

Homework 5

IE 7374 Data Management for Analytics

Readings:

Chapter 11, NoSQL Databases (study the review questions and scenario conclusion)

Problem 1. Problems and exercises, 11.3E

Problem 2. Problems and exercises, 11.5E

Problem 3. Based on the MongoDB “Restaurants database” (on “people” collection) in the online playground:

- Write a query using a map-reduce pipeline to get the **average rating for each restaurant_id**.
- Write a query to get the **average rating for each restaurant_id** without using a map-reduce pipeline.
- Write a query using a map-reduce pipeline to get the **highest (maximum) rating for each restaurant_id**.
- Write a query to get the **highest (maximum) rating for each restaurant_id** without using a map-reduce pipeline.

Problem 4. Based on the MongoDB “Restaurants database” (on “restaurant” collection) in the online playground:

Notice that **two duplicate data regard as two data**.

As an example, for location ‘Fallowfield’, you may count 2 for below two identical data.

```
{ "_id" : 3, "address" : "317 Wilmslow Road", "location" : "Fallowfield", "name" : "23rd Street"  
{ "_id" : 4, "address" : "317 Wilmslow Road", "location" : "Fallowfield", "name" : "23rd Street"
```

- Write a query using a map-reduce pipeline to get the **count of restaurants for each location**.
- Write a query to get the **average rating for each location** without using a map-reduce pipeline.
- Write a query to get the **type of food with highest average rating** without using a map-reduce pipeline.

Problem 5. Based on the Cypher “Beer database” in the online playground

- Write a Cypher query to get a list of beer types that Seppe liked
- Write a Cypher query to get a list of beer types that both Seppe and Bart liked

- c. Write a Cypher query to get the name, year, type of the beers that were brewed by Brouwerij Lupus and Seppe liked
- d. Write a Cypher query to get a list of people's names that liked the beers Seppe liked

Problem 6. Based on the Cypher “Book club database” in the online playground

- a. Find a count of all Bart Baesens's friends, but **only if at least one FRIENDS_OF relation exists**. Output his name and count of friends as NumberOfFriends.
- b. Find a list of books that Bart Baesens likes but Seppe vanden Broucke doesn't.
- c. Find all friends that have more than 1 book in common?
- d. Return the genres the friends of Seppe vanden Broucke like together with how many times they liked a book of that genre. Sort by how many times decreasing.

Problem 7. Based on the MongoDB “Restaurant” database, answer the following questions:

- a. On the “people” collection, display all rate that Seppe made.
- b. On the ‘people’ collection, construct a query to determine the average rating given by "Wilfried" across all his reviews.
- c. On the ‘restaurants’ collection, write a MongoDB query to find all restaurants with a rating of at least 5 and serving either "Pizza" or "Curry". Display only the name, location, and type_of_food.
- d. On the ‘restaurants’ collection, write a query that lists all distinct types of food along with the number of restaurants offering that type of food. Sort the results by the number of restaurants in descending order.
- e. On the ‘restaurants’ collection, write a MongoDB aggregation query that calculates the average rating for each type of food (type_of_food). Sort the results by average rating in descending order.

Problem 8. Based on the MongoDB “**wine**” database, answer the following questions:

- a. On the “**products**” collection, display all products that have are of type either "rose" or "sparkling"
- b. On the “**products**”, write a query to count the products in wine database that have available quantity between 50 and 90 (both inclusive)

- c. On the "**supplies**" collection, write a query to determine the total count of supplies each of the suppliers has made for which the delivery_period is not null. Also, display the result
- d. Construct a MongoDB query on the "**products**" collection to accomplish the following tasks:
 - 1. Retrieve all products whose name does not contain the term 'Chateau'.
 - 2. Group these products by their wine type.
 - 3. Calculate the total sum of available quantities for each wine type, ignoring any null values