

# csc343 winter 2022

## assignment #1: relational algebra

### due February 13th, 4 p.m.

## goals

This assignment aims to help you learn to:

- read a relational schema and analyze instances of the schema
- read and apply integrity constraints
- express queries and integrity constraints of your own
- think about the limits of what can be expressed in relational algebra

Your assignment must be typed to produce a PDF document **a1.pdf** (hand-written submissions are not acceptable). You may work on the assignment in groups of 1 or 2, and submit a single assignment for the entire group on **MarkUs**. Partners do not have to be in the same section, but must be in CSC343 at St. George this semester. You must establish your group well before the due date by submitting an incomplete, or even empty, submission. Teaching staff are likely to get quite grumpy if you try to form, or dissolve, a group close to the due date!

## background

You will be working on a schema for a database to track covid-19 vaccinations.<sup>1</sup> Vaccines batches are tracked from the factory that produces them. Their arrival in time Canada, and in the province or territory they are to be administered in, is recorded. The manufacturer records minimum intervals for follow-up doses, as well as the number of doses in a vaccination sequence. There is a unique identifier for each vial of vaccine.

Patients, vaccine administrators and attendants are each recorded, both to follow up on subsequent doses (where required by the manufacturer), and to track vaccine efficacy and safety. Each patient's covid status at the time of vaccination is recorded, and the time of the latest subsequent infection is recorded. Patients are observed by the attendants for at least 15 minutes after vaccination, and any bad reactions are treated and recorded. Vaccine attendants and administrators are assumed to be patients.

---

<sup>1</sup>Details may not be accurate, and you should **not** rely on any of this for medical advice.

## relations

- Batch(bID, mID, productionDate, vialCount)  
Vaccine batch *bID*, manufacturer *mID*, was produced on *productionDate*, with *vialCount* vials in this batch.
- Vial(vID, bID, thawTime, dose\_count)  
Vial *vID* from batch *bID* removed from cold storage at *thawTime*, with *dose\_count* doses remaining. The number of doses remaining is decremented after each dose is used.<sup>2</sup>
- Manufacturer(mID, name, thawMax, intervalMin, sequence\_length, duration)  
Manufacturer *mID*, with company *name*, *thawMax* maximum hours vaccine is usable after being removed from cold storage, *intervalMin* minimum days to next dose, *sequence\_length* number of vaccinations to be fully vaccinated, *duration* numbers of days of protection after vaccination.
- Tracking(bID, *canadaDate*, *locationDate*, locationName)  
Batch *bID* arrived in Canada on *canadaDate*, shipped to province or territory *locationName* on *locationDate*.
- Vaccination(pID, date, vID, adID, atID, reaction, covidStatus)  
Patient *pID* vaccinated on *date* from vial *vID*. The dose was administered by *adID*, the patient was attended by *atID*. At vaccination time the patient had infection status *covidStatus* and reaction to vaccine *reaction*.
- Patient(pID, latestPositiveTest)  
Patient *pID* had most recent positive Covid-19 test on *latestPositiveTest* (00:00:00, January 1st, 1970 if this has never happened).
- Staff(sID, pID, specialty)  
Medical staff *sID* is also patient *pID*, and has medical specialty *speciality*.

Note that there are 3 parts to work on in this assignment: **our constraints**, **queries**, and **your constraints**. Make sure you complete all 3.

## our constraints

**TASK:** For one point each, give a one-sentence explanation of what each constraint implies.

- $\Pi_{pID}Staff - \Pi_{pID}Patient = \emptyset$
- $(\Pi_{adID}Vaccination \cup \Pi_{atID}Vaccination) \subseteq \Pi_{sID}Staff$
- $\Pi_{specialty}Staff \subseteq \{'RN', 'RPN', 'MD', 'Pharmacist'\}$
- $\Pi_{pID}Vaccination \subseteq \Pi_{pID}Patient$
- $\Pi_{bID}Vial - \Pi_{bID}Batch = \emptyset$
- $\Pi_{covidStatus}Vaccination \subseteq \{'positive', 'negative'\}$
- $\Pi_{reaction}Vaccination \subseteq \{'true', 'false'\}$
- $\Pi_{mID}Batch \subseteq \Pi_{mID}Manufacturer$

---

<sup>2</sup>A timestamp of 00:00:00, January 1st, 1970 is recorded for any events that have not happened (yet).

