# **Apache Hive Installation on Ubuntu**

****

**This include**

1. **Installing hadoop**

Prerequisites : Hadoop 3.1.2 is already installed on Ubuntu

<https://phoenixnap.com/kb/install-hadoop-ubuntu>

1. **Installing hive**

Refer the available

How to Install Apache Hive on Ubuntu {Step-by-Step Guide}.html

For complete installation steps

<https://phoenixnap.com/kb/install-hive-on-ubuntu>

Apart from that need add entry in  hive-site.xml

<https://stackoverflow.com/questions/27099898/java-net-urisyntaxexception-when-starting-hive>

1. **Install Mysql as metastore for hive**

Default metastore used derby DB, which allow only single instance of hive session, so suggested to use mysql for hive metastore

Fore steps to follow refer

<https://issues.apache.org/jira/browse/HIVE-16800>

For configuration and user creation refer

<https://dzone.com/articles/how-configure-mysql-metastore>

No need to do the step

mysql> SOURCE $HIVE\_HOME/scripts/metastore/upgrade/mysql/hive-schema-0.14.0.mysql.sql;

When run hive/bin/schematool -dbType mysql -initSchema

It will auto create all the required schema for the metastore in configured mysql db

Note: there can be issues during installation, need check

# Steps for docker:

If have docker image for ubuntu.

Then all the installation can be done in docker image

A new user hadoop can be created for installing the hadoop and hive

To dowload ubuntu image and start ubuntu docker container

docker pull ubuntu

docker run -it -v /home/nandu/dockershare/:/home/dockershare ubuntu

Perform the above steps in docker to install hadoop, hive and mysql

Create a start\_hadoop.sh file under / directory in docker with following content

/etc/init.d/ssh start

/etc/init.d/mysql start

su - hadoop -c "/home/hadoop/hadoop-3.2.1/sbin/start-dfs.sh"

su - hadoop -c "/home/hadoop/hadoop-3.2.1/sbin/start-yarn.sh"

su - hadoop -c "/home/hadoop/hadoop-3.2.1/bin/hdfs dfsadmin -safemode leave"

su - hadoop

Check the docker process

docker ps -a

Commit the changes done in container

docker commit -m "ubuntu hadooop hive mysql" -a user\_name

Docker\_image\_id new\_docker\_image\_name

Can Save the container to file for back an also use it incase of failure.

docker save -o path\_docker\_imagname

docker load -i path\_docker\_imagname

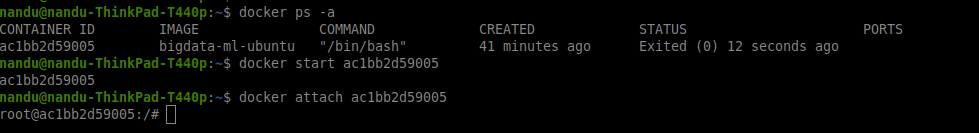
Run the new docker image

docker run -it -v /home/nandu/dockershare/:/home/dockershare new\_docker\_image\_name

