COMP-421 Database Systems, Winter 2022

Project 1: Database Design and Data Modelling

Due Date Feb 01, 12:00 noon

In the programming project of this course, you will develop and build a database application for a **real-world** domain. Step-by-step, you will:

p1: design a model.
p2: create a schema and implement a database using DB2, populate your database with data, maintain, query and update your data
p3: develop application programs, and implement a user-friendly interface. The interface can be very simple, so no requirement for web-programming, etc. You will only use a standard programming language (Java) in the last project deliverable.

In the first deliverable of the project, you are only concerned with developing a data model for your application.

The COVID-19 outbreak in recent years has exposed several issues with the healthcare systems across the world. Governments are exploring ways in which they can reduce the strain on the hospital systems by offloading non-critical healthcare to other institutions.

A great example of this is the Midwifery (*Sage-femme*) program run by the Quebec Ministry of Health. This program allows pregnant couples to receive support and basic health care services through a personalized midwife (who is medically trained and certified) and not be caught up in the bureaucratic rigmarole and intimidation of a hospital-based environment.

The ministry has hired you to develop a database (and associated application) to manage the Midwifery services across the province. Will you play a pivotal role in bringing affordable and personable health care to the expecting parents in Quebec?

To constrain to the scope of the course, we have simplified many real-word situations and procedures in the below description. Any semblance to reality is merely an accident. I also encourage you to lookup some of the terms and concepts to explore a bit of the world beyond your CS background.

Application Requirement

A midwife is associated with either a community clinic or a birthing center (the latter has additional facilities to give birth at the center, unlike the former). The system you build should also keep track of the midwife's name, email, phone number, and practitioner id.

Parent(s) sign up for the program through a website. The sign up process involves first submitting the personal information of the mother (name, Quebec health care card identification, date of birth, address, phone number, email, current profession) as well as expected time frame for birth (year-month). Optionally, the same personal information of the second parent can also be recorded in the system (which minimally would be their name, date of birth, phone number and profession, the other attributes being optional - but the system should nevertheless be able to store them if provided). For simplicity, we will assume that the second parent is the biological father. Parents who qualify are then invited by their local facility to attend an online information session at some future date and time depending on the availability of such sessions. The information session is hosted by one of the midwives. The system should also record the date and time of the session, its language and the couples (even if it is only one of the parents - we do not distinguish) who registered for it, attended, etc.

From the parents who express interest with the program after the information session, an institution selects some of the couples (due to various reasons that we will not get into the details) and assigns them to a specific midwife

for that pregnancy (could be different from the one who did the information session). We also keep track of whether this is the first pregnancy, second pregnancy, etc., (of the couple, not just the mother - this is an over simplification). The expectation is that the midwife will support the couple throughout the pregnancy and the first couple of months of postpartum.

From here on, the midwife sets up regular appointments with the parents, which may be infrequent (like once in a month) in the early months of pregnancy and possibly a couple of times a week or more towards the end of the pregnancy and immediately after giving birth or as needed. The system will have to keep track of these appointments (date and time) and also any associated notes (observations) from the midwife. Make sure each note (even within the same pregnancy) can be uniquely tracked and store the date and time when it was recorded in the system. Multiple notes are possible within an appointment, in which case they would have different timestamps.

In some of these appointments, the midwife may also either takes any necessary medical test samples (such as blood for testing the iron content in the mother's blood), which they will send to labs or writes a referral for any tests (such as ultrasound) that will have to be conducted at a lab by the parents. The system will have to keep track of these tests such as the type of test (*blood iron, first trimester ultrasound*, etc.), the date the test was prescribed, the date the sample was taken (if applicable), the date the actual lab work was done, as well as the result from the test, and the name, id and the phone number of the technician who conducted the actual lab work. The test results are updated by the technician directly on to the system.

In one of the initial appointments, the midwife, upon discussion with the mother, will set a more accurate expectation on the due date based on her *last menstrual period*. Later, a *dating ultrasound* test is conducted to do another estimate of the expected due date. If these two dates vary, the midwife upon discussion with the mother will choose one of them as the final estimated due date for the pregnancy. The system must still maintain all of these different dates and information.

At some point in the pregnancy, the mother's blood type information (may or may not be the outcome of one of the tests) is recorded. Father's blood type is also recorded if already known through some medically reliable mechanism (like from a blood donor card).

Further, at some point, the number of babies and their gender is also determined (through some of the tests, but this information may come from different tests at different points in time). At some point, their blood type, date of birth (and time), etc., also would get added. The name of the child may also get included if the parents have already chosen one.

