HADOOP
COURSE CONTENT

Faculty:

Fee:

Duration:

Demo Time:

Batch Time:

- ✓ **What is BigData?**
  - o **Sources**
  - o **Characteristics**

- ✓ **What is Hadoop?**

- ✓ **Relation between Hadoop and Bigdata?**

- ✓ **Challenges of Big Data**

- ✓ **How to overcome the Bigdata challenges?**

- ✓ **How Hadoop is going to solve the Bigdata issues?**

- ✓ **olutions for Bigdata Problem**

- ✓ **Traditional Database systems Vs Hadoop**

- ✓ **Hadoop Ecosystems**

- ✓ **Real time scope and oppurtunities of Hadoop**

Hadoop Processes

HDFS (Hadoop Distributed File system)

- ✓ **Traditional File System Vs Distributed File System**
- ✓ **Evolution of HDFS**
- ✓ **Significance of HDFS in Hadoop**
- ✓ **History of Hadoop**
- ✓ **Storage daemons of Hadoop**
- ✓ **NameNode**
- ✓ **DataNode**
- ✓ **Secondary NameNode**
- ✓ **HDFS Architecture**
- ✓ **Storage options in Hadoop**
  - o **What is Block?**
  - o **How to configure blocksize?**
  - o **Can we change the block size?**
- ✓ **Replication Mechanism in Hadoop**
  - o **Failover mechanism**
  - o **Can we change the Replication factor in Hadoop?**
  - o **Reliability assurance in Hadoop**
- ✓ **Metadata maintenance in Hadoop**
  - o **Automatic metadata maintenance**

- ✓ **Design and drawbacks of HDFS**
- ✓ **Basic Unix Commands**
- ✓ **Accessing HDFS**
  - o **HDFS Commands**
- ✓ **Difference between Hadoop versions on high Availability of NameNode**

MapReduce

- ✓ **MapReduce Importance in Hadoop**

- ✓ **Processing Daemons of Hadoop**
  - o **Job Tracker**
  - o **Task Tracker**
  - o **Communication between Jobtracker and Tasktracker**
- ✓ **Components**
  - o **InputSplit and its importance**
  - o **Difference between Input split and Block size**
  - o **Input Split and Mappers**
- ✓ **MapReduce Programming Flow**
- ✓ **Phases involved in a MR flow**
- ✓ **MapReduce datatypes Vs primitive datatypes**
- ✓ **Writing a MapReduce Program**
  - o **Mapper Code**
  - o **Reducer Code**
  - o **Driver Code**
- ✓ **Input and Output Formats in MapReduce**
- ✓ **Examples and Assignments on MapReduce Programs**
- ✓ **Partitioner in MapReduce**
  - o **Importance of Partitioner**
- ✓ **Combiner in MapReduce**
  - o **Combiner importance in MapReduce**
  - o **Performance improvement with Combiner**

YARN

- ✓ **YARN(Yet Another Resource Negotiator) Importance**
- ✓ **YARN(Yet Another Resource Negotiator) Components**
  - o **Resource Manager**
  - o **Application Manager**
  - o **Application Master**
  - o **Node Manager**
  - o **Container**

- ✓ **YARN Architecture**
- ✓ **Difference between MapReduce and YARN**
- ✓ **Speculative Execution**
- ✓ **Interview questions and Answers on MapReduce**

Metadata Components
- ✓ **FSImage**
- ✓ **Fstime**
- ✓ **edits**
- ✓ **version**

Zookeeper

- ✓ **Zookeeper Datamodel**
- ✓ **Distributed Coordination**

Hadoop1 Vs Hadoop2

- ✓ **Difference between Hadoop Versions**
- ✓ **Difference between MR processes**
- ✓ **Interview oriented Important points**

Hive

- ✓ **Introduction to Hive**
- ✓ **Hive Importance in Hadoop**
- ✓ **Important points on Hive**
- ✓ **Design of Hive**
- ✓ **Hive Architecture**
  - o **Driver**
  - o **Compiler**
  - o **Optimizer**
  - o **Semantic analyzer**

- ✓ **Metastore in Hive**
    - o **Importance of Hive Metastore**
    - o **Internal and External Metastore in Hive**
    - o **Configuring External metastore**

