

一、数据探查

```
[student4@emr-header-1 hive]$ hadoop fs -ls /data/hive/
Found 10 items
-rw-r----- 2 student1 hadoop 171432 2022-04-03 13:56 /data/hive/movie
drwxr-x--x - hdfs hadoop 0 2022-04-03 22:29 /data/hive/movies
drwxr-x--x - student3 hadoop 0 2022-04-01 17:32 /data/hive/movies.dat
drwxr-x--x - student1 hadoop 0 2022-04-03 14:06 /data/hive/pandi
drwxr-x--x - student1 hadoop 0 2022-04-01 14:41 /data/hive/rating
drwxr-x--x - student hadoop 0 2022-04-03 16:45 /data/hive/ratings
drwxr-x--x - student3 hadoop 0 2022-04-02 17:00 /data/hive/ratings.dat
drwxr-x--x - student1 hadoop 0 2022-04-03 16:49 /data/hive/user
drwxr-x--x - student hadoop 0 2022-04-03 22:27 /data/hive/users
drwxr-x--x - anonymous hadoop 0 2022-04-02 15:59 /data/hive/users.dat
```

```
[student4@emr-header-1 hive]$ hadoop fs -ls /data/hive/users/
Found 1 items
-rw-r----- 2 student1 hadoop 134368 2022-04-03 22:27 /data/hive/users/users.dat
[student4@emr-header-1 hive]$ hadoop fs -ls /data/hive/movies/
Found 1 items
-rw-r----- 2 student1 hadoop 171432 2022-04-03 22:29 /data/hive/movies/movies.dat
[student4@emr-header-1 hive]$ hadoop fs -ls /data/hive/ratings/
Found 1 items
-rw-r----- 2 student1 hadoop 24594131 2022-04-03 16:45 /data/hive/ratings/ratings.dat
[student4@emr-header-1 hive]$
```

```
[student4@emr-header-1 hive]$ hadoop fs -cat /data/hive/users/* | head
22/04/03 22:39:10 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false
se, remoteHostTrusted = false
1::F::1::10::48067
2::M::56::16::70072
3::M::25::15::55117
4::M::45::7::02460
5::M::25::20::55455
6::F::50::9::55117
7::M::35::1::06810
8::M::25::12::11413
9::M::25::17::61614
10::F::35::1::95370
cat: Unable to write to output stream.
```

```
[student4@emr-header-1 hive]$ hadoop fs -cat /data/hive/movies/* | head
22/04/03 22:40:44 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false
se, remoteHostTrusted = false
1::Toy Story (1995)::Animation|Children's|Comedy
2::Jumanji (1995)::Adventure|Children's|Fantasy
3::Grumpier Old Men (1995)::Comedy|Romance
4::Waiting to Exhale (1995)::Comedy|Drama
5::Father of the Bride Part II (1995)::Comedy
6::Heat (1995)::Action|Crime|Thriller
7::Sabrina (1995)::Comedy|Romance
8::Tom and Huck (1995)::Adventure|Children's
9::Sudden Death (1995)::Action
10::GoldenEye (1995)::Action|Adventure|Thriller
cat: Unable to write to output stream.
```

```

cat: Unable to write to output stream.
[student4@emr-header-1 hive]$ hadoop fs -cat /data/hive/ratings/* | head
22/04/03 22:41:20 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
1::1193::5::978300760
1::661::3::978302109
1::914::3::978301968
1::3408::4::978300275
1::2355::5::978824291
1::1197::3::978302268
1::1287::5::978302039
1::2804::5::978300719
1::594::4::978302268
1::919::4::978301368
cat: Unable to write to output stream.

```

二、 确认 Hive 数据表已创建

```

hive> show tables;
OK
leo_movie
leo_rating
leo_user
mm_movies
mostrate_rau
movies
see2seebase
student_sc_th_txt
t_movie
t_movie_chloe
t_movie_jz
t_movie_lm
t_movie_xz
t_movie_ygs
t_movie_zanglei
t_movie_zdd
t_rating
t_rating_j
t_rating_lm
t_rating_jz
t_rating_ygs
t_rating_zanglei
t_rating_zdd
t_rating_chloe
t_user
t_users
t_user_chloe
t_user_jz
t_user_lm
t_user_xz
t_user_ygs
t_user_zanglei
t_user_zdd
terry_movie
terry_ratings
terry_user
test_partition
Time taken: 0.439 seconds, Fetched: 37 row(s)