Midway through the pregnancy, a "backup" midwife is also assigned, who can take over if the primary is not available (also helps the primary during the actual birth). This person is capable of doing all the same activities (such as taking samples for tests, notes, etc.) as the primary and will also have some appointments, etc., to help the couple familiarize with the backup midwife as well. Around this time, the decision on the location of birth is also made depending on the couple's preferences and circumstances. The location can be at home or one of the birthing centers (need not be the place where the midwife is working). Community clinics do not have birthing facilities.

The system should be able to figure out to which pregnancy/individual a note, test, etc., applies.

The system should also keep track of the basic contact information of the healthcare institutions such as name, phone number, email, address and optionally their website.

Your Task

1. (20 Points) An ER schema/diagram including your data requirements. Be careful not to forget to underline key attributes, indicate the types of relationship sets, etc. Make sure that your ER diagram has large enough fonts, etc. so that the TA can read it properly. You may use a page size larger than letter format in your drawing software. **The ER diagram you submit should be a searchable PDF** (If you export your drawing as PDF in most of the drawing softwares like Powerpoint, Google draw, etc., you should get a searchable PDF). This helps TAs quickly lookup some of the key information without having to manually look for it in the entire diagram. **Do not hand draw and scan!!** Turn in the diagram as **ER.pdf**. **If you hand draw or turn in a non-searchable PDF, that will be a 10 point penalty!** Use only the ER drawing notations that we discussed and covered in the lecture slides. **Points will be deducted for not following this** (we do not care about colours, but if you use them, be consistent). Use sensible names for your entity sets, attributes and relationships so that the TA can easily relate your model to the requirement. You may use sensible abbreviations of the name (e.g *hcardid* to indicate health care card identification), etc. to reduce clutter in your diagram.

The above application description may not be comprehensive in capturing every minute details. You are free to make assumptions that do not defy common sense as long as it does not go against what is already explicitly

provided. Please DO NOT add more requirements into the system. You will end up making your life difficult further down. Do not make "assumptions" because you do not know how to solve a modeling problem. If you make some assumptions which you think may not be that "intuitive", please provide a PDF file **project.pdf** that contains those, so that the bureaucrats do not get on your case. Make sure to write them under a section heading **Assumptions**.

Further, if there are any requirements and constraints within the application that you cannot depict in the ER diagram (there will be quite a few), you need to point them out now, so that your developer (your future self) can keep an eye out for them when writing the application code. Indicate them under a section **Restrictions** in **project.pdf**.

2. (5 Points) Use the method for translating an ER diagram to relations described in class and depict each resulting relation in the form `Relationname(`<u>`attr1`</u>`, attr2, attr3,...)` underlining the key attributes. Indicate when attributes are foreign keys to other relations by writing something like "attr3 foreign key referencing relation X" beside the relation. Are there opportunities to combine relations without introducing redundancy? If so, indicate which, and if not, tell us there are none. Write all of this under a section **Relational Translation** in **project.pdf**. Also indicate any ER model aspects that your relational model does not capture in this section. Again, this submission should be typed in and searchable and not hand written.

Further, remember that a significant aspect of this question is about testing your ability to translate an ER to corresponding relational model. If you have prior experience with relational systems so you wrote a relational model that can store the correct information, but it is not a translation of your ER, you will get no points.

Note: DO NOT write SQL create statements instead of the relational model. **You will not be given ANY points**. You do not have to decide on the data types either.

Submission Format: Your submission will contain **two** separate documents:

1. **ER.pdf** : The ER diagram of your project. This is a PDF file (not drawn on paper). Make sure that the ER diagram you submit is readable (such as font size, etc) and searchable (similar to how you can search for a word in the lecture slides and find it). You may use a page size larger than letter for your ER diagram page if that ensures it is readable. Make sure that the thick line vs thin lines, dotted lines, etc., are very clear in the diagram without having to zoom into it at 200x.

2. **project.pdf** Rest of your project document (assumptions, restrictions, relational translation, etc.), should be in a PDF file (again a searchable PDF).

Please turn in your submission in mycourses under project 1. Submissions to wrong folders may not get graded. **You are responsible to download and verify that your submission is correct (not corrupted or incorrect file, etc).** There will be no accommodations if you decided to ignore doing due diligence (you will get a 0). Project grades also influences the pass/fail criteria of the course. Please review the course outline for any of these details.

Approach

Read the requirement part 2-3 times to get a general idea of what you are dealing with. Sit down with a pencil and paper (old school style), draw out just the entity sets and keys in it that evidently stands out in the requirements. Draw the relationships between them. Now go back to the requirements and see what is it that is missing in your ER model and make one more pass. Now you have a decent base ER model to work with. You can start drawing it in your drawing software and tune/adjust as required, add additional attributes that are part of the entity sets, etc.

