Tristan Shi

CSCI-UA.0060

Professor Prager

23 Feburary 2025

Normalization Assignment #3

1. The following is a handwritten Parts Order Form for a local Supply company. You can assume that employees help any and all different customers, i.e. there is no special relationship between specific employees and customers. A cage code is the identifier of the cages (or shelves) that the inventory is stored in.

Happy Supplies
Parts Warehouse

Customer Name:	*Jeff Peterson*	**Date:**	*7/1/2024*
Customer Number:	*HG54587*	**Time:**	*10:30am*
		Employee:	*D. Harrison*
Customer Type:	*Consumer*		

Part Number	Name	Type	Cage Code	Quantity Ordered	Unit Price
10654	Float Control	Plumbing	G413	4	12
10456	Modulator	Electrical	H433	3	7
10776	Hose Assembly	Plumbing	G413	7	9
10657	Float Assembly	Plumbing	G413	5	10

a. State any assumptions that you make about the data and the attributes shown

i. Only one employee is needed to help each customer

ii. quantity_ordered relates to how many of each item the customer bought, not the inventory the store has ordered

b. Describe and illustrate the process of normalizing the attributes shown in this form.

i. 1NF: This table is in 1NF because there are no repeating values and unique names in every column. The primary key for this table is a composite key

composed of **date, time, customerID, and part_number.** Additionally, for clarity, I would add to all the “part” associated attributes the word “part” as a prefix.

1NF

customer_id	**date**	**time**	**part_numb er**	customer_name	employee	customer_type	part_name	part_type	cage_code	quantity_ordered	unit_price
HG54587	7/1/2024	10:30	10654	Jeff Peterson	D. Harrison	consumer	Float Control	Plumbing	G413	4	12
HG54587	7/1/2024	10:30	10456	Jeff Peterson	D. Harrison	consumer	Modulator	Electrical	H433	3	7
HG54587	7/1/2024	10:30	10776	Jeff Peterson	D. Harrison	consumer	Hose Assembly	Plumbing	G413	7	9
HG54587	7/1/2024	10:30	10657	Jeff Peterson	D. Harrison	consumer	Float Assembly	Plumbing	G413	5	10

ii. 2NF: To be in second normal form, every column must rely on the primary key (no partial dependencies). However, since we have a four-part composite key, there are a lot of partial dependencies. **customer_type** and **customer name** are only dependent on **customer_id (PK),** necessitating a new table just for customer info. **part_name, part_type, cage_code,** and **unit_price** are only dependent on **part_number (PK)** so that data is moved to a separate table. Finally, both **quantity_ordered** and **employee** are directly related to the four-part composite key we defined in **1NF**.

2NF

Orders

customer_id	**date**	**time**	**part_number**	quantity_ordered	employee
HG54587	7/1/2024	10:30	10654	4	D. Harrison
HG54587	7/1/2024	10:30	10456	3	D. Harrison

HG54587	7/1/2024	10:30	10776	7	D. Harrison
HG54587	7/1/2024	10:30	10657	5	D. Harrison

Customer

customer_id	customer_name	customer_type
HG54587	Jeff Peterson	consumer

Part

part_number	part_name	part_type	cage_code	unit_price
10654	Float Control	Plumbing	G413	12
10456	Modulator	Electrical	H433	7
10776	Hose Assembly	Plumbing	G413	9
10657	Float Assembly	Plumbing	G413	10

iii. 3NF: A table is in 3NF if there are no transitive dependencies, i.e. no column that is dependent on something other than the primary key. This table is already in 3NF, as none of the columns in the table are directly related to anything but their primary key. Every single column in every table is directly related only to their primary key.

3NF

Orders

customer_id	**date**	**time**	**part_number**	quantity_ordered	employee
HG54587	7/1/2024	10:30	10654	4	D. Harrison
HG54587	7/1/2024	10:30	10456	3	D. Harrison
HG54587	7/1/2024	10:30	10776	7	D. Harrison
HG54587	7/1/2024	10:30	10657	5	D. Harrison

Customer

customer_id	customer_name	customer_type
HG54587	Jeff Peterson	consumer

Part

part_number	part_name	part_type	cage_code	unit_price
10654	Float Control	Plumbing	G413	12
10456	Modulator	Electrical	H433	7
10776	Hose Assembly	Plumbing	G413	9
10657	Float Assembly	Plumbing	G413	10

2. The data shown below is used by the Panacea Mental Health Corporation to track its therapists. Therapists may work at a number of different branches, but they only see patients at one specific branch on any given day. A patient is given an appointment at a specific time and date at a particular branch with one therapist. Patients may have multiple appointments in any given day and with multiple different therapists.

staffNo	therapistName	patNo	patName	appointment		branchNo
				date	time	
S1011	Fred Smith	P100	Lily White	9/12/2022	10:00	M15
S1011	Fred Smith	P105	Jill Baker	9/12/2022	12:00	M15
S1024	Heidi Pierce	P108	Andy McKee	9/12/2022	10:00	Q10
S1024	Heidi Pierce	P108	Andy McKee	9/14/2022	14:00	Q10
S1032	Richard Levin	P105	Jill Baker	9/14/2022	16:30	M15
S1032	Richard Levin	P110	Jimmy Winter	9/15/2022	18:00	B13

a. State any assumptions you make about the data and the attributes shown in this table.

i. Patients and therapists meet at the same branch

ii. A patient does not always meet with the same staff member (otherwise staff_number might be dependent on patient number).

b. Describe and illustrate the process of normalizing the table to 3NF relations.

i. 1NF: This table does not start in 1NF because the appointment table contains multiple values in the **appointment** column, containing both **date AND time**. I would resolve this by splitting this into two columns, **appointment_date** and **appointment_time**. The primary key for this table is

a composite key of **date, time, patient_number,** and **staff_number**. These four values are enough to define a unique row.

