

1. select * from temperature_data
where zipcode > 300000
and
zipcode < 399999;

```
> select * from temperature_data
> where zipcode > 300000
> and
> zipcode < 399999;
OK
1990-03-10      381920  15.0
1991-01-10      302918  22.0
1990-02-12      384902   9.0
1991-03-10      381920  16.0
1990-01-10      302918  23.0
1991-02-12      384902  10.0
1993-03-10      381920  16.0
1994-01-10      302918  23.0
1991-02-12      384902  10.0
1991-03-10      381920  16.0
1990-01-10      302918  23.0
1991-02-12      384902  10.0
Time taken: 0.719 seconds, Fetched: 12 row(s)
```

2. select from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy') year ,
max(temperature) maxtemp from temperature_data
group by from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy');

```
OK
1990      23.0
1991      22.0
1993      16.0
1994      23.0
Time taken: 40.406 seconds, Fetched: 4 row(s)
```

3. select from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy') year ,
max(temperature) maxtemp,
count(from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy')) cnt
from temperature_data
group by from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy')
having count(from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy')) >= 2;

```
Total MapReduce CPU Time Spent: 5 seconds 930 msec
OK
1990      23.0      7
1991      22.0      9
1993      16.0      2
1994      23.0      2
Time taken: 45.484 seconds, Fetched: 4 row(s)
hive>
```

4. create view temperature_data_vw

as

```
select from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy') year ,  
max(temperature) maxtemp,  
count(from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy')) cnt  
from temperature_data  
group by from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy')  
having count(from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy')) >= 2;
```

```
hive>  
>  
> create view temperature_data_vw  
> as  
> select from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy') year ,  
> max(temperature) maxtemp,  
> count(from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy')) cnt  
> from temperature_data  
> group by from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy')  
> having count(from_unixtime(unix_timestamp(date,'MM-dd-yyyy'),'yyyy')) >= 2;  
OK  
Time taken: 0.297 seconds
```

5. INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/output'

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '|'

SELECT * FROM temperature_data_vw;

```
[acadgild@localhost output]$ cat 000000_0  
1990|23.0|7  
1991|22.0|9  
1993|16.0|2  
1994|23.0|2  
[acadgild@localhost output]$
```