

Group members

Jiale Sun :JIS96@pitt@edu

Kyuri Song: kys15@pitt@edu

Song Tang: sot13@pitt@edu

Questions:

1)What genre is the movie CopyCat in?

```
db.movies.find({"Title": /Copycat/ }).pretty()
```

```
[> db.movies.find({"Title": /Copycat/}).pretty()
{
  "_id" : ObjectId("58a12862ca7267baaae5b8f6"),
  "MovieID" : "22",
  "Title" : "Copycat (1995)",
  "Genres" : "Crime|Drama|Horror|Mystery|Thriller\n"
}
```

the Genres are Crime, Drama, Horror, Mystery, and Thriller.

2)what genre has the most movies?

the genre which has the most movies is Drama, the following diagram is query language.

```
> action=db.movies.find({"Genres": { $in: ['Action']}}).count()
1473
> adventure=db.movies.find({"Genres": { $in: ['Adventure']}}).count()
1025
> animation=db.movies.find({"Genres": { $in: ['Animation']}}).count()
286
> children=db.movies.find({"Genres": { $in: ["Children's"]}}).count()
0
> comedy=db.movies.find({"Genres": { $in: ["Comedy"]}}).count()
3703
> crime=db.movies.find({"Genres": { $in: ["Crime"]}}).count()
1118
> documentary=db.movies.find({"Genres": { $in: ["Documentary"]}}).count()
482
> drama=db.movies.find({"Genres": { $in: ["Drama"]}}).count()
5339
> fantasy=db.movies.find({"Genres": { $in: ["Fantasy"]}}).count()
543
> filmnoir=db.movies.find({"Genres": { $in: ["Film-Noir"]}}).count()
148
> horror=db.movies.find({"Genres": { $in: ["Horror"]}}).count()
1013
> musical=db.movies.find({"Genres": { $in: ["Musical"]}}).count()
436
> mystery=db.movies.find({"Genres": { $in: ["Musical"]}}).count()
436
> mystery=db.movies.find({"Genres": { $in: ["Mystery"]}}).count()
509
> romance=db.movies.find({"Genres": { $in: ["Romance"]}}).count()
1685
> scifi=db.movies.find({"Genres": { $in: ["Sci-Fi"]}}).count()
754
> thriller=db.movies.find({"Genres": { $in: ["Thriller"]}}).count()
1706
> war=db.movies.find({"Genres": { $in: ["War"]}}).count()
511
> western=db.movies.find({"Genres": { $in: ["Western"]}}).count()
275
~
```

3)what tags did user 146 use to describe the movie "2001: A Space Odyssey"

```
> db.movies.find({"Title": /A Space Odyssey/}).pretty()
{
  "_id" : ObjectId("58a12863ca7267baaae5bc6c"),
  "MovieID" : "924",
  "Title" : "2001: A Space Odyssey (1968)",
  "Genres" : "Adventure|Sci-Fi\n"
}
```

we query the MovieID of "2001: A Space Odyssey" the movieID is 924.
and then we query the tags in tags collections

```
[> db.tags.find({'$and': [{"UserID":"146"}, {"MovieID":"924"}]}).pretty()
{
  "_id" : ObjectId("58a5a905ca72672ed24d0e33"),
  "UserID" : "146",
  "MovieID" : "924",
  "Tags" : "Arthur C. Clarke",
  "Timestamp" : "1220880014\n"
}
{
  "_id" : ObjectId("58a5a905ca72672ed24d0e34"),
  "UserID" : "146",
  "MovieID" : "924",
  "Tags" : "artificial intelligence",
  "Timestamp" : "1195559992\n"
}
{
  "_id" : ObjectId("58a5a905ca72672ed24d0e35"),
  "UserID" : "146",
  "MovieID" : "924",
  "Tags" : "based on a book",
  "Timestamp" : "1162263470\n"
}
> █
```

the tags for movies "2001: A Space Odyssey" of user 146 are "Arthur C. Clarks", "artificial intelligence", and "based on a book"

4)What are the top 5 movies titles with the highest avg rating?

