HW1 -Part4 Report

* Write a Java (could be a console app - will only run once to import the data into

MongoDB) program to read the following file, and insert into 3 different collections (movies, ratings, tags).

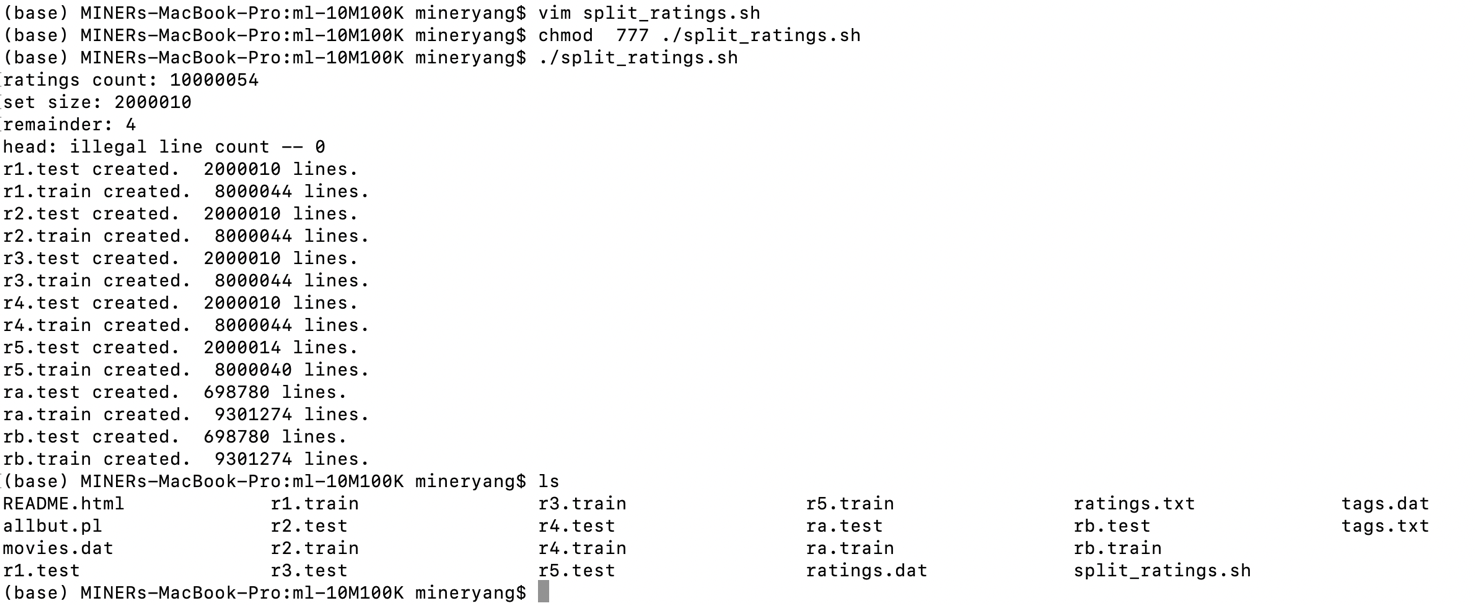
1. Write Java app (IDE NetBeans 8.2)

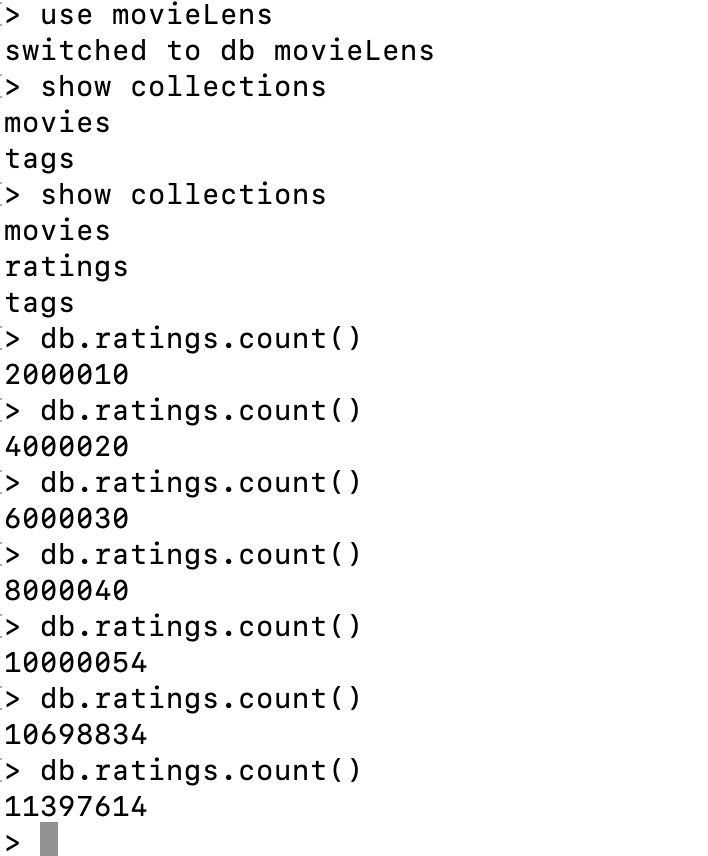
\*Project1: TXTread applied to read records from movies.txt into MongoDB

