MONASH University
Information Technology

GROUP ASSIGNMENT COVER SHEET

Student ID Number	Surname	Given Names
30864941	*Lee*	Andrew

* Please include the names of all other group members.

Unit name and code	FIT3176
Title of assignment	FIT3176 Group Assignment - Sem 2/2023 (Weight: 10%)
Lecturer/tutor	Farah Kabir

Tutorial day and time	Tuesday 5PM	**Campus:** Clayton

Is this an authorised group assignment? ☐ **Yes** ☐ No

Has any part of this assignment been previously submitted as part of another unit/course? ☐ Yes ☐ **No**

Due Date 10th of November 2023	**Date submitted** 10th of November 2023

All work must be submitted by the due date. If an extension of work is granted this must be specified with the signature of the lecturer/tutor.

Extension granted until (date) ......10th of November 2023**..........................** **Signature of lecturer/tutor**
.... Monash Special Considerations Team**.................................**

Please note that it is your responsibility to retain copies of your assessments.

Intentional plagiarism or collusion amounts to cheating under Part 7 of the Monash University (Council) Regulations

Plagiarism: Plagiarism means taking and using another person's ideas or manner of expressing them and passing them off as one's own. For example, by failing to give appropriate acknowledgement. The material used can be from any source (staff, students or the internet, published and unpublished works).

Collusion: Collusion means unauthorised collaboration with another person on assessable written, oral or practical work and includes paying another person to complete all or part of the work.

Where there are reasonable grounds for believing that intentional plagiarism or collusion has occurred, this will be reported to the Associate Dean (Education) or delegate, who may disallow the work concerned by prohibiting assessment or refer the matter to the Faculty Discipline Panel for a hearing.

Student Statement:

- I have read the university's Student Academic Integrity Policy and Procedures.
- I understand the consequences of engaging in plagiarism and collusion as described in Part 7 of the Monash University (Council) Regulations http://adm.monash.edu/legal/legislation/statutes
- I have taken proper care to safeguard this work and made all reasonable efforts to ensure it could not be copied.
- No part of this assignment has been previously submitted as part of another unit/course.
- I acknowledge and agree that the assessor of this assignment may for the purposes of assessment, reproduce the assignment and:
    - i. provide to another member of faculty and any external marker; and/or
    - ii. submit it to a text matching software; and/or
    - iii. submit it to a text matching software which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking.
- I certify that I have not plagiarised the work of others or participated in unauthorised collaboration when preparing this assignment.

Signature ....Andrew Lee.. Date……10th of November 2023………………………………

    * delete (iii) if not applicable

Signature _____Andrew Lee_____Date:_____10/11/2023_____

Contribution Declaration Form
(to be completed by all team members)

Please fill in the form with the contribution from each student towards the assignment.

1 NAME AND CONTRIBUTION DETAILS

Student ID	Student Name	Contribution Percentage
30864941	Andrew Lee	100%

2 DECLARATION

We declare that:
- The information we have supplied in or with this form is complete and correct.
- We understand that the information we have provided in this form will be used for individual assessment of the assignment.

3 SIGNATURE

Signatures

Andrew Lee	

Day   Month   Year

Date   10 / 11 / 2023

Analysing MovieLens 20M dataset with polyglot persistence data model

Group 34

Andrew Lee

Short introduction of the data model name, purpose and motivation

The database that has been selected is called "MovieLens 20M".

What is the data model about?

- Real world Emulation, the dataset depict multicomponent of interactions between user and movie content
- Holistic Data Mapping, systematic mapping with user preferences,ratings,links and tags.

Why is the data model needed?

- It encapsulates a wide array of attributes revolving user engagement and content personalisation.
- To concisely evaluate pros and cons of each selected database technologies in CAPS theory

Why did we select this data model?

- Rich volume of data as it contains over 20 million ratings, which will give valuable insights in user preferences
- Deployment of polyglot persistence model in a realistic scenario

Short introduction of the data sources with accessible source links

Where did you find the data?