First of all create a collection to store the movies ID and average of each movies ratings

```
> answer=db.Rating.group(
... {
...   key: { "MovieID": 1},
...   initial: { total: 0, count: 0 },
...   reduce: function( curr, result) {
...     result.total += curr.Rating;
...     result.count++;
...   },
...   finalize: function(result) {
...     result.avg = result.total/result.count;
...     delete result.total;
...     delete result.count;
...   }
... })
[
  {
    "MovieID" : "122",
    "avg" : 2.861318407960199
  },
  {
    "MovieID" : "185",
    "avg" : 3.1252086811352253
  },
  {
    "MovieID" : "231",
    "avg" : 2.936950310906952
  },
  {
    "MovieID" : "292",
    "avg" : 3.4184136858475895
  },
  {
    "MovieID" : "316",
    "avg" : 3.3493527080581242
  },
  {
    "MovieID" : "329",
    "avg" : 3.3362714170841836
  },
  {
    "MovieID" : "355",
    "avg" : 2.480152814014163
  },
  {
    "MovieID" : "356",
    "avg" : 4.0135821458629595
  },
  {
    "MovieID" : "362",
    "avg" : 3.4477091633466137
  },
  {
    "MovieID" : "364",
    "avg" : 3.7485933625786765
  },
  {
    "MovieID" : "370",
    "avg" : 2.9600487210718636
  },
  {
    "MovieID" : "377",
    "avg" : 3.5289498062994777
  }
]
```

and then insert it into a new collection

```
> db.moviesavg.insert(answer)
BulkWriteResult({
  "writeErrors" : [ ],
  "writeConcernErrors" : [ ],
  "nInserted" : 10677,
  "nUpserted" : 0,
  "nMatched" : 0,
  "nModified" : 0,
  "nRemoved" : 0,
  "upserted" : [ ]
})
```

and find the top 5 ratings and movies ID

```
> db.moviesavg.find().sort({ avg: -1}).limit(5)
[ { "_id" : ObjectId("58ab456950dc10e2ed56e1c1"), "MovieID" : "53355", "avg" : 5 }
  { "_id" : ObjectId("58ab456950dc10e2ed56e1c3"), "MovieID" : "51209", "avg" : 5 }
  { "_id" : ObjectId("58ab456950dc10e2ed56e1da"), "MovieID" : "33264", "avg" : 5 }
  { "_id" : ObjectId("58ab456950dc10e2ed56e1ff"), "MovieID" : "42783", "avg" : 5 }
  { "_id" : ObjectId("58ab456950dc10e2ed56e224"), "MovieID" : "64275", "avg" : 5 }
```

and then find the movie's title of top 5 movies ratings

```
[> db.movies.find( { $or:[ {"MovieID":"53355"}, {"MovieID":"51209"}, {"MovieID":"33264"}, {"MovieID":"42783"}
, {"MovieID":"64275"} ]}).pretty()
{
  "_id" : ObjectId("58a12865ca7267baaae5dc5c"),
  "MovieID" : "33264",
  "Title" : "Satan's Tango (Sátántangó) (1994)",
  "Genres" : "Drama\n"
}
{
  "_id" : ObjectId("58a12865ca7267baaae5dda1"),
  "MovieID" : "42783",
  "Title" : "Shadows of Forgotten Ancestors (1964)",
  "Genres" : "Drama|Romance\n"
}
{
  "_id" : ObjectId("58a12865ca7267baaae5df2e"),
  "MovieID" : "51209",
  "Title" : "Fighting Elegy (Kenka erejii) (1966)",
  "Genres" : "Action|Comedy\n"
}
{
  "_id" : ObjectId("58a12865ca7267baaae5df98"),
  "MovieID" : "53355",
  "Title" : "Sun Alley (Sonnenallee) (1999)",
  "Genres" : "Comedy|Romance\n"
}
{
  "_id" : ObjectId("58a12866ca7267baaae5e257"),
  "MovieID" : "64275",
  "Title" : "Blue Light, The (Das Blaue Licht) (1932)",
  "Genres" : "Drama|Fantasy|Mystery\n"
}
```

the top 5 movies are "Satan's Tango (1994)", "Shadows of Forgotten Ancestors (1964)", "Fighting Elegy (1966)", "Sun Alley (1999)" and "Blue Light, The (Das Blaue Licht) (1932)".

