****

**NoSQL Assignment**

**By**

**Group 6**

**Mr. Wasin Heesawat 6288077**

**Ms. Intr-orn Lertsupakitsin 6288089**

**Mr. Krissanapong Palakham 6288102**

**Mr. Pongsakorn Piboonpongpun 6288107**

**Presented to**

**Dr. Wudhichart Sawangphol**

**A Report Submitted in Partial Fulfillment of**

**the Requirements for**

**ITCS495 Special Topics in Database and Intelligent Systems**

**Faculty of Information and Communication Technology**

**Mahidol University**

**Semester 1/2022**

Table of Contents

[I. Data Selection 3](#_Toc115437355)

[A. Purposes of data 3](#_Toc115437356)

[B. Data Model 3](#_Toc115437357)

[C. Motivation 3](#_Toc115437358)

[II. NoSQL Technology Selection 4](#_Toc115437359)

[A. Main Features 4](#_Toc115437360)

[B. Tools and Services 4](#_Toc115437361)

[III. NoSQL Implementation 7](#_Toc115437362)

[A. Database Creation 7](#_Toc115437363)

[C. Testing Data Queries 8](#_Toc115437364)

[D. NoSQL Service Information 10](#_Toc115437365)

[IV. References 11](#_Toc115437366)

# Data Selection

## Purposes of data

The purpose that we select this data is we see the insight of tv shows and movies on Netflix whether type, genres, and rating.

## Data Model

* + 1. **Show\_id** – This column contains the id of the show on Netflix (Format: stxxx).
    2. **Type –** This column contains the type of show on Netflix.
    3. **Title –** This column contains the title name of the show on Netflix.
    4. **Director –** This column contains the name of the director of each show on Netflix.
    5. **Cast –** This column contains the name of the actor or actress of each show on Netflix.
    6. **Country –** This column contains the country that produces the show which is published on Netflix.
    7. **Date\_added –** This column contains day, month, and year of the show that was published on Netflix.
    8. **Realease\_year –** This column contains the year of the show that was published on Netflix.
    9. **Rating –** This column contains the type of movie rating of each show on Netflix.
    10. **Duration –** This column contains the time duration of the show on Netflix.
    11. **Genres –** This column contains the genres of the show on Netflix.
    12. **Description –** This column contains the description of the show on Netflix.

## Motivation

We do it for people who like movies and want to see information about movies of interest such as who is the director or where the movie is made and can show users the type of movie they like or are interested in.

# NoSQL Technology Selection

## Main Features

The main feature used in this assignment is MongoDB technology. MongoDB is an open-source NoSQL Database that provides a document-oriented database regarding JSON-like documents. [1] It can perform data without any data structures to query faster than RDBMS. Additionally, MongoDB is appropriate for CAP theorem-based hierarchical data storage (Consistency, Availability, and Partition tolerance) and also has main features including [2]

* **Schema-less Database** – A collection with the ability to store various document kinds is provided by a schema-less database. Users may quickly insert a variety of data formats, regardless of the size, content, or number of fields. Unlike the RDBMS, a document cannot be comparable to another document.
* **Scalability** – With the use of the Shard key, a vast amount of data can be divided into manageable sections because MongoDB can distribute data across various servers. The distribution of those data units across Shards, which are spread across various Physical Servers, is also uniform.
* **Replication** – By making multiple copies of the data and delivering each copy to a different server, MongoDB offers high availability and redundancy. If one server fails, users may still access the data from the other servers.
* **Indexing** – There is less time required to get and search the data in each document because every item within documents is indexed in both primary and secondary indices.