- I found the dataset through grouplens, a research group in the department of computer science and engineering at University of Minnesota, they publish open source datasets that is easily accessible to the public at no given cost.
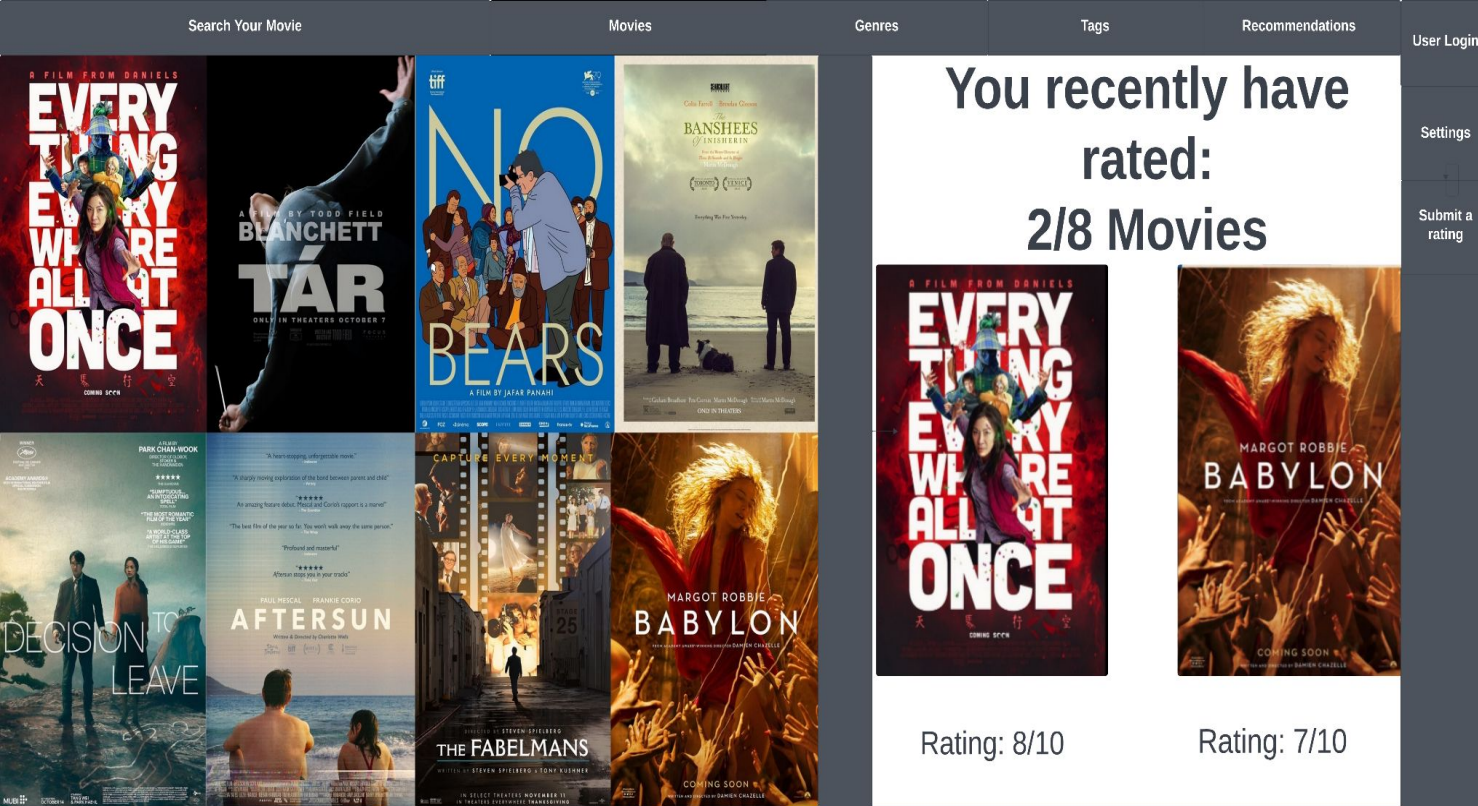
What is the link to your data source?

- https://grouplens.org/datasets/movielens

Is your data source easily accessible (no account login/google drive share etc.)

- The data source does not require any form of logins or google drive, as it is downloadable to the public through the domain website.

Mockup of User Interface

User Login

Settings

Submit a rating

You recently have rated:
2/8 Movies

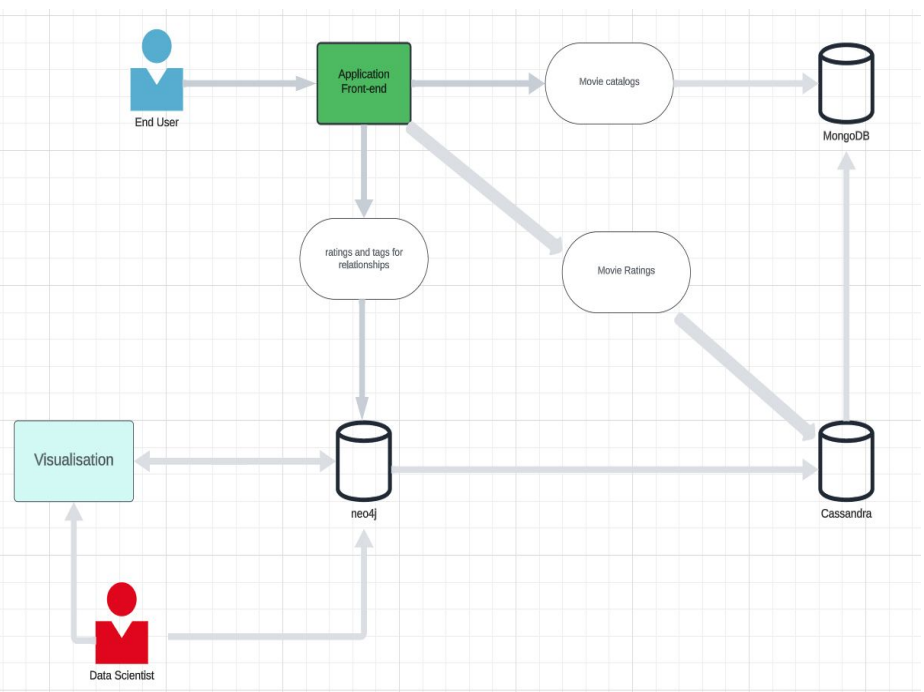Rating: 8/10   Rating: 7/10

Landing page:

- Searchbar

- Movies

- Genres

- Tags

- Recommendations

- User Login

- Settings

- Rating

Design of Polyglot persistence Data Model with Transaction Management

How will you manage transactions?

A high level overview of polyglot deployment:

Legend:

Short justifications on why specific databases were used to store specific data

Neo4j
- Atomicity
- Consistency
- Isolation
- Durability
- Graph Relations
-

Cassandra

- It is basically available
- Eventual consistency
- Soft State

MongoDB
- Atomicity
- Consistency
- Isolation
- Durability

MongoDB
- will be used to store static data

Cassandra
- is used to handle large write operations

Neo4j will
- then be used to create recommendations

User will be
- Recommended Movies based on previous ratings

- Similar movies

- Popular movies

- Expandable to other forms of data analysis

Any references to support your justifications

Data Consistency Models: ACID vs. BASE Explained

https://neo4j.com/blog/acid-vs-base-consistency-models-explained/

Week 12 – NoSQL Transactions (continued) and Polyglot Persistence

https://lms.monash.edu/mod/resource/view.php?id=11900090