If container alredy created then can run and attach to it

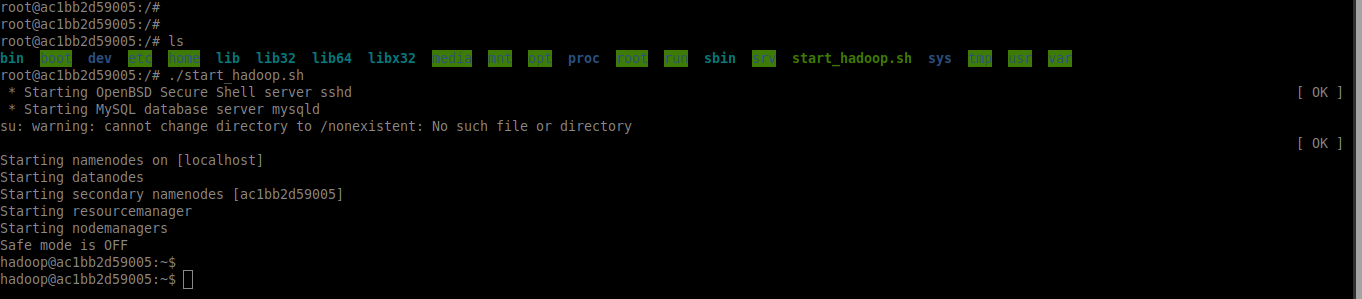
docker start containter\_id

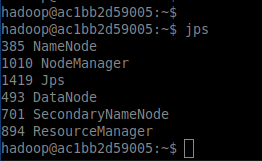
docker attach container\_id



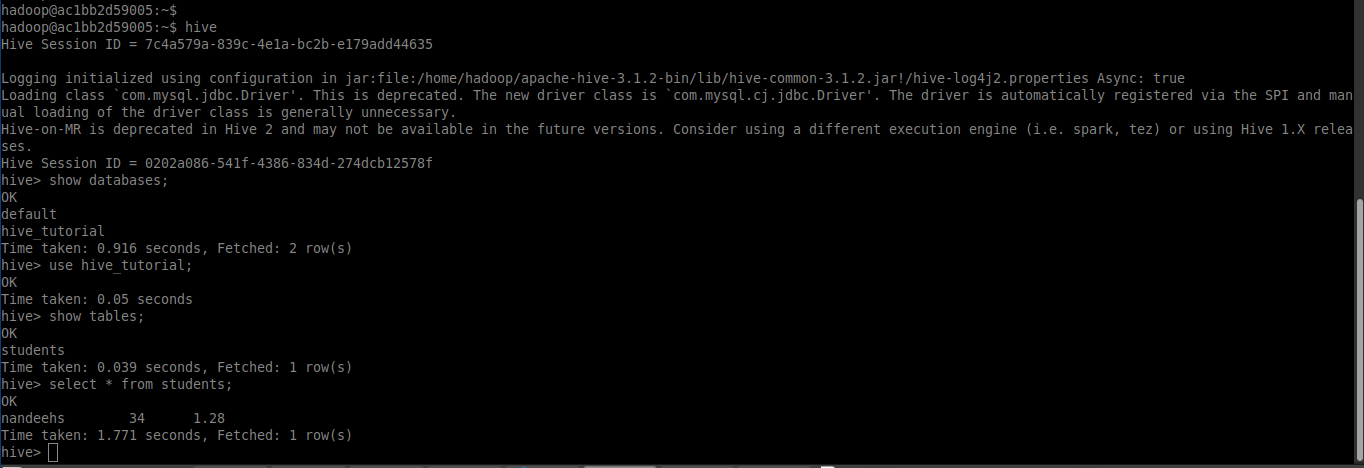
After login to container run the file start\_hadoop.sh to start ssh, mysql and hadoop services

Can use jps to check the hadoop process running.

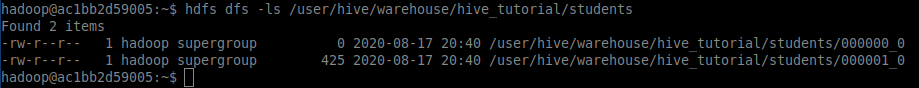




Login to hive and perform the operations.



Can see the newly created hive data in hdfs



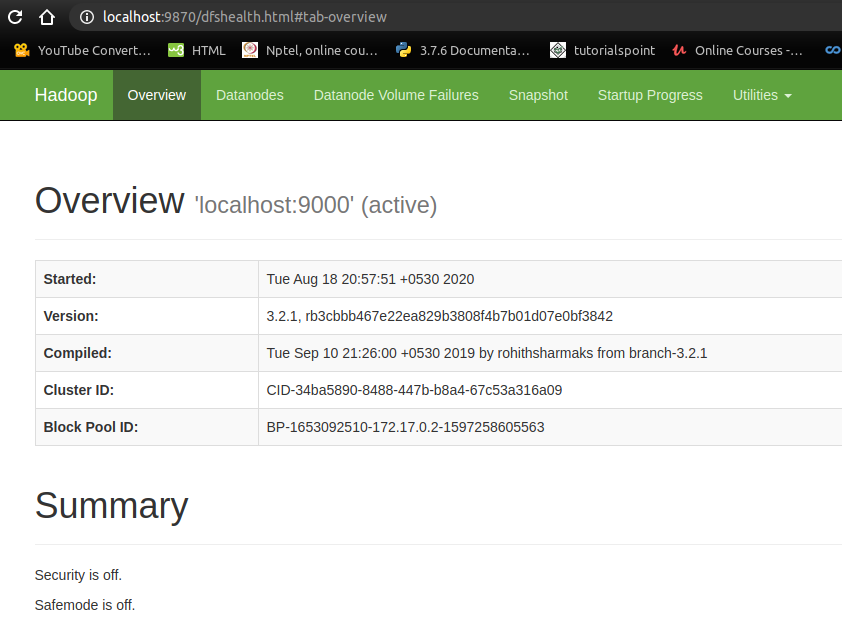
Extended:

