

Name:-Preksha Patel

Sapid:-60004210126

Branch:-Computer Engineering

Div:-C2 ; Batch:-1

BDI-Experiment 4

BDI

SDS	Page No.
Date	

Name:- Preksha A. Patel
Sapid:- 60004210126
Branch:- Computer Engineering
Div:- C2, Batch:-1

Experiment no. 04
HIVE COMMANDS

Aim :- Execution of HIVE commands in Hortonworks sandbox

Theory :-

- 1) show databases ;
This command will display a list of databases present in the HIVE metastore.
- 2) use away-db ;
This command uses the database named "away-db" for all the further operations performed.
- 3) create table employee (eid int, ename varchar(20), address varchar(20), emobile bigint)
partitioned by (load_date date),
clustered by (eid) into 3 buckets,
stored as orc, + properties ('transactional'='true');
This query is used to create a table named employee with columns and clustered by 'eid' into 3 buckets. The table is stored in ORC format with transactional support enabled. This query will create an employee table with specified schema and properties.
- 4) Insert into employee partition
(load_date = '2024-02-14') values
(018, "Tina", "mike park", 123456578);
To insert data into 'employee' table — partitioned by 'load_date'. This query inserts a single row into 'employee' table and assigns it to the partition "load_date" = "2024-02-14".

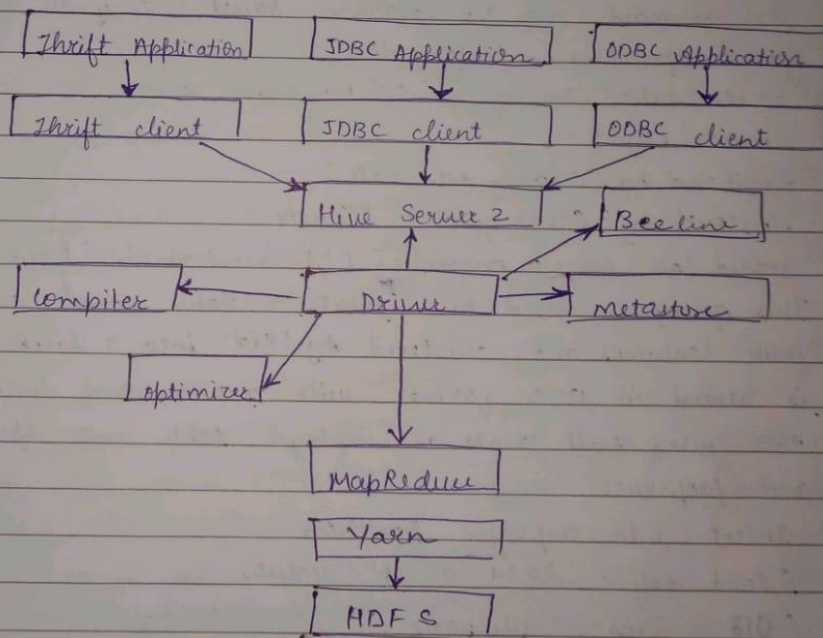
PTO

5) select * from employee ;

This query will retrieve all the columns and rows from the 'employee' table.

Hive is a data warehousing infrastructure based on Hadoop, facilitating querying and analysis of large datasets stored in Hadoop's distributed storage. It uses a language similar to SQL, called HiveQL. Hive translates queries into MapReduce or jobs, providing similar SQL interface for big data processing and analysis.

HIVE ARCHITECTURE :-



Conclusion :-

Therefore, we have executed Hive commands successfully in Hortonworks sandbox.