## Tools and Services

There are several tools, applications, and services that could be used to develop the NoSQL script. However, we came up with three technologies that are important and famous technologies providing built-in platforms, online document databases, and cloud services. The three technologies are included as follows:

* **NoSQL Booster for MongoDB**

[1] NoSQL Booster is a cross-platform utility for MongoDB that has a graphical user interface (GUI) and numerous built-in. [3] NoSQL Booster is pleased with its capacity to automatically complete collection names, techniques, keywords, and properties. Additionally, it offers a number of capabilities and could construct a script query more quickly including:

* + **Chaining Fluent Query Interface** – NoSQL Booster provides a fluent query builder API that allows query construction using chaining syntax.
  + **Re-Schema Tool** – The MongoDB Collection Schema can be updated using the GUI provided by the re-schema tool. Users who update by making a few clicks are essentially creating a code.
  + **One Click Grouping and Filtering** – The query results can be grouped using the One-Click grouping tool to compute statistics for the selected field.
  + **Test Data Generator** - Test Data Generator can produce arbitrary phony data for testing needs. False data can imitate real data because the owner of the data cannot always control its distribution and usage. Users could thereby guarantee statistical significance without any limitations.
* **MongoDB Atlas**

With MongoDB Atlas, a managed service cloud database, they can easily install, maintain, and repair their deployments on the cloud-based service provider of their choosing (AWS, Azure, and GCP). Employing MongoDB Atlas is the most effective way to set up, execute, and scale MongoDB on the cloud. A MongoDB database may be launched fast and easily using Atlas using a few clicks. [4] The MongoDB Atlas has a number of functionalities, including: [5]

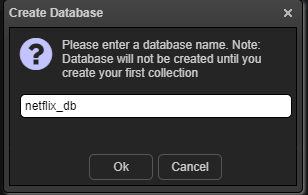
* + **Built-in Automation Features** – Users of MongoDB Atlas can improve their competitiveness by using automation tools for database administration.
  + **Easier Scalability –** Applications and systems must respond to unforeseen requirements as well as the company's typical growth rate. A NoSQL database called MongoDB uses a method called Sharding to provide horizontal scalability.
  + **Strong Security Measures –** The approach used by MongoDB Atlas is multilayered. Authentication, authorization, encryption, and other security capabilities are available in MongoDB Atlas. These attributes strengthen security and help stop data leaks.
  + **Extensive Analytics and Monitoring –** Built-in tracking and monitoring features in MongoDB Atlas offer helpful details about the functionality of your database. Additionally, it receives real-time insights into its server clusters and creates customized alerts, enabling users to address problems before they have an impact on performance or the end-user experience.
* **Google Cloud Platform (GCP)**

Google cloud platform or GCP is the public cloud that has various services such as Big Data, IoT, Data storage or etc. It can assist the user in creating a server or application that uses MongoDB and Google Cloud to simplify working with data and relieve the application teams and developers of operational burdens. [6] There are several features in GCP as follows:

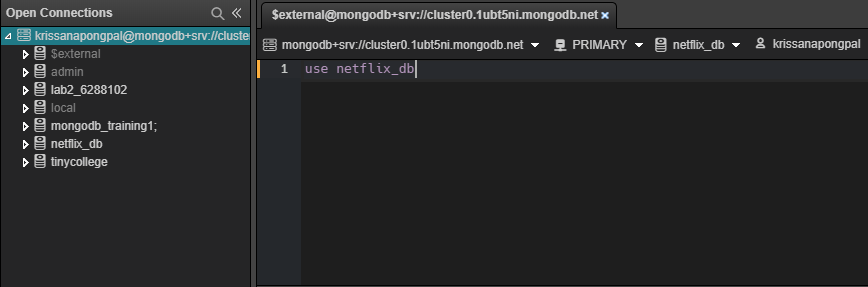
* + **Fault Tolerance and High Availability –** GCP makes sure that applications are not just consistently and conveniently accessible, but also very resilient to difficulties. The availability of GCP's services and products is also intended to assist with application deployment concerns.
  + **Scalability and Elasticity –** The built-in services of Google Cloud allow for automatic application scaling and load distribution. In addition, Users can automatically add or delete instances and set a policy that scales up or down the number of instances based on measured load and target utilization using Managed Instance Groups, a feature of GCP that allows you to manage identical instances as a single entity in a single zone.
  + **Disaster Recovery –** GCP has a robust global network, redundant points of presence throughout the world, a scalable service infrastructure, a highly secure network, and is generally privacy compliant, all of which help it meet an application's SLA needs.

