

Assignment 2: MongoDB

MongoDB (help link: <https://docs.mongodb.com/>)

We have provided a data set which consists of 10 million ratings and 100,000 tag applications applied to 10,000 movies by 72,000 users.

Design data structures to represent the data (you might have one document that stores all information aggregated under user object or one document that stores information about aggregate under movie object, or split into two or tree aggregations).

Your models should be able to answer the following queries below.

Create a data store. Extract, transform and load the provided data set into your data storage.

Answer the following questions.

Questions:

- 1) what genre is the movie CopyCat in?
- 2) what genre has the most movies?
- 3) what tags did user 146 use to describe the movie "2001: A Space Odyssey"
- 4) what are the top 5 movies with the highest avg rating?
- 5) write 3 different queries of your choice to demonstrate that your data storage is working.

Hints:

Find "Install MongoDB Community Edition" and follow the instructions for your OS

The [readme](#) file contains a description of the data files.

In each of the case, you will need to write a program in a programming language of your choice (one that is, of course, supported by your OS).

For your information, Java and Python are supported by most of the OSs.

If you cannot answer one of the 3 specific queries, please write the reasons why you cannot do that. Maybe it is the limit of the data model or data store or query functionality.

You should submit a zip file with your program and a brief document explaining what you did, queries that you asked, and important parts of the results.