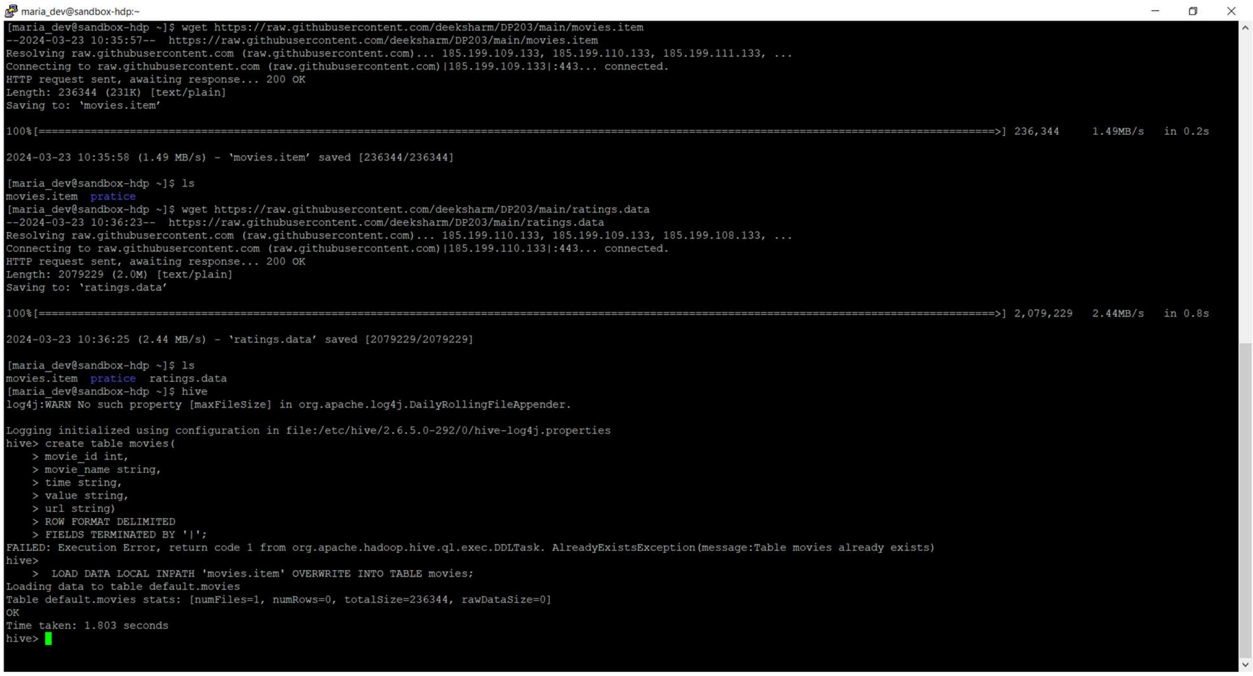
HADOOP ASSIGNEMENT-2

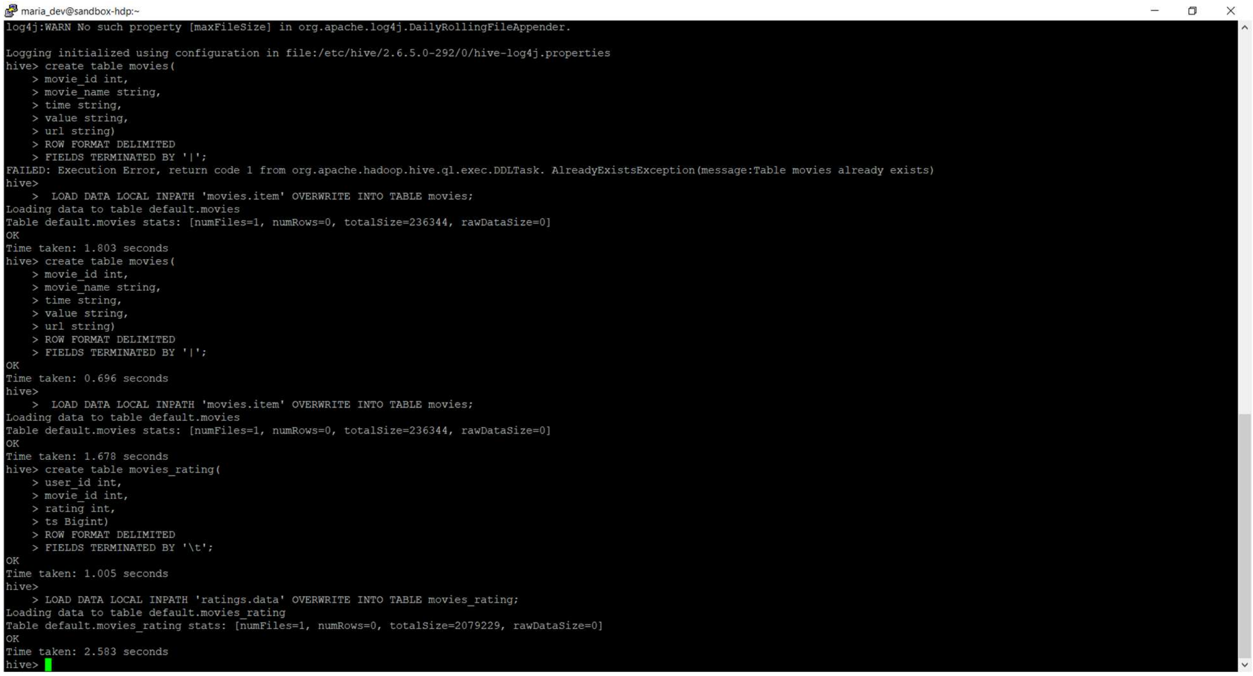


```
[maria_dev@sandbox-hdp ~]$ wget https://raw.githubusercontent.com/deeksharm/DP203/main/movies.item
--2024-03-23 10:35:57--  https://raw.githubusercontent.com/deeksharm/DP203/main/movies.item
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.109.133, 185.199.110.133, 185.199.111.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.109.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 236344 (231K) [text/plain]
Saving to: 'movies.item'

100%[==========================================================================================================================================>] 236,344     1.49MB/s   in 0.2s

2024-03-23 10:35:58 (1.49 MB/s) - 'movies.item' saved [236344/236344]

[maria_dev@sandbox-hdp ~]$ ls
movies.item  pratice
[maria_dev@sandbox-hdp ~]$ wget https://raw.githubusercontent.com/deeksharm/DP203/main/ratings.data
--2024-03-23 10:36:23--  https://raw.githubusercontent.com/deeksharm/DP203/main/ratings.data
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.110.133, 185.199.109.133, 185.199.108.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.110.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2079229 (2.0M) [text/plain]
Saving to: 'ratings.data'

100%[==========================================================================================================================================>] 2,079,229   2.44MB/s   in 0.8s

2024-03-23 10:36:25 (2.44 MB/s) - 'ratings.data' saved [2079229/2079229]

[maria_dev@sandbox-hdp ~]$ ls
movies.item  pratice  ratings.data
[maria_dev@sandbox-hdp ~]$ hive
log4j:WARN No such property [maxFileSize] in org.apache.log4j.DailyRollingFileAppender.

Logging initialized using configuration in file:/etc/hive/2.6.5.0-292/0/hive-log4j.properties
hive> create table movies(
    > movie_id int,
    > movie_name string,
    > time string,
    > value string,
    > url string)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY '|';
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table movies already exists)
hive>
    >  LOAD DATA LOCAL INPATH 'movies.item' OVERWRITE INTO TABLE movies;
Loading data to table default.movies
Table default.movies stats: [numFiles=1, numRows=0, totalSize=236344, rawDataSize=0]
OK
Time taken: 1.803 seconds
hive>
```

- I fetch "**movie.items**" from website into local using command "**wget <url>".**

- I fetch "**rating.data**" from website into local using command "**wget<url>".**

