FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO
Mestrado Integrado em Engenharia Informática e Computação

Database Technology

Assignment nr. 3

Version 2 – Teaching service distribution

NoSQL Assignment

Goals

Build a small DB using two distinct NoSQL approaches, a document database (MongoDB) and a graph database (Neo4j) and compare both approaches and with the relational approach.

Work groups

The assignment should be executed by a team of one or two elements.

Subject

The situation is about recording the teaching service of professors giving classes in courses. Its description goes along the relational model of a database that you will use as source to populate the NoSQL models you will design. The tables of the relational database are available in the user GTD10 in the Oracle server (connection: BD, user: <user>, password: <pass>, host: oraalu.fe.up.pt, port: 1521, SID: ALU).
There are courses (table XUCS), described by a code (codigo), a designation (designacao), an acronym (sigla_uc) and a program (curso).
Courses have occurrences in several years. Each occurrence is recorded by a row in the table XOCORRENCIAS, with information on the course code (codigo), academic year (ano_letivo), period of classes (periodo, that may be A-annual, 1S- first semester, 1T- first trimester, etc.), number of enrolled students (inscritos), students with distributed assessment (com_frequencia), number of approved (aprovados), course goals (objetivos) and content (conteudo), and department in charge (departamento).
Each occurrence may have one or more class types (T-theoretic, P-practical, L-laboratory, TP-theoretic/practical, OT- tutorial guidance). Each class type for an occurrence is recorded on table XTIPOSAULA with the number of similar classes (turnos), the number of hours per week for each class (horas_turno), and in some cases the number of weekly classes (n_aulas).
The table XDSD records the teaching service distribution, in each semester, for each professor. More specifically, it records, for each class type of an occurrence, how many weekly hours are assigned to that professor. If a professor is teaching, in a single class, more than one course at the same time, for example from different programs, the weight of that course, in the perspective of the professor, may be less than 1 and recorded in attribute fator. Otherwise, the attribute fator will be 1. From the program perspective, the attribute fator is ignored. The attribute ordem enables listing the set of professors of a specific course occurrence in a specific order.
The professors are recorded in the table XDOCENTES with a number (nr), a name (nome), an acronym (sigla), a category code (categoria), a given name (proprio), a family name (apelido), and a status (estado: A-active, NA-non active, R-retired).

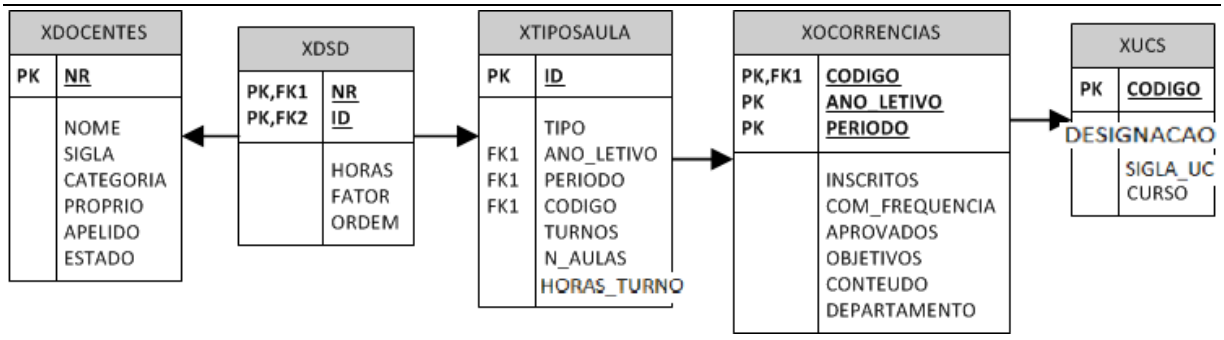

Figure 1 – Relational model for the case Teaching Service.

The tables are in Oracle user GTD10.

TASKS

The following tasks should be performed, described in a report and presented in a specific session:

1) Design a Mongo document model for the Teaching Service example, explaining the decisions made. The information available under this model should be equivalent to the information in the relational database.
2) Migrate the data from the Oracle user GTD10 into the NoSQL database. One possibility is to write a PL/SQL package able to extract the data using an appropriate SQL query and to produce the appropriate Mongo method calls to populate the NoSQL database designed in 1). Other methods may be used. Document the procedure in the report, anyway.
3) Design a Neo4j graph model for the Teaching Service example.
4) Migrate the data from the Oracle user GTD10 into the Neo4j database.
5) Prepare Mongo and Neo4j queries for the following questions:
    a. How many class hours of each type did the program 233 got in year 2004/2005?
    b. Which courses (show the code, total class hours required, total classes assigned) have a difference between total class hours required and the service actually assigned in year 2003/2004?
    c. Who is the professor with more class hours for each type of class, in the academic year 2003/2004? Show the number and name of the professor, the type of class and the total of class hours times the factor.
    d. Which is the average number of hours by professor by year in each category, in the years between 2001/2002 and 2004/2005?
    e. Which is the total hours per week, on each semester, that a hypothetical student enrolled in every course of a single curricular year from each program would get.
    f. Ask the database a query you think is interesting.
6) Compare the Mongo, Neo4j and Oracle implementations from the viewpoints of data size, processing time, and query easiness.