\*Project2: TXTreadTag applied to read records from tags.txt into MongoDB

\*Project3: TXTreadRating applied to read records from ratings.dat into MongoDB

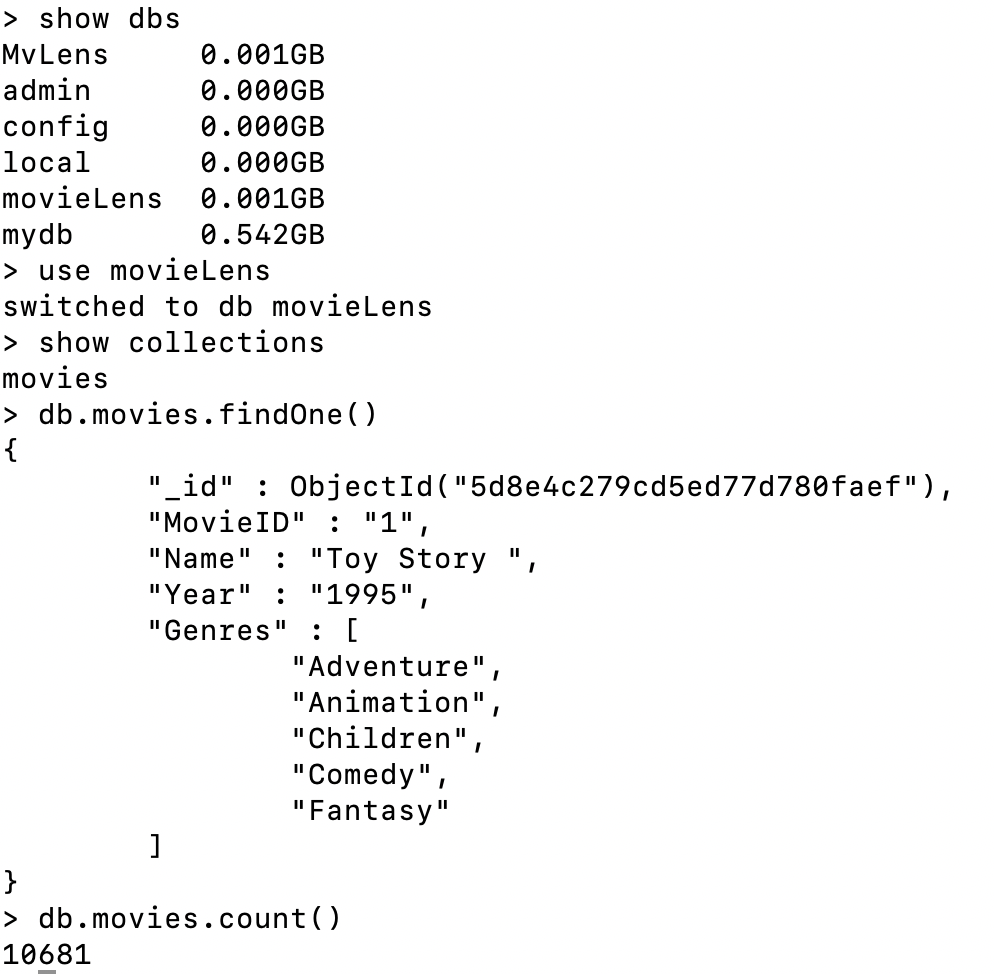
Since ratings.dat is large and GC overhead limited, can’t be read at once. So, I use split\_ratings.sh file to separate ratings.dat into 7 small pieces and import one by one.