1) Show databases

```
hive> show databases;  
OK  
default  
foodmart  
xademo  
Time taken: 1.332 seconds, Fetched: 3 row(s)
```

2) Create database

```
hive> create database gandhar;  
OK
```

3) Use database

```
hive> use gandhar;  
OK  
Time taken: 0.398 seconds
```

4) Create table

```
hive> create table student(id int, name varchar(20), branch varchar(20), mobile int)  
> PARTITIONED BY(load_data date)  
> CLUSTERED BY(id) INTO 3 BUCKETS  
> STORED AS ORC TBLPROPERTIES('transactional' = 'true');  
OK  
Time taken: 1.34 seconds
```

5) Inserting values into table

```
hive> insert into student partition (load_data='2023-04-09') values ('101', 'Manav', 'Computer', 123456);  
Query ID = root_20240214035518_c605d715-45d8-4860-8fdc-bee34ea0838e  
Total jobs = 1  
Launching Job 1 out of 1  
  
Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)  
  
-----  
VERTICES    STATUS    TOTAL    COMPLETED    RUNNING    PENDING    FAILED    KILLED  
-----  
Map 1 ..... SUCCEEDED      1          1          0          0          0          0  
-----  
VERTICES: 01/01 [=====>>>] 100% ELAPSED TIME: 7.70 s  
-----  
Loading data to table gandhar.student partition (load_data=2023-04-09)  
Partition gandhar.student{load_data=2023-04-09} stats: [numFiles=1, numRows=0, totalSize=858, rawDataSize=0]  
OK  
Time taken: 12.551 seconds
```



```
hive> insert into student partition (load_data='2023-04-10') values ('102', 'Aditya', 'Computer', 123233);
Query ID = root_20240214035726_51e6bad9-c967-42da-a2eb-6e197a56bbfe
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01 [=====>>>] 100% ELAPSED TIME: 5.29 s
-----
Loading data to table gandhar.student partition (load_data=2023-04-10)
Partition gandhar.student{load_data=2023-04-10} stats: [numFiles=1, numRows=0, totalSize=860, rawDataSize=0]
OK
Time taken: 7.061 seconds
```

6) Display data

```
hive> select * from student;
OK
101      Manav      Computer      123456  2023-04-09
102      Aditya     Computer      123233  2023-04-10
Time taken: 0.27 seconds, Fetched: 2 row(s)
```

7) Order by

```
hive> select id, name from student order by id desc;
Query ID = root_20240214035923_a770cf77-452d-4dfc-be6f-3416ca85c82e
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      6          6          0          0          0          0
Reducer 2 .....  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 02/02 [=====>>>] 100% ELAPSED TIME: 41.30 s
-----
OK
102      Aditya
101      Manav
Time taken: 42.677 seconds, Fetched: 2 row(s)
```

8) Group by

```
hive> select branch, count(*) from student group by branch;
Query ID = root_20240214040307_c873d793-12e3-4eb4-a85e-964fd5115edf
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      6          6          0          0          2          0
Reducer 2 .....  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 02/02 [=====]>>] 100% ELAPSED TIME: 18.93 s
-----
OK
Computer      2
Mechanical    1
Time taken: 20.002 seconds, Fetched: 2 row(s)
```

9) Update tuple

```
hive> update student set name='Sean Fargose' where id='103';
Query ID = root_20240214040454_eeb9d679-be52-440b-b57a-755e1787dd22
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      6          6          0          0          2          0
Reducer 2 .....  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 02/02 [=====]>>] 100% ELAPSED TIME: 17.70 s
-----
Loading data to table gandhar.student partition (load_data=null)
  Time taken for load dynamic partitions : 467
  Loading partition {load_data=2023-04-10}
    Time taken for adding to write entity : 1
Partition gandhar.student{load_data=2023-04-10} stats: [numFiles=3, numRows=0, totalSize=2630, rawDataSize=0]
OK
Time taken: 22.433 seconds
```

