****

**Join us as beginner, walk out as an Expert**

**By Mr Suraz...**

**Our Teaching Strategy.**

* Focus on each and every concept of Hadoop.
* Map-Reduce will be taught for more than 45 sessions of at least 1 hour each
* More than 45 POCS only for Map-Reduce which illustrate different concepts.
* Free 25 Recorded session on Core Java to brush up your Knowledge on core Java
* Free First 18 Hadoop Recorded Sessions to make you comfortable with Hadoop setup and architecture & basic concept
* Tips and Trick while programming Map-reduce.
* Flow control of Each Program will be explained very clearly...
* Each Student will create their own Hadoop Cluster Setup
* Working on Latest and greatest Versions
* Hive ,Pig, Scoop, HBase, Oozie in details
* Complete guidelines on Cloudera Certifications.
* Special sessions on Eclipse tools to understand how to use them effectively.

# Big-Data and Hadoop(15 Hours)

* 1. Introduction to big data and Hadoop
  2. Hadoop Architecture
  3. Installing Ubuntu with Java 1.8 on VM Workstation 11
  4. Hadoop Versioning and Configuration
  5. Single Node Hadoop 1.2.1 installation on Ubuntu 14.4.1
  6. Single Node Hadoop 2.7.3 installation on Ubuntu 16.04
  7. Multi Node Hadoop 2.7.3 installation on Ubuntu 16.04
  8. Linux commands and Hadoop commands
  9. Cluster architecture and block placement
  10. Modes in Hadoop
      1. Local Mode
      2. Pseudo Distributed Mode
      3. Fully Distributed Mode
  11. Hadoop Daemon
      1. Master Daemons(Name Node, Secondary Name Node, Job Tracker)
      2. Slave Daemons(Job tracker, Task tracker)
  12. Task Instance
  13. Hadoop HDFS Commands
  14. Accessing HDFS
      1. CLI Approach
      2. Java Approach
  15. Installing and using Hadoop 2.X

## Map-Reduce(Using New API)(20 Hours)

* 1. Understanding Map Reduce Framework
  2. Inspiration to Word-Count Example
  3. Developing Map-Reduce Program using Eclipse Luna
  4. HDFS Read-Write Process
  5. Map-Reduce Life Cycle Method
  6. Serialization(Java)
  7. Data-types
  8. Comparator and Comparable(Java)
  9. Custom Output File
  10. Analysing Temperature dataset using Map-Reduce
  11. Custom Partitioner & Combiner
  12. Running Map-Reduce in Local and Pseudo Distributed Mode.

# Advanced Map-Reduce(25 Hours)

* 1. Enum(Java)
  2. Custom and Dynamic Counters
  3. Running Map-Reduce in Multi-node Hadoop Cluster
  4. Custom Writable
  5. Site Data Distribution
     1. Using Configuration
     2. Using DistributedCache
     3. Using stringifier
  6. Input Formatters
     1. NLine Input Format
     2. XML Input Format
     3. DB Input Format
     4. Sequence File Format
     5. Avro File Format
  7. Sorting
     1. Primary Reverse Sorting
     2. Secondary Sorting
  8. Joins
     1. Map-side Joins
     2. Reduce side Joins
  9. Compression Technique
     1. Gzip
     2. snappy
     3. bzip2
     4. deflate
  10. Processing Multiple Line using Map-Reduce
  11. Processing XML File using Map-Reduce
  12. TokenMapper
  13. Testing MapReduce with MR Unit
  14. Working with NYSE DataSets
  15. Running Map-Reduce in Cloudera Box

# HIVE(21 hours)

* 1. Hive Introduction & Installation
  2. Data Types in Hive
  3. Commands in Hive
  4. Exploring Internal and External Table
  5. Partitions
  6. Bucketing
  7. Complex data types(Array,Map,Structure)
  8. UDF in Hive
     1. Built-in UDF
     2. Custom UDF
  9. Thrift Server
  10. Java to Hive Connection
  11. Joins in Hive
  12. Working with HUE
  13. Bucket Map-side Join
  14. More commands
      1. View
      2. SortBy
      3. Distribute By
      4. Lateral View
  15. Working with Beeline
  16. Configure MySQL instead of Derby
  17. Working with HUE
  18. Performing update and delete in Hive
  19. Running Hive in Cloudera
  20. NYSE dataset Assignment in Hive
  21. Movie Rating Assignment in Hive

# SQOOP(6 hours)

* 1. Sqoop Installations and Basics
  2. Importing Data from Oracle to HDFS
  3. Advance Imports
  4. Working with sqoop and Hive
  5. Exporting Data from HDFS to Oracle
  6. Sqoop Metastore
  7. Real time use-case
  8. Running Sqoop in Cloudera
  9. Assignments

# PIG(18 Hours)

* 1. Installation and Introduction
  2. WordCount in Pig
  3. NYSE in Pig
  4. Working With Complex Datatypes
  5. Pig Schema
  6. Miscellaneous Command
     1. Group
     2. Filter
     3. Order
     4. Distinct
     5. Join
     6. Flatten
     7. Co-group
     8. Union
     9. Illustrate
     10. Explain
  7. UDFs in Pig
  8. Parameter Substitution and DryRun
  9. Processing XML file using Pig
  10. Pig Macros
  11. Testing Pig Scripts using PigUnit.
  12. Running Pig in Cloudera
  13. Assignments

# Hbase (9 Hours)

* 1. HBase Introduction & Installation
  2. Exploring HBase Shell
  3. Hbase Architecture
  4. HBase Storage Techinique
  5. HBasing with Java
  6. CRUD with HBase
  7. Map-Reduce HBase Integration
  8. Filters in Hbase
  9. Assignments

# OOZIE (6 Hours)

* 1. Installing Oozie
  2. Running Map-Reduce Program with Oozie
  3. Running Pig and Sqoop with Oozie
  4. Integrating Map-reduce,Pig,Hive with Oozie
  5. Running Coordinator Jobs
     1. Based on Particular time
     2. Based on Data Availability

1. **Project Works**

Working on Amazon dataset with advance map-reduce concept, Integrated with HBase, scheduled through oozie workflow.

1. **Side Topics**
2. MySQL Installation on Linux
3. Oracle Installation on Linux
4. Some assignments on ElasticSearch
5. Working with Maven
6. Using Junits
7. Eclipse Debugging
8. Java Best practices