Assignment 4

Note: This assignment should be **solved in groups** with up to 3 people. Please upload only one solution of a group to Moodle but check the crosses in the campus system individually. make sure to name all group members in your submission by including a file group.txt containing the names of all group members.

Download the data set aau.zip from the moodle course. The archive contains data about courses and curricula as well as result lists of exams.

Exercise 1 – Create A DWH

Reason about the files contained in the file aau.zip. Create a DWH schema with grades as facts, including the following dimensions:

Lecturer: {Name, Rank (Univ Ass, Postdoc-Ass, Prof, Ass Prof, …), Title (DI, DR,…) } → Department → University
Course: {Course, Type (VO, VC, UE,…), ECTS, Level} → Department* → UniversityName
Time**: Day → Month → Semester → Year
Student: Name
StudyPlan: {StudyplanTitle, Degree (Bachelor/Master), Branch (Technical Studies/Economics)}

Note: The arrow indicates hierarchy levels. The sets contain multiple attributes at the same hierarch level.

a) Design a suitable DWH schema in form of relational tables.
b) Load the data from the JSON files into your schema.

*A course is assigned to the budget of some department; this is not necessarily the department where the lecturer is assigned.
*** The level of a course can differ from the level of a student. Students may already take some master's courses during their bachelor program.
** For the time dimension: Only fill the dimension table with the dates of actual exams.

Hints:

- for a) It's up to you to implement either a star or a snowflake schema.

- for b) You may load the files in postgres and then write queries to transform and load the data into your schema. Alternatively, you may first upload the files to mongo-db, transform the result with queries and then load the result into postgres. Other solutions like JS/python/java/MapForce are also welcome. Also online tools might help: https://sqlify.io/convert/json/to/postgres

- In case of data quality problems, you may need to fix the wrong values and document the changes and assumptions. Please document!

- For the sake of simplicity, you do not need to create surrogate keys to replace existing keys.

Upload: Please upload a complete script for the data import. If manual steps are required, document them / provide the changed files.

Exercise 2 – DWH Querying

The data should now be shown as a pivot table with the dimensions of student, StudyPlan, and lecturer showing the average grades.

Write a query in SQL returning all required data to fill the pivot table and supporting OLAP operations on the pivot table without issuing an additional query.

Hint: Use a suitable SQL99 group by operator and show how the groupings relate to the pivot table.

It is sufficient to only provide the required data in form of an SQL result. You may sketch how the SQL result set relates to the cells of a pivot table.

Exercise 3 – Data Integration (see Video Lecture in Moodle)

a) Briefly explain the three main phases of Data Integration
b) Why are Precision and Recall insufficient for assessing matching systems in isolation?
c) Provide an example for a global as view mapping and a local as view mapping for the following schema:

   Global Schema (**PK**, *FK*):
   Person(**id**, givenName, lastName, job, age)

   Local Schema (**PK**, *FK*):
   MyPerson(**svnr**, givenName, lastName, age)
   MyPobs (***svnr***,jobtitle)

d) Discuss the benefits and drawbacks of local as view vs global as view regarding rewriting complexity, change of data sources, and constraints over data sources.