# Hadoop-Hive Project Report

## Overview

This report provides a comprehensive analysis of the data using Hive queries. Each query addresses specific aspects of the dataset, and the outputs validate the results.

## Hive Queries and Outputs

### 1. Top-rated movies:

**Query:**

SELECT movieId, AVG(rating) as avg_rating

FROM movie_ratings

GROUP BY movieId

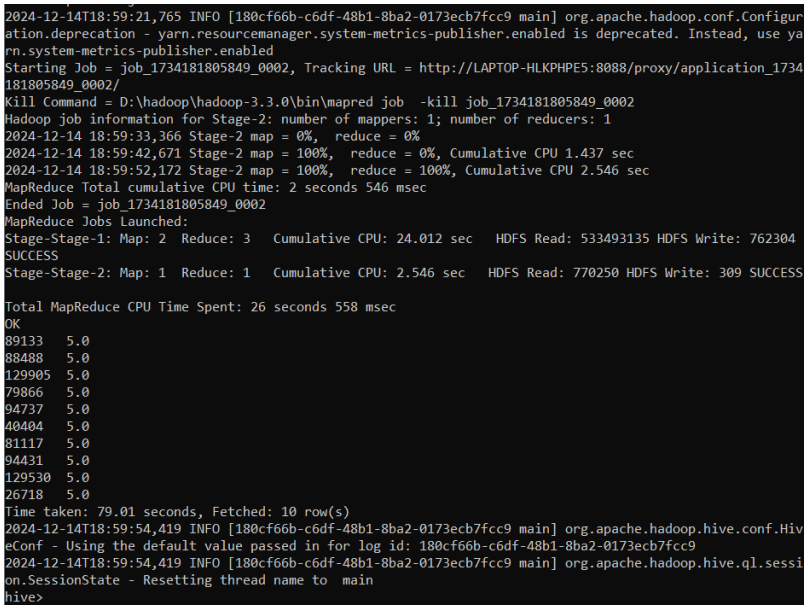
ORDER BY avg_rating DESC

LIMIT 10;

**Description:** This query retrieves the top 10 movies with the highest average ratings.

**Output:** *(Screenshot of the query result showcasing the top-rated movies).*

```
hive> SELECT movieId, AVG(rating) as avg_rating
    > FROM movie_ratings
    > GROUP BY movieId
    > ORDER BY avg_rating DESC
    > LIMIT 10;
2024-12-14T18:58:35,295 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passe
d in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T18:58:35,295 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread nam
e to 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main
2024-12-14T18:58:35,712 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.common.F
ileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/aakas/180cf66b-c6df-48
b1-8ba2-0173ecb7fcc9/hive_2024-12-14_18-58-35_313_6461569921446321087-1/-mr-10001/.hive-staging_hive_202
4-12-14_18-58-35_313_6461569921446321087-1
Query ID = aakas_20241214185835_7003b617-d001-4a19-8d8a-da0bbd1714ac
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
2024-12-14T18:58:36,128 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.conf.Configur
ation.deprecation - mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.r
eplication
2024-12-14T18:58:37,342 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.conf.Configur
ation.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use ya
rn.system-metrics-publisher.enabled
2024-12-14T18:58:37,641 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.conf.Configur
ation - resource-types.xml not found
Starting Job = job_1734181805849_0001, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0001/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0001
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 3
2024-12-14 18:58:47,455 Stage-1 map = 0%,  reduce = 0%
2024-12-14 18:59:04,389 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 10.516 sec
2024-12-14 18:59:19,160 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 24.012 sec
MapReduce Total cumulative CPU time: 24 seconds 12 msec
Ended Job = job_1734181805849_0001
Launching Job 2 out of 2
2024-12-14T18:59:21,329 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.conf.Configur
ation.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use ya
rn.system-metrics-publisher.enabled
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
```




## 2. Movies with the Highest Number of Ratings

**Query:**

```
SELECT movieId, COUNT(*) as rating_count
FROM movie_ratings
GROUP BY movieId
ORDER BY rating_count DESC
LIMIT 10;
```

