Aim: Implement various operations on Hive (Create, Insert, Update)

Objective: Study, Understand and Implement various operations on data using Hive.

Theory:

Hive:

Apache Hive is a data warehouse and an ETL tool which provides an SQL-like interface between the user and the Hadoop distributed file system (HDFS) which integrates Hadoop. It is built on top of Hadoop. It is a software project that provides data query and analysis. It facilitates reading, writing and handling wide datasets that stored in distributed storage and queried by Structure Query Language (SQL) syntax. It is not built for Online Transactional Processing (OLTP) workloads. It is frequently used for data warehousing tasks like data encapsulation, Ad- hoc Queries, and analysis of huge datasets. It is designed to enhance scalability, extensibility, performance, fault-tolerance and loose-coupling with its input formats.

Implementation:

1. At the most we need to start the cluster using the start-all.sh script further check all daemons are running using the jsp command

2. Once we have the cluster running we also need to download the tar.gz file which we will be using to install the hive in our system. To install the hive we will be just extracting the data in the file tar -xzfs <tar file name>

3. Now folder will be visible for Apache-hive in which we have the bin where the executable for the hive is present which will be using to get the hive shell

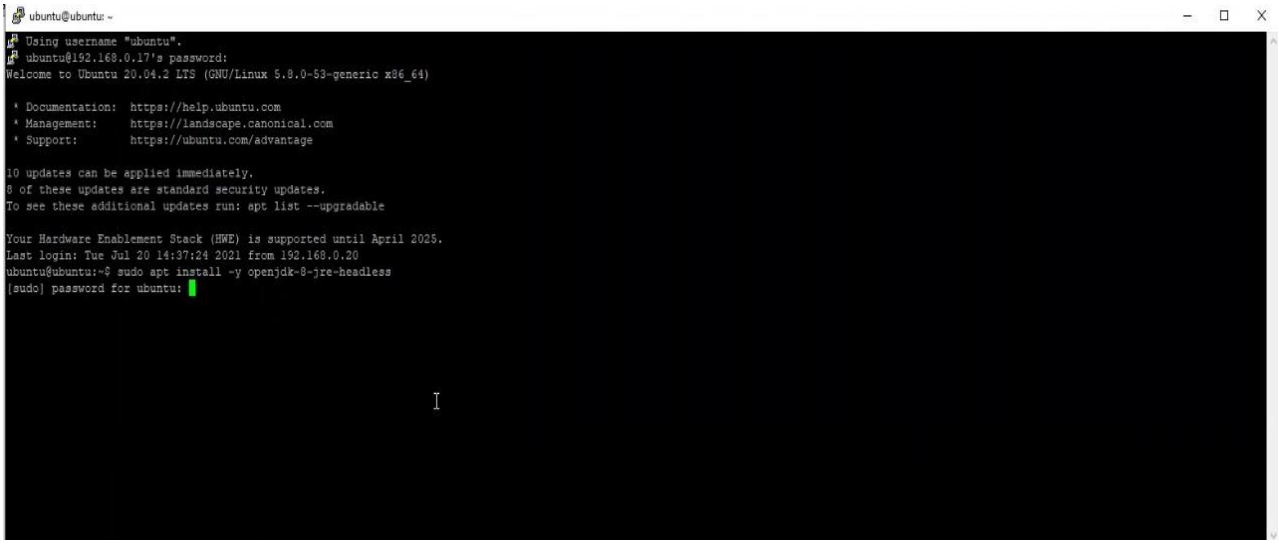
4. Once we get the hive shell ready we will be using the HQL command first we will be creating our database using the use command which will switch us to the new database and if not existed it will create and then switch.

