Fall 2022
Due: Friday Nov. 4, 2011 at 11:59 p.m.

This assignment is in 4 parts.

Part 1: Convert the ER Model below into a relational database.

Part 2: Create a Postgres database schema containing tables and attributes created in Part 1. It must be emphasized it is part of the project that the tables be part of a Postgres *schema*. Make sure to enforce primary key and foreign key constraints if they exist. Place your commands for creating the schema into a script file called "create" which should be e-mailed according to the instructions below. The pg_dump command could be very useful here. The script will be graded for functionality by sending it as input to psql, i.e. `psql < create`.

Part 3: Write a program to read some data from some data files and load the data into the Postgres database that you created in Part 2. The datafile format is described below. The file name to be read should be obtained from the user through command line arguments. The source code file should be named "load" with the appropriate suffix ("load.c", "load.cc", "load.java" or "load.py"). See below for more instructions concerning programming language and libraries.

Part 4: Write a program which will query the user for a troop number. The program should query the database and print out the total number of boxes (and dollar amount) sold by the troop for each cookie type they sell. This program is described more fully below. The name of the executable for this program should be "query".

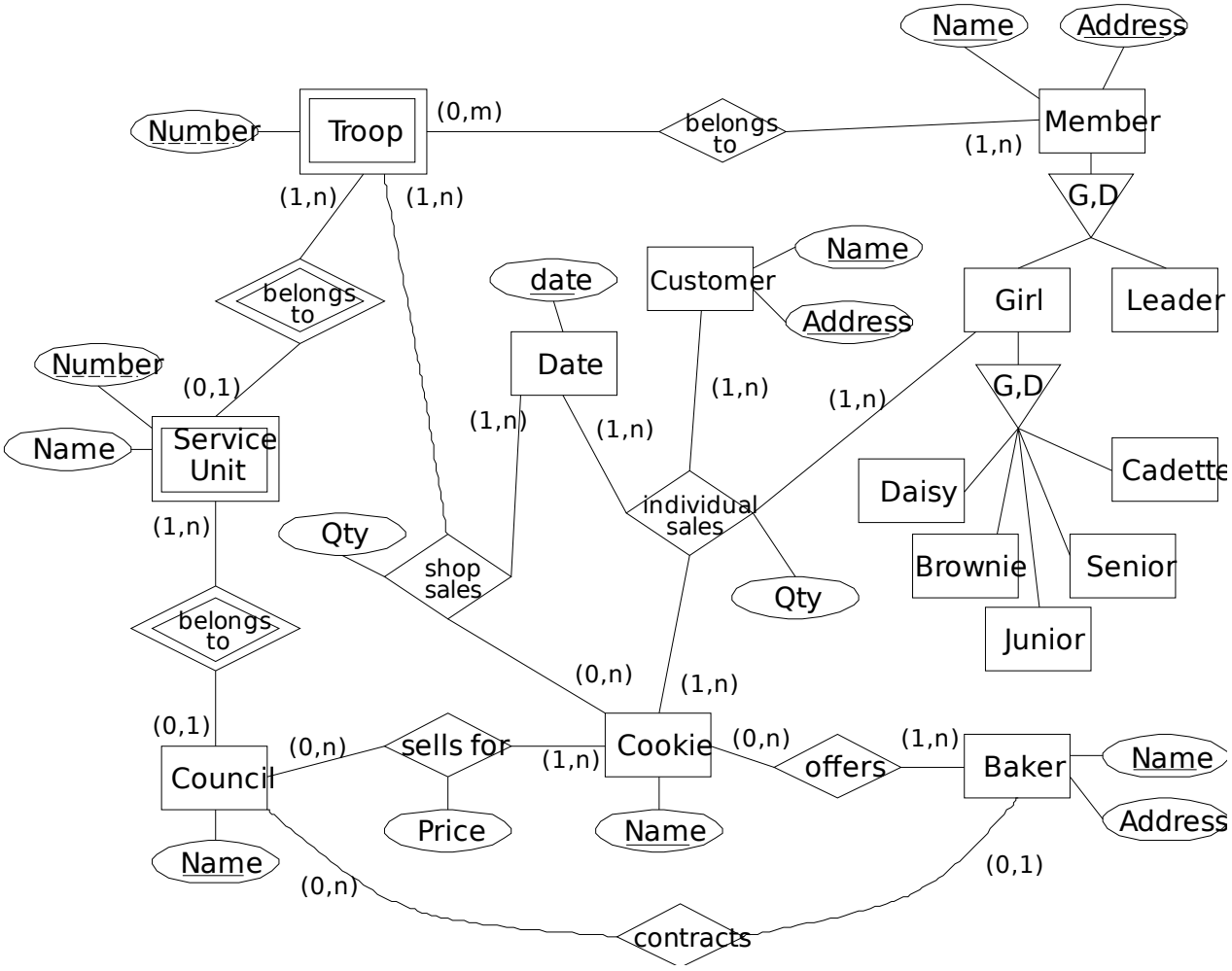
Submission Instructions: Place your schema creation script, your program source code files (and associated files) into one directory on your turing account. Do NOT send data files. Move up from that directory and type the following magic incantation (assuming the directory in question is called assign3)

```
tar -cvf - assign3 | gzip -c9 > assign3.tgz
```

Spaces are important. This will create a file, assign3.tgz that should be e-mailed to `cs511_2022_3@cs.niu.edu`.

