308-612A Database Programming Principles

Supplementary Questions and Exercises

Relations

1. A construction project consists of various ACTIVITIES, each with a DURATION and a set of PREREQUISITE activities which must be completed before it can begin. EAch activity has associated PERSONNEL with the skills required for that activity, and each person has an ASSIGNMENT to specific activities. Using the capitalized words above as attributes, devise a relation or set of relations which reflects all this in a sensible way. State your assumptions and give sample data sufficient to show all possibilities.

2. A "Leontief" or "input-output" matrix is used by economists to record transfers between sectors of a national economy. In an economy with sectors such as *Raw Materials*, *Manufacturing*, *Communications*, *Services* and *Import-Export*, \$10 billion might represent the transfer from *Raw Materials* to *Manufacturing* in a given year.

   (a) Give a relational representation of this model of the economy of a single nation for one year. Show functional dependences and indicate what normal form your relation(s) is (are) in. Use one of the table, graph or matrix forms to show sample data.

   (b) Give a relational representation of this model extended to several years and several nations. (Read (c), below, before proceeding.) Show functional dependences.

   (c) Give a relational representation as in (b) but remove the *Import-Export* sector from each nation. Instead, give an integrated representation in which each sector of each national economy connects with each other sector of each other national economy. Show functional dependences.

3. A choir membership list contains the following attributes: NAME, ADDRESS and PHONE number of the member, with two dates, APSTARTDATE and APENDDATE, giving the range of dates during which the member had the address and phone number (this is for archival purposes: to hold reunions and to apply for cultural grants, it is useful to have past addresses); a WORKPHONE number with corresponding WSTARTDATE and WENDDATE; a choir ROLE (e.g., director, accompanist, soprano); and two further dates, STARTDATE and ENDDATE, giving the range of dates during which the individual was a member of the choir. Design a database consisting of one or more relations to represent this data. Show functional dependences, keys, sample data and any assumptions you make. If you get more than one relation, can they be joined without loss of information to form a single meaningful relation?

4. Using point form (rather than attempting to write an essay), say as much as you can about normal forms of relations.

5. The following attributes take part in a geographical database listing the populations of towns and the areas of their countries: YEAR (of the census), TOWN NAME, COUNTRY NAME, AREA, POPULATION, LATITUDE, LONGITUDE (of the town).

   Design a database consisting of one or more relations to represent this data. Show functional dependences, keys, sample data, and any assumptions you make. If you get more than one relation, can they be joined without loss of information to form a single meaningful relation?

6. State, and discuss the significance of, the principles of abstraction and closure. In your discussion, make *specific* reference to the relational and domain algebras. Use point form rather than attempting to write an essay.

7. A *hypertext* is a set of text documents connected by *links*, so that selecting a predetermined word in a text can retrieve another text which explains the word or its topic. Links can be of two types, a *new window* type, which opens a new window for the target text, and an *expansion* type, which expands the word by replacing it with the target text.

   Design a database consisting of one or more relations on the attributes *document, word, seq, linkType, linkFrom, linkTo*, where *seq, linkFrom* and *linkTo* give sequence numbers. Show enough data to illustrate more than one document (but use abbreviations and keep the examples small). Show keys and functional dependences. State your assumptions.

8. In a Human Resources Information System for Ideal University, a "person" has a unique *employee number*, a *name* and an *address*. There are two kinds of person: a "pensioner" has a *termination date*, and an *accumulated amount* of capital on which to base an annuity; an "employee" has a *start date* and a *source of funds*. Employees can be "academic", "student" or "support": academics have a *salary*; students have an *amount per pay* and a *termination date*; and support staff have *wages* and *hours worked*.

   Design a database consisting of one or more relations on the attributes given by the italicised words and phrases above. State your assumptions and show enough data to illustrate them clearly. Use abbreviations and keep the examples small. Show keys and functional dependences.

9. A minerals company owns about thirty mines and wishes to record data on technologies used for refilling mined-out sections. There are nine such technologies, using one of three different *filltypes* (rock, sand, or tailing), coming in one of three *concentrations* (low: $\approx 67\%$; high: $\approx 77\%$; paste: $\approx 82\%$). Each technology also uses one of half a dozen pumps, described by *pumptype* (centrifugal or reciprocating), *capacity* (tons per hour), *pressure*, *manufacturer*, and *cost*. Each technology works best for a given *LHratio* (length to height) and produces a certain *flowrate*. Each mine has a *name*.

   Including the above ten italicized labels as attributes, design a relation or relations to support the data, minimizing redundancy. You may need additional attributes. State assumptions explicitly. Show keys and functional dependences. Show a *few* but *appropriate* tuples of sample data, which illustrate assumptions and dependences.

