DATABASE DESIGN II – 1DL400 – 2022

Assignment 3 (Lab 3)
NoSQL

Examination
This assignment aims to familiarize you with NoSQL and the differences of NoSQL with traditional Relational Databases. In order to complete this assignment, you may need to do research beyond the material given in the class or provided on Studium. Proceeding with your own research is a skill you also need to develop, since computer science is based on fast evolving technologies.

The assignment must be submitted before the indicated deadline at **Studium**. Please notice that you are not expected to finish this assignment during the lab: it may take longer. You must submit the result of this assignment on **Studium** as one single PDF file per group. Please consider that all the submitted assignments will be printed and corrected on A4 paper, so indicate group name, group participants with ID (personnummer, if you have one).

In this assignment, you need to do some tasks with MongoDB. So, we first explain how to install and use MongoDB. Then we give two sets of tasks you need to do.

Instructions for running MongoDB:
The official documentation on the MongoDB website is quite detailed. You can follow the appropriate links for your operating system. When prompted to select download, **select MongoDB Version 4.2.23** for your operating system.

Windows:
https://docs.mongodb.com/manual/tutorial/install-mongodb-on-windows/
Follow the instructions to install MongoDB **not** as a service

MacOS:
https://docs.mongodb.com/manual/tutorial/install-mongodb-on-os-x/

Linux:
https://docs.mongodb.com/manual/administration/install-on-linux/
Select the appropriate distribution (and later version) of your OS

All of these instructions conclude with you having access to a client terminal on which you can access your MongoDB database. Hereon we will refer to this as the client terminal.

Question 1. In this exercise, you will have to work with a small MongoDB database and apply on it search queries. Also, you will be asked to compare it with relational databases.

Instructions how to generate the database: Download the file "mdb_statements.txt" from the assignment folder from Studium, then copy and paste the content of the file on the MongoDB client terminal. These statements will generate a database and a collection and insert in it three documents.

Complete the following tasks:
a. Give search queries that can do the following
   I.   Find a restaurant in Manhattan
   II. Find a restaurant with Italian cuisine
   III. Find documents with zip code "10075"!
   IV. Find a restaurant with either Italian cuisine or having zip code "11225"

You can find more information about *find()* statements in the https://www.mongodb.com/basics/get-started .

b. Propose a relational database schema for this collection of documents, i.e. propose the set of relations and their attributes according to the relational model. The aim of this exercise is to help you to understand the difference with respect to data modelling of NoSQL and relational databases.

c. What advantages (efficiency, data modelling, etc.) with regards to data modelling can you observe with the NOSQL data model over the relational data model?

Question 2. In this question, we will investigate the benefit of using indexes on a database and see how speed is improved. For this purpose, we will give you a large file on which you need to experiment.

Instructions how to generate the database:
- Download the file "items.json" provided on studium and place it in an accessible directory (e.g. "C:\Users**<username>**\Desktop\items.json" for Windows).
- In a new terminal window (separate from the client terminal), insert the following command to import the file (modify **the path to reflect the directory where you stored "items.json"**). <u>NOTE:</u> You should find the following command in plain text by downloading the "q2_statements.txt" file from Studium. There have been problems observed when copying from a pdf file.
mongoimport --db testDB --collection items --type json --file "Replace\This\With\The\Path\to\items.json"
(If mongoimport is not in the PATH, you have to write the full path in order to be able to execute the command.)
- Execute "use testDB" to continue working on the modified database.

Complete the following tasks:
a. Give a search query that can find items with "price" between 2500 and 2550.
b. Repeat your query followed by ".explain("executionStats")". This statement will give you details about the execution of the statement such as time required to run this statement. Note (1) the value of "executionTimeMillis", (2) the amount of results returned, and (3) the value of "totalDocsExamined".
c. Then, create an Index on the attribute "price" (you can find information about indexes at https://docs.mongodb.com/manual/).
d. Repeat task b. Notice the difference in performance. Examine the value of "totalDocsExamined" instead of the "ExecutionTime" since variations on your computer's workload and other factors will heavily influence execution time.
e. Try to modify the size of the price range and compare the performance difference. Explain why the price selectivity affects it. (If you need to delete an index, read about the function "db.collection.dropIndex()")