Review your ER model. In essence, you are trying to read the model and figure out what it can do. Is there any important "information" that your system is not able to capture? (e.g. Who is the mother of a particular baby?) - this is a very bad situation. Take actions to address it. Does your model results in duplicate information - try to avoid this when you can through proper modeling techniques, if that is not possible, or if you have sensible reasons (as discussed in class) to keep it as such, document it in your **restrictions** so that your developer is aware of this. Can your model result in inconsistent information being stored? (e.g a midwife who is not involved with the appointment being able to add notes to it) - again if a proper modeling technique does not exist to solve this problem, then you should document it.

Visualize the data as it flows it into the system (as time progresses). Which entity sets and relationships store that data? Are there constraints in your model which will prevent some information being stored because some other additional information as required by your constraint will be available only in the future?

It is beneficial if you split the entire exercise over multiple days instead of cramming into one day, so that you have a little bit of time to think about it at the back of your mind as you keep refining your model. Good ideas do not pop out when you are under duress. Try to "immerse" yourself into the various roles in the system and imagine how the data gets captured in your model.

Remember, your ER model will not be able to capture all of the requirements. The challenge is to recognize this for specific situations. Do not over-complicate the model. It does not necessarily solve the problems, but might add more. A simple model should be preferred over a complex model with the same number of limitations (i.e., do not try to throw every technique you learned into the model if it is unnecessary).

An ideal implementation of this requirement should result in approx 15 entity sets and a similar number of relationships. Use that as a guidance to keep track of your modeling adventures.

Additional Assumptions

(a) A person may parent children with different individuals.

(b) Midwives can be assumed to be employed at a single place.

(c) Do not worry about "changes", e.g. - midwives changing employers, retiring, parents switching midwives, clinics, etc.

(d) There could be multiple babies involved in a pregnancy.

(e) All births in the same pregnancy will happen at the same location. Though the babies themselves are born one after another.

(f) Tests are meant only for the mother or the baby.

(g) Health care card information, practitioner id, and email are considered unique to an individual, none of the other attributes should be assumed to be unique. You may have to take this into accommodation when designing keys and then potentially add artificial keys. Make sure to include a note in the **project.pdf** listing all the artificial keys.

(h) Although it is possible that blood types of an individual might change in extreme circumstances, we will assume that it remains the same.

(i) In reality midwives will have to transfer care to a hospital if the pregnancy is determined to be medically complicated making natural birth risky. We will ignore this aspect.

(j) A midwife can be primary for one pregnancy and a backup for another (i.e. every midwife is capable of either roles).

(k) For the purposes of home birth, "birth location" is considered as the address of the mother.

(l) There could be more than one info session running at the same schedule.

(m) A couple could also end up attending more than one information session but they should attend at least one before they are assigned a midwife.

(n) We do not have to store any other information that pertains only to the lab tech (as to who employs them, etc.). Only what is required to connect them to the given requirements.

(o) The system should be able to keep track of components from different pregnancies, etc., properly and not just rely on "time" attributes. For example, was that dating ultrasound from the first pregnancy or the second one?

(p) Make sure you do not miss any associations (direct or indirect) between the data - e.g. "who wrote this note and who are the concerned parents?"

Questions ?

Please use **Ed** for any clarifications you need (`projects → p1`). Do not email the instructor or TAs as this leads to a lot of duplicate questions and responses (not an efficient system). Such emails will not receive any replies.

The only emails you are allowed to write is to your grading TA if you have questions on the feedback you received (as they are personal in nature) - do not post them in Ed, your grading TA may not see it.

Please check the pinned post "P1 general clarifications" in Ed before you post a new question. It might have been already addressed there, in which case we will not address it again.

Questions about general clarifications must be marked public (as other students will also benefit from this and may even have a valid response). TAs and Instructors upon their discretion may toggle any private posts into public mode for the benefit of the student population at large.

Do keep in mind that clarifications are limited to the nature of the application requirement (for example you want to make an assumption that a midwife may not be immediately assigned when an information session is advertised and only later, as it is not explicitly stated in the requirement). Do not ask for clarification on whether your ER solution/approach is correct or not. Remember, the bureaucrats know only of their application needs, not much about software development.

There will be specific office hours for the project that will be announced closer to the due date.

Extensions and Late submissions

- Remember, your submission is **due on 12:00 noon**. There is no place for excuses.

- A maximum of 2 days of late submission is allowed with a penalty of 20% of the achieved grade per day (rounded up, even for a minute).

- Penalty waivers are granted only for medically documented emergencies and under any circumstances **will not be granted unless requested 24 hours before the due**. I expect you to be better organized and get the job done ahead and leave the last 24 hours only for one last final check.