# Assignment 3 – MongoDB

**Group**: 23

**Students**: Dries Costermans, Tibo Van den Eede, Marco Zoso

**Introduction**

In this assignment, we created collections, cleaned and inserted data, and fetched the data to answer some questions. We set up the MongoDB database which required cleaning and inserting the data into defined tables. Next, we wrote queries to the database to gain knowledge of the dataset. Tibo did the data insertion and query 9. Marco and Dries did the other queries.

**Discussion**

We tried to follow the instructions as it was explained on the assignment sheet. We learned to work with a NoSQL database and the difference between working with a MySQL database.

**MySQL vs MongoDB**

The main difference between those two types of databases is that relational has a more rigid and well-defined structure than non-relational. This has obviously an effect on the corresponding query languages; indeed SQL is better for smaller amounts of data, while MongoDB allows the user to handle bigger databases.

Overall, we preferred to use SQL, but this opinion is a bit biased since we all had good knowledge of it and, for this reason, we found the queries way easier in this language. At the same time, we think that the challenge established by the MongoDB assignment allowed us to learn how to use this query language properly; a desirable skill in the era of big data.

# Results of queries:

1. How many users, activities and trackpoints are there in the dataset (after it is inserted into the database).

Afbeelding met tekst

Automatisch gegenereerde beschrijving

2. Find the average number of activities per user.



3. Find the top 20 users with the highest number of activities.

Afbeelding met tafel

Automatisch gegenereerde beschrijving

4. Find all users who have taken a taxi.

Afbeelding met tekst

Automatisch gegenereerde beschrijving

5. Find all types of transportation modes and count how many activities that are tagged with these transportation mode labels. Do not count the rows where the mode is null.

Afbeelding met tekst

Automatisch gegenereerde beschrijving

6. a) Find the year with the most activities.



b) Is this also the year with the most recorded hours?



Just like we found out in the previous assignment, 2008, which is the year with the most activities, is not the year with the most recorded hours because 2009 has more. With regard to query 6b, it is worth specifying that the difference between data times in MongoDB returns the result in milliseconds; for this reason, we had to convert it into hours by dividing it by 3600000.

7. Find the total distance (in km) walked in 2008, by user with id=112.



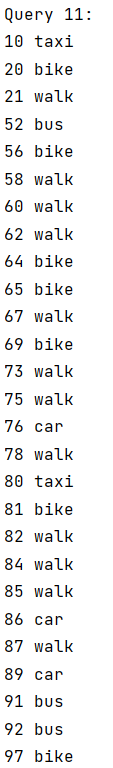
8. Find the top 20 users who have gained the most altitude meters.

9. Find all users who have invalid activities, and the number of invalid activities per user

10. Find the users who have tracked an activity in the Forbidden City of Beijing.

11. Find all users who have registered transportation\_mode and their most used transportation\_mode.

Afbeelding met tekst

Automatisch gegenereerde beschrijving

**Feedback**

Last assignment we didn’t see a big overlap with the course topic, this assignment was better in our opinion because it was more related to the course topic. We understand it is useful to make the switch from a SQL database to a NoSQL database and the consequences of this.