1)Create database(hadoopproject), check it and use it

create database hadoopproject;

hive> show databases;
OK
assignment
default
hadoopproject
weekend
Time taken: 6.049 seconds

hive> use hadoopproject;
OK

---

2) create tables into database and load data

CREATE TABLE IF NOT EXISTS channel(tv_show string, channel_name string)
row format delimited
fields terminated by ',';

DESCRIBE channel;
OK
channel_name string
tv_show string

load data local inpath'/home/training/projectdata/genchanA.txt' into table channel;
load data local inpath'/home/training/projectdata/genchanB.txt' into table channel;
load data local inpath'/home/training/projectdata/genchanC.txt' into table channel;

CREATE TABLE IF NOT EXISTS viewers(tv_show string, number_views int)
row format delimited

fields terminated by ',';

DESCRIBE viewers;
OK
tv_show   string
number_views   int

load data local inpath'/home/training/projectdata/gennumA.txt' into table viewers;
load data local inpath'/home/training/projectdata/gennumB.txt' into table viewers;
load data local inpath'/home/training/projectdata/gennumC.txt' into table viewers;

PS : I cahnged the name of the data set files for simplicity
it was originally join2_genchanA.txt   for example

---
3) check the tables are created and data is correct

Select * from Channel limit 5;
Select count(*)from Channel;
result =600

Select * from viewers limit 5;
Select count(*)from viewers;
result =6000

---
4)Join tables and check the data is correct

SELECT *
FROM viewers FULL OUTER JOIN channel

```
ON  viewers.tv_show=channel.tv_show
limit 5;
```

---

5) Probelm statements queries

Probelm Stetment 1:Toatal number of shows on ABC channel

```
SELECT SUM(number_views)
FROM viewers
FULL OUTER JOIN channel
ON channel.tv_show = viewers.tv_show
where (channel_name=='ABC')
group by channel_name
;
```

Result= 1115974

Probelm Stetment 2:Toatal number of shows on BAT channel

```
SELECT SUM(number_views)
FROM viewers
FULL OUTER JOIN channel
ON channel.tv_show = viewers.tv_show
where (channel_name=='BAT')
group by channel_name
;
```

Result==5099141

Probelm Stetment 3:The aired shows on Zoo, NOX , ABC channel

```
SELECT * from channel
where (channel_name=='ABC') OR (channel_name=='NOX') OR
(channel_name=='ZOO');
```

PS : ZOO channel is not in data set

Result = (please check the attached screen shots)

Probelm Stetment 4 : The Most viewed show on ABC channel

```
SELECT SUM(number_views), channel.tv_show
FROM viewers
FULL OUTER JOIN channel
ON channel.tv_show = viewers.tv_show
where (channel_name=='ABC')
group by channel.tv_show
ORDER BY number_views;
```

Result =Hourly_Talking 108163