**Description:** Identifies the top 10 movies that have received the highest number of ratings.

**Output:** *(Screenshot of the query result showing movies ranked by the number of ratings)*

**PTO**

```
hive> SELECT movieId, COUNT(*) as rating_count
    > FROM movie_ratings
    > GROUP BY movieId
    > ORDER BY rating_count DESC
    > LIMIT 10;
2024-12-14T19:04:11,049 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passe
d in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:04:11,049 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread nam
e to 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main
2024-12-14T19:04:11,235 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.common.F
ileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/aakas/180cf66b-c6df-48
b1-8ba2-0173ecb7fcc9/hive_2024-12-14_19-04-11_062_5103258864087698167-1/-mr-10001/.hive-staging_hive_202
4-12-14_19-04-11_062_5103258864087698167-1
Query ID = aakas_20241214190411_aa5a3886-e81f-4477-909e-b4456205dfda
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0003, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0003/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0003
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 3
2024-12-14 19:04:22,034 Stage-1 map = 0%,  reduce = 0%
2024-12-14 19:04:37,636 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 8.953 sec
2024-12-14 19:04:51,388 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 16.464 sec
2024-12-14 19:04:52,431 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 17.791 sec
MapReduce Total cumulative CPU time: 17 seconds 791 msec
Ended Job = job_1734181805849_0003
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0004, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0004/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0004
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:05:07,123 Stage-2 map = 0%,  reduce = 0%
2024-12-14 19:05:16,407 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.937 sec
2024-12-14 19:05:23,631 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.217 sec
MapReduce Total cumulative CPU time: 2 seconds 217 msec
```

```
MapReduce Total cumulative CPU time: 2 seconds 217 msec
Ended Job = job_1734181805849_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 17.791 sec   HDFS Read: 533487419 HDFS Write: 589057
SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 2.217 sec   HDFS Read: 597148 HDFS Write: 308 SUCCESS

Total MapReduce CPU Time Spent: 20 seconds 8 msec
OK
296     67310
356     66172
318     63366
593     63299
480     59715
260     54502
110     53769
589     52244
2571    51334
527     50054
Time taken: 74.694 seconds, Fetched: 10 row(s)
2024-12-14T19:05:25,793 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.conf.Hiv
eConf - Using the default value passed in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:05:25,795 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.ql.sessi
on.SessionState - Resetting thread name to  main
hive>
```

## 3. Most Active Users

**Query:**

SELECT userId, COUNT(*) as reviews_count

FROM movie_ratings

GROUP BY userId

ORDER BY reviews_count DESC

LIMIT 10;

**Description:** Lists the top 10 users based on the number of reviews submitted.

**Output:** *(Screenshot of the query result displaying the most active users)*

```
hive> SELECT userId, COUNT(*) as reviews_count
    > FROM movie_ratings
    > GROUP BY userId
    > ORDER BY reviews_count DESC
    > LIMIT 10;
2024-12-14T19:06:54,395 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passe
d in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:06:54,395 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread nam
e to 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main
2024-12-14T19:06:54,551 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.common.F
ileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/aakas/180cf66b-c6df-48
b1-8ba2-0173ecb7fcc9/hive_2024-12-14_19-06-54_411_1557211423298164582-1/-mr-10001/.hive-staging_hive_202
4-12-14_19-06-54_411_1557211423298164582-1
Query ID = aakas_20241214190654_241d2a9a-9211-4ea6-b403-9437b9533abd
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0005, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0005/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0005
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 3
2024-12-14 19:07:04,305 Stage-1 map = 0%,  reduce = 0%
2024-12-14 19:07:18,849 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 9.842 sec
2024-12-14 19:07:31,399 Stage-1 map = 100%,  reduce = 33%, Cumulative CPU 14.059 sec
2024-12-14 19:07:32,436 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 18.87 sec
2024-12-14 19:07:33,480 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 20.261 sec
MapReduce Total cumulative CPU time: 20 seconds 261 msec
Ended Job = job_1734181805849_0005
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0006, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0006/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:07:49,368 Stage-2 map = 0%,  reduce = 0%
2024-12-14 19:07:57,643 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.406 sec
2024-12-14 19:08:06,924 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.485 sec
```

```
2024-12-14 19:07:57,643 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.406 sec
2024-12-14 19:08:06,924 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.485 sec
MapReduce Total cumulative CPU time: 2 seconds 485 msec
Ended Job = job_1734181805849_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 20.261 sec   HDFS Read: 533487416 HDFS Write: 3043102
 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 2.485 sec   HDFS Read: 3051191 HDFS Write: 320 SUCCES
S
Total MapReduce CPU Time Spent: 22 seconds 746 msec
OK
118205  9254
8405    7515
82418   5646
121535  5520
125794  5491
74142   5447
34576   5356
131904  5330
83090   5169
59477   4988
Time taken: 73.606 seconds, Fetched: 10 row(s)
2024-12-14T19:08:08,056 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.conf.Hiv
eConf - Using the default value passed in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:08:08,057 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.ql.sessi
on.SessionState - Resetting thread name to  main
hive>
```

