CS 61: Database Systems | **Fall 2022**

PSET 1 — October 2, 2022

Prof. Palmer | ***Student: Amittai Siavava***

Credit Statement

I worked on these problems alone, with reference to class notes and slides used in class.

Problems

1. (10 points) Consider the relation structure shown below:

BUILDING_CODE	ROOM_CODE	TEACHER_LNAME	TEACHER_FNAME	TEACHER_INITIAL	DAYS_TIME
KOM	204E	Williston	Horace	G	MWF 8:00-8:50
KOM	123	Cordoza	Maria	L	MWF 8:00-8:50
LDB	504	Patroski	Donald	J	TTh 1:00-2:15
KOM	34	Hawkins	Anne	W	MWF 10:00-10:50
JKP	225B	Risell	James		TTh 9:00-10:15
LDB	301	Robertson	Jeanette	P	TTh 9:00-10:15
KOM	204E	Cordoza	Maria	I	MWF 9:00-9:50
LDB	504	Williston	Horace	G	TTh 1:00-2:15
KOM	34	Cordoza	Maria	L	MWF 11:00-11:50
LDB	504	Patroski	Donald	J	MWF 2:00-2:50

(a) Identify and discuss the serious data redundancy problems.

The main data redundancy in the `TEACHER_LNAME` and `TEACHER_FNAME` fields:
(a) “Cordoza Maria” occurs 3 times.
(b) “Willstone Horace” occurs 2 times.
(c) “Patroski Donald” occurs 2 times.

This redundancy requires that any update to a row (tuple) containing any of those three be reflected across every other rows containing the same. Otherwise, the data will be inconsistent. Indeed, we see “Cordoza Maria” having two initials: “L” and “I”. It is likely that one initial is outdated, but the change was not broadcast to all rows containing the “Cordoza Maria” combination for `TEACHER_LNAME` and `TEACHER_FNAME`.

The relation structure should be normalized so that each update needs to happen only at a single place.

(b) If this relation is the only one describing the teaching schedule, what problems might arise if the KOM building was condemned and every row referring to that building was deleted?

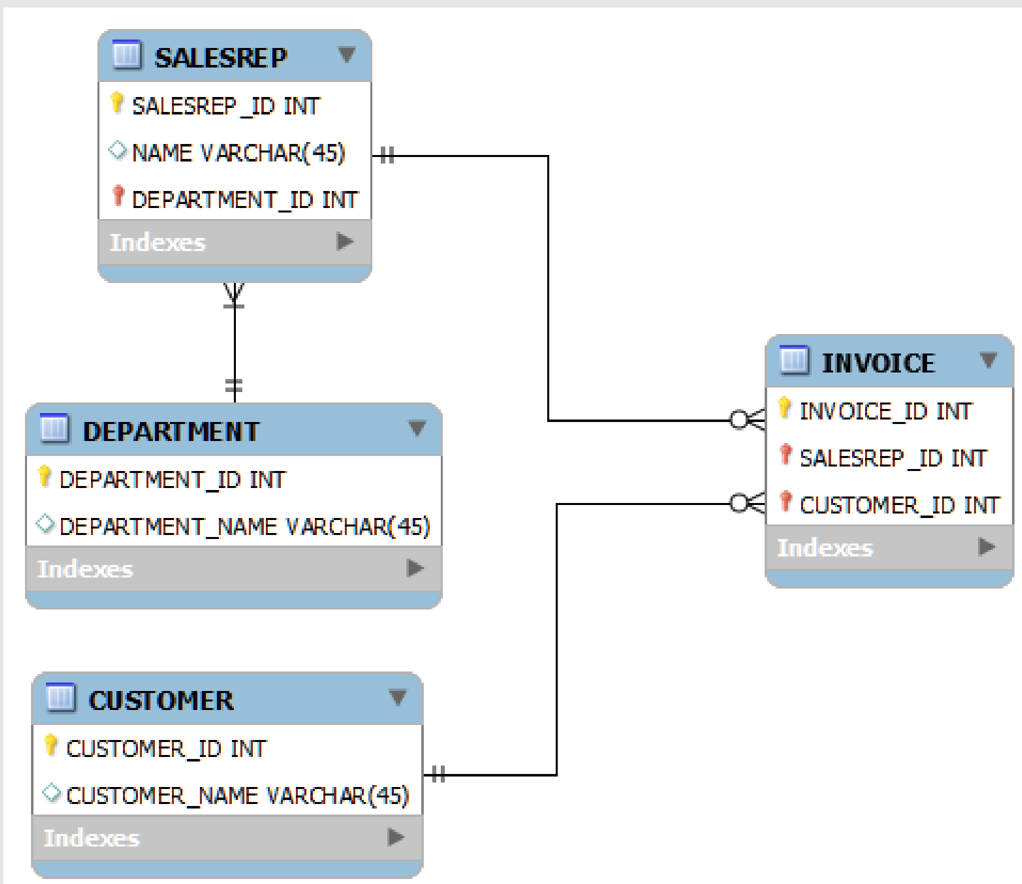
Deleting all rows (tuples) referring to the KOM building would result int he loss of all data for Horace Williston, Cordoza Maria, and Hawkins Anne. Their existence in the system, including details such as their `TEACHER_INITIAL` and `DAYS_TIME` would be lost despite not being dependent on the `BUILDING_CODE`. A better data model would extract those attributes into separate tables to have atomic data.

