Data Modeling and Databases II (DMD2)

Assignment 1. Innopolis University, Spring 2020

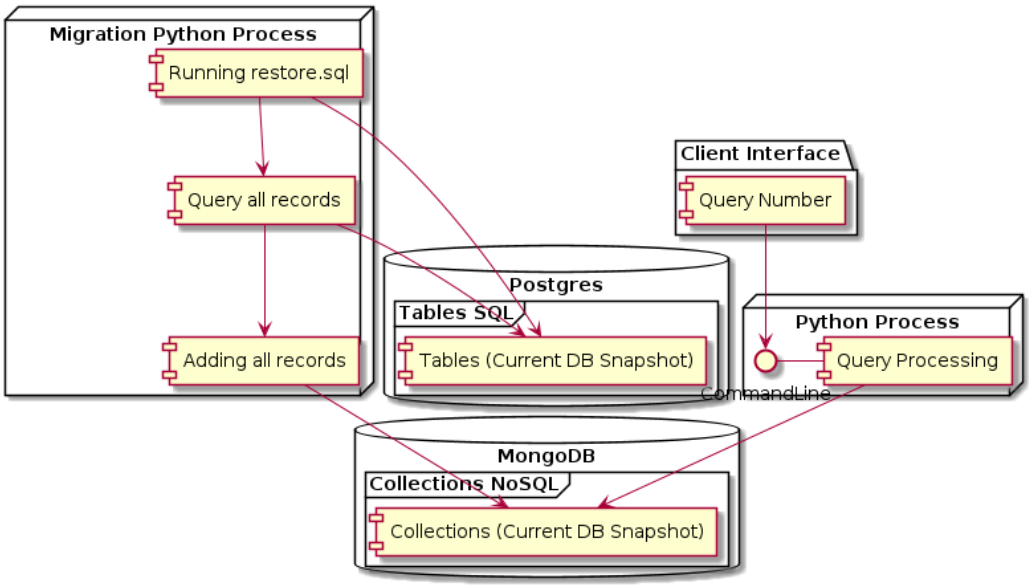
Name: Hany Hamed
Group Number: BS18-06
Professor: Alexey Kanatov
GitHub Repository: here (It is public after the submission Deadline passed)

Note: The report is more than 1 page, due to the resolution of the image of the component diagram and the cover page.

Contents

Part1: Component Diagram

- Migration Process of the DB from Postgres to MongoDB:
  1. Running restore.sql after modifying the paths of .dat files.
  2. Running "migration.py" script to get(Query) from Postgres all the tables' names as collection names, the columns names as the dictionaries' keys and the records as the dictionaries' value and each dictionary is being added to MongoDB as record to a specific collection.



Component Diagram for the architecture including the migration process

Part2: Adjustments

There was no adjustment has been done to the schema of the DB while migration. But, there was some problems in the types that MongoDB supports which requires changing to conform the supported types, such as: datetime (converted to datetime supported by MongoDB), memoryview (converted to Bytes then String), Decimal (converted to Decimal128 supported by MongoDB).

Part3: Performance

It has been noticed that with less data to be used using project, the queries will be faster as we include only the data that is required to be queried for the client.

1) It was fast. It takes in average = 0.095s
2) It takes in average = 12s, as it get all the different combinations of a tuple of 2 actors and the number of times they co-acted. Using aggregate queries speed up the queries through the project and group.
3) It takes in average = 34.5s, as they count all the times that the film with their categories is rented.
4) It takes in average = 11.3s to fully recommend 10 films according to the history of the customer in renting the films based on the best 5 categories (Frequent 5 categories in the films) that the customer rents
5) It takes in average = 13s, as it execute BFS on the results to get the separation degrees.