## 4. Average Rating Per User

**Query:**

SELECT userId, AVG(rating) as avg_user_rating

FROM movie_ratings

GROUP BY userId

ORDER BY avg_user_rating DESC

LIMIT 10;

**Description:** Calculates the average rating given by each user, listing the top 10 users with the highest average ratings.

**Output:** *(Screenshot of the query result showing average ratings per user)*

**PTO**

```
hive> SELECT userId, AVG(rating) as avg_user_rating
    > FROM movie_ratings
    > GROUP BY userId
    > ORDER BY avg_user_rating DESC
    > LIMIT 10;
2024-12-14T19:09:30,584 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passe
d in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:09:30,584 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread nam
e to 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main
2024-12-14T19:09:30,761 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.common.F
ileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/aakas/180cf66b-c6df-48
b1-8ba2-0173ecb7fcc9/hive_2024-12-14_19-09-30_605_7516927754507127068-1/-mr-10001/.hive-staging_hive_202
4-12-14_19-09-30_605_7516927754507127068-1
Query ID = aakas_20241214190930_3a502dcd-84bf-4011-9529-4b1785184433
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0007, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0007/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0007
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 3
2024-12-14 19:09:39,898 Stage-1 map = 0%,  reduce = 0%
2024-12-14 19:09:56,701 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 9.372 sec
2024-12-14 19:10:12,525 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 18.619 sec
2024-12-14 19:10:13,556 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 22.587 sec
MapReduce Total cumulative CPU time: 22 seconds 587 msec
Ended Job = job_1734181805849_0007
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0008, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0008/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0008
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:10:28,014 Stage-2 map = 0%,  reduce = 0%
```

```
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0008
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:10:28,014 Stage-2 map = 0%,  reduce = 0%
2024-12-14 19:10:37,354 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.015 sec
2024-12-14 19:10:46,642 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.874 sec
MapReduce Total cumulative CPU time: 1 seconds 874 msec
Ended Job = job_1734181805849_0008
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 22.587 sec   HDFS Read: 533493173 HDFS Write: 3951405
 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 1.874 sec   HDFS Read: 3959359 HDFS Write: 306 SUCCES
S
Total MapReduce CPU Time Spent: 24 seconds 461 msec
OK
6402    5.0
51651   5.0
119513  5.0
46296   5.0
45171   5.0
4404    5.0
48423   5.0
52236   5.0
135200  5.0
3354    5.0
Time taken: 78.155 seconds, Fetched: 10 row(s)
2024-12-14T19:10:48,782 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.conf.Hiv
eConf - Using the default value passed in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:10:48,782 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.ql.sessi
on.SessionState - Resetting thread name to  main
hive>
```