10. A real estate company needs to store the following data on houses, plus photographs which are stored digitally and can be shown to clients. For each house we record *street, number, living area, building evaluation, land area, land evaluation*, and a history of sales in the past five years: *date* and *price*.

    Each house has a set of photographs, which are linked by buttons: every house has a shot of the outside front, with possible buttons leading to views of the sides and of the inside front hall. The front hall photo may have buttons leading to other rooms or upstairs. Every photo has a button leading back to the view of the outside front.

    In order that the houses can be displayed on a map, we store the *longitude* and *latitude* of each end of each street and the range of house numbers on it. (All streets are assumed to be straight.)

    Design (a) relation(s) to capture all this data in the most flexible way. You may need to add attributes to those named in italics, above. You may assume that there is a

photo type of attribute, but that it requires a lot of storage, and you do not want photo duplication.

Be explicit about any assumptions. Show dependences and keys. Show sample data.

11. A university maintains a follow-up service which seeks input from its graduates every five years after graduating, and from the employers of those graduates.

    Each grad is sent a questionnaire with the following items, some of which are already filled in from university records: *sid, name, degree, year, department*, and, for each course the grad took, *cno, title, prof*, a field, *useful*, which says (`yes` or `no`) whether the grad has used the course material on the job in the past five years, and, if so, whether it was *valuable* or not, and a list of *topics* which the course should have taught but did not.

    The grad's company is sent information about graduates working for them and their training, and asked for lists of missing *skills* for each university department that it would like to see in future graduates.

    Design (a) relation(s) to capture all this data in the most flexible way, both the data sent by the university and the data returned. Do this from the point of view of the university: data about different graduates and companies does not have to be stored in different relations.

    You may need to add attributes to those in italics above. Be explicit about any assumptions. Show dependences and keys. Show sample data.

12. A world-wide database for travellers records information on 500,000 restaurants, 100 restaurant chains, and 1000 restaurant districts.

    Some, but certainly not all, of the restaurants belong to chains. Some, but certainly not all, of the restaurants are in restaurant districts.

    The attributes should include *RestName, Address, TC* (whether the restaurant is in town or in the country), *Cuisine* (e.g., Afghanistani, Greek, Hamburger), *PriceRange* (low/medium/high), *ChainName* and *HeadQuarters* of the chain, *District, City* of the district, and *South, West, North* and *East* streets bounding the district.

    Design (a) relation(s) to capture all this data in the most flexible way,

    You may need to add attributes to those in italics above. Be explicit about any assumptions. Show dependences and keys. Show sample data.

13. A city maintains data on properties with the following attributes. *Street, Number, LandArea*, and *ConstructionCode* are permanent attributes of a property. *Owner, LandEvaluation*, and *BuildingEvaluation* are for current values. *DateOfSale*, and *SalePrice* record the history of sales (say about one sale every ten years).

    Design a relational database to capture all the data in the most flexible way. You may need to add attributes to those in italics above. Be explicit about any assumptions. Show dependences and keys. Show sample data.

14. An international database keeps track of population by recording, for each person, dates and countries of birth and death, and dates and countries of migration (that is, country of emmigration and country of immigration; moves within a country are not recorded). The data is recorded so as to permit calculations of total population of a country for any year, births or deaths per capita for any year, and similar statistics for migration. (It can also determine *anniversaries*—e.g., for any given day in the year, who was born on that day.)

Design a relational database to capture all the data in the most flexible way. You will need to devise suitable attributes from the above description. Be explicit about any assumptions. Show dependences and keys. Show sample data.

15. In the international agriculture market, countries produce foods in various quantities, as measured by various units, e.g., bushels of wheat, kilos of potatoes, dozens of eggs, head of cattle. Producing each food requires resources, which can be measured by the amount of grain (wheat, corn, rice) required to produce a unit of that food, e.g., eight pounds of grain per pound of beef. Countries import and export foods in trade. The populations of countries can be used to calculate production, imports, and exports *per capita*.

    Design a database of attributes and relations to capture the above data in a flexible way. (For instance, it should be possible to calulate whether a country is a net importer or exporter, or what its *per capita* consumption of any food is.) Be explicit about any assumptions. Show dependences and keys. Show sample data.

16. A system of nature trails consists of trails of different "colours" (e.g., the red trail, the yellow trail, etc.). Trails also have names, such as "The Eagles' Nest" or "The Hemlock Ridge". Most trails are loops, starting and ending at the same point, but some are link tails, crossing from one point in one trail to some point in another. Along each trail are occasional points or regions of interest, such as a porcupine nest or a stretch of nettles. These are described with a phrase, or a longer text, or a diagram or photograph. Some of these parts of a trail have more than one associated description, such as "maple sugar bush" and "steep gradient". Some of the regions of interest are quite extensive in size.

    Design a relational database to integrate all this data into a trail map in the most flexible way. You will need to devise suitable attributes from the above description. Be explicit about any assumptions. Show dependences and keys. Show sample data.

