

HADOOP DEVELOPMENT RESEARCH COURSE CONTENT

- 1. 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
- 2. Map Reduce Architecture
 - Exploring JobTracker/TaskTracker
 - How a client submits a Map-Reduce job
 - Exploring Mapper/Reducer/Combiner
 - Shuffle: Sort & Partition
 - Input/output formats
 - Job Scheduling (FIFO, Fair Scheduler, Capacity Scheduler) Exploring the Apache MapReduce Web UI
- 3. 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
- Indexing in HDFS
- Hands-On Exercise
- 4. Hadoop Administrative Tasks
 - Routine Administrative Procedures
 - Understanding dfsadmin and mradmin Block Scanner, Balancer
 - Health Check & Safe mode
 - Data Node commissioning/decommissioning
 - Monitoring and Debugging on a production
 - Cluster Name Node Backup and Recovery
 - ACL (Access control list) Upgrading Hadoop
- 5. HBase Architecture
 - Introduction to Hbase
 - HBase vs. RDBMS
 - Exploring HBase Master & region server
 - Column Families and Regions
 - Basic Hbase shell commands
- 6. Hive Architecture
 - Introduction to Hive
 - HBase vs Hive
 - Installation of Hive
 - HQL (Hive query language)
 - Basic Hive commands
- 7. Pig Architecture
 - Introduction to Pig
 - Installation of Pig on your system
 - Basic Pig commands
 - Hands-On Exercise

- 8. 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
- 9. Mini Project / POC (Proof of Concept)
 - Facebook-Hive POC
 - Usages of Hadoop/Hive @ Facebook
 - Static & dynamic partitioning