Hadoop Administration - COURSE CONTENT

Ch1: Introduction to Hadoop

- ☐ The amount of data processing in today's life
- ☐ What Hadoop is why it is important?
- ☐ Hadoop comparison with traditional systems
- ☐ Hadoop history
- ☐ Hadoop main components and architecture

Ch2: Hadoop Distributed File System (HDFS)

- ☐ HDFS overview and design
- ☐ HDFS architecture
- ☐ HDFS file storage
- ☐ Component failures and recoveries
- ☐ Block placement
- ☐ Balancing the Hadoop cluster

Ch3: Planning your Hadoop cluster

- ☐ Planning a Hadoop cluster and its capacity
- ☐ Hadoop software and hardware configuration
- ☐ HDFS Block replication and rack awareness
- ☐ Network topology for Hadoop cluster

Ch4: Hadoop Deployment

- ☐ Different Hadoop deployment types
- ☐ Hadoop distribution options

- ☐ Hadoop competitors
- ☐ Hadoop installation procedure
- ☐ Distributed cluster architecture

Ch5: Working with HDFS

- ☐ Ways of accessing data in HDFS
- ☐ Common HDFS operations and commands
- ☐ Different HDFS commands
- ☐ Internals of a file read in HDFS
- ☐ Data copying with 'distcp'

Ch6: Map-Reduce Abstraction

- ☐ What MapReduce is and why it is popular
- ☐ The Big Picture of the MapReduce
- ☐ MapReduce process and terminology
- ☐ MapReduce components failures and recoveries
- ☐ Working with MapReduce

Ch7: Hadoop Cluster Configuration

- ☐ Hadoop configuration overview and important configuration file
- ☐ Configuration parameters and values
- ☐ HDFS parameters MapReduce parameters
- ☐ Hadoop environment setup
- ☐ 'Include' and 'Exclude' configuration files

Ch8: Hadoop Administration and Maintenance

- ☐ Namenode/Datanode directory structures and files

- ☐ File system image and Edit log

- ☐ The Checkpoint Procedure

- ☐ Namenode failure and recovery procedure

- ☐ Safe Mode

- ☐ Metadata and Data backup

- ☐ Potential problems and solutions / what to look for

- ☐ Adding and removing nodes

Ch9: Hadoop Monitoring and Troubleshooting

- ☐ Best practices of monitoring a Hadoop cluster

- ☐ Using logs and stack traces for monitoring and troubleshooting

- ☐ Using open-source tools to monitor Hadoop cluster

Ch10: Job Scheduling

- ☐ How to schedule Hadoop Jobs on the same cluster

- ☐ Default Hadoop FIFO Schedule

- ☐ Fair Scheduler and its configuration

Ch11: Hadoop Multi Node Cluster Setup

- ☐ Running Map Reduce Jobs on Amazon Ec2

- ☐ Hadoop Multi Node Cluster Setup using Amazon ec2 - Creating 4 node cluster setup

- ☐ Running Map Reduce Jobs on Cluster

- ☐ High Availability Fedration, Yarn and Security