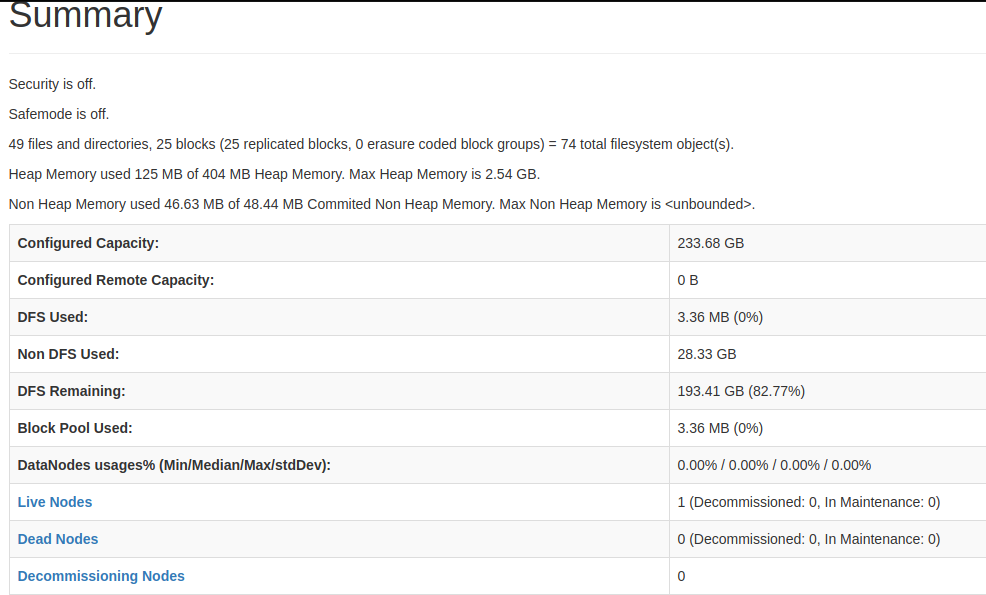
Can start the docker container with port mapping

docker run -it -p 9870:9870 -p 9864:9864 -p 8088:8088 -p 3306:3306 -v /home/nandu/dockershare/:/home/dockershare bigdata-ml-ubuntu