# NoSQL Implementation

## Database Creation

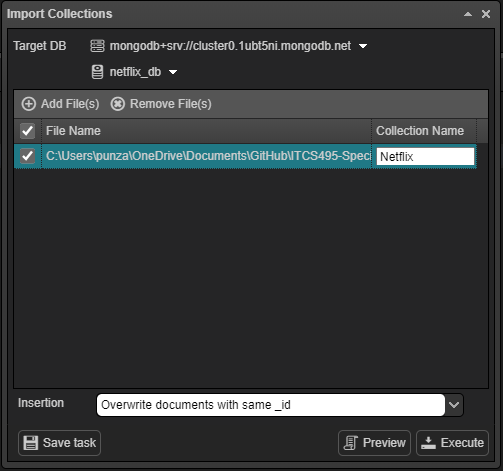
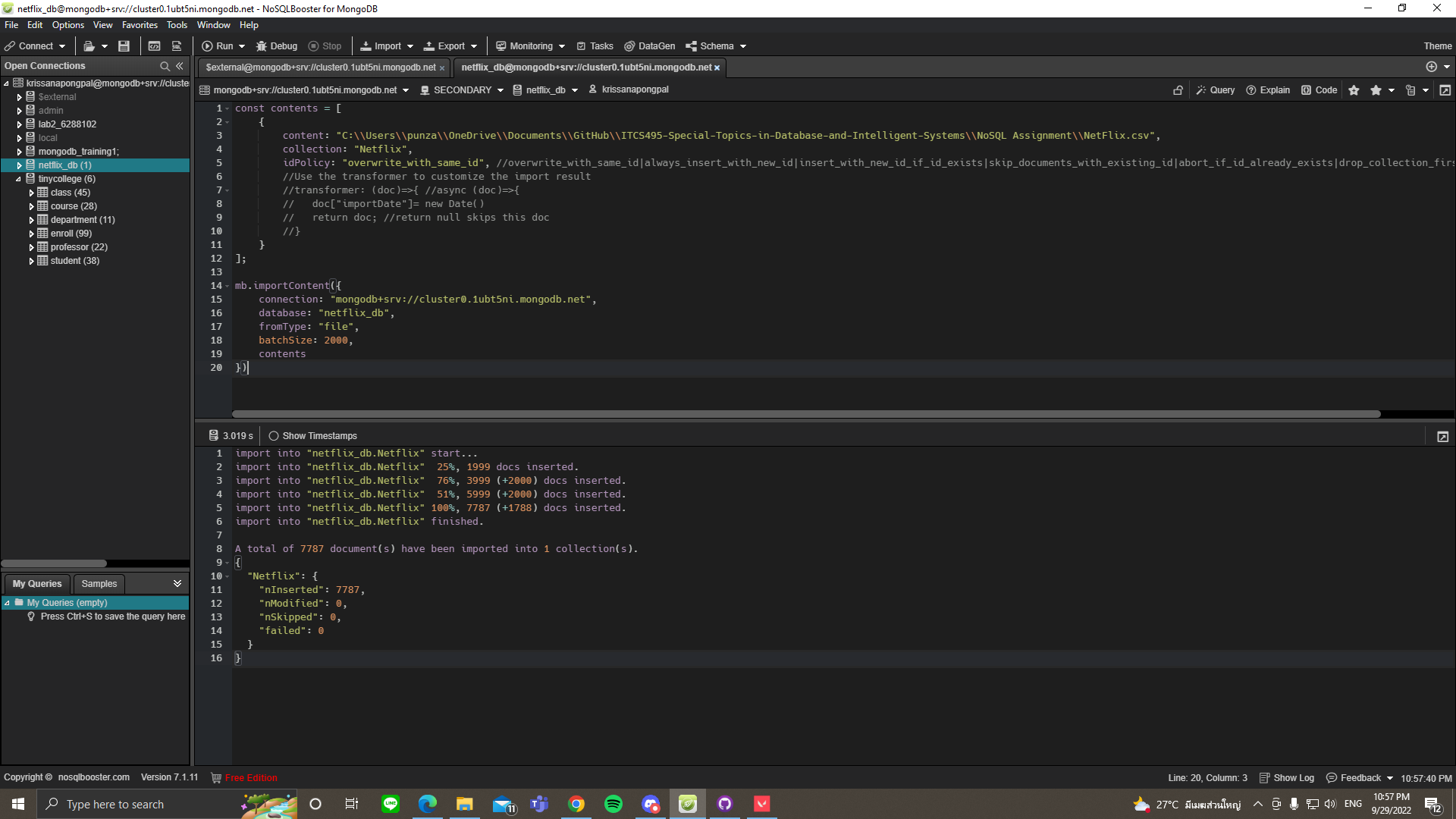
****

