

## Hadoop Course Content

### **Hadoop Developer/Admin Training Course Content**

#### **Hadoop Architecture**

- ❖ Introduction to Hadoop
- ❖ Parallel Computer 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 the HDFS Apache Web UI
- ❖ NameNode architecture
- ❖ (EditLog, FsImage, location of replicas)
- ❖ Secondary NameNode architecture
- ❖ DataNode architecture

#### **MapReduce Architecture**

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

#### **Hadoop Developer Tasks**

- ❖ Writing a Map-Reduce programme
- ❖ Reading and writing data using Java
- ❖ Hadoop Eclipse integration
- ❖ Mapper in details
- ❖ Reducer in details
- ❖ Using Combiners
- ❖ Reducing Intermediate Data with Combiners
- ❖ Writing Partitioners for Better Load Balancing
- ❖ Sorting in HDFS
- ❖ Searching in HDFS
- ❖ Hands-On Exercise

#### **Hadoop Administrative Tasks**

- ❖ Writing a Map-Reduce programme
- ❖ Reading and writing data using Java
- ❖ Hadoop Eclipse integration
- ❖ Mapper in details
- ❖ Reducer in details
- ❖ Using Combiners
- ❖ Reducing Intermediate Data with Combiners
- ❖ Writing Partitioners for Better Load Balancing
- ❖ Sorting in HDFS
- ❖ Searching in HDFS
- ❖ Hands-On Exercise

#### **HBase Architecture**

- ❖ Routine Administrative Procedures
- ❖ Understanding dfsadmin and mradmin
- ❖ Block Scanner, Balancer
- ❖ Health Check & Safe mode
- ❖ Monitoring and Debugging on a production cluster
- ❖ NameNode Back up and Recovery
- ❖ DataNode commissioning/decommissioning
- ❖ ACL (Access control list)
- ❖ Upgrading Hadoop

#### **Hive Architecture**

- ❖ Introduction to Hive
- ❖ HBase vs Hive
- ❖ Installation of Hive on your system
- ❖ HQL (Hive query language )
- ❖ Basic Hive commands
- ❖ Hands-on-Exercise

#### **PIG Architecture hadoop**

- ❖ Introduction to Pig
- ❖ Installation of Pig on your system
- ❖ Basic Pig commands
- ❖ Hands-On Exercise

#### **Sqoop Architecture**

- ❖ 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 Exercise

#### **Mini Project / POC ( Proof of Concept )**

- ❖ Facebook-Hive POC
- ❖ Usages of Hadoop/Hive @ Facebook
- ❖ Static & dynamic partitioning
- ❖ UDF ( User defined functions )