

## EXPERIMENT 8: *Performing CRUD Operations on Apache Hive*

**Name: Adwait Purao**

**UID : 2021300101**

**Division: COMPS B**

### **1) Create Table cricket\_matches**

a) Query :

```
CREATE TABLE cricket_matches (  
    match_id INT,  
    team1 STRING,  
    team2 STRING,  
    match_date DATE,  
    venue STRING,  
    result STRING  
);
```

```
hive> CREATE TABLE cricket_matches (  
    >     match_id INT,  
    >     team1 STRING,  
    >     team2 STRING,  
    >     match_date DATE,  
    >     venue STRING,  
    >     result STRING  
    > );  
OK  
Time taken: 0.85 seconds  
hive> █
```

### **2) Insert into table cricket\_matches**

a) Query :

```
INSERT INTO cricket_matches VALUES  
(1, 'India', 'Australia', '2024-03-25', 'Melbourne Cricket Ground', 'India  
won by 50 runs'),  
(2, 'England', 'New Zealand', '2024-03-26', 'Lord''s Cricket Ground', 'New  
Zealand won by 3 wickets'),  
(3, 'South Africa', 'Pakistan', '2024-03-27', 'Wanderers Stadium', 'Match
```

```
tied');
```

```
hive> INSERT INTO cricket_matches VALUES
> (1, 'India', 'Australia', '2024-03-25', 'Melbourne Cricket Ground', 'India won by 50 runs'),
> (2, 'England', 'New Zealand', '2024-03-26', 'Lord's Cricket Ground', 'New Zealand won by 3 wickets'),
> (3, 'South Africa', 'Pakistan', '2024-03-27', 'Wanderers Stadium', 'Match tied');
Query ID = hadoop_20240327001139_4d44613d-1621-4069-ae03-9d7400bcd0b5
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Job running in-process (local Hadoop)
2024-03-27 00:11:47,677 Stage-1 map = 0%, reduce = 0%
2024-03-27 00:11:50,740 Stage-1 map = 100%, reduce = 0%
Ended Job = job_local700889455_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://localhost:9000/user/hive/warehouse/cricket_matches/.hive-staging_hive_2024-03-27_00-11-39_691_8784142399305
678855-1/-ext-10000
Loading data to table default.cricket_matches
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 0 HDFS Write: 302 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 12.113 seconds
hive>
```

### 3) Select Data from table cricket\_matches

a) Query :

```
SELECT * FROM cricket_matches;
```

```
hive> SELECT * FROM cricket_matches;
OK
1      India      Australia      2024-03-25      Melbourne Cricket Ground      India won by 50 runs
2      England New Zealand      2024-03-26      Lords Cricket Ground      New Zealand won by 3 wickets
3      South Africa      Pakistan      2024-03-27      Wanderers Stadium      Match tied
Time taken: 0.291 seconds, Fetched: 3 row(s)
hive>
```

### 4) Alter Table cricket\_matches

a) Query :

```
ALTER TABLE cricket_matches
ADD COLUMNS (winning_team STRING);
```

```
hive> ALTER TABLE cricket_matches
> ADD COLUMNS (winning_team STRING);
OK
Time taken: 0.288 seconds
hive> █
```

### 5) Update Table cricket\_matches

a) Query :

```
INSERT OVERWRITE TABLE cricket_matches
```

```

SELECT match_id, team1, team2, match_date, venue, result,
CASE
    WHEN result LIKE '%India%' THEN 'India'
    WHEN result LIKE '%New Zealand%' THEN 'New Zealand'
    WHEN result LIKE '%South Africa%' THEN 'South Africa'
    ELSE 'Unknown'
END AS winning_team
FROM cricket_matches;

```

```

hive> INSERT OVERWRITE TABLE cricket_matches
> SELECT match_id, team1, team2, match_date, venue, result,
> CASE
>     WHEN result LIKE '%India%' THEN 'India'
>     WHEN result LIKE '%New Zealand%' THEN 'New Zealand'
>     WHEN result LIKE '%South Africa%' THEN 'South Africa'
>     ELSE 'Unknown'
> END AS winning_team
> FROM cricket_matches;
Query ID = hadoop_20240327001450_c3a20609-3193-46d1-a1fd-689ad6e9c5cb
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Job running in-process (local Hadoop)
2024-03-27 00:14:52,187 Stage-1 map = 100%, reduce = 0%
Ended Job = job_local1274122013_0002
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://localhost:9000/user/hive/warehouse/cricket_matches/.hive-staging_hive_2024-03-27_00-14-50_326_8901504066214
068905-1/-ext-10000
Loading data to table default.cricket_matches
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 524 HDFS Write: 630 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 2.331 seconds
hive>

```

## 6) Delete Data from table cricket\_matches

a) Query :

```

CREATE TABLE cricket_matches_new AS
SELECT *
FROM cricket_matches
WHERE match_id <> 1;

DROP TABLE cricket_matches;

ALTER TABLE cricket_matches_new RENAME TO cricket_matches;

```

```

hive> CREATE TABLE cricket_matches_new AS
> SELECT *
> FROM cricket_matches
> WHERE match_id <= 1;
Query ID = hadoop_20240327001528_ffc425d3-98ad-4274-9ef4-43bc1262f0e1
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Job running in-process (local Hadoop)
2024-03-27 00:15:30,853 Stage-1 map = 0%, reduce = 0%
2024-03-27 00:15:31,892 Stage-1 map = 100%, reduce = 0%
Ended Job = job_local95671708_0003
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://localhost:9000/user/hive/warehouse/.hive-staging_hive_2024-03-27_00-15-28_848_5880250105311322040-1/-ext-10002
Moving data to directory hdfs://localhost:9000/user/hive/warehouse/cricket_matches_new
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 852 HDFS Write: 881 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 4.133 seconds
hive>
> DROP TABLE cricket_matches;
OK
Time taken: 1.34 seconds
hive>
> ALTER TABLE cricket_matches_new RENAME TO cricket_matches;
OK
Time taken: 0.119 seconds
hive>

```

## 7) Show all tables in Database

a) Query :

```
show tables;
```

```

hive> show tables;
OK
cricket_matches
Time taken: 0.042 seconds, Fetched: 1 row(s)
hive>

```

## 8) Load Data from CSV file

cricket\_players.csv :

```

1,Player1,India
2,Player2,Australia
3,Player3,England

```

a) Query :

i) First Create Table

```

CREATE TABLE cricket_players (
    player_id INT,
    player_name STRING,
    country STRING
);

```

ii) Then copy the csv file to hadoop file system

```
hadoop fs -put /path/to/players.csv /user/hive/warehouse/cricket_players/
```

iii) Load Data into hive table

```
LOAD DATA INPATH '/user/hive/warehouse/cricket_players/players.csv' INTO  
TABLE cricket_players;
```

```
hive> CREATE TABLE cricket_players (  
  >   player_id INT,  
  >   player_name STRING,  
  >   country STRING  
  > );  
OK  
Time taken: 1.513 seconds  
hive> LOAD DATA INPATH '/user/hive/warehouse/cricket_players/players.csv' INTO TABLE cricket_players;  
Loading data to table default.cricket_players  
OK  
Time taken: 1.349 seconds
```