

Advanced Databases & noSQL (INFDEV026B)

Assignment 2

Context

A multinational company wants to update its information system by changing the database storing information about employees and their professional positions in the company itself. The company will store each employee in the database, who is defined by his bsn, name, and surname. The company might need to record multiple addresses for an employee, since it might require that he/she moves temporarily to a new address if he/she is working on a project abroad. Each address is identified by the country, the city, the street, the number, and a postal code. You can assume that the combination of street number and postal code is unique. The company needs to know at any moment which address is the residence of the employee. Each employee in the company can work in different positions at the same time. For example, John Smith might work as a Project leader and as a Software Developer. Each employee works a different amount of hours for each position. A position is identified by a name (for example Project leader), a description, and an hour fee (the salary per hour). For each employee the company stores his/her degrees, which are identified by a course, a school, and a level (bachelor, master of science, PhD, etc.). Note that the same person might have, for example, a bachelor in Computer Science and a master of science in Automation Engineering. Each position is connected to a project. For example, an employee can be Project Leader for the project Cryptographic Software Development for CIA and Software Developer for Physics Engines for computer games. Note that some positions are not necessarily connected to a project, for example an Employee manager participates in no project. A project is recorded with a unique id, a budget, and the total amount of allocated hours. The company needs to store the headquarters it owns. Each headquarters is identified by a unique building name, the number of rooms (offices, conference rooms, etc.), and the monthly rent. For each employee, the company records the main headquarters he/she works for. Projects are also located into headquarters, which might be different from the main headquarters of an employee working in that project. Each headquarters has an address as well, in the same format as the one of an employee.

Task description

Create an application (in Java or C#) that inserts random data inside the relational database created in the previous assignment. Strings defining names should be picked randomly inside a specified lists of names. Furthermore, the

numerical data should have likely values, for example the age of an employee cannot be 1500, or the our fee 100000. Insert at least 10000 employees, and 100 projects. Your application should use a fixed random seed which can be set to generate the random data, in order to able to debug the application.

Implement the following queries using the map-reduce paradigm in either C# or Java, and output the data on either a GUI or the shell. Re-implement the function Map and Reduce from scratch, do not use LINQ or other pre-built libraries.

1. An employee is underworking if he is working less than 5 hours on his projects, he is working normally if he is working between 5 and 20 hours, and he is overworking if he is working more than 20 hours. Output the number of employees that are overworking per project.
2. Find the total working hours of all employees and average working hours per employee.
3. Find the total fee of an employee. This number is obtained by multiplying the working hours on all the projects by the hour fee of his position.

You are allowed to use only the map-reduce functions seen in class.

Hand-in

The implementation of the assignment should include:

- The code to generate the database.
- The source code of the program to populate the database.