10) Delete tuple

```
hive> delete from student where name = 'Sean Fargose';
Query ID = root_20240214040618_3c56ccb4-60c3-4e31-9724-2851a88f0f63
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      6          6          0          0          2          0
Reducer 2 .....  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 02/02 [=====]>>] 100% ELAPSED TIME: 11.35 s
-----
Loading data to table gandhar.student partition (load_data=null)
  Time taken for load dynamic partitions : 274
  Loading partition {load_data=2023-04-10}
    Time taken for adding to write entity : 0
Partition gandhar.student{load_data=2023-04-10} stats: [numFiles=4, numRows=0, totalSize=3163, rawDataSize=0]
OK
Time taken: 13.157 seconds
```

11) Alter table name

```
hive> alter table student rename to stud;
OK
Time taken: 0.526 seconds
```

```
hive> show tables;
OK
stud
values__tmp__table__1
values__tmp__table__2
values__tmp__table__3
values__tmp__table__4
values__tmp__table__5
values__tmp__table__6
Time taken: 0.164 seconds, Fetched: 7 row(s)
```

12) Add column

```
hive> alter table stud add columns(cgpa double);
OK
Time taken: 0.357 seconds
```

13) Change column name

```
hive> alter table stud change name sname String;
OK
Time taken: 0.352 seconds
```

14) Updating values in new column

```
hive> update stud set cgpa=6.9 where id = 101;
Query ID = root_20240214041257_0cd1a0ab-5a3b-4785-952b-6a4c0d42d65d
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      6          6          0          0          0          0
Reducer 2 .....  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 02/02 [======>>>] 100% ELAPSED TIME: 29.99 s
-----
Loading data to table gandhar.stud partition (load_data=null)
  Time taken for load dynamic partitions : 524
  Loading partition {load_data=2023-04-09}
    Time taken for adding to write entity : 1
Partition gandhar.stud{load_data=2023-04-09} stats: [numFiles=2, numRows=0, totalSize=1816, rawDataSize=0]
OK
Time taken: 34.957 seconds
```



```
hive> select * from stud;
OK
101      Manav    Computer      123456  6.9      2023-04-09
102      Aditya   Computer      123233  NULL     2023-04-10
Time taken: 0.355 seconds, Fetched: 2 row(s)
```

15) Aggregate functions (sum, max, min, count)

```
hive> select sum(cgpa) from stud;
Query ID = root_20240214041532_c0f8cc72-7a85-4e43-b433-2484a13f7801
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      6          6          0          0          4          0
Reducer 2 .....  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 19.06 s
-----
OK
14.3
Time taken: 21.446 seconds, Fetched: 1 row(s)
```

```
hive> select max(cgpa) from stud;
Query ID = root_20240214041636_1981e772-dae1-4272-90d2-e1f00f62bbc9
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      6          6          0          0          3          0
Reducer 2 .....  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 11.90 s
-----
OK
7.4
Time taken: 12.967 seconds, Fetched: 1 row(s)
```

```
hive> select min(cgpa) from stud;
Query ID = root_20240214041730_0cfd0127-19c1-4a41-bee7-7e7916f03c42
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED    6         6         0         0         4         0
Reducer 2 .....  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 10.37 s
-----
OK
6.9
Time taken: 11.066 seconds, Fetched: 1 row(s)
```

```
hive> select count(cgpa) from stud;
Query ID = root_20240214041859_205a0283-b3ed-4dc4-8471-dc0d53ad1288
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0001)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED    6         6         0         0         2         0
Reducer 2 .....  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 12.95 s
-----
OK
2
Time taken: 13.864 seconds, Fetched: 1 row(s)
```

16) Arithmetic functions

```
hive> select * from stud where cgpa < 9.0;
OK
101    Manav    Computer    123456  6.9    2023-04-09
102    Aditya    Computer    123233  7.4    2023-04-10
Time taken: 9.406 seconds, Fetched: 2 row(s)
```

```
hive> select * from stud where cgpa >= 7;
OK
102    Aditya    Computer    123233  7.4    2023-04-10
Time taken: 0.772 seconds, Fetched: 1 row(s)
```