Firstly, we create the current database that we are working with. We will set the name of the database as “netflix\_db”. After that, we click the “Ok” button, the MongoDB will create the database.



After MongoDB creates the database, the database will show on the left of the program. Then, we call to use the "netflix\_db" command to use the database we are working with. Finally, we will change the database type to “netflix\_db”.

* 1. **Data Import**

** **

****

After we create the database, we import the collection name “Netflix” to “netflix\_db” as a target DB and select “Overwrite documents with same\_id” for the insertion. Finally, select the execute to show up in the Netflix collection and continuously work with it.

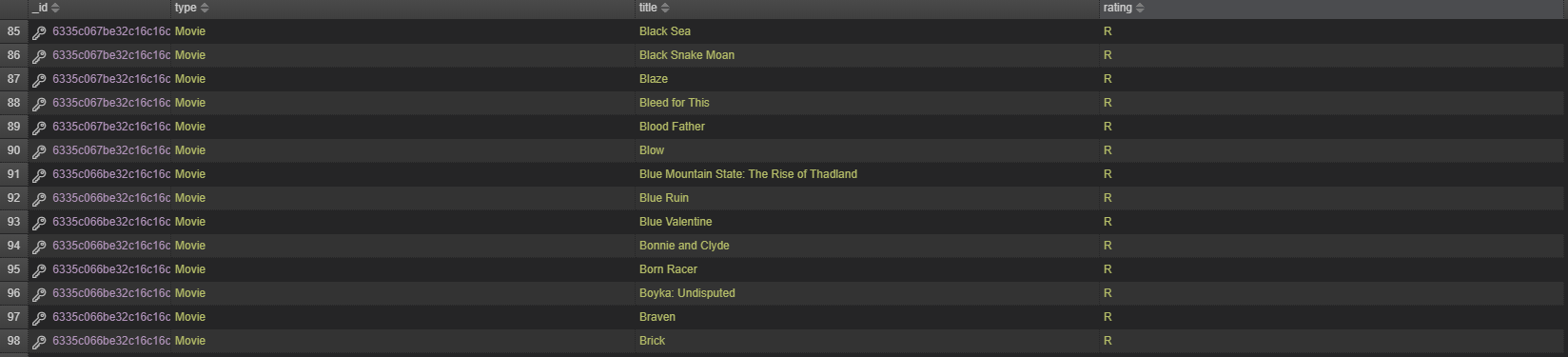
## Testing Data Queries

* **Query the Restricted movies that show title, type, rating, and title show in ascending order**

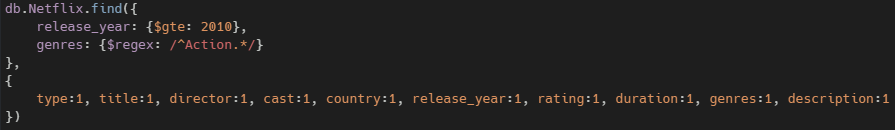


For the first query, we want to find the restricted movies and show the title, type, rating, and title show in ascending order. We write a find command that the rating is ‘R’ and use the projection to specify the title, type, and rating. We add a “sort function” to sort the tile in ascending order.