```

```

Time taken: 0.199 seconds, Fetched: 10 row(s)
hive> select * from t_user limit 10;
OK
1      F      1      10      48067
2      M      56      16      70072
3      M      25      15      55117
4      M      45      7       02460
5      M      25      20      55455
6      F      50      9       55117
7      M      35      1       06810
8      M      25      12      11413
9      M      25      17      61614
10     F      35      1       95370
Time taken: 1.223 seconds, Fetched: 10 row(s)
hive> select * from t_movie limit 10;
OK
1      Toy Story (1995)      Animation|Children's|Comedy
2      Jumanji (1995)      Adventure|Children's|Fantasy
3      Grumpier Old Men (1995) Comedy|Romance
4      Waiting to Exhale (1995) Comedy|Drama
5      Father of the Bride Part II (1995) Comedy
6      Heat (1995)      Action|Crime|Thriller
7      Sabrina (1995)      Comedy|Romance
8      Tom and Huck (1995)      Adventure|Children's
9      Sudden Death (1995)      Action
10     GoldenEye (1995)      Action|Adventure|Thriller
Time taken: 0.112 seconds, Fetched: 10 row(s)
hive> select * from t_rating limit 10;
OK
1      1193      5      978300760
1      661       3      978302109
1      914       3      978301968
1      3408      4      978300275
1      2355      5      978824291
1      1197      3      978302268
1      1287      5      978302039
1      2804      5      978300719
1      594       4      978302268
1      919       4      978301368
Time taken: 0.114 seconds, Fetched: 10 row(s)
hive>

```

三、 解题目一

```

Select u.age age, avg(r.rate) avgrate
From t_user u, t_rating r
Where u.userid = r.userid
And r.movieid = 2116
Group by u.age;

```

```
hive> Select u.age age, avg(r.rate) avgrate
> From t_user u, t_rating r
> Where u.userid = r.userid
> And r.movieid = 2116
> Group by u.age;

Query ID = student4_20220403231211_0374da7b-e6ef-4e35-9626-5a0e1f1e573f
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1645699879292_2023)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 2 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 3 ..... container  SUCCEEDED    2         2         0         0         0         0
-----
VERTICES: 03/03 [=====>>] 100% ELAPSED TIME: 9.58 s
-----
OK
1      3.2941176470588234
18     3.3580246913580245
25     3.436548223350254
35     3.2278481012658227
45     2.8275862068965516
50     3.32
56     3.5
Time taken: 16.588 seconds, Fetched: 7 row(s)
hive> █
```

四、 解题目二

```
Select u.sex sex, m.moviename name, avg(r.rate) avgrate, count(*) total
From t_user u, t_rating r, t_movie m
Where u.userid = r.userid
And m.movieid = r.movieid
And u.sex = 'M'
Group by u.sex,m.moviename
Having total >= 50
Order by avgrate desc
Limit 10;
```

```

hive> Select u.sex sex, m.moviename name, avg(r.rate) avgrate, count(*) total
> From t_user u, t_rating r, t_movie m
> Where u.userid = r.userid
> And m.movieid = r.movieid
> And u.sex = 'M'
> Group by u.sex,m.moviename
> Having total >= 50
> Order by avgrate desc
> Limit 10;

```

Query ID = student4_20220403234225_91762b7b-0b05-4429-ae19-872bdc31dab3

Total jobs = 1

Launching Job 1 out of 1

Tez session was closed. Reopening...

Session re-established.

Session re-established.

Status: Running (Executing on YARN cluster with App id application_1645699879292_2037)

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Map 5	container	SUCCEEDED	2	2	0	0	0	0
Map 2	container	SUCCEEDED	1	1	0	0	0	0
Reducer 3	container	SUCCEEDED	2	2	0	0	0	0
Reducer 4	container	SUCCEEDED	1	1	0	0	0	0

VERTICES: 05/05 [=====] 100% ELAPSED TIME: 12.32 s

OK

M	Sanjuro (1962)	4.639344262295082	122
M	Godfather, The (1972)	4.583333333333333	3480
M	Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954)	4.576628352490421	1044
M	Shawshank Redemption, The (1994)	4.560625	3200
M	Raiders of the Lost Ark (1981)	4.520597322348094	3884
M	Usual Suspects, The (1995)	4.518248175182482	2740
M	Star Wars: Episode IV - A New Hope (1977)	4.495307167235495	4688
M	Schindler's List (1993)	4.49141503848431	3378
M	Paths of Glory (1957)	4.485148514851486	404
M	Wrong Trousers, The (1993)	4.478260869565218	1288

Time taken: 20.142 seconds Fetched: 10 row(s)