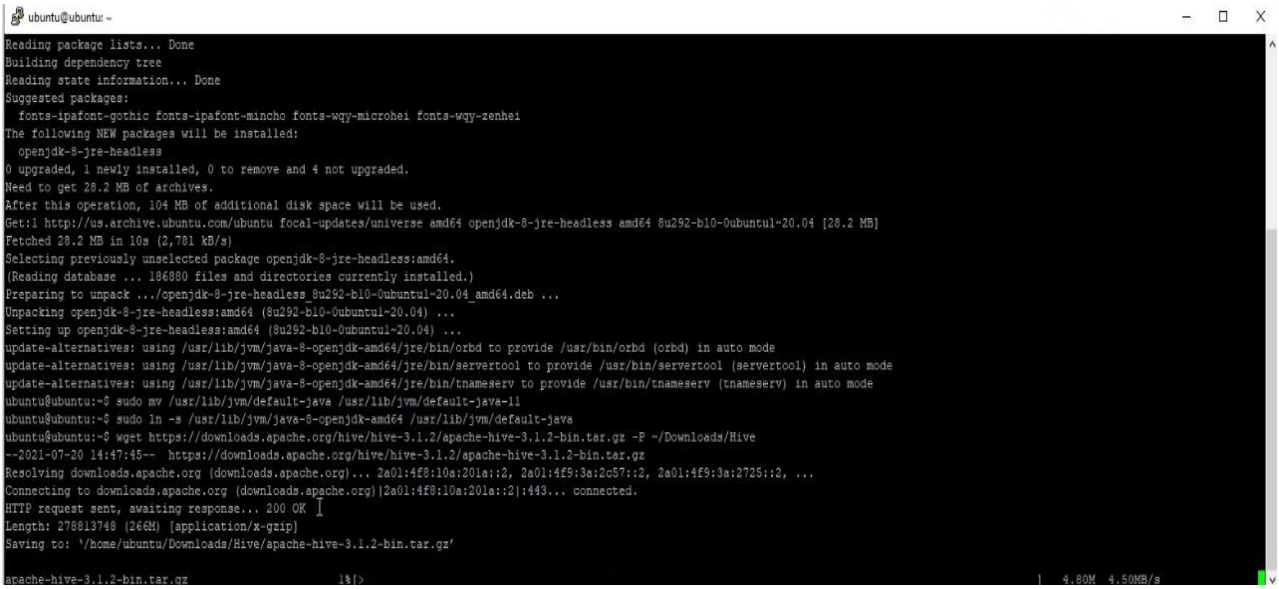
$create database <database_name>

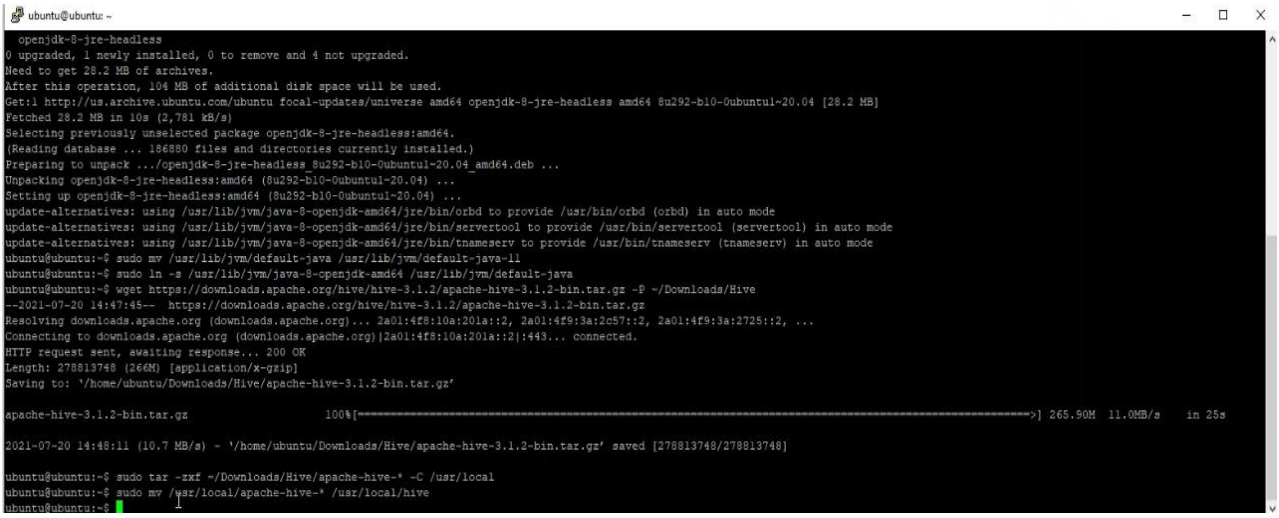
$use <database_name>

5. We can ask to show the tables using the $show tables command further we can also create the table using the $create table <table_name> (<column_name> <type of value>, …) query.

6. We can now insert necessary values in the table using the insert values command.


ubuntu@ubuntu: ~
Using username "ubuntu".
ubuntu@192.168.0.17's password:
Welcome to Ubuntu 20.04.2 LTS (GNU/Linux 5.8.0-53-generic x86_64)
* Documentation: https://help.ubuntu.com
* Management: https://landscape.canonical.com
* Support: https://ubuntu.com/advantage
10 updates can be applied immediately.
8 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable
Your Hardware Enablement Stack (HWE) is supported until April 2025.
Last login: Tue Jul 20 14:37:24 2021 from 192.168.0.20
ubuntu@ubuntu:~$ sudo apt install -y openjdk-8-jre-headless
[sudo] password for ubuntu:


