Assignment #1:

* 1. Create "ratings" managed table
  2. Load the data from ~/futurence\_hadoop-pyspark/labs/dataset/movie/ratings.csv
  3. Display the ratings data
  4. Display rating-wise count.

-------------------------------------------------------------------------------------------------------------------------------------

CREATE TABLE ratings(user\_id int, movie\_Id int,rating varchar(3),timestp varchar(20))

. . . . . . . . > ROW FORMAT DELIMITED

. . . . . . . . > FIELDS TERMINATED BY ','

. . . . . . . . > LINES TERMINATED BY '\n'

. . . . . . . . > STORED AS TEXTFILE;

LOAD DATA LOCAL INPATH '~/futurence\_hadoop-pyspark/labs/dataset/movie/ratings.csv ' OVERWRITE INTO TABLE ratings;

SELECT \* FROM ratings;

SELECT rating, count(movie\_id) FROM ratings GROUP BY rating;