**Result:**



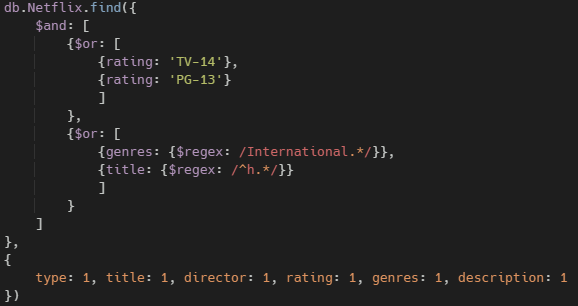
* **Query movies after or equal to 2010 that show type, title, director, cast, country, rating, duration, genres, description, and genres is action.**



For the second query, we want to find the action movies that have the release year after 2010 and show the result as type, title, director, cast, country, rating, duration, genres, and description. We write a find command that release\_year is greater than or equal to 2010 and genres use regex for find pattern matching that we find the pattern is “Action” and use the projection to specify type, title, director, cast, country, rating, duration, genres, description.

**Result: **

* **Query movies’ type, title, director, rating, genres, and description that rating is TV-14 or PG-13, also genre is international both movies and TV shows, or title contains “S” at the first letter**



The third query is created in order to find the Netflix movies and TV shows that rating “TV-14” or “PG-13” representing content for teens. Moreover, the query not only finds a rating but also contains “International” genres or titles containing “H” at the first letter. We start by using “$and” in order to retrieve both ratings and others. In the first code for “$and”, we use “$or” in order to choose ratings for both “TV-14” and “PG-13”. [7] In the second code for “$and”, we also use “$or” in order to choose genres that contain “International” or title that contains “H” at the first letter by using “$regex” for both scripts. In addition, we also conduct the projection that we set it to 1 in order to represent an only type, title, director, rating, genres, and description from queries.

**Result:**



## NoSQL Service Information

We conduct this assignment by using the MongoDB services. However, in order to access the MongoDB databases, users are required to access via NoSQL Booster for MongoDB by following these steps:

1. Open the NoSQL Booster for MongoDB on your computer
2. Open a connection and connect from URI and insert “mongodb+srv://krissanapongpal:za4DwPWLW560O13X@[cluster0.1ubt5ni.mongodb.net/test](http://cluster0.1ubt5ni.mongodb.net/test)”
3. Using a database named “netflix\_db” and collection “Netflix” in order to retrieve a data

# References

|  |  |
| --- | --- |
| [1] | R. Tiwari, "NoSQLBooster For MongoDB Simplified: A Comprehensive Guide 101," 16 February 2022. [Online]. Available: https://hevodata.com/learn/nosqlbooster/#con. [Accessed 29 September 2022]. |
| [2] | A. Saini, "What is MongoDB – Working and Features," Geeksforgeeks, 6 June 2021. [Online]. Available: https://www.geeksforgeeks.org/what-is-mongodb-working-and-features/. [Accessed 29 September 2022]. |
| [3] | Scalegrid, "Which is the Best MongoDB GUI?," Scalegrid, 15 August 2019. [Online]. Available: https://scalegrid.io/blog/which-is-the-best-mongodb-gui/. [Accessed 29 September 2022]. |
| [4] | MongoDB, "MongoDB Atlas Tutorial," MongoDB, [Online]. Available: https://www.mongodb.com/basics/mongodb-atlas-tutorial. [Accessed 29 September 2022]. |
| [5] | Abraham, "Top 10 features of MongoDB Atlas," Fosslinux, 27 January 2022. [Online]. Available: https://www.fosslinux.com/50299/features-of-mongodb-atlas.htm. [Accessed 29 September 2022]. |
| [6] | MongoDB, "MongoDB on Google Cloud," MongoDB, [Online]. Available: https://www.mongodb.com/mongodb-on-google-cloud. [Accessed 29 September 2022]. |
| [7] | Netflix, "Maturity ratings for TV shows and movies on Netflix," Netflix, [Online]. Available: https://help.netflix.com/th/node/2064/us. [Accessed 29 September 2022]. |