ubuntu@ubuntu: ~
Reading package lists... Done
Building dependency tree
Reading state information... Done
Suggested packages:
fonts-ipafont-gothic fonts-ipafont-mincho fonts-wqy-microhei fonts-wqy-zenhei
The following NEW packages will be installed:
openjdk-8-jre-headless
0 upgraded, 1 newly installed, 0 to remove and 4 not upgraded.
Need to get 28.2 MB of archives.
After this operation, 104 MB of additional disk space will be used.
Get:1 http://us.archive.ubuntu.com/ubuntu focal-updates/universe amd64 openjdk-8-jre-headless amd64 8u292-b10-0ubuntu1~20.04 [28.2 MB]
Fetched 28.2 MB in 10s (2,781 kB/s)
Selecting previously unselected package openjdk-8-jre-headless:amd64.
(Reading database ... 186880 files and directories currently installed.)
Preparing to unpack .../openjdk-8-jre-headless_8u292-b10-0ubuntu1~20.04_amd64.deb ...
Unpacking openjdk-8-jre-headless:amd64 (8u292-b10-0ubuntu1~20.04) ...
Setting up openjdk-8-jre-headless:amd64 (8u292-b10-0ubuntu1~20.04) ...
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/orbd to provide /usr/bin/orbd (orbd) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/servertool to provide /usr/bin/servertool (servertool) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/tnameserv to provide /usr/bin/tnameserv (tnameserv) in auto mode
ubuntu@ubuntu:~$ sudo mv /usr/lib/jvm/default-java /usr/lib/jvm/default-java-11
ubuntu@ubuntu:~$ sudo ln -s /usr/lib/jvm/java-8-openjdk-amd64 /usr/lib/jvm/default-java
ubuntu@ubuntu:~$ wget https://downloads.apache.org/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz -P ~/Downloads/Hive
--2021-07-20 14:47:45-- https://downloads.apache.org/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 2a01:4f8:10a:201a::2, 2a01:4f9:3a:2c57::2, 2a01:4f9:3a:2725::2, ...
Connecting to downloads.apache.org (downloads.apache.org)|2a01:4f8:10a:201a::2|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 278813748 (266M) [application/x-gzip]
Saving to: '/home/ubuntu/Downloads/Hive/apache-hive-3.1.2-bin.tar.gz'
apache-hive-3.1.2-bin.tar.gz 1%[>] 4.80M 4.50MB/s


ubuntu@ubuntu: ~
openjdk-8-jre-headless
0 upgraded, 1 newly installed, 0 to remove and 4 not upgraded.
Need to get 28.2 MB of archives.
After this operation, 104 MB of additional disk space will be used.
Get:1 http://us.archive.ubuntu.com/ubuntu focal-updates/universe amd64 openjdk-8-jre-headless amd64 8u292-b10-0ubuntu1~20.04 [28.2 MB]
Fetched 28.2 MB in 10s (2,781 kB/s)
Selecting previously unselected package openjdk-8-jre-headless:amd64.
(Reading database ... 186880 files and directories currently installed.)
Preparing to unpack .../openjdk-8-jre-headless_8u292-b10-0ubuntu1~20.04_amd64.deb ...
Unpacking openjdk-8-jre-headless:amd64 (8u292-b10-0ubuntu1~20.04) ...
Setting up openjdk-8-jre-headless:amd64 (8u292-b10-0ubuntu1~20.04) ...
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/orbd to provide /usr/bin/orbd (orbd) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/servertool to provide /usr/bin/servertool (servertool) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/tnameserv to provide /usr/bin/tnameserv (tnameserv) in auto mode
ubuntu@ubuntu:~$ sudo mv /usr/lib/jvm/default-java /usr/lib/jvm/default-java-11
ubuntu@ubuntu:~$ sudo ln -s /usr/lib/jvm/java-8-openjdk-amd64 /usr/lib/jvm/default-java
ubuntu@ubuntu:~$ wget https://downloads.apache.org/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz -P ~/Downloads/Hive
--2021-07-20 14:47:45-- https://downloads.apache.org/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 2a01:4f8:10a:201a::2, 2a01:4f9:3a:2c57::2, 2a01:4f9:3a:2725::2, ...
Connecting to downloads.apache.org (downloads.apache.org)|2a01:4f8:10a:201a::2|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 278813748 (266M) [application/x-gzip]
Saving to: '/home/ubuntu/Downloads/Hive/apache-hive-3.1.2-bin.tar.gz'
apache-hive-3.1.2-bin.tar.gz 100%[==========>] 265.90M 11.0MB/s in 25s
2021-07-20 14:48:11 (10.7 MB/s) - '/home/ubuntu/Downloads/Hive/apache-hive-3.1.2-bin.tar.gz' saved [278813748/278813748]
ubuntu@ubuntu:~$ sudo tar -zxf ~/Downloads/Hive/apache-hive-* -C /usr/local
ubuntu@ubuntu:~$ sudo mv /usr/local/apache-hive-* /usr/local/hive
ubuntu@ubuntu:~$

Conclusion/ outcome: Thus we have successfully performed and understood the create, and insert, operations using the HQL in Apache hive.