

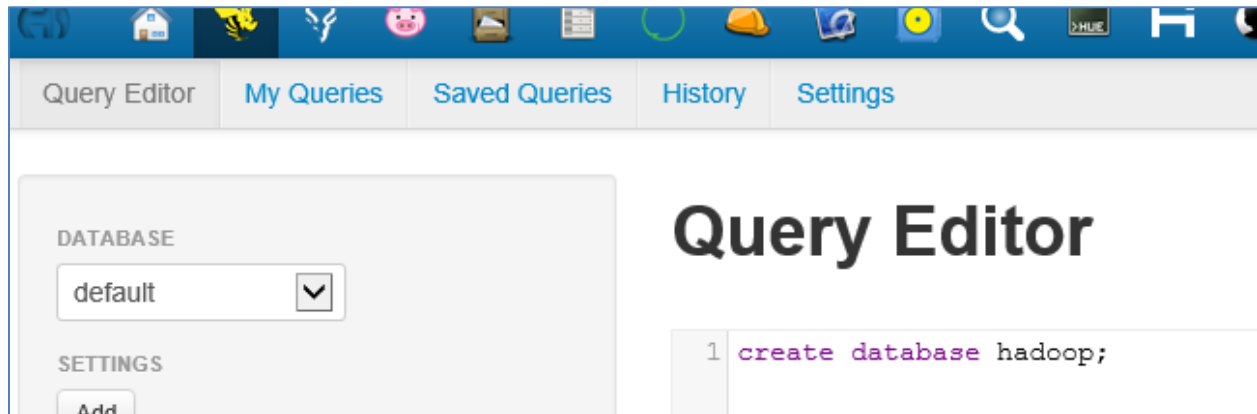
Hive Project Report

Tasks List

- Creating Databases
- Creating Table
- Importing Data from file
- Data movement from one table to another
- Use of Where clause
- Use of Function with Group By
- Use of Order By
- Use of Join
- Creating External Table
- Dropping Table
- Output Format

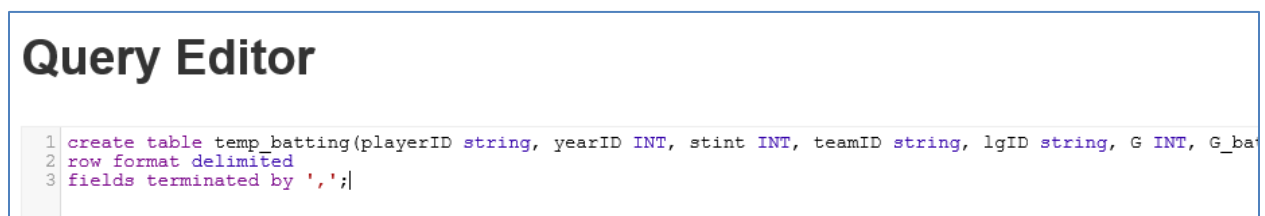
Creating Database

Create database <database_name>;

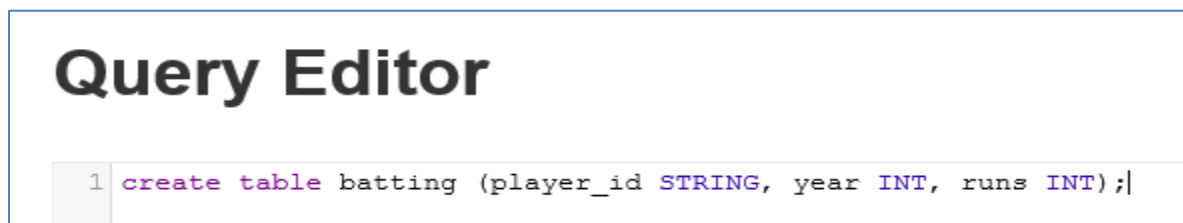


Creating Tables

create table temp_batting(playerID string, yearID INT, stint INT, teamID string, lgID string, G INT, G_batting INT, AB INT, R INT, H INT, 2B INT, 3B INT, HR INT, RBI INT, SB INT, CS INT, BB INT, SO INT, IBB INT, HBP INT, SH INT, SF INT, GIDP INT, G_old INT)
row format delimited
fields terminated by ',';



create table batting (player_id STRING, year INT, runs INT);



Importing Data from file

load data inpath '/user/root/Batting.csv' into table temp_batting;

Query Editor

```
1 load data inpath '/user/root/Batting.csv' into table temp_batting;|
```

Data movement from one table to another

insert into table batting
select playerid, yearid, r from temp_batting;

Query Editor

```
1 insert into table batting  
2 select playerid, yearid, r  
3 from temp_batting;|
```

Use of Where clause

select * from batting where player_id='aardsda01';

Query Editor

```
1 select * from batting where player_id='aardsda01';|
```

Use of Function with Group By

SELECT year, max(runs) FROM batting GROUP BY year;

```
1 SELECT year, max(runs) FROM batting GROUP BY year;
```

SELECT player_id, max(runs) FROM batting GROUP BY player_id;

```
1 SELECT player_id, max(runs) FROM batting GROUP BY player_id;
```

Use of Order By

SELECT player_id, max(runs) as max_runs FROM batting GROUP BY player_id
order by max_runs desc;

Query Editor

```
1 SELECT player_id, max(runs) as max_runs FROM batting GROUP BY player_id order by max_runs desc;
```

Use of Join

Find Player of the Year

```
SELECT a.year, a.player_id, a.runs from batting a  
JOIN (SELECT year, max(runs) runs FROM batting GROUP BY year ) b  
ON (a.year = b.year AND a.runs = b.runs);
```

Query Editor

```
1 SELECT a.year, a.player_id, a.runs from batting a  
2 JOIN (SELECT year, max(runs) runs FROM batting GROUP BY year ) b  
3 ON (a.year = b.year AND a.runs = b.runs);
```

Creating External Table Dropping Table

Please refer to the videos attached

Output Format

As shown below:-

Query Results: **Unsaved Query**

DOWNLOADS
[Download as CSV](#)
[Download as XLS](#)
[Save](#)

MR JOBS (2)
[job_201501171612_0011](#)
[job_201501171612_0012](#)

ResultsQueryLogColumns

	player_id
0	hamilbi01
1	ruthba01
2	brownto01
3	oneilti01
4	gehrilo01

I need .csv file for each select query