- $\Pi_{bID}Tracking - \Pi_{bID}Batch = \emptyset$
- $\Pi_{vID}Vaccination - \Pi_{vID}Vial = \emptyset$

## queries

**TASK:** For 4 points each, write relational algebra expressions for each of the queries below. You must use notations from this course and operators:

$$\pi, \sigma, \rho, \bowtie, \bowtie_{condition}, \times, \cap, \cup, -, =$$

You may also use constants:

$$\text{today (for current date)} \quad \emptyset \text{ (for the empty set)}$$

In your queries pay attention to the following:

- All relations are sets, and you may only use relational algebra operators covered in Chapter 2 of the course text.
- Do not make assumptions that are not enforced by our constraints above, so your queries should work correctly for any database that obeys our schema and constraints.
- Other than constants such as 23 or "Moderna", a select operation only examines values contained in a tuple, not aggregated over an entire column.
- Your selection conditions can use arithmetic operators, such as  $+, -, \times, /, \leq, =, \neq, \geq, >, <$ . You can use logical operators such as  $\vee, \wedge$ , and  $\neg$ , and treat dates and numeric attributes as numbers that you can perform arithmetic on. Difference of two dates produces a floating point number of days, and adding/subtracting a float to a date produces a new date the appropriate number of days later/earlier. Dates are comparable.
- Provide good comments to explain your intentions. Use meaningful variable names that include attributes in brackets, e.g. `VaccinePatients(pID) := ...`
- Allow the return of multiple tuples if that is appropriate for your query.

There may be a query or queries that cannot be expressed in the relational algebra you have been taught so far, in which case just write "cannot be expressed." The queries below are not in any particular order.

1. Rationale: Let's see how well we're doing.

**Query:** Find pID of all patients who have received a dose of a two-dose vaccine, followed by any other vaccine after the minimum interval of the former vaccine, and who are currently within the duration of protection of some vaccine.

**Query:** Find the specialties of every staff who has administered vaccines from every batch that was used in British Columbia after April 2021.

2. Rationale: Let's see how badly we're doing.

**Query:** Find pID of all patients whose latest positive test is after the duration of their latest vaccine expired.

**Query:** Find  $sID$  of all staff who administered a vaccination from a vial that had thawed longer than recommended by the manufacturer or administered a vaccine earlier than the minimum interval from an earlier vaccine.

**Query:** Find  $vID$  of all vials that had 2 doses or fewer used by the time they had exceeded the maximum time recommended by the manufacturer after thawing.

**Query:** Find  $vID$  of all vials that had all their doses used by the time they had exceeded the maximum time recommended by the manufacturer after thawing.

3. Rationale: Trace exposures.

**Query:** Staff  $sID_1$  is exposed to covid-positive staff  $sID_2$  if one or more of (a), (b), or (c) occurred:

- (a) staff  $sID_2$  administered or attended staff  $sID_1$ 's vaccination,
- (b) staff  $sID_1$  administered or attended staff  $sID_2$ 's vaccination,
- (c) or if some staff exposed to  $sID_2$  administered or attended  $sID_1$ 's, or had a vaccination administered or attended by  $sID_1$ . vaccination.

Find  $sID$  of all staff exposed to the covid-positive staff with  $sID$  42.

4. Rationale: Quality control.

**Query:** Find the staff who gave the most recent vaccine that had a reaction. Keep ties.

**Query:** Find all patients who did not have a positive covid status when they were first vaccinated, but did have a positive test at some later date.

## your constraints

**TASK:** For two points each, derive a relational algebra expression of the form  $R = \emptyset$ , where  $R$  may be derived in several steps, by assigning intermediate results to a variable. If the constraint cannot be expressed in the relational algebra you have been taught, write "cannot be expressed."

1. No batch is from two different manufacturers.
2. Every manufacturer has produced at least one vial.
3. Every manufacturer's vaccine has arrived in Canada.
4. All staff receive at least two doses.

## submissions

Submit **a1.pdf** on **MarkUs**. One submission per group, whether a group is one or two people. You declare a group by submitting an empty, or partial, file, and this should be done well before the due date. You may always replace such a file with a better version, until the due date.

Double check that you have submitted the correct version of your file by downloading it from MarkUs.

## marking

We mark your submission for correctness, but also for good form:

- For full marks you should add comments to describe the *data*, rather than *technique*, of your queries. These may help you get part marks if there is a flaw in your query.
- Please use the assignment operator, “:=” for intermediate results, and use meaningful names. Ensure you include the attribute names in brackets e.g. `VaccinePatients(pID) := ...`
- Name relations and attributes in a manner that helps the reader remember their intended meaning.
- Format the algebraic expressions with line breaks and formatting that help make the meaning clear.