GirlScout ER Model – Entities, Relationships, and Attributes

The ER model helps maintain a database of cookies sold by girl scouts.



Each entity and relationship below contains a description of its role in the project, along with necessary attributes as needed.

Entities

Cookie
Every year, the Girl Scouts hold a fund raising activity by selling cookies. There are a small number of cookie types that are available each year, about 6. The names of the cookies are distinct, and serve as an effective identifier. The price of the cookies is set by each Girl Scout council selling them, and may vary from council to council. All mentions of cookie quantities in this problem refer to boxes of cookies, not individual cookie counts.

Baker
There are only two bakers currently used by the Girl Scouts. Between them, they handle the entire United States. Each baker determines what cookies will be offered. Each Girl Scout council contracts with a baker to provide the cookies for its cookie sales for the year.

Troop
The basic unit of the Girl Scout organization is the troop. Each troop consists of a number of girls and their leaders. Each troop consists of girls who are mostly the same rank (see below). Each troop is identified by a number, but that number is unique only within the larger organization of a council. The number is not unique across the country.

Service Unit
A service unit consists of a number of troops. Each service unit has a number, which is unique within a council. It also has a name, which is also unique within the council that the service unit belongs to.

Council
A council consists of a number of service units. Each council has a unique name.

Girl
Girls are uniquely identified by their name and address. Each girl in a troop has a rank which is determined roughly by age. The ranks are (in order of increasing age): Daisy, Brownie, Junior, Cadette, and Senior. (There are technically more, but that will do.) Each girl only belongs to one troop.

Leader
Leaders are also uniquely identified by name and address. Leaders will sometimes be a leader in more than one troop. (Usually when they have multiple daughters of different ages.)

Customer
The customers in this entity are those that the girls contact when they go collecting orders for cookies. Each customer is uniquely identified by name and address.

Relationships

Individual Sales
At a time determined by the council to which they belong, Girl Scouts will go out and collect orders for cookies. People sign up to purchase whatever amount they desire. No money is collected at that time. After the orders are tallied, they are sent to the baker, and about a month later, back come boxes and boxes of cookies. At that time, the girls go back out to their customers, deliver the cookies, and collect the money.

Shop Sales
A troop will also have cookie shops, where they go out as a troop to a public place and sell cookies to people who happen by. As the cookies are sold and money collected on location, no specific information about the customers is needed.

Data Files

Some data files will be provided for this assignment. Like most real world data, the files are not in an ideal format for the data base. Your loading program will have to read and process the data file records.

The file consists of cookie sales data. One line represents one piece of sales data. Each line consists of fields separated with a vertical bar character, '|'. For simplicity, you may assume that this character will not occur in the data file unless it is a field separator . The lines are very long. The fields, in order, are

- Council name
- Service Unit name
- Service Unit number
- Troop number
- Leader name
- Leader address
- Girl name
- Girl address

- Girl rank
- Customer Name
- Customer Address
- Cookie name
- Cookie price per box
- Baker name
- Baker address
- Quantity sold
- Date sold

If the sale represents a shop sale rather than an individual sale, the girl name, girl address, customer name, and customer address fields will all be empty.

Program(s)

Your program to load data files should not be hardwired to a specific set of files. The user should be allowed to determine through command line arguments, what files are to be loaded. The name of the program should be "`load`".

A separate program should ask the user for a council name and a troop number. The program should print out the total amounts (box count and dollar amount) for each cookie sold by the troop. The name of the program should be "`query`".

Additional Implementation Points

- Numeric values should be stored as integers or floating point numbers, not strings.
- The database should be implemented as a schema.
- Do NOT submit data files with your assignment.

Programming Languages and Libraries

Assignments will be graded on the department's Unix system and are expected to run from the command line. Accepted implementation languages for parts 3 and 4 of this assignment are C, C++, Java, and Python. As of this writing, the language compilers/interpreters and associated libraries available for implementation are

Language	Compiler/ Interpreter	Postgres Library
C/C++	gcc/g++ 10.2	libpq
Java	OpenJDK 11	postgresql JDBC
Python	Python 3.9.2	psycopg2

Hints

A function that takes a line of input in the given format and splits it up into its component values is immensely valuable.

When reading an input file, it's very handy to keep a running count of the the lines read. Then when an error occurs, you can print out the file line number along with your error message. This makes it easy to find the special case that caused your program to fail.