Hadoop Developer/
Admin Training Course Content

AmeriCloud
SOLUTIONS

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 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

Hadoop Developer Tasks
- Writing a map-reduce program
- Mapper in details
- Reducer in details
- Using Combiners (Advanced Topic)
- Sorting in HDFS
- Searching in HDFS
- Indexing in HDFS
- Hands-On Exercise

Hadoop Administrative Tasks
- 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
- HQL (Hive query language)
- Basic Hive commands
- Hands-On Exercise

Pig Architecture
- Introduction to Pig
- Basic Pig commands
- Using Pig shell
- Pig Latin concepts
- Design of data pipelines
- Hands-On Exercise

Sqoop Architecture
- Introduction to Sqoop
- Installation of Sqoop
- Import data from RDBMS to HDFS
- Hands-On Exercise

Oozie
- Introduction to Oozie
- Installation
- Write Oozie workflow
- Deploy and run Oozie workflow

Scala Crash course
- Quick tour of Scala
- Scala functions
- Hands-on exercises

Introduction to Spark
- Spark deconstructed
- Spark essentials
- Hands-on Spark
- SparkSQL
- Hands-on SparkSQL
- Case studies