COMS 363 Fall 2022

Practice Problems for Exam 1
Topic: Relational Data Model

Suppose our client tells us real-world constraints that can be formally captured by the following functional dependencies.

{pid}→{pname}
{sid}→{sname}
{sid}→{address}
{sid,pid}→{cost}

Suppose three design choices are given to you.

Design A has three tables: parts, suppliers, catalog

parts(pid int, pname varchar(30), PRIMARY KEY (pid));

suppliers (sid int, sname varchar(30), address varchar(50), PRIMARY KEY (sid));

sid=1, sname="Pak"

catalog (sid int, pid int, cost decimal(10,0), PRIMARY KEY (sid,pid), FOREIGN KEY (pid) REFERENCES parts (pid), FOREIGN KEY (sid) REFERENCES suppliers(sid));

Design B has two tables: suppliers and parts2.

suppliers (sid int, sname varchar(30), address varchar(50), PRIMARY KEY (sid));

parts2 (pid int, pname varchar(30), sid **int not null**, cost decimal(10,0), PRIMARY KEY (pid), foreign key(sid) references suppliers(sid));

Design C has one table suppliers_parts.

suppliers_parts (sid int, sname varchar(30), address varchar(50), pid int, pname varchar(30), cost decimal(10,0), primary key(sid, pid))

sid	sname	address	pid	pname	cost
1	Amazon	Ames	202	Oxygen Sensor	50
5	Walmart	Des Moines	102	Wiper Blades	60
5	Walmart	Des Moines	201	Catalytic Converter	45
5	Walmart	Des Moines	202	Oxygen Sensor	65
5	Walmart	Des Moines	301	Ignition Coil	90

Figure 1. Instance of suppliers_parts

1. Can data in all the rows in suppliers_parts in Design C be stored in relations in Design A? If your answer is no, give an example showing the rows in the relation in Fig. 1 that cannot be stored in relations in Design A. If your answer is yes, provide a reason why. The reason must not be specific to the instances in Fig. 1.
Note that we have learned how to store the same data across multiple tables and how to obtain the data from multiple through join operations or outer join operations.

Yes/No (2 points): Yes

Reason (6 points)

In the catalog table of Design A, sid and pid attributes form the primary key. Therefore, a supplier can supply multiple parts and a part can be sold by multiple suppliers. This is same as in Design C. Furthermore, in Design A sid is a foreign key to suppliers(sid) and pid is a foreign key to parts(pid).
Joining the three tables based on the foreign key of the referencing table(s) and the primary key of the referenced table(s) together gives the same information as in the suppliers_parts table. No attributes and values of rows in the suppliers_parts table are missing.

2. Can data in all the rows in Design C be stored in relations in Design B? If your answer is no, give an example showing the rows in the relation in Fig. 1 that cannot be stored in relations in Design B. If your answer is yes, give a reason why. The reason must not be specific to the instance in Fig. 1.

Answer: yes or no (2 points):

No

Reason: (6 points)

In Design B, each part must be supplied by exactly one supplier due to the not null constraint on the sid attribute. Therefore, if the first row in Figure 1 with pid of 202 is already stored in parts2, we cannot store the forth row with pid of 202 in parts2 since pid is the primary key of parts2. No two rows can have the same pid value.

Parts2

202, Oxygen Sensor, 1, 50

Cannot store this row in Parts2.

202, Oxygen Sensor, 5, 65

3. Does a relational DBMS allow insertion of information about a new supplier who has not supplied any part in the suppliers_parts relation in Design C? This new supplier is given a sid value that is not already used in the suppliers_parts relation. Provide the reason to support your answer.

Answer (Yes/No): (2 points)

No

Reason to support your answer (4 points):

Since pid is part of the primary key for suppliers_parts, its value cannot be null.

4. Describe at least three problems caused by redundancy in Design C and give an example to illustrate each problem using the data in Figure 1.

Problem 1: Insertion anomalies: It may not be possible to store some information unless some other information is stored as well. See pages 1-7 of WK3_RelationalDatabaseDesign PowerPoint file. Since sid and pid form the primary key, we cannot insert data about a new supplier such as name and address without knowing the part that this new supplier supplies. Similarly, we cannot insert data about a new part without knowing the supplier that supplies the part.

Example: We cannot insert a new supplier with the sid value of 6 and the sname value of "New" without knowing the pid of the part this supplier supplies.

Problem 2: Deletion anomalies: It may not be possible to delete some information without losing some other information as well. In other words, we lose more data that we intend to. When we remove all the parts a supplier supplies, we also lose the supplier's name and address.

Example: When deleting all the parts that Walmart supplies, we also lose data that sid of 5 is Walmart and the address of Walmart.

Problem 3: Update anomalies: If one copy of such repeated data is updated, an inconsistency is created unless all copies are similarly updated.

Example: Des Moines is repeated four times (row 2 to row 5) for the sid value of 5 (Walmart supplier). Any update to the values of the sname and address attributes with the same sid value needs to be done similarly on all these rows. Otherwise, an inconsistency occurs. For instance, if the sname attribute value of row 2 is updated to "Walmart Co.", the sname values of rows 3-5 must also be updated to the same "Walmart Co."; otherwise, we have an inconsistency about which is the true name of this supplier.