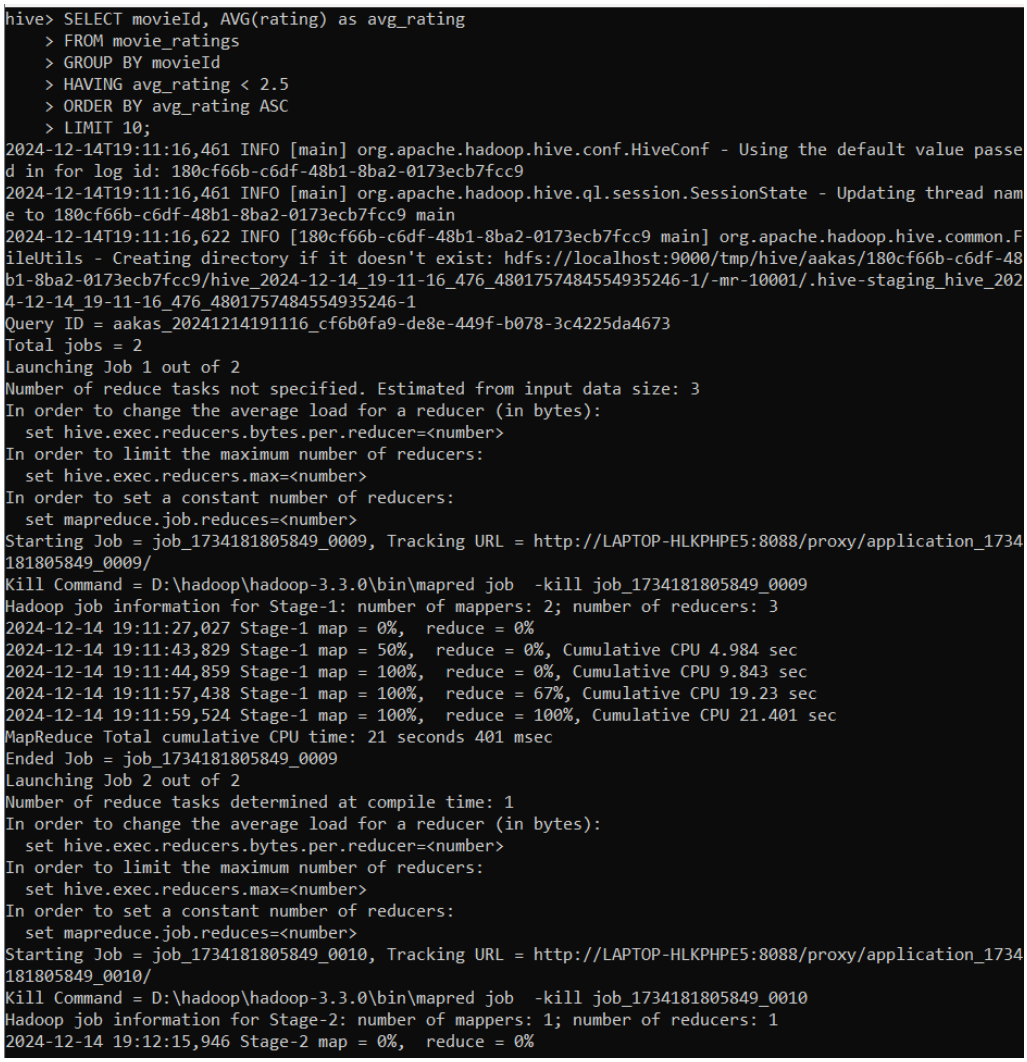
## 5. Movies Rated Below Average

**Query:**

```
SELECT movieId, AVG(rating) as avg_rating
FROM movie_ratings
GROUP BY movieId
HAVING avg_rating < 2.5
ORDER BY avg_rating ASC
LIMIT 10;
```