- I list the files in the local by using command "**ls**".



```
log4j:WARN No such property [maxFileSize] in org.apache.log4j.DailyRollingFileAppender.

Logging initialized using configuration in file:/etc/hive/2.6.5.0-292/0/hive-log4j.properties
hive> create table movies(
    > movie_id int,
    > movie_name string,
    > time string,
    > value string,
    > url string)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY '|';
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table movies already exists)
hive>
    >  LOAD DATA LOCAL INPATH 'movies.item' OVERWRITE INTO TABLE movies;
Loading data to table default.movies
Table default.movies stats: [numFiles=1, numRows=0, totalSize=236344, rawDataSize=0]
OK
Time taken: 1.803 seconds
hive> create table movies(
    > movie_id int,
    > movie_name string,
    > time string,
    > value string,
    > url string)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY '|';
OK
Time taken: 0.696 seconds
hive>
    >  LOAD DATA LOCAL INPATH 'movies.item' OVERWRITE INTO TABLE movies;
Loading data to table default.movies
Table default.movies stats: [numFiles=1, numRows=0, totalSize=236344, rawDataSize=0]
OK
Time taken: 1.678 seconds
hive> create table movies_rating(
    > user_id int,
    > movie_id int,
    > rating int,
    > ts Bigint)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY '\t';
OK
Time taken: 1.005 seconds
hive>
    > LOAD DATA LOCAL INPATH 'ratings.data' OVERWRITE INTO TABLE movies_rating;
Loading data to table default.movies_rating
Table default.movies_rating stats: [numFiles=1, numRows=0, totalSize=2079229, rawDataSize=0]
OK
Time taken: 2.583 seconds
hive>
```

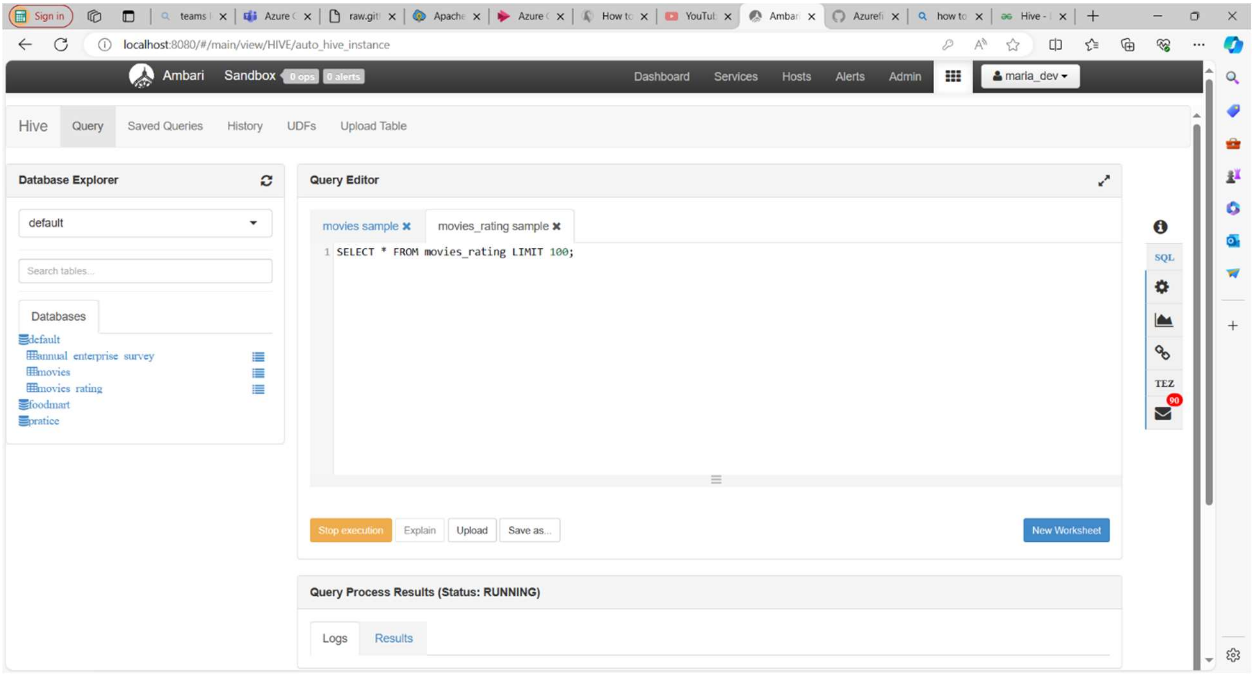
- I switch into hive service by using command “**hive**”.
- I created one “**movies table”** using normal SQL DDL command .

-
  ```
  create table movies(
  movie_id int,
  movie_name string,
  time string,
  value string,
  url string)
  ROW FORMAT DELIMITED
  FIELDS TERMINATED BY '|';
  ```
- I created one “**movies_rating table**” using normal SQL DDL command.

-
  ```
  create table movies_rating(
  user_id int,
  movie_id int,
  rating int,
  ts Bigint)
  ROW FORMAT DELIMITED
  FIELDS TERMINATED BY '\t';
  ```
- I load recorders into that table “**movies**” by using command.

-
  ```
  LOAD DATA LOCAL INPATH 'movies.item' OVERWRITE INTO TABLE movies;
  ```
- I load recorders into that table “**movies_rating**” by using command.