- ✓ **Mysql configuration with Hive**
- ✓ **Hive and Hadoop Integration**
- ✓ **Datatypes in Hive**
    - o **Hive Complex datatypes**
    - o **array**
    - o **map**
    - o **struct**
    - o **union**
- ✓ **Examples to work on complex datatypes**
- ✓ **Datamodel of Hive**
- ✓ **Tables**
    - o **Managed Tables**
    - o **Temporary Tables**
    - o **External Tables**
- ✓ **Hive Optimization techiniques**
- ✓ **Hive Partitions**
    - o **Static Partition**
    - o **Dynamic Partition**
- ✓ **Usecases on Hive Partitions**
- ✓ **Bucketing concept**
    - o **Logic of Bucketization**

- ✓ **Usecases on Bucketisation**
- ✓ **User Defined Functions(UDFs) in Hive**
    - o **UDFs**
    - o **UDAFs**
    - o **UDTFs**

- ✓ **Serializer Deserializer(SerDe) in Hive**
- ✓ **Hive Storage formats**
- ✓ **Processing different files in Hive**
- ✓ **Joins in Hive**
- ✓ **Interview questions and answers**
  - ***Pig***
- ✓ **Introduction to Pig**
- ✓ **Pig datatypes**
- ✓ **Bag,Tuple,Map,Filed**
- ✓ **Schema Design of Pig**
- ✓ **Pig Installation(local and MR modes)**

- ***HBase***
- ✓ **Introduction to NoSQL**
- ✓ **Hbase Introduction**
- ✓ **Difference between RDBMS and HBase**
- ✓ **Hbase Components**
  - o **Column Family**
  - o **Qualifiers**
  - o **Row Key**
  - o **values**
- ✓ **Hbase Architecture**
  - o **Hmaster**
  - o **Region Server**
  - o **Zookeeper**
- ✓ **Installation**
- ✓ **Hbase CRUD operations**
  - o **create**
  - o **scan/get**
  - o **put**
  - o **delete/delete all/drop**

- ✓ **Filters**
- ✓ **Bulk Loading in Hbase**
- ✓ **Alter and HELP Commands**
- ✓ **Interview questions and Answers**

Phoenix

- ✓ **Introduction to Phoenix**
- ✓ **Installation**
- ✓ **Integration with Hbase**
- ✓ **Difference between Hbase and Phoenix**
- ✓ **Phoenix examples**

Sqoop

- ✓ **Introduction to Sqoop**
- ✓ **Sqoop import commands**
- ✓ **Sqoop export commands**
- ✓ **Sqoop Mappers**
- ✓ **Connection with RDBMS**
- ✓ **Integration with Hive**
- ✓ **Examples on Hive and Hbase import commands**

Mongo DB

- ✓ **Introduction to Mongo DB**
- ✓ **Mongo DB Installation**
- ✓ **Examples on Mongo DB**

Cassandra

- ✓ **Introduction to Cassandra**
- ✓ **Cassandra Installation**
- ✓ **Examples on Cassandra**

Oozie

- ✓ **Introduction to oozie**
- ✓ **Basics and Workflow**
- ✓ **Interview questions and Answers**

Spark and Scala:

- ✓ **Introduction to Spark**
- ✓ **Spark Versions**
- ✓ **Evolution and need of Spark**
- ✓ **Spark components**
  - o **Spark core**
  - o **Spark SQL**
  - o **Spark Streaming**
  - o **Spark MLLib**
  - o **Spark GraphX**
- ✓ **Resilient Distributed Dataset(RDD) and its features**
  - o **Transformations**
  - o **Actions**
- ✓ **Dataframe**

Scala

- ✓ **What is Scala?**
- ✓ **Java Vs Scala**
- ✓ **Scala Collections**
  - o **Mutable**
  - o **Immutable**
- ✓ **Scala Functions**
  - o **Named Functions**
  - o **Anonymoue Functions**
  - o **Curried Functions**

Admin Topics

**Installation of Hadoop,Hive,Hbase,Sqoop,Spark/scala etc., hadoop
Ecosystems.**

UNIX Basics

To the extent required for operations for Hadoop (Complimentary)

Highlights of the Course:

- ✓ **Subject oriented learning**
- ✓ **Clear understanding of basics**
- ✓ **Discussions on Interview scenario questions while covering topics**
- ✓ **Exclusive Access to a variety of latest interview questions and answers**
- ✓ **Soft copies for the important topics will be provided as discussed in the class**
- ✓ **Hand-outs will be given which would serve as a knowledge-check**
- ✓ **Assistance in Resume preparation**