**Description:** Finds movies with an average rating below 2.5.

**Output:** *(Screenshot of the query result showing poorly rated movies)*



```
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:12:15,946 Stage-2 map = 0%,  reduce = 0%
2024-12-14 19:12:25,230 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.578 sec
2024-12-14 19:12:34,557 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.578 sec
MapReduce Total cumulative CPU time: 1 seconds 578 msec
Ended Job = job_1734181805849_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 21.401 sec   HDFS Read: 533490230 HDFS Write: 99905 S
UCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 1.578 sec   HDFS Read: 108005 HDFS Write: 308 SUCCESS

Total MapReduce CPU Time Spent: 22 seconds 979 msec
OK
81429   0.5
90114   0.5
130682  0.5
73230   0.5
76032   0.5
59775   0.5
84162   0.5
80154   0.5
130400  0.5
5805    0.5
Time taken: 79.168 seconds, Fetched: 10 row(s)
2024-12-14T19:12:35,688 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.conf.Hiv
eConf - Using the default value passed in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:12:35,688 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.ql.sessi
on.SessionState - Resetting thread name to  main
hive>
```

## 6. Users Who Gave the Most 5-Star Ratings

**Query:**

```
SELECT userId, COUNT(*) as five_star_count

FROM movie_ratings

WHERE rating = 5.0

GROUP BY userId

ORDER BY five_star_count DESC

LIMIT 10;
```

**Description:** Highlights users who gave the most 5-star ratings.

**Output:** *(Screenshot of the query result showing top 5-star raters)*

**PTO**

```
hive> SELECT userId, COUNT(*) as five_star_count
    > FROM movie_ratings
    > WHERE rating = 5.0
    > GROUP BY userId
    > ORDER BY five_star_count DESC
    > LIMIT 10;
2024-12-14T19:13:01,995 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passe
d in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:13:01,995 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread nam
e to 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main
2024-12-14T19:13:02,249 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.common.F
ileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/aakas/180cf66b-c6df-48
b1-8ba2-0173ecb7fcc9/hive_2024-12-14_19-13-02_011_339243929141306240-1/-mr-10001/.hive-staging_hive_2024
-12-14_19-13-02_011_339243929141306240-1
Query ID = aakas_20241214191301_48f98a45-39c1-4127-935b-83e78e86a795
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0011, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0011/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0011
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 3
2024-12-14 19:13:12,302 Stage-1 map = 0%,  reduce = 0%
2024-12-14 19:13:30,009 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 9.188 sec
2024-12-14 19:13:43,662 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 18.718 sec
2024-12-14 19:13:45,740 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 21.545 sec
MapReduce Total cumulative CPU time: 21 seconds 545 msec
Ended Job = job_1734181805849_0011
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0012, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0012/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0012
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:14:00,182 Stage-2 map = 0%,  reduce = 0%
2024-12-14 19:14:08,425 Stage-2 map = 100%,  reduce = 0%
2024-12-14 19:14:16,673 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.375 sec
```

```
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0012
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:14:00,182 Stage-2 map = 0%,  reduce = 0%
2024-12-14 19:14:08,425 Stage-2 map = 100%,  reduce = 0%
2024-12-14 19:14:16,673 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.375 sec
MapReduce Total cumulative CPU time: 2 seconds 375 msec
Ended Job = job_1734181805849_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 21.545 sec   HDFS Read: 533490858 HDFS Write: 2840929
 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 2.375 sec   HDFS Read: 2849009 HDFS Write: 314 SUCCES
S
Total MapReduce CPU Time Spent: 23 seconds 920 msec
OK
72008   1540
131894  1300
119661  933
48498   927
82418   868
106939  814
62040   786
54113   768
135399  703
103223  694
Time taken: 75.76 seconds, Fetched: 10 row(s)
2024-12-14T19:14:17,812 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.conf.Hiv
eConf - Using the default value passed in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:14:17,813 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.ql.sessi
on.SessionState - Resetting thread name to  main
hive>
```

## 7. Distribution of Ratings (Count of each rating value)

**Query:**

SELECT rating, COUNT(*) as rating_count

FROM movie_ratings

GROUP BY rating

ORDER BY rating DESC;