5) Write 3 different queries of your choice to demonstrate that your data storage is working.

a) find all the rating information for a specified movie called "Heat(1995)"

```
db.movies.aggregate(  
  [  
    {  
      $lookup: {  
        from:"ratings",  
        localField:"MovieID",  
        foreignField:"MovieID",  
        as:"rating_docs" }  
      },  
    { $out:"movies_ratings" } ] ).pretty()  
  
db.movies_ratings.find({"Title":"Heat (1995)"}).pretty()
```

```
> db.movies_ratings.find({"Title":"Heat (1995)"}).pretty()  
{  
  "_id" : ObjectId("58a12862ca7267bbaae5b8e6"),  
  "MovieID" : "6",  
  "Title" : "Heat (1995)",  
  "Genres" : "Action|Crime|Thriller\\n",  
  "rating_docs" : [  
    {  
      "_id" : ObjectId("58a51125ca726723c2ec01e1"),  
      "UserID" : "8",  
      "MovieID" : "6",  
      "Rating" : "4",  
      "Timestamp" : "1116547028\\n"  
    },  
    {  
      "_id" : ObjectId("58a51126ca726723c2ec0a4f"),  
      "UserID" : "24",  
      "MovieID" : "6",  
      "Rating" : "3",  
      "Timestamp" : "868254301\\n"  
    },  
    {  
      "_id" : ObjectId("58a51126ca726723c2ec0b8f"),  
      "UserID" : "33",  
      "MovieID" : "6",  
      "Rating" : "3",  
      "Timestamp" : "849544240\\n"  
    },  
    {  
      "_id" : ObjectId("58a51126ca726723c2ec0eae"),  
      "UserID" : "36",  
      "MovieID" : "6",  
      "Rating" : "5",  
      "Timestamp" : "1049770748\\n"  
    },  
    {  
      "_id" : ObjectId("58a51127ca726723c2ec18ad"),  
      "UserID" : "57",  
      "MovieID" : "6",  
      "Rating" : "4",  
      "Timestamp" : "868259640\\n"  
    },  
    {  
      "_id" : ObjectId("58a51128ca726723c2ec1ad6"),  
      "UserID" : "64",  
      "MovieID" : "6",  
      "Rating" : "3",  
      "Timestamp" : "857914274\\n"  
    },  
    {  
      "_id" : ObjectId("58a51128ca726723c2ec1ec1"),  
      "UserID" : "73",  
      "MovieID" : "6",  
      "Rating" : "5",  
      "Timestamp" : "974298142\\n"  
    },  
    {  
      "_id" : ObjectId("58a51128ca726723c2ec2611"),  
      "UserID" : "90",  
      "Rating" : "5",  
      "Timestamp" : "974298142\\n"  
    }  
  ]  
}
```


b) what movies has the lowest average ratings and what are the tags for this movies??
find the movieID of the lowest average ratings movie

```
[> db.moviesavg.find().sort({ avg: 1}).limit(1).pretty()
{
  "_id" : ObjectId("58ab456950dc10e2ed56e018"),
  "MovieID" : "5805",
  "avg" : 0.5
}
```

and then look up the movie Title

```
> db.movies.find({"MovieID": "5805"}).pretty()
{
  "_id" : ObjectId("58ab64b8ca7267edf818965b"),
  "MovieID" : "5805",
  "Title" : "Besotted (2001)",
  "Genres" : [
    "Drama"
  ]
}
```

the movie is "Besotted (2001)"

c) how many ratings for this movies "Man of the House (1995)"?

```
[> db.movies.find({"Title": "Man of the House (1995)"}).pretty()
{
  "_id" : ObjectId("58ab64b7ca7267edf818811f"),
  "MovieID" : "274",
  "Title" : "Man of the House (1995)",
  "Genres" : [
    "Comedy"
  ]
}
```

```
[> db.Rating.find({"MovieID": "274"}).count()
732
```

there are 732 ratings is for movie "Man of the House (1995)"