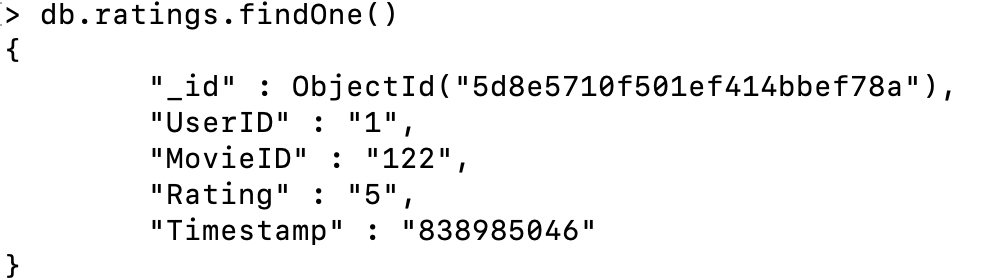
1. Movies

In movies dataset, have format like MovieID::Title::Genres. When reading file, I separate Title into Name field and Year Field. And there are several kinds of genres in the Genres fields, so they are be stored in a List. Structure shown as below. There are 10681 documents in total in the movies-collection.



1. Ratings

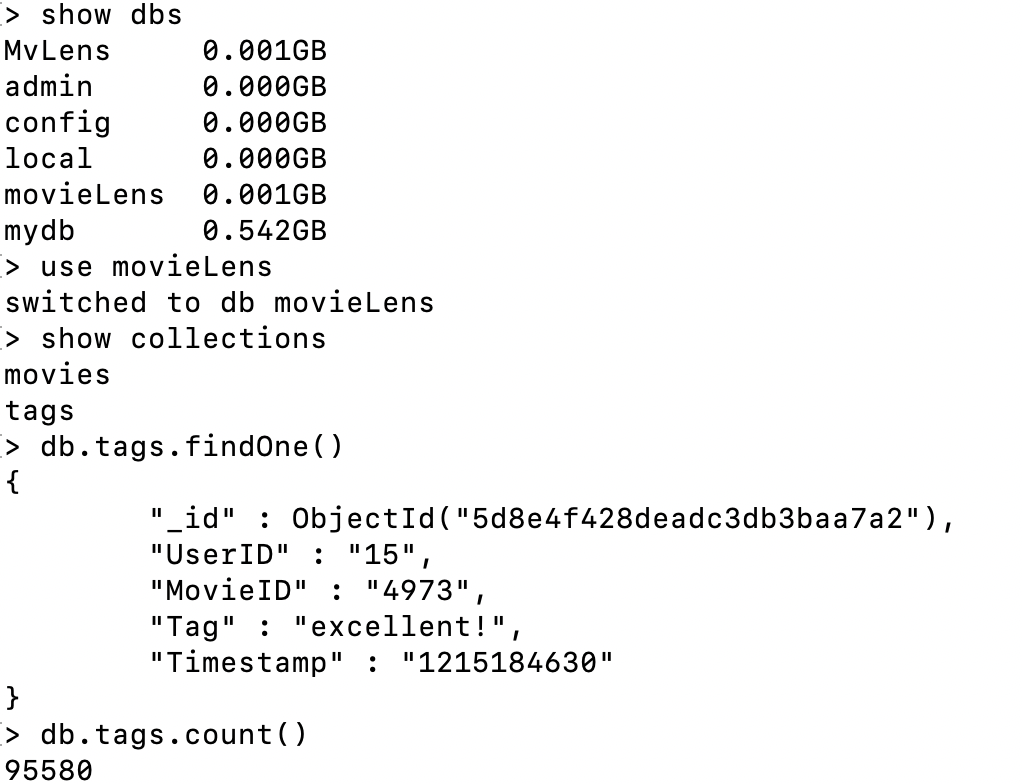
In rating dataset have format like UserID::MovieID::Rating::Timestamp as shown below. There are 11397614 documents in the ratings-collection.