17) Inner Join


```
hive> select * from customer c join orders o
  > on (c.id = o.customer_id);
Query ID = root_20240214050915_351c4b53-2867-4f5c-87a0-7595be7b0ce8
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0003)

-----
VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED    9         9         0         0         0         0
Map 2 .....  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 02/02 [=====]>>] 100% ELAPSED TIME: 47.69 s
-----
OK
2      Mike    27      Bhopal  35000   2023-03-11      101    2      20000
2      Mike    27      Bhopal  35000   2023-03-11      102    2      25000
Time taken: 53.275 seconds, Fetched: 2 row(s)
```

18) Left Outer Join

```
hive> select * from customer c left outer join orders o
  > on (c.id = o.customer_id);
Query ID = root_20240214051122_5a19dcc9-a374-4602-b029-f9a688a01e32
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0003)

-----
VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED    9         9         0         0         0         0
Map 2 .....  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 02/02 [=====]>>] 100% ELAPSED TIME: 17.39 s
-----
OK
1      Ross    25      Mumbai  25000   2023-03-09      NULL    NULL    NULL
3      Albin    24      Pune    50000   2023-03-10      NULL    NULL    NULL
2      Mike    27      Bhopal  35000   2023-03-11      101    2      20000
2      Mike    27      Bhopal  35000   2023-03-11      102    2      25000
Time taken: 20.524 seconds, Fetched: 4 row(s)
```

19) Right Outer Join

```
hive> select * from customer c right outer join orders o
  > on (c.id = o.customer_id);
Query ID = root_20240214051357_d26cdc0c-794e-436e-ac45-417c1d596300
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0003)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED    9         9         0         0         0         0
Map 2 .....  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 02/02 [=====>>>] 100% ELAPSED TIME: 9.80 s
-----
OK
2      Mike    27      Bhopal  35000   2023-03-11    101     2      20000
2      Mike    27      Bhopal  35000   2023-03-11    102     2      25000
NULL   NULL    NULL   NULL    NULL    104     4      10000
Time taken: 11.452 seconds, Fetched: 3 row(s)
```

20) Full Outer Join

```
hive> select * from customer c full outer join orders o
  > on (c.id = o.customer_id);
Query ID = root_20240214051525_b841e829-7006-4052-ad48-bc62f066f587
Total jobs = 1
Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1707879688667_0003)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED    9         9         0         0         4         0
Map 3 .....  SUCCEEDED    1         1         0         0         0         0
Reducer 2 ..... SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 03/03 [=====>>>] 100% ELAPSED TIME: 38.94 s
-----
OK
1      Ross    25      Mumbai  25000   2023-03-09    NULL    NULL    NULL
2      Mike    27      Bhopal  35000   2023-03-11    101     2      20000
2      Mike    27      Bhopal  35000   2023-03-11    102     2      25000
3      Albin   24      Pune    50000   2023-03-10    NULL    NULL    NULL
NULL   NULL    NULL   NULL    NULL    104     4      10000
Time taken: 40.32 seconds, Fetched: 5 row(s)
```

21) Views in HIVE

```
hive> create view if not exists customer_vw  
  > as select * from customer where address="Mumbai";  
OK  
Time taken: 0.618 seconds
```

```
hive> select * from customer_vw;  
OK  
1      Ross    25      Mumbai  25000   2023-03-09  
Time taken: 0.267 seconds, Fetched: 1 row(s)
```

```
hive> alter view customer_vw as select * from customer;  
OK  
Time taken: 0.396 seconds  
hive> select * from customer_vw;  
OK  
1      Ross    25      Mumbai  25000   2023-03-09  
3      Albin   24      Pune    50000   2023-03-10  
2      Mike    27      Bhopal  35000   2023-03-11  
Time taken: 0.239 seconds, Fetched: 3 row(s)
```

```
hive> drop view if exists customer_vw;  
OK  
Time taken: 0.575 seconds
```