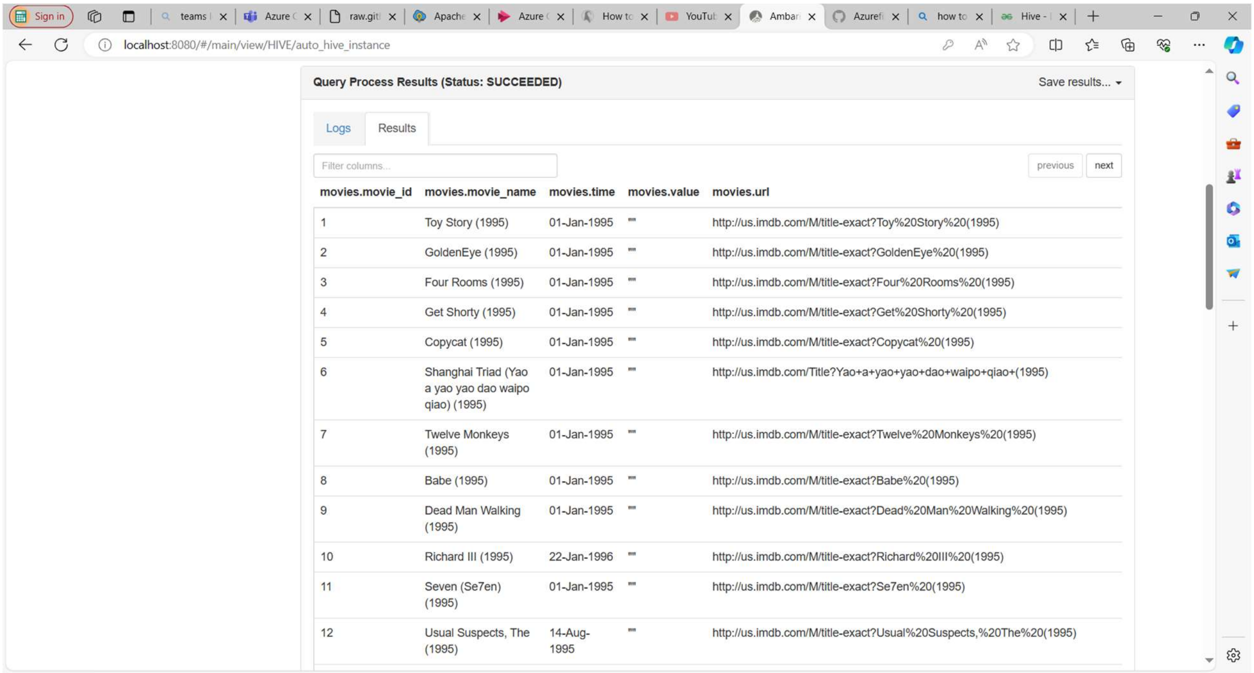
-
  ```
  LOAD DATA LOCAL INPATH 'ratings.data' OVERWRITE INTO TABLE movies_rating;
  ```



```
SELECT * FROM movies_rating LIMIT 100;
```

Query Process Results (Status: RUNNING)