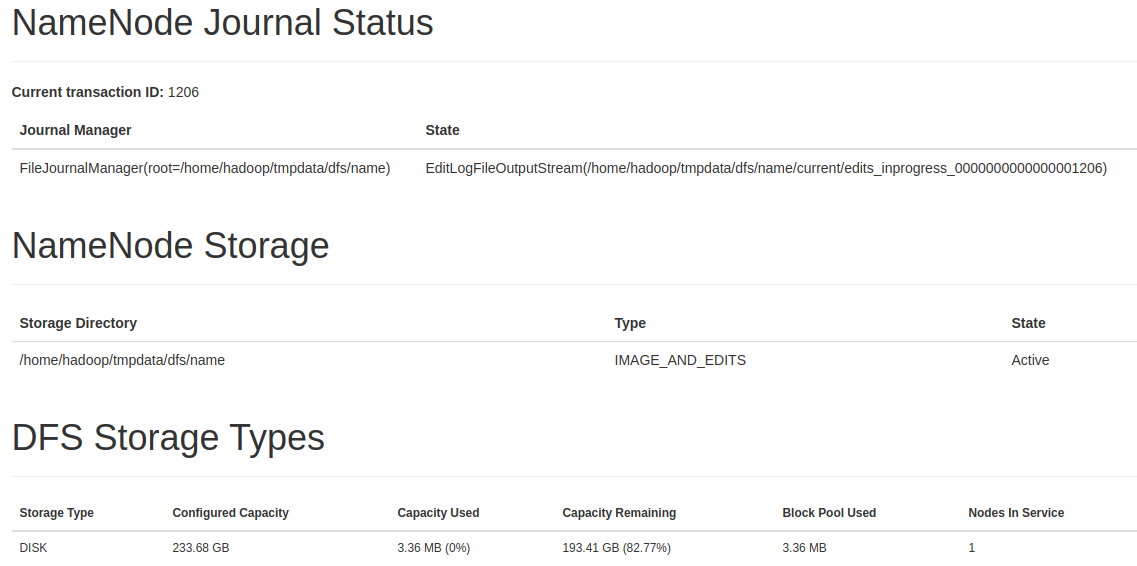
After this can access the hadoop pages

 Hadoop NameNode UI: <http://localhost:9870>

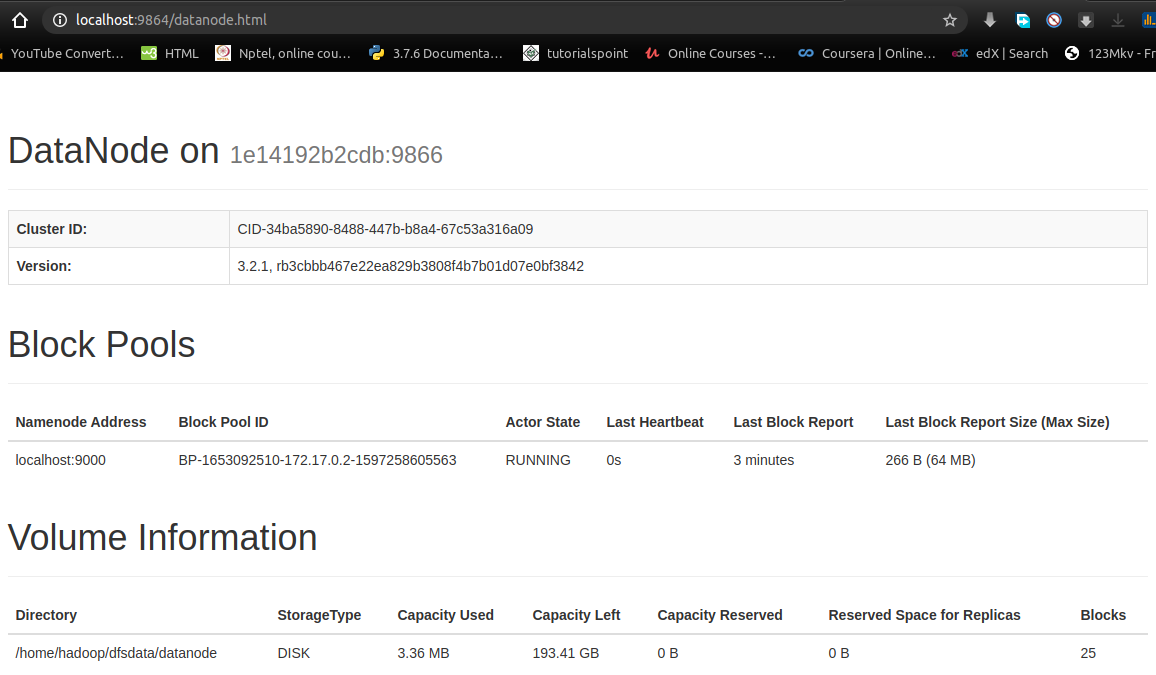








DataNodes information : http://localhost:9864

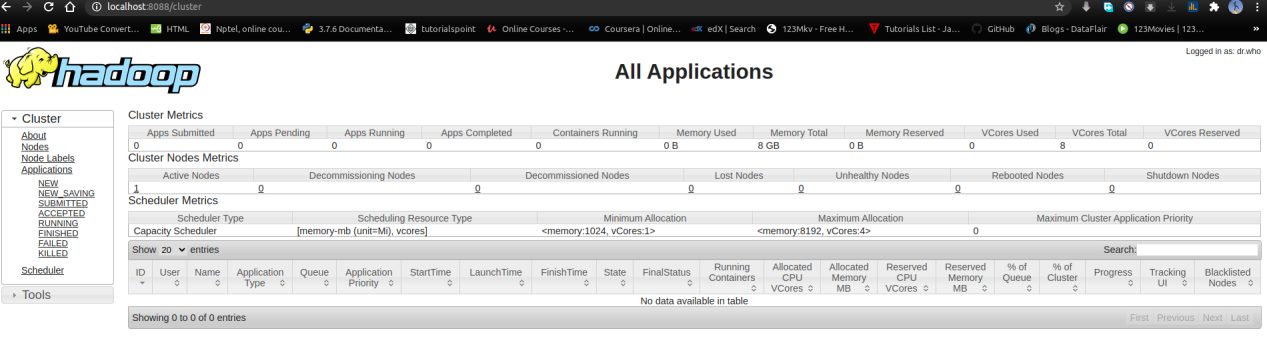


YARN Resource Manager : http://localhost:8088

Note: make sure to set the yarn.resourcemanager.hostname in yarn-site.xml as 0.0.0.0

If not then this page will not be accessible

<https://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-common/yarn-default.xml>



Can connect to mysql databases using dbweaver

Note: in the file /etc/mysql/mysql.conf.d/mysqld.cnf set bind-address to 0.0.0.0 and restrt the mysql service in docker to connect

