prerequisites and Requirements

Hadoop Training Course Prerequisites

- Basic Programming, Data Structure, Operating System
- Familiarity with Unix
- Core Java (OOPS Concepts, Collections , Exceptions) — For Map-Reduce Programming
- SQL Query knowledge

Hardware and Software Requirements

- Any Linux flavour OS (Ex: Ubuntu 14.04 is preferred) with 4 GB RAM (minimum), 100 GB HDD
- Java 1.8
- Open-SSH server & client
- MYSQL Database
- Eclipse IDE
- Virtual Box (To use Linux OS along with Windows OS)

WOIR Software will help in doing all the essential software setup. Please note that this assistance will be provided remotely. We will require persons who can execute our instructions for installing necessary software in your lab.