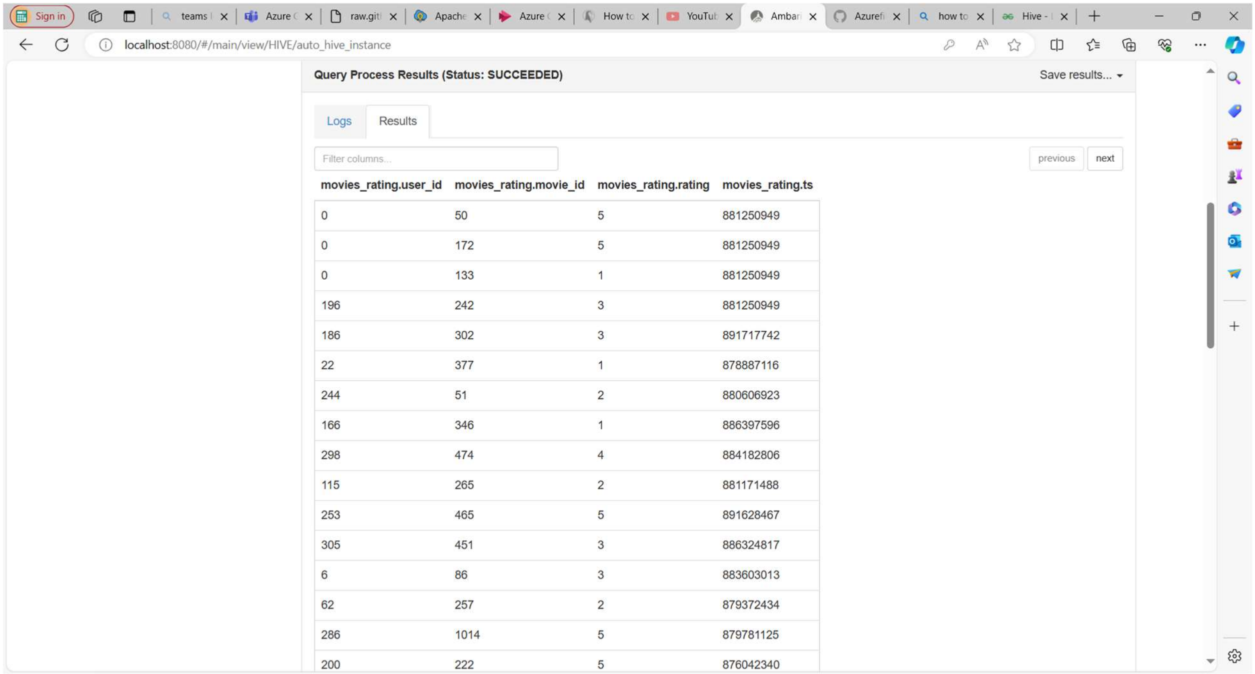
I run query in Query editor of hive service and check whether recorders get inserted correctly or not.



Query Process Results (Status: SUCCEEDED)

