

Data225 DB Systems for Analytics

This is a lab group project.

Lab 2

In this assignment you will do something very similar to that of lab 1 but in the context of NoSQL. This lab assignment focuses on the migration of The Movie Database (TMDb) analysis database to MongoDB cluster. The TMDb dataset, containing metadata for 45,000 movies, including user ratings, provides an opportunity to explore various aspects of the film industry. The migration to a mongo cluster aims to leverage the benefits of a cloud-native approach, emphasizing flexibility, scalability, and resilience.

Dataset: You can find the dataset for this lab assignment [here](#). You are required to use these files for your project:

1. credits.csv
2. keywords.csv
3. links.csv
4. movies_metadata.csv
5. ratings_small.csv

You can ignore the links_small.csv and ratings.csv files for this lab.

The following steps can be similar to Lab1 but try to repeat those steps in relation to NoSQL DB. Your task is to create a comprehensive project report detailing the migration process, system requirements, conceptual database design and functional analysis. The project proposal should clearly define the application system that your team is proposing and the results of your requirements analysis and conceptual design. The proposal does not have to be long, three to five typed pages plus supporting diagrams and the like is sufficient. Be concise, but thorough.

Additionally, you are required to upload all relevant code to a private GitHub repository and share access with the ISAs.

Perform necessary pre-processing steps on the TMDb dataset to ensure data cleanliness and suitability for analysis

The project report should include the following sections:

1. Problem Statement:
 - Clearly define the application system your team is proposing, including

limitations.

- Explain why a database is necessary for the system.
- Justify the need for migrating the TMDB Analysis database to the MDB cluster.

2. Solution Requirements:

- Outline the requirements for the solution.
- Describe the system's functionalities and limitations.
- Explain how users will interact with the system.

3. Conceptual Database Design:

- Discuss the database requirements for the application system.
- Document structure for your database.
- Explain the denormalization process and how primary and foreign keys are represented.

4. Functional Analysis:

- Discuss the functional components of the proposed application system.
- Detail the database interactions for each functional component, including:
 - Document structure
 - Access privileges
 - NoSQL codes/ queries (> 10)
 - Logging mechanisms
 - Aggregation (> 5)

5. Database Migration to AWS:

- Provide a step-by-step guide on migrating the TMDB Analysis database to the MDB cluster.
- Demonstrate the connectivity to Mongo cluster using Python.
- Include screenshots and code snippets to validate the successful migration.

Submission:

Submit a comprehensive Project Report, including supporting diagrams and code snippets. Ensure that all relevant code is uploaded to a private GitHub repository, and access is shared with the evaluators.

Grading Criteria:

1. Clearly define project requirements and specifications.

2. Present a well defined denormalization structure and procedure.
3. Specify the document structure, access privileges.
4. Include at least ten NoSQL codes/queries. (> 10) & Aggregation queries (> 5)
5. Demonstrate connectivity to MongoDB in cloud (<https://mlab.com/>
(Links to an external site.) or Mongoddb atlas).
6. Implement logging of database activities.
7. NoSQL performance Measurement and compare with MySQL.
8. Visualization of MongoDB data.
9. Project Report.