



Computer Sciences Department
CS 3810 - Principles of Database Systems – Spring 2021

Database Project 02

Deadline: March 28th, 11:59pm

Overview

In this assignment you are given a dataset of job positions and required skills for jobs in the IT industry. Your task is to create a database from the given dataset and use the database to answer a few queries using SQL. The dataset is named `job_skills.csv` and is available for download at https://github.com/thyagomota/21SCS3810/tree/main/db02_skills. Below is one line of `job_skills.csv` to illustrate the file's structure.

```
355:database administrator - data modeling,147:sql;175:windows  
server;1438:performance tuning;1706:data  
modeling;1918:rdbms;2747:dba;3014:dbms;3847:performance  
monitoring
```

The line above shows the skills for a "database administrator - data modeling" position (id = 355). This position identifies 8 skills:

- sql (id = 147),
- windows server (id = 175),
- performance tuning (id = 1438),
- data modeling (id = 1706),
- rdbms (id = 1918),
- dba (id = 2747),
- dbms (id: 3014),
- performance monitoring (id = 3847).

The number of skills per job position varies in the dataset. You can assume that the job position ids are unique (within other job positions) and that the skill ids are unique (within other skills).

Data Model

You are expected to come up with a database design (from the given dataset) that is normalized up to the 3NF. Your database design should be presented using an SQL script file based on the `job_skills.sql` template (also available at https://github.com/thyagomota/21SCS3810/tree/main/db02_skills). There should be 2 database users:



Computer Sciences Department
CS 3810 - Principles of Database Systems – Spring 2021

- `job_skills` (with only `SELECT` capabilities on all tables) and
- `job_skills_admin` (with full capabilities on all tables).

Data Load

You should implement a data load program in Java that is able to read the given dataset and populate all of the tables of your database. The data load program should use the `job_skills_admin` user and all of the connection parameters (server, database, user, and password) should be obtained from a `config.properties` file. The data load program should be named `JobSkills.java`. A template for this file is also available at https://github.com/thyagomota/21SCS3810/tree/main/db02_skills.

Your `config.properties` file should follow the exact format below. You should not upload this file on Canvas.

```
server=<server localhost>
database=skills
user=job_skills_admin
password=<password>
```

Queries

Once the database is created and all of its tables are populated, you should write SQL statements to answer the following queries:

- a) what is the total number of job positions?
- b) what is the total number of skills?
- c) which job position titles have the word "database"?
- d) provide an alphabetical list of all job position titles that require "sql" or "mysql" as a skill.
- e) which skills "database analyst"-like positions have that "database admin"-like positions don't?
- f) list the top 20 skills required by job positions having the word "database" in their titles.

Your queries should be saved in `job_skills.sql`.

Reflection



Computer Sciences Department

CS 3810 - Principles of Database Systems – Spring 2021

Answer the following question when you submit your project: from the skills table, which ones do you think you have developed in 3810 so far? You can submit your reflection as a Canvas or SQL (file) comment.

Deliverables

- `job_skills.sql`
- `JobSkills.java`

Rubric

+5 database creation

+15 tables creation

+5 user access

+30 data load script

+5 query a)

+5 query b)

+5 query c)

+10 query d)

+10 query e)

+10 query f)

-5 didn't submit a reflection