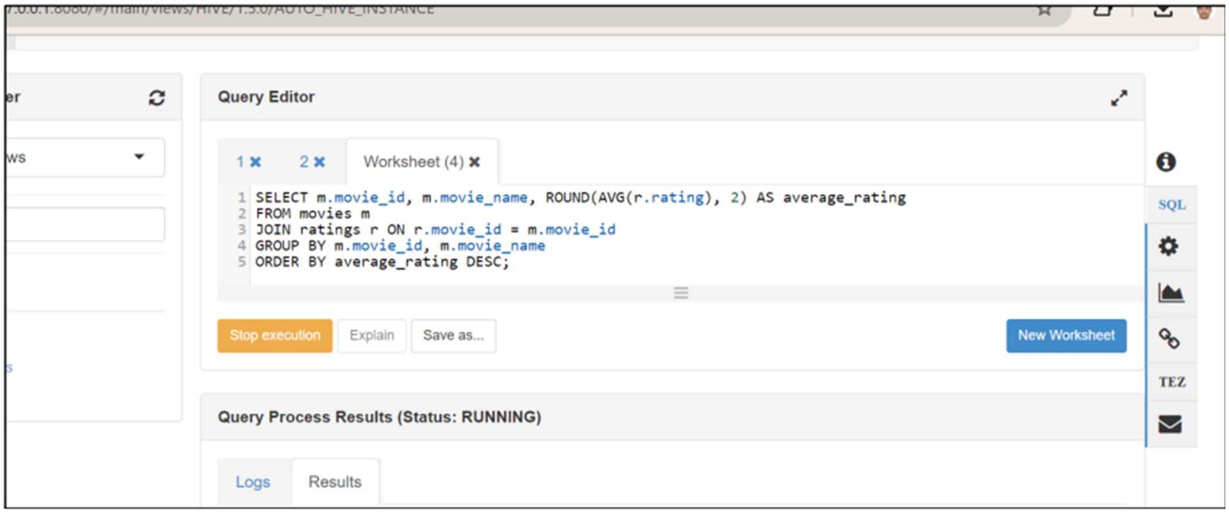
movies.movie_id	movies.movie_name	movies.time	movies.value	movies.url
1	Toy Story (1995)	01-Jan-1995	""	http://us.imdb.com/M/title-exact?Toy%20Story%20(1995)
2	GoldenEye (1995)	01-Jan-1995	""	http://us.imdb.com/M/title-exact?GoldenEye%20(1995)
3	Four Rooms (1995)	01-Jan-1995	""	http://us.imdb.com/M/title-exact?Four%20Rooms%20(1995)
4	Get Shorty (1995)	01-Jan-1995	""	http://us.imdb.com/M/title-exact?Get%20Shorty%20(1995)
5	Copycat (1995)	01-Jan-1995	""	http://us.imdb.com/M/title-exact?Copycat%20(1995)
6	Shanghai Triad (Yao a yao yao dao waipo qiao) (1995)	01-Jan-1995	""	http://us.imdb.com/Title?Yao+a+yao+yao+dao+waipo+qiao+(1995)
7	Twelve Monkeys (1995)	01-Jan-1995	""	http://us.imdb.com/M/title-exact?Twelve%20Monkeys%20(1995)
8	Babe (1995)	01-Jan-1995	""	http://us.imdb.com/M/title-exact?Babe%20(1995)
9	Dead Man Walking (1995)	01-Jan-1995	""	http://us.imdb.com/M/title-exact?Dead%20Man%20Walking%20(1995)
10	Richard III (1995)	22-Jan-1996	""	http://us.imdb.com/M/title-exact?Richard%20III%20(1995)
11	Seven (Se7en) (1995)	01-Jan-1995	""	http://us.imdb.com/M/title-exact?Se7en%20(1995)
12	Usual Suspects, The (1995)	14-Aug-1995	""	http://us.imdb.com/M/title-exact?Usual%20Suspects,%20The%20(1995)

Result of above SQL Query for “**movietable**” which is executed on query editor in hive service.


localhost:8080/#/main/view/HIVE/auto_hive_instance
Query Process Results (Status: SUCCEEDED)
Save results...
Logs
Results
Filter columns...
previous
next

movies_rating.user_id	movies_rating.movie_id	movies_rating.rating	movies_rating.ts
0	50	5	881250949
0	172	5	881250949
0	133	1	881250949
196	242	3	881250949
186	302	3	891717742
22	377	1	878887116
244	51	2	880606923
166	346	1	886397596
298	474	4	884182806
115	265	2	881171488
253	465	5	891628467
305	451	3	886324817
6	86	3	883603013
62	257	2	879372434
286	1014	5	879781125
200	222	5	876042340

Result of above SQL Query for “**movie_rating table**” which is executed on query editor in hive service.



```
SELECT m.movie_id, m.movie_name, ROUND(AVG(r.rating), 2) AS average_rating
FROM movies m
JOIN ratings r ON r.movie_id = m.movie_id
GROUP BY m.movie_id, m.movie_name
ORDER BY average_rating DESC;
```

