**Assignment Hive 2**

Objective: using above two created data sets, create score values for each player for their batting and bowling performance based on given rules. Finally calculate **"cost_per_score"** for each player using their cost from player_cost.csv file.

**Rule for batting score:**

1) One point for each run.
2) 2 points for each six.
3) Pace rate : (runs scored – balls faced).

**Rule for bowling score:**

1) 20 points for each wicket
2) Pace rate: (1.5*balls bowled – runs conceded).

This assignment will use union all, group by, sort by, inbuilt functions and views.

Solution

Hive script:

```
create table id
(id int, name String)
row format delimited
fields terminated by ','
stored as textfile;

insert overwrite table id
select * from
(select bt.id, bt.name from batting bt
union all
select bl.id, bl.name from bowling bl) UnionResult;

select * from id;

insert overwrite table id
select distinct i.id, i.name from id i;

create table bat_score
(id int, score float)
row format delimited
fields terminated by ','
stored as textfile;

insert overwrite table bat_score
select bt.id, (bt.runs + 2*bt.sixes + bt.runs - bt.runs*100/bt.strike_rate) from batting bt
order by bt.id asc;

create table bowl_score
(id int, score float)
row format delimited
fields terminated by ','
stored as textfile;

insert overwrite table bowl_score
select bl.id, (20*bl.wickets + 1.5*6*bl.overs - bl.runs) from bowling bl
order by bl.id asc;

create table total_score
(id int, score float)
row format delimited
fields terminated by ','
stored as textfile;

insert overwrite table total_score
select * from
(select bt.id, bt.score from bat_score bt
```

```
union all
select bl.id, bl.score from bowl_score bl) UnionResult;

insert overwrite table total_score
select ts.id, sum(ts.score) from total_score ts
group by ts.id;

create table cost
(id int, cost int)
row format delimited
fields terminated by ','
stored as textfile;

load data local inpath '/home/training/Desktop/hive/player_cost.csv'
overwrite into table cost;

create table final
(id int, name String, score float, cost int, cost_per_score float)
row format delimited
fields terminated by ','
stored as textfile;

create view first_join as
select a.id, a.name, b.score from id a join total_score b
on (a.id = b.id);

show tables;

describe first_join;

insert overwrite table final
select a.id, a.name, a.score, b.cost, b.cost/a.score from first_join a join cost b
on (a.id = b.id);

select * from final
sort by cost_per_score;
```

```
hive> select * from final
    > sort by cost_per_score;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201405251253_0019, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201405251253_0019
Kill Command = /usr/lib/hadoop/bin/hadoop job  -Dmapred.job.tracker=localhost:8021 -kill job_201405251253_0019
2014-05-25 15:31:24,532 Stage-1 map = 0%,  reduce = 0%
2014-05-25 15:31:27,545 Stage-1 map = 100%,  reduce = 0%
2014-05-25 15:31:34,591 Stage-1 map = 100%,  reduce = 33%
2014-05-25 15:31:35,595 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_201405251253_0019
OK
31      M Vijay 336.99097       500000  1483.7192
44      M Vohra 188.99878       500000  2645.5198
4       DR Smith        761.99646       5000000 6561.7104
3       DA Warner       722.99316       5500000 7607.264
18      LMP Simmons     673.9885        5500000 8160.3765
22      MS Dhoni        733.9911        6000000 8174.4863
49      STR Binny       498.79523       4500000 9021.738
```

Output:

```
hive> show tables;
OK
bat_score
batting
bowl_score
bowling
cost
final
first_join
id
total_score
Time taken: 0.036 seconds
hive>
```

**Contents of directory /user/hive/warehouse**

Goto : /user/hive/warehouse go

Go to parent directory

| Name | Type | Size | Replication | Block Size | Modification Time | Permission | Owner | Group |
|---|---|---|---|---|---|---|---|---|
| case_ipl.db | dir | | | | 2014-05-25 15:29 | rwxr-xr-x | training | supergroup |
| trial.db | dir | | | | 2014-05-20 00:31 | rwxr-xr-x | training | supergroup |

**Contents of directory /user/hive/warehouse/case_ipl.db**

Goto : /user/hive/warehouse/case_ go

Go to parent directory

| Name | Type | Size | Replication | Block Size | Modification Time | Permission | Owner | Group |
|---|---|---|---|---|---|---|---|---|
| bat_score | dir | | | | 2014-05-25 14:42 | rwxr-xr-x | training | supergroup |
| batting | dir | | | | 2014-05-25 13:15 | rwxr-xr-x | training | supergroup |
| bowl_score | dir | | | | 2014-05-25 14:48 | rwxr-xr-x | training | supergroup |
| bowling | dir | | | | 2014-05-25 13:44 | rwxr-xr-x | training | supergroup |
| cost | dir | | | | 2014-05-25 15:09 | rwxr-xr-x | training | supergroup |
| final | dir | | | | 2014-05-25 15:29 | rwxr-xr-x | training | supergroup |
| id | dir | | | | 2014-05-25 14:26 | rwxr-xr-x | training | supergroup |
| total_score | dir | | | | 2014-05-25 15:02 | rwxr-xr-x | training | supergroup |

**Contents of directory /user/hive/warehouse/case_ipl.db/batting**

Goto : /user/hive/warehouse/case_ go

Go to parent directory

| Name | Type | Size | Replication | Block Size | Modification Time | Permission | Owner | Group |
|---|---|---|---|---|---|---|---|---|
| team=Bangalore | dir | | | | 2014-05-25 13:14 | rwxr-xr-x | training | supergroup |
| team=Chennai | dir | | | | 2014-05-25 13:14 | rwxr-xr-x | training | supergroup |
| team=Delhi | dir | | | | 2014-05-25 13:11 | rwxr-xr-x | training | supergroup |
| team=Hyderabad | dir | | | | 2014-05-25 13:15 | rwxr-xr-x | training | supergroup |
| team=Kolkata | dir | | | | 2014-05-25 13:15 | rwxr-xr-x | training | supergroup |
| team=Mumbai | dir | | | | 2014-05-25 13:15 | rwxr-xr-x | training | supergroup |
| team=Punjab | dir | | | | 2014-05-25 13:15 | rwxr-xr-x | training | supergroup |
| team=Rajasthan | dir | | | | 2014-05-25 13:15 | rwxr-xr-x | training | supergroup |