

Assignment #1:

Create "ratings" managed table

Load the data from

~/futurence_hadoop-pyspark/labs/dataset/movie/ratings.csv

Display the ratings data

Display rating wise count

```
CREATE TABLE IF NOT EXISTS RATINGS (userId int, movieId int, rating float,
timestampRating int)
COMMENT 'Rating Details'
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
STORED AS TEXTFILE;
```

```
LOAD DATA LOCAL INPATH
'/home/uttam/futurence-datengg-bootcamp/dataset/ratings.csv' OVERWRITE
INTO TABLE RATINGS;
```

```
SELECT * FROM RATINGS;
```

```
Select rating, count(*) From Ratings where rating IS NOT NULL Group By
rating;
```

```
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.31 sec HDFS Read: 2497468 HDFS Write: 301 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 310 msec
OK
+-----+-----+
| rating | _c1 |
+-----+-----+
| 0.5 | 1370 |
| 1.0 | 2811 |
| 1.5 | 1791 |
| 2.0 | 7551 |
| 2.5 | 5550 |
| 3.0 | 20047 |
| 3.5 | 13136 |
| 4.0 | 26818 |
| 4.5 | 8551 |
| 5.0 | 13211 |
+-----+-----+
10 rows selected (20.706 seconds)
```

Assignment #2:

Create "weather" external table under /user/training/weather

Load the data from ~/futurence_hadoop-pyspark/labs/dataset/weather to
/user/training/weather

Display the weather data

Display Max, Min weather

Display month wise Max and Min weather

```

CREATE EXTERNAL TABLE IF NOT EXISTS weather (WBANNO int, LST_DATE DATE,
CRX_VN FLOAT, LONGITUDE FLOAT,LATITUDE FLOAT,T_DAILY_MAX FLOAT,T_DAILY_MIN
FLOAT,T_DAILY_MEAN FLOAT, T_DAILY_AVG FLOAT,P_DAILY_CALC FLOAT, SOLARAD_DAILY
String,SUR_TEMP_DAILY_TYPE FLOAT, SUR_TEMP_DAILY_MAX FLOAT,
SUR_TEMP_DAILY_MIN FLOAT,SUR_TEMP_DAILY_AVG FLOAT,RH_DAILY_MAX FLOAT,
RH_DAILY_MIN FLOAT,RH_DAILY_AVG FLOAT,SOIL_MOISTURE_5_DAILY FLOAT,
SOIL_MOISTURE_10_DAILY FLOAT, SOIL_MOISTURE_20_DAILY  FLOAT,
SOIL_MOISTURE_50_DAILY FLOAT, SOIL_MOISTURE_100_DAILY FLOAT,
SOIL_TEMP_5_DAILY FLOAT, SOIL_TEMP_10_DAILY FLOAT, SOIL_TEMP_20_DAILY FLOAT,
SOIL_TEMP_50_DAILY FLOAT, SOIL_TEMP_100_DAILY  FLOAT)
COMMENT 'Weather Details'
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '#'
LINES TERMINATED BY '\n'
LOCATION '/user/training/weather';

```

```

LOAD DATA LOCAL INPATH
'/home/uttam/futureense-datengg-bootcamp/dataset/weather_data_formatted.txt'
OVERWRITE INTO TABLE weather;

```

```
select * from weather;
```

```
select max(t_daily_max), min(t_daily_min) from weather;
```

```

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.74 sec HDFS Read: 52352 HDFS Write: 109 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 740 msec
OK
+-----+-----+
| _c0 | _c1 |
+-----+-----+
| 36.0 | -7.9 |
+-----+-----+

```

```

select month(lst_date) as `Month`, max(t_daily_max) as `Max Temperature`,
min(t_daily_min) as `Min Temperature` from weather group by month(lst_date);

```

month	max temperature	min temperature
1	26.5	-7.9
2	26.6	-3.5
3	29.1	-3.2
4	30.8	8.0
5	31.1	14.3
6	33.6	0.0
7	36.0	19.8

7 rows selected (17.073 seconds)