2. (20 points)
Create a Crow's Foot ERD to include the following business rules for the ProdCo company:

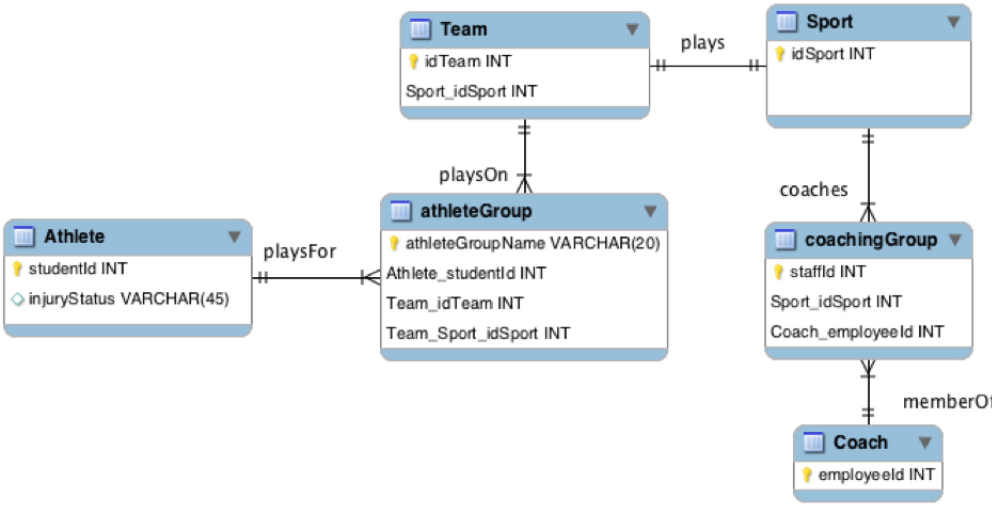
- Each sales representative writes many invoices.
- Each invoice is written by one sales representative.
- Each sales representative is assigned to one department.
- Each department has at least one sales representative.
- Each customer can generate many invoices.
- When a new customer is added that customer may or may not have any invoices.
- Each invoice is generated by one customer.

In MySQL Workbench, I couldn't find an option to specify "one-to-many" without specifying the lower limit of either 0 or 1. So I deduced what further constraints might hold:

- New sales representatives might not have an invoice yet, so I made that relationship zero-or-many on the side of the invoice.
- I deduced that each invoice must have a customer, so I made that relationship one-and-only-one on the side of the customer.



3. (20 points)
Write the business rules that are reflected in the ERD shown in the figure below. *[Remember, business rules generally* ***DO NOT*** *include "helper" relations.]*



(a) Each athlete plays for at least one athlete group.
(b) Each team plays on at least one athlete group.
(c) Each team plays a single sport.
(d) Each sport is coached by at least one coaching group.
(e) Each coach is a member of at least one coaching group.

4. (25 points)
(From Tannenbaum) Given the following relational model:

```
employee (personName, street, city)
works (personName, companyName, salary)
company (companyName, city)
manages (personName, managerName)
```

Give relational algebra expressions for the following queries:

(a) Find the names of all employees who work for "First Bank Corporation".

$$\Pi_{personName}(\sigma_{companyName=''FirstBankCorporation''}(works))$$

(b) Find the names and cities of residence of all employees who work for "First Bank Corporation".

$$\Pi_{personName,\ city}(employee \bowtie (\sigma_{companyName=''FirstBankCorporation''}(works)))$$

(c) Find the names, street addresses, and cities of residence of all employees who work for "First Bank Corporation" and earn more than $10,000.

$$\Pi_{personName,\ street,\ city}\ ($$
$$\sigma_{companyName=''FirstBankCorporation''\ \wedge\ salary\ >\ 10000}\ (employee\ \bowtie\ works)$$
$$)$$

(d) Find the names of all employees in this database who live in the same city as the company for which they work.

$$\Pi_{personName}\ (\sigma_{MM}(employee\ \bowtie\ works)\ \bowtie_{MM.city=company.city}\ company)$$

(e) Assume the companies may be located in several cities. Find all companies located in every city in which "Small Bank Corporation" is located.

$$\Pi_{companyName}\ (company \div \Pi_{city}\ (\sigma_{companyName\ =\ ''SmallBankCorporation''}\ (company)))$$

5. (25 points)
Consider the two tables $T1$ and $T2$:

T1:

a	b	c
5	Q	3
10	R	7
15	Q	11

T2:

x	y	z
25	R	13
10	R	5
10	S	11

Show the results of the following relational algebra expressions:

(a) $T1 \bowtie_{T1.a=T2.x} T2$

a	b	c	x	y	z
10	R	7	10	R	5
10	R	7	10	S	11

(b) $\Pi_z(T2) - \Pi_{T2.z}(\sigma_{T2.z>T2B.z}(T2 \times \sigma_{T2B}(T2)))$

z
5

(c) $T1 \cup T2$

a	b	c
5	Q	3
10	R	7
15	Q	11
25	R	13
10	R	5
10	S	11

(d) $T1 \times T2$

a	b	c	x	y	z
5	Q	3	25	R	13
5	Q	3	10	R	5
5	Q	3	10	S	11
10	R	7	25	R	13
10	R	7	10	R	5
10	R	7	10	S	11
15	Q	11	25	R	13
15	Q	11	10	R	5
15	Q	11	10	S	11

(e) $T1 \bowtie_{T1.a<T2.x} T2$

a	b	c	x	y	z
5	Q	3	25	R	13
5	Q	3	10	R	5
5	Q	3	10	S	11
10	R	7	25	R	13
15	Q	11	25	R	13

(f) $\Pi_{T1.a,T1.c,T2.x}(\sigma_{T1.a=T2.x \wedge (T1.a>7 \vee T1.c<7)} (T1 \times T2))$

a	c	x
10	7	10
10	7	10