**Day4 : HADOOP**

\*\*GOOGLE - slashbigdata (blog for all this – go and search there hadoop and select first one)

Starting with Hadoop :

$cd hadoop2

$ls

$cd etc

$ls

$cd hadoop

$pwd

$copy the java path

$vi hadoop-env.sh

-change the path of export Java home =/usr/java/jdk1.8.0

**#entry for data node**

$ vi hdfs-site.xml - show system is either name/data node

-<configuration>

<property>

<name>

Dfs.datanode.data.dir

</name>

<value>/home/ec2-user/omendra</value>

</property>

</configuration>

~

**#entry for name node**

$ vi hdfs-site.xml - show system is either name/data node

-<configuration>

<property>

<name>

Dfs.namenode.name.dir

</name>

<value>/home/ec2-user/omendra</value>

</property>

</configuration>

\*there is only one name node in a cluster

**#now copy the ipv4 name from instance and paste in the machine**

**So that we can accesss from anwhere:**

$sudo hostnamectl set-hostname ec2-13-234-33-51.ap-south-1.compute.amazonaws.com

**#now fix the file name for the ip in /etc/hosts**

172.31.40.50 ec2-13-234-33-51.ap-south-1.compute.amazonaws.com

**#now give the name of namecode in core-xml**

Bcoz every datanode has duty of giving info to name node in **every 3sec.** that we are active for receiving data known as **heartbeat**

Otherwise name node do not send any data to datanode

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://ec2-34-197-238-12.compute-1.amazonaws.com:9001</value>

</property>

</configuration>

**To make folder in namenode (only namenode)**

$hdfs namenode -format

$hadoop-daemon.sh start namenode

**To make folder in datanode (only namenode)**

$hadoop-daemon.sh start datanode

Now check the ls we get the folder named in the file