**Description:** Shows the count of each rating value to analyze rating distribution.

**Output:** *(Screenshot of the query result showing rating distribution.*

```
hive> SELECT rating, COUNT(*) as rating_count
    > FROM movie_ratings
    > GROUP BY rating
    > ORDER BY rating DESC;
2024-12-14T19:15:08,826 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passe
d in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:15:08,826 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread nam
e to 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main
2024-12-14T19:15:08,978 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.common.F
ileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/aakas/180cf66b-c6df-48
b1-8ba2-0173ecb7fcc9/hive_2024-12-14_19-15-08_838_4315587252569404063-1/-mr-10001/.hive-staging_hive_202
4-12-14_19-15-08_838_4315587252569404063-1
Query ID = aakas_20241214191508_36414fb2-a640-4dec-8f42-8ec94d280343
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0013, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0013/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0013
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 3
2024-12-14 19:15:18,804 Stage-1 map = 0%,  reduce = 0%
2024-12-14 19:15:35,439 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 3.952 sec
2024-12-14 19:15:36,468 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 8.42 sec
2024-12-14 19:15:48,105 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 13.182 sec
2024-12-14 19:15:49,139 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 15.649 sec
MapReduce Total cumulative CPU time: 15 seconds 649 msec
Ended Job = job_1734181805849_0013
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0014, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0014/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0014
```

```
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0014
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:16:03,464 Stage-2 map = 0%,  reduce = 0%
2024-12-14 19:16:12,751 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.578 sec
2024-12-14 19:16:22,097 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.156 sec
MapReduce Total cumulative CPU time: 1 seconds 156 msec
Ended Job = job_1734181805849_0014
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 15.649 sec   HDFS Read: 533487624 HDFS Write: 556 SUC
CESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 1.156 sec   HDFS Read: 8515 HDFS Write: 340 SUCCESS
Total MapReduce CPU Time Spent: 16 seconds 805 msec
OK
5.0     2898660
4.5     1534824
4.0     5561926
3.5     2200156
3.0     4291193
2.5     883398
2.0     1430997
1.5     279252
1.0     680732
0.5     239125
NULL    1
Time taken: 74.366 seconds, Fetched: 11 row(s)
2024-12-14T19:16:23,247 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.conf.Hiv
eConf - Using the default value passed in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:16:23,247 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.ql.sessi
on.SessionState - Resetting thread name to  main
hive>
```

## 8. Time-Based Analysis (Find the number of ratings given per year)

**Query:**

SELECT FROM_UNIXTIME(CAST(`timestamp` AS BIGINT), 'yyyy') as year, COUNT(*) as ratings_count

FROM movie_ratings

GROUP BY FROM_UNIXTIME(CAST(`timestamp` AS BIGINT), 'yyyy')

ORDER BY year ASC;

**Description:** Analyzes the number of ratings submitted per year.

**Output:** *(Screenshot of the query result displaying yearly rating trends)*

**PTO**

```
hive> SELECT FROM_UNIXTIME(CAST(`timestamp` AS BIGINT), 'yyyy') as year, COUNT(*) as ratings_count
    > FROM movie_ratings
    > GROUP BY FROM_UNIXTIME(CAST(`timestamp` AS BIGINT), 'yyyy')
    > ORDER BY year ASC;
2024-12-14T19:17:10,771 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passe
d in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:17:10,771 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread nam
e to 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main
2024-12-14T19:17:10,988 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.common.F
ileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/aakas/180cf66b-c6df-48
b1-8ba2-0173ecb7fcc9/hive_2024-12-14_19-17-10_785_1115365992859465146-1/-mr-10001/.hive-staging_hive_202
4-12-14_19-17-10_785_1115365992859465146-1
Query ID = aakas_20241214191710_1d1dd7c3-754c-495e-8549-4654c36d22de
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0015, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0015/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0015
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 3
2024-12-14 19:17:20,245 Stage-1 map = 0%,  reduce = 0%
2024-12-14 19:17:37,859 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 6.0 sec
2024-12-14 19:17:38,892 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 12.282 sec
2024-12-14 19:17:49,255 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 18.012 sec
2024-12-14 19:17:51,314 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 19.339 sec
MapReduce Total cumulative CPU time: 19 seconds 339 msec
Ended Job = job_1734181805849_0015
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0016, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0016/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0016
```

