Overview: Embedding SQL within another programming language is nice for applications that require more complex calculations or manipulations than plain SQL can handle. For Java, the JDBC API is a common mechanism by which programs can exchange data with DBMSes.

Multiple times per year, the Florida Department of Education’s Florida Standards Assessments (FSA) tests students in a subset of subjects and reports a per–school summary of the results in Microsoft Excel .xls files. It’s easy to get a spreadsheet program (such as Excel or LibreOffice) to output a spreadsheet in CSV (Comma Separated Value) format, which is a good format for reading into a program, and from there into a database.

Assignment: For this assignment, you will need to do the following, not necessarily in this order:

1. Get the Excel files for the FSA Geometry EOC State Report of Schools (not Districts!) results for the spring administrations from the years 2017, 2018, 2019, and 2021 from the fldoe.org website (see the Data section, below, for the URL).

2. Using a spreadsheet program, convert their content to four separate CSV files (or another text file format (e.g., TSV) that appeals to you), one per year.

3. “Scrub” the files to make them consistent and suitable for importing into an Oracle database. (See Data for a potential labor–sharing option.)

4. Within your Oracle database, create four tables with an appropriate schema, one table per year, and use the GRANT command to make your tables readable by others; see the “Other Requirements and Hints” section, below.

5. Copy the results data from the files into the relations. (We expect that you will write a small Java program for this, but that’s not a requirement.)

6. Write an application in Java 16, using JDBC, that offers the user a text menu of four questions that can be asked of the database’s content. The questions are:

   (a) How many schools are listed with different school names in 2021 than they were listed in 2017? To be counted, a school must appear in both years.

   (b) Considering only the years 2019 and 2021 (pre– and mid–pandemic), and only the schools for which level percentage data is available for both years, what are the district names and school names of the schools that reported both decreases in percentages in both Levels 4 & 5 and increases in percentages in both Levels 1 & 2 in 2021 as compared to 2019? For example, if a school’s Levels 1 through 5 percentages in 2019 were 31, 22, 23, 14, and 10, and in 2021 were 33, 23, 25, 11, and 8, that school would appear in the results, because both Levels 4 and 5 decreased (14 < 11 & 10 > 8), and both Levels 1 and 2 increased (31 < 33 & 22 < 23).

   (c) For a year and a case–insensitive district name given by the user, if the district has at least 10 schools in the report, list in descending order the five schools with the highest combined level 4 and 5 percentages. Display those schools’ numbers, names, the two individual percentages, and their sum. If the district has less than 10 schools, print “I’m sorry, but district [name] has just [#] schools.” Replace [name] and [#] with the given district name and quantity of schools.

   (d) A question of your choice, subject to these restrictions: The question must use two or more of the four relations and must be constructed using at least one piece of information gathered from the user. Try imagining yourself as a parent of a student and think about what you might like to know.

(Continued...)
**Data:** The Excel files with the Florida assessment data for the years 2017 through 2021 (excepting 2020) can be downloaded starting from this FL DOE web page:

https://www.fldoe.org/accountability/assessments/k-12-student-assessment/results/

You will need to perform some ‘scrubbing’ of the data. For example, we are interested in only the school data, not the metadata at the tops of the spreadsheets, so don’t try to store that. Another example: Many schools are missing all of the data except the numbers of students. Mark missing data with NULLs. You are likely to think of additional needed adjustments as you explore the data. These sorts of inconsistencies can be eliminated with a combination of creative (or brute-force!) editor skills, small data cleaning programs of your own creation, and/or basic SQL statements. It’s always tedious, but rarely difficult. (And see the Other Requirements and Hints section, below, too.)

**Output:** Your application is to display the output of a question in a clear, easy-to-read format. You’re writing your own program, so you are not restricted to SQL’s output format.

**Hand In:** You are required to submit a .tar file of your well-documented application program file(s) — including any code written to automate the data scrubbing and/or loading processes, although that code may be minimally documented — via turnin to the folder cs460p3. Name your main application program’s main() class **Prog3**, so that we don’t have to guess which file to compile, but split up your code over additional files as appropriate for good code modularity.

**Want to Learn More?**

- The Florida Statewide Assessments Portal: https://fsassessments.org/

**Other Requirements and Hints:**

- Because we will be grading your program on lectura using Oracle, it needs to run on lectura and use Oracle.
- If you wish to share any necessary data conversion and “scrubbing” chores with a few of your classmates, that’s fine. Stop collaborating when you start coding your JDBC application. In your documentation, be sure to credit those who helped you with the data organization. DO NOT share scripts and the like on Piazza or elsewhere; we don’t want one helpful person doing all of the dirty work for the entire class!
- It is OK to share query results on Piazza. Doing so can help you discover, for example, that your query isn’t finding everything it should be finding.
- Make certain that your database tables are accessible to us (by GRANTing us SELECT privileges) and that your relations are prefixed with “yourNetID.” in your applications’s queries so that we can execute your program against your database, just as my tables for Homework #3 were accessible to you with the “mccann.” prefix. The form of the GRANT command is:

  GRANT SELECT ON tablename TO PUBLIC;

  If you need to delete and re-create your tables, you’ll need to re-issue the GRANT commands on the new tables.
- Remember to test your queries on relatively small sample tables before trying them on the full tables. Finding errors in small amounts of data is much easier than in large amounts.