1. Tags

In tags dataset have format like UserID::MovieID::Tag::Timestamp

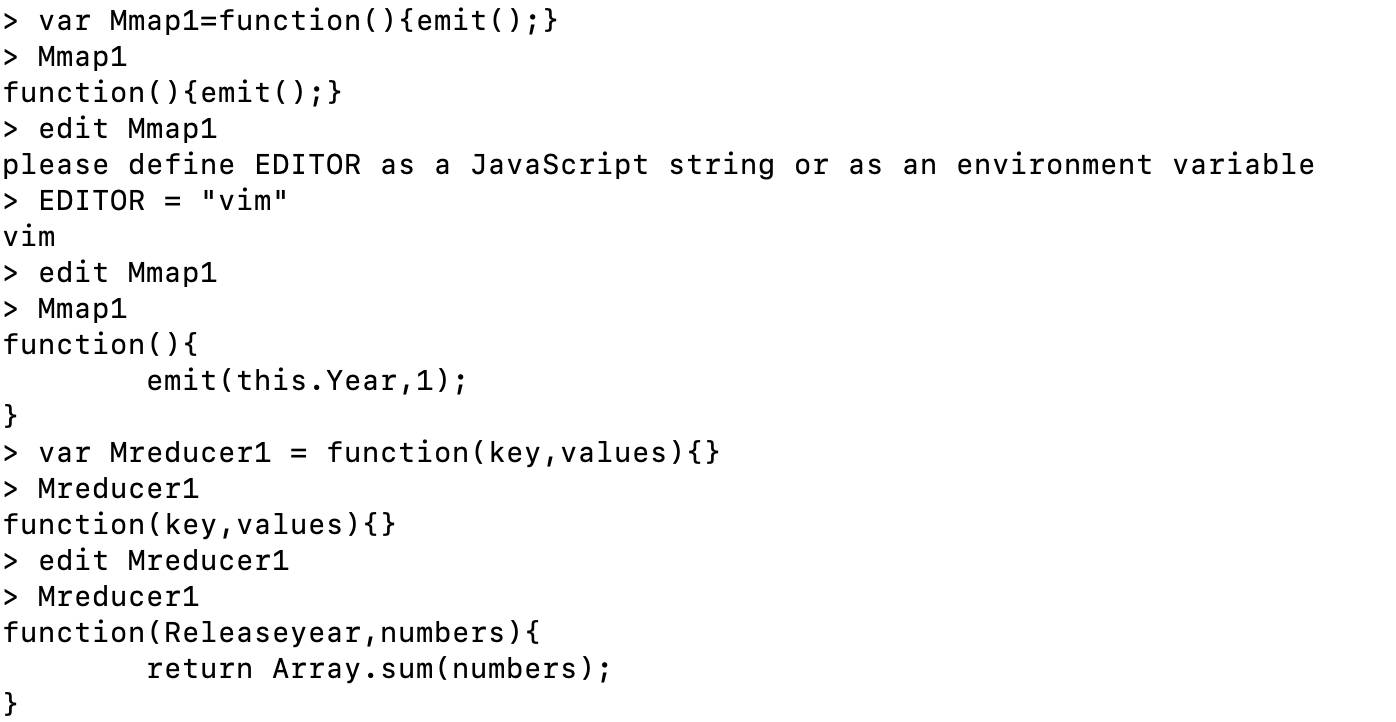
as shown below. There are 95580 documents in the tags-collection.



* Write a MapReduce to do the followings:

1. Number of Movies released per year (Movies Collection)

**Process:**



**Result**:

A close up of a piece of paper

Description automatically generated

A close up of text on a white background

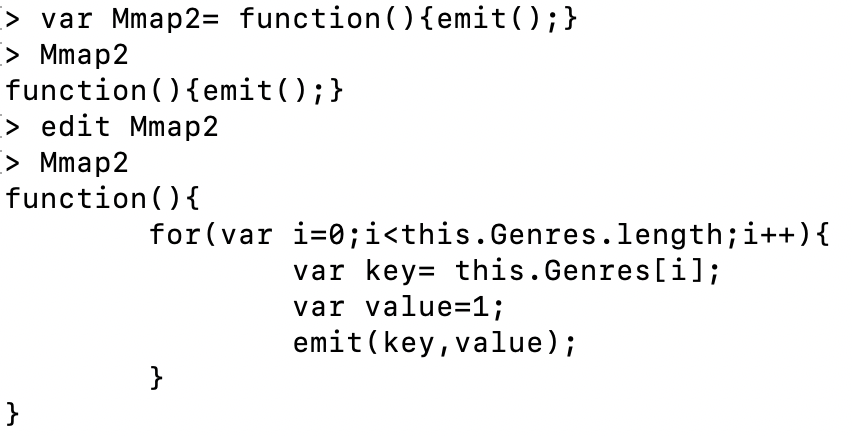
Description automatically generated

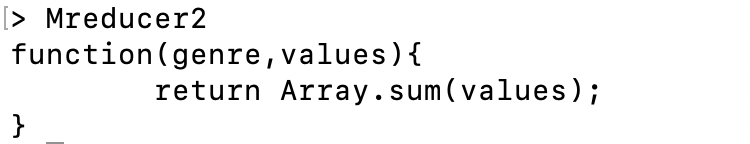
A close up of text on a white background

Description automatically generated

1. Number of Movies per genre (Movies Collection)

**Process:**





A picture containing screenshot

Description automatically generated

**Result:**

**A close up of text on a white background

Description automatically generated**

1. Number of Movies per rating (Ratings Collection)

**Process:**

A screenshot of a cell phone

Description automatically generated

**Result:**

**A close up of text on a white background

Description automatically generated**

1. Number of times each movie was tagged (Tags Collection)

**Process**:

A screenshot of a cell phone

Description automatically generated

**Result**:(part of)

A close up of a piece of paper

Description automatically generated