Hadoop Administrator

SevenMentor
PVT.LTD

www.sevenmentor.com

Session 1

a) Why Hadoop
Discussion about the drawbacks of traditonal RDBMS and why Hadoop is better than traditional RDBMS.
b) Introduction to Hadoop
Discussion about HDFS Architecture. Namenode, Secondary Namenode, Resource Manager, Data node and Node manager.

Session 2

a) Introduction to AWS Cloud
Brief discussion about cloud technologies in Industry. Why cloud is better than baremetal and VM's. Discussion about various components which we are going to learn through the course duration (AWS EC2, AWS S3, EMR, VPC, Snapshots, AMI, IAM)
b) Basic Networking concepts
Discussion about basic networking concepts that are going to be required in Hadoop.
c) Introduction to Linux
Discussion on why Linux skillset is important for Hadoop. Overview of Linux and practising basic commands.

Session 3

a) AWS Cloud (AWS EC2)
Hands on practise on AWS EC2. Deployment of an instance using AWS EC2 and connecting to that instance using terminal and putty.
b) Hands on Practise of Linux Concepts
When connected to the instance deployed through AWS Ec2, hands on practise of linux commands will be done to understand linux concepts.

Session 4

a) Single Node Architecture (Hadoop 1x)
Deployment of Single Node Hadoop Cluster using command line interface and discussing hadoop daemons.
Accessing Namenode UI, Resource Manager UI
Running a mapreduce job.

Session 5

a) Multi Node Architecture (Hadoop 1x)
Deployment of Multi Node Hadoop Cluster using command line interface and discussing hadoop daemons.
Setting up password less login architecture.
Accessing Namenode UI, Resource Manager UI

Session 6

a) AWS S3 and EMR
Creating a S3 bucket and storing the data into it.
Creating a cluster in EMR and processing the data stored in S3 bucket, after processing storing the result back into the S3 bucket.

Session 7

a) Hadoop Eco-system (Hive, Flume, Sqoop and Pig)
Installing, configuring and using Apache Hive on hadoop cluster.
Installing, configuring and using Apache Pig on hadoop cluster.
Installing, configuring and using Apache Flume on hadoop cluster.
Installing, configuring and using Apache Sqoop on hadoop cluster.

Session 8

a) Hadoop 2x
Why hadoop 2x is better than hadoop 1x
Discussion on Prerequisites of hadoop 2x
Deploying Single Node hadoop 2x cluster

Session 9

Multi node hadoop 2x cluster, (AWS: Snapshots and AMI)
Creating an image of hadoop 2x prerequistes on AWS cloud
Deploying hadoop 2x multi node architecture using that image.

Session 10

a) Hortonworks cluster
Deploying hortonworks cluster using Ambari
Performing basic admin tasks on Hortonworks cluster
b) Cloudera cluster
Deploying Cloudera cluster using Cloudera Manager
Performing basic admin tasks on Cloudera cluster

Session 11

a) Cloudera cluster
Deploying Cloudera cluster using Cloudera Manager
Performing basic admin tasks on Cloudera cluster

Session 12

a) Cloudera cluster Admin Tasks
Deploying Cloudera cluster using Cloudera Manager
Accessing Namenode UI, Resource Manager UI, submitting job, allocating resources to job.
Enabling and understanding concepts of High Availability concets in cloudera cluster.

SevenMentor PVT.LTD

Session 13

a) Cloudera Director
Deploying Cloudera Manager using Cloudera Director.
Discussing on shell scripts to deploy Cloudera Director.

Session 14

Enabling Kerberos using MIT
Understanding hadoop insecurities
Understanding basic concepts of kerberos
Understanding and creating a VPC in AWS

Session 15

Enabling Kerberos in Cloudera cluster using Active Directory
Understanding roles of sentry in a cloudera cluster.

Session 16

a) Capacity Planning
Brief discussion about how a cluster should be planned according to the requirements of the organization.
b) HBase
Brief discussion about HBase concepts and deploying HBase in cloudera cluster.
c) Kafka
Brief discussion about kakfa concepts and deploying kafka in cloudera cluster.
d) Project Discussion