17. The *quad-edge* representation of spatial data represents a map as two sets of cycles of edges: the counterclockwise cycles of the edges around each vertex, and the counterclockwise cycles of the edges around each polygon.

    (Think of the pieces of a jigsaw puzzle, each as a polygon with straight sides: the sides are the edges, and the points where the edges meet each other are vertexes; some of these vertexes have only two edges meeting there, some have three or more.)

    Each edge is directed, and has an origin and a destination vertex; each has a polygon to its left and one to its right. (We can pretend that the outside edges of the whole map (or jigsaw puzzle) bound a fictitious external polygon.) Thus, each edge must appear four times: once in the cycle for each of these points, and once in the cycle for each of these polygons.

    The vertexes each have x and y coordinates. Since this representation makes vertexes and polygons symmetrical to each other, in the sense that each is a cycle of edges, each polygon also should be permitted to have coordinates, which could be those of its centre of mass, or of its circumcentre.

    Design a relational database to capture all these properties. Make up a small example. and use it to populate your relation(s). Indicate functional dependences and keys, and *state your assumptions* explicitly.

18. The *quad-edge* representation of spatial data represents a map as two sets of cycles of edges: the counterclockwise cycles of the edges around each vertex, and the counterclockwise cycles of the edges around each polygon.

(Think of the pieces of a jigsaw puzzle, each as a polygon with straight sides: the sides are the edges, and the points where the edges meet each other are vertexes; some of these vertexes have only two edges meeting there, some have three or more.)

Each edge is directed, and has an origin and a destination vertex; each has a polygon to its left and one to its right. (We can pretend that the outside edges of the whole map (or jigsaw puzzle) bound a fictitious external polygon.) Thus, each edge must appear four times: once in the cycle for each of these points, and once in the cycle for each of these polygons.

The vertexes each have x and y coordinates. Since this representation makes vertexes and polygons symmetrical to each other, in the sense that each is a cycle of edges, each polygon also should be permitted to have coordinates, which could be those of its centre of mass, or of its circumcentre.

Design a relational database to capture all these properties. Make up a small example. and use it to populate your relation(s). Indicate functional dependences and keys, and *state your assumptions* explicitly.

19. Design both a nested relation and the corresponding set of flat relations to accommodate bibliographic data such as the following.

    - Alfred North Whitehead (Sc.D., F.R.S.) and Bertrand Russell (M.A., F.R.S.) wrote "Principia Mathematica" in 1910.
    - Panini wrote "Ashtadhyayi" (First book on Sanskrit grammar) in 560 B.C.

    Use separate entries for each part of each name and for each degree or qualification. State any assumptions you make. Show sample data, including the above. Show keys and functional dependences. Comment on any inadequacies in the relational formalism this exercise reveals.

20. The Research Grants subsystem of a university accounting system must keep track of *Grantee*s, their *Address*es, granting *Agency*(ie)s and their *Address*es, receipts and expenses for each *Account*, including *Date*, *Amount*, *Categ*ory (e.g., personnel, equipment, supplies), and details of the payee for expenses and of the source for receipts.

    Keeping redundancy and null values to a minimum, and inventing additional attributes as necessary, design relation(s) with attributes including all the above italicised words for a database which will permit monthly reporting to grantees of all details above, including year-to-date, and annual reporting to granting agencies of summaries by category.

    State your assumptions explicitly, using keys and functional dependences, and illustrated with as much, but no more, sample data as necessary.

21. A family tree database must store data on people and their relationships: name, spouse(s), children, date of birth, date(s) of marriage(s) and divorce(s), date of death, etc. Taking into account unorthodoxies, such as adoptions and that children may be born or adopted without a marriage having been formalized, design a flat (first normal form) relational database to represent this data as flexibly as possible. Be explicit about any assumptions. Show dependences and keys. Show sample data.

22. A hypothetical state election system must 1) register up to 3 million voters before the election and 2) coordinate 5000 voting machines on election day.

    The registration process captures name, address, citizenship, age, and a choice from seven political parties plus "other" plus "decline to state".

The voting machines each provide the system with an average of 600 ballots, each with votes for Federal President, Federal Senator, 2 to 5 Federal Representatives, 1 to 4 State Senators, and 1 to 4 State Representatives. Ballots contain no information that might identify the voter, but the voting machines check voter IDs against the registration data and maintain independent statistics on how many eligible voters showed up.

Design a flat (first normal form) relational database to represent this data as flexibly as possible. Be explicit about any assumptions. Show dependences and keys. Show sample data.

The author gratefully acknowledges support from the taxpayers of Québec and of Canada who have paid his salary and research grants while this work was developed at McGill University, and from his students (who built the implementations and investigated the data structures and algorithms) and their funding agencies.