



Hadoop 1.x.x

- Introduction to Hadoop
- Parallel Computing vs Distributed Computing
- How to install Hadoop on your system
- How to install Hadoop cluster on multiple machines
- Hadoop daemons introduction: NameNode, DataNode, JobTracker, TaskTracker
- Exploring HDFS (Hadoop Distributed File System)
- Exploring Apache HDFS web UI
- Namenode architecture (FS Image, Replica placement)
- Secondary Namenode architecture
- Datanode architecture

YARN (Hadoop 2.x.x)

- Introduction to YARN (Hadoop 2.x.x)
- Hadoop 1 Vs Hadoop 2
- Hadoop 2 installation
- Copy data from local file system to HDFS
- Execute Hadoop job on YARN
- Exploring HDFS/YARN/Job history UI
- Hands-On Exercise

Hadoop Administrative Tasks

- Routine Administrative Procedures
- Understanding dfsadmin and mradmin
- Block Scanner, HDFS Balancer
- Health Check & Safe mode
- Monitoring and Debugging on Hadoop cluster
- Namenode backup and recovery
- Datanode commissioning/decommissioning
- ACL (Access Control List)
- Upgrading Hadoop

MapReduce Architecture

- Exploring JobTracker/TaskTracker
- How to run a Map-Reduce job
- Exploring Mapper/Reducer/Combiner
- Shuffle: Sort & Partition
- Input/output formats
- Exploring Apache MapReduce web UI



Hadoop Developer Tasks

- Hadoop Eclipse integration
- Reading and writing data using Java
- How to write a Map-Reduce Job
- Mapper/Reducer in details
- Searching in HDFS
- Sorting in HDFS

HBase

- Introduction to HBase
- Installation of HBase on your system
- Exploring HBase Master & RegionServers
- Exploring Zookeeper
- Column Families and Qualifiers
- Basic HBase shell commands.
- Hands-On Exercise

Hive

- Introduction to Hive
- HBase vs Hive
- Installation of Hive on your system
- HQL (Hive query language)
- Basic Hive commands
- Hands-On Exercise
- Static and Dynamic partitioning
- Bucketing in hive
- Different File Formats (TEXT vs RC vs ORC)
- Text File processing in hive
- Parsing Json and Xml files in hive
- Handling complex datatypes in hive
- Joins in Hive
- UDF

Pig

- Introduction to Pig
- Installation of Pig on your system
- Basic Pig commands
- Hands-On Exercise
- Text file processing in pig

- Json and xml file processing in xml



- Built-in pig APIs Ex:- FixedWidthLoader
- UDF

Sqoop

- Introduction to Sqoop
- Installation of Sqoop on your system
- Import/Export data from RDBMS to HDFS
- Import/Export data from RDBMS to HBase
- Import/Export data from RDBMS to Hive
- Hands-On Exercises

Spark/Scala

- Introduction to Spark
- Hadoop vs Spark
- Installation of Spark on your system
- Introduction to Scala (Scalable programming language)
- Basic Scala commands
- Scala Building blocks
- Classes, Objects and Traits in Scala
- Collections in Scala
- Hands-On Exercise

Additional Topics

- **Apache Flume**
- **Apache Kafka**
- **Mongo DB**

Mini Project / POC (Proof of Concept)

- Facebook-Hive POC
- Usages of Hadoop/Hive @ Facebook
- Static & Dynamic partitioning
- UDF (User defined functions)
- Project usecases
- Hands-On Exercise