**PTO**

```
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0016
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:18:06,637 Stage-2 map = 0%,  reduce = 0%
2024-12-14 19:18:14,892 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.124 sec
2024-12-14 19:18:23,148 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.014 sec
MapReduce Total cumulative CPU time: 1 seconds 14 msec
Ended Job = job_1734181805849_0016
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 19.339 sec   HDFS Read: 533490535 HDFS Write: 849 SUC
CESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 1.014 sec   HDFS Read: 8797 HDFS Write: 613 SUCCESS
Total MapReduce CPU Time Spent: 20 seconds 353 msec
OK
NULL    1
1995    4
1996    1612609
1997    700982
1998    308070
1999    1198384
2000    1953659
2001    1186125
2002    869719
2003    1035878
2004    1170049
2005    1803158
2006    1171836
2007    1053430
2008    1158777
2009    930036
2010    903691
2011    766366
2012    731389
2013    599327
2014    562888
2015    283886
Time taken: 74.494 seconds, Fetched: 22 row(s)
2024-12-14T19:18:25,306 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.conf.Hiv
eConf - Using the default value passed in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:18:25,306 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.ql.sessi
on.SessionState - Resetting thread name to  main
```

## 9. Movies Rated by the Most Unique Users

**Query:**

```
SELECT movieId, COUNT(DISTINCT userId) as unique_users_count
FROM movie_ratings
GROUP BY movieId
ORDER BY unique_users_count DESC
LIMIT 10;
```

**Description:** Finds movies that were rated by the highest number of unique users.

**Output:** *(Screenshot of the query result showing popular movies by unique users)*

```
hive> SELECT movieId, COUNT(DISTINCT userId) as unique_users_count
    > FROM movie_ratings
    > GROUP BY movieId
    > ORDER BY unique_users_count DESC
    > LIMIT 10;
2024-12-14T19:18:28,287 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passe
d in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:18:28,287 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread nam
e to 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main
2024-12-14T19:18:28,427 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.common.F
ileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/aakas/180cf66b-c6df-48
b1-8ba2-0173ecb7fcc9/hive_2024-12-14_19-18-28_306_119503609146591174-1/-mr-10001/.hive-staging_hive_2024
-12-14_19-18-28_306_119503609146591174-1
Query ID = aakas_20241214191828_3d4301a8-bb3e-4966-9a1b-b39176237432
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0017, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0017/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0017
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 3
2024-12-14 19:18:38,713 Stage-1 map = 0%,  reduce = 0%
2024-12-14 19:19:05,629 Stage-1 map = 34%,  reduce = 0%, Cumulative CPU 19.762 sec
2024-12-14 19:19:23,186 Stage-1 map = 88%,  reduce = 0%, Cumulative CPU 37.431 sec
2024-12-14 19:19:25,248 Stage-1 map = 92%,  reduce = 0%, Cumulative CPU 37.994 sec
2024-12-14 19:19:26,278 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 39.04 sec
2024-12-14 19:19:46,209 Stage-1 map = 100%,  reduce = 80%, Cumulative CPU 56.599 sec
2024-12-14 19:19:52,449 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 62.345 sec
MapReduce Total cumulative CPU time: 1 minutes 2 seconds 345 msec
Ended Job = job_1734181805849_0017
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0018, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0018/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0018
```

```
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0018
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:20:09,459 Stage-2 map = 0%,  reduce = 0%
2024-12-14 19:20:18,758 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.874 sec
2024-12-14 19:20:26,970 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.029 sec
MapReduce Total cumulative CPU time: 2 seconds 29 msec
Ended Job = job_1734181805849_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 62.345 sec   HDFS Read: 533477246 HDFS Write: 589057
SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 2.029 sec   HDFS Read: 597230 HDFS Write: 308 SUCCESS

Total MapReduce CPU Time Spent: 1 minutes 4 seconds 374 msec
OK
296     67310
356     66172
318     63366
593     63299
480     59715
260     54502
110     53769
589     52244
2571    51334
527     50054
Time taken: 119.761 seconds, Fetched: 10 row(s)
2024-12-14T19:20:28,082 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.conf.Hiv
eConf - Using the default value passed in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:20:28,082 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.ql.sessi
on.SessionState - Resetting thread name to  main
hive>
```