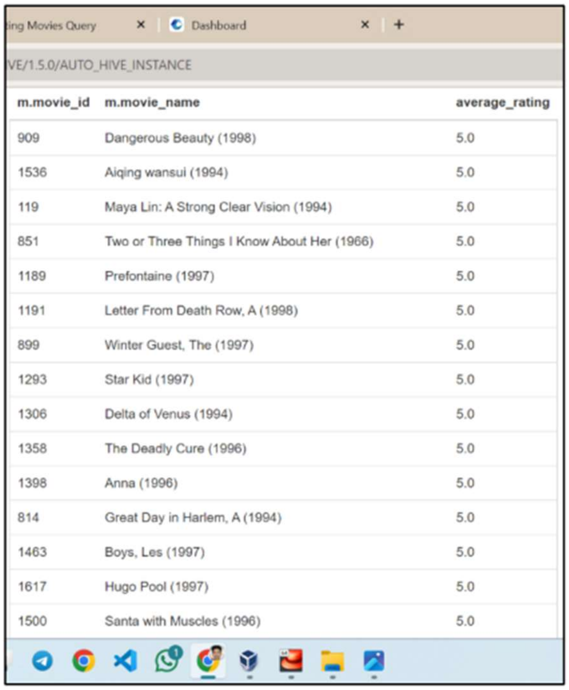
SELECT m.movie_id,m.movie_name,ROUND(AVG(r.rating),2) As average_rating

FROM movies m

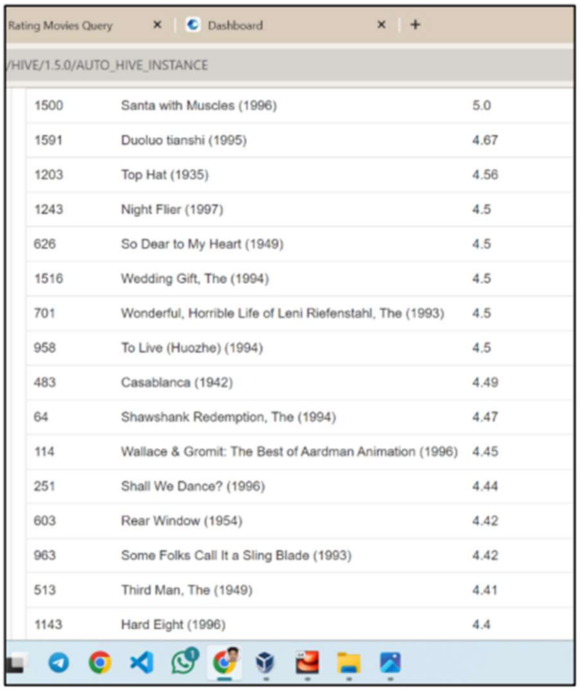
JOIN ratings r ON r.movie_id=m.movie_id

GROUO BY m.movie_id,m.movie_name

ORDER BY average_rating desc;



m.movie_id	m.movie_name	average_rating
909	Dangerous Beauty (1998)	5.0
1536	Aiqing wansui (1994)	5.0
119	Maya Lin: A Strong Clear Vision (1994)	5.0
851	Two or Three Things I Know About Her (1966)	5.0
1189	Prefontaine (1997)	5.0
1191	Letter From Death Row, A (1998)	5.0
899	Winter Guest, The (1997)	5.0
1293	Star Kid (1997)	5.0
1306	Delta of Venus (1994)	5.0
1358	The Deadly Cure (1996)	5.0
1398	Anna (1996)	5.0
814	Great Day in Harlem, A (1994)	5.0
1463	Boys, Les (1997)	5.0
1617	Hugo Pool (1997)	5.0
1500	Santa with Muscles (1996)	5.0



1500	Santa with Muscles (1996)	5.0
1591	Duoluo tianshi (1995)	4.67
1203	Top Hat (1935)	4.56
1243	Night Flier (1997)	4.5
626	So Dear to My Heart (1949)	4.5
1516	Wedding Gift, The (1994)	4.5
701	Wonderful, Horrible Life of Leni Riefenstahl, The (1993)	4.5
958	To Live (Huozhe) (1994)	4.5
483	Casablanca (1942)	4.49
64	Shawshank Redemption, The (1994)	4.47
114	Wallace & Gromit: The Best of Aardman Animation (1996)	4.45
251	Shall We Dance? (1996)	4.44
603	Rear Window (1954)	4.42
963	Some Folks Call It a Sling Blade (1993)	4.42
513	Third Man, The (1949)	4.41
1143	Hard Eight (1996)	4.4