

Hadoop Syllabus:

- About Hadoop , Installation
- Architecture – HDFS , MapReduce, Terminology
- Sample MapReduce Job – WordCount
- Block, Split
- Combiner , Custom Combiner
- Setup & Clean ()
- Partitioner logic ,Custom Partitioner
- Sequence file
- N line input format along with Record Reader
- Input/Output Format – Composite, Multiple
- Cluster setup (demo) ,
- Joins , Distributed Cache
- Compressions and HAR
- Custom Data types ,Custom Input Format
- Counters and Fair, Capacity, FIFO scheduler

PIG

- Installation , Architecture ,
- Datatypes (scalar , complex) ,
- Running Pig (interactive , Batch)
- Pig Operators – Local, Store,Dump,Distinct, Filter, ForEach, generate , Limit, Union ,
join, order by, Describe
- Group by , Avg Default UDFs available (Built in function) REG EX
EXPLAIN
- Parallel processing
- Custom UDF, How to use your custom UDF in your script

Hive

- Installation , Hive Services , Architecture , Comparing Hive to traditional Databases
- Relational Data Analysis – (data types (primitive,complex)databases-
tables,create,alter,delete)
- Hive Schema & Data storage Loading data into Hive views Storing query results
(store)
- Text processing - Built in functions , string functions , regular expressions
- Managed vs External Tables

- Optimization : Partitioning , bucketing , indexing data
- Extending Hive : Custom UDFs,Custom SerDes

HBase

- Introduction – history and evolution Installaton –standalone (distributed) – starting

hbase webUI

- Hbase shell introduction – storing , reading data with shell
- Data Model, Physical Model and Hbase Distributed Model.
- Hbase java client Read and Write paths

ZOOKEEPER

OOZIE

HBASE

SCOOP

FLUME

YARN