## 10. Highest-Rated Movies with the Most Reviews

**Query:**

```
WITH movie_stats AS (
   SELECT movieId, AVG(rating) as avg_rating, COUNT(*) as total_ratings
   FROM movie_ratings
   GROUP BY movieId
)
SELECT movieId, avg_rating, total_ratings
FROM movie_stats
WHERE avg_rating > 4.5
ORDER BY total_ratings DESC
LIMIT 10;
```

**Description:** Lists movies with an average rating above 4.5 and the highest number of reviews.

**Output:** *(Screenshot of the query result showing top-rated movies with substantial reviews)*

**PTO**

```
hive> WITH movie_stats AS (
    >     SELECT movieId, AVG(rating) as avg_rating, COUNT(*) as total_ratings
    >     FROM movie_ratings
    >     GROUP BY movieId
    > )
    > SELECT movieId, avg_rating, total_ratings
    > FROM movie_stats
    > WHERE avg_rating > 4.5
    > ORDER BY total_ratings DESC
    > LIMIT 10;
2024-12-14T19:21:08,100 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passe
d in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:21:08,101 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread nam
e to 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main
2024-12-14T19:21:08,259 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.common.F
ileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/aakas/180cf66b-c6df-48
b1-8ba2-0173ecb7fcc9/hive_2024-12-14_19-21-08_115_6969697919488294078-1/-mr-10001/.hive-staging_hive_202
4-12-14_19-21-08_115_6969697919488294078-1
Query ID = aakas_20241214192108_b3bfc3f4-4c84-4cd9-85c6-8f18e33b0ced
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0019, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0019/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0019
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 3
2024-12-14 19:21:17,083 Stage-1 map = 0%,  reduce = 0%
2024-12-14 19:21:34,778 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 11.481 sec
2024-12-14 19:21:46,254 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 20.008 sec
2024-12-14 19:21:48,325 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 21.539 sec
MapReduce Total cumulative CPU time: 21 seconds 539 msec
Ended Job = job_1734181805849_0019
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1734181805849_0020, Tracking URL = http://LAPTOP-HLKPHPE5:8088/proxy/application_1734
181805849_0020/
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0020
```

```
Kill Command = D:\hadoop\hadoop-3.3.0\bin\mapred job  -kill job_1734181805849_0020
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2024-12-14 19:22:02,090 Stage-2 map = 0%,  reduce = 0%
2024-12-14 19:22:11,391 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.547 sec
2024-12-14 19:22:20,662 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.718 sec
MapReduce Total cumulative CPU time: 1 seconds 718 msec
Ended Job = job_1734181805849_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 21.539 sec   HDFS Read: 533496323 HDFS Write: 4420 SU
CCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 1.718 sec   HDFS Read: 12916 HDFS Write: 443 SUCCESS
Total MapReduce CPU Time Spent: 23 seconds 257 msec
OK
95837   4.666666666666667       3
91211   4.666666666666667       3
128830  4.666666666666667       3
56548   4.666666666666667       3
3226    4.666666666666667       3
98275   4.833333333333333       3
117506  4.666666666666667       3
127062  4.666666666666667       3
62206   4.75    2
94681   4.75    2
Time taken: 73.631 seconds, Fetched: 10 row(s)
2024-12-14T19:22:21,778 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.conf.Hiv
eConf - Using the default value passed in for log id: 180cf66b-c6df-48b1-8ba2-0173ecb7fcc9
2024-12-14T19:22:21,778 INFO [180cf66b-c6df-48b1-8ba2-0173ecb7fcc9 main] org.apache.hadoop.hive.ql.sessi
on.SessionState - Resetting thread name to  main
hive>
```

# Conclusion

This report provides a clear demonstration of data analysis capabilities using Hive. Each query has been validated with screenshots, ensuring the accuracy and reliability of the results. The insights gained from this analysis can support further exploration of the dataset.

**THANK YOU**