1NF

pat_number	**appointment_date**	**appointment_time**	patient_name	branch_number	therapist_name	**staff_number**
p100	9/12/2022	10:00	Lily White	m15	Fred Smith	s1011
p105	9/12/2022	12:00	Jill Baker	m15	Fred Smith	s1011
p108	9/12/2022	10:00	Andy McKee	q10	Heidi Pierce	s1024
p108	9/14/2022	14:00	Andy McKee	q10	Heidi Pierce	s1024
p105	9/14/2022	16:30	Jill Baker	m15	Richard Levin	s1032
p110	9/15/2022	18:00	Jimmy Winter	b13	Richard Levin	s1032

ii. 2NF: To be in 2NF, there must be no partial dependencies (i.e. one column not depending on the primary key. **patient_name** is only dependent on the **pat_number (PK)**, necessitating the creation of a new table called **patient** with only **patient_name** and **patient_number (PK)**. Additionally, **therapist name** and **branch_number** are only dependent on **staff_number (PK)**, so I would create a separate table **staff_number (PK), branch_number,** and **therapist name.**

2NF

Staff

branch_number	therapist_name	**staff_number**
m15	Fred Smith	s1011
m15	Fred Smith	s1011

q10	Heidi Pierce	s1024
q10	Heidi Pierce	s1024
m15	Richard Levin	s1032
b13	Richard Levin	s1032

Appointment

pat_number	**appointment_date**	**appointment_time**	**staff_number**
p100	9/12/2022	10:00	s1011
p105	9/12/2022	12:00	s1011
p108	9/12/2022	10:00	s1024
p108	9/14/2022	14:00	s1024
p105	9/14/2022	16:30	s1032
p110	9/15/2022	18:00	s1032

Patient

pat_number	patient_name
p100	Lily White
p105	Jill Baker
p108	Andy McKee
p108	Andy McKee
p105	Jill Baker
p110	Jimmy Winter

iii. To be in third normal form, we must make sure that there are no transitive dependencies (i.e. every column only relates to the primary key in each table and not to each other.) In the staff table, both the branch and name of the staff are directly related to the staff number. In the appointment table, none of the values are related to each other and thus have no transitive

dependencies. Finally, for the patient table, the patient's name is directly related to the patient_number. This previous set of tables used for 2NF is also in 3NF.

3NF

Staff

branch_number	therapist_name	**staff_number**
m15	Fred Smith	s1011
m15	Fred Smith	s1011
q10	Heidi Pierce	s1024
q10	Heidi Pierce	s1024
m15	Richard Levin	s1032
b13	Richard Levin	s1032

Appointment

pat_number	**appointment_date**	**appointment_time**	**staff_number**
p100	9/12/2022	10:00	s1011
p105	9/12/2022	12:00	s1011
p108	9/12/2022	10:00	s1024
p108	9/14/2022	14:00	s1024
p105	9/14/2022	16:30	s1032
p110	9/15/2022	18:00	s1032

Patient

pat_number	patient_name
p100	Lily White
p105	Jill Baker

p108	Andy McKee
p108	Andy McKee
p105	Jill Baker
p110	Jimmy Winter

3. The Maid Better temp agency supplies help to Event Management companies within the New York area. Below is the data that the company uses to track employee hours against different contracts. The Employee Number (eNo) is unique for each member of staff. Each contract only applies to one event. There may be different contracts for an event depending upon different service needs.

eNo	contractNo	hours	eName	eventNo	eventLoc
1135	C1024	16	Smith J	H25	Queens
1057	C1024	24	Hocine D	H25	Queens
1068	C1025	28	White T	H4	Yonkers
1135	C1025	15	Smith J	H4	Yonkers
1135	C1026	10	Smith J	H25	Queens

a. **State any assumptions you make about the data and the attributes shown in this table.**
   i. Every maid can only take one contract at a time
   ii. Events do not change locations
b. **Describe and illustrate the process of normalizing the table to 3NF relations.**
   i. 1NF: To convert this table to first normal form, we must first look for duplicate/repeating data values. There are no repeating columns and no repeating rows, as well as only one data value per cell, meaning this table is already in 1nf. The primary key for this table is a composite key of **employee number** and **contract number**, as these two values are enough to uniquely identify every row in this table.

1NF

employee_number	**contract_number**	hours	employee_name	event_number	event_location
1135	C1024	16	Smith J	H25	Queens
1057	C1024	24	Hocine D	H25	Queens
1068	C1025	28	White T	H4	Yonkers
1135	C1025	15	Smith J	H4	Yonkers
1135	C1026	10	Smith J	H25	Queens

ii. 2NF: To convert this dataset into 2NF, we must eliminate partial dependencies, by making sure that each attribute relates only to the primary key. However, **employee_name** relates only to **employee_number (PK)**, while **event_number** and **event_location** only relate to **contract_number (PK)**. Meanwhile, **hours** (worked) relates to both **employee_number (PK)** and **contract_number (PK)**. Therefore, I would add an **employee** table with only **employee_number (PK)** and **employee_name**, a **contract** table with **contract_number (PK)**, **event_number,** and **event_name,** as well as a **work** table with **employee_number (PK), contract_number (PK),** and **hours.**

2NF

work

employee_number	**contract_number**	hours
1135	C1024	16
1057	C1024	24
1068	C1025	28
1135	C1025	15
1135	C1026	10

employee

employee_number	employee_name
1135	Smith J
1057	Hocine D
1068	White T
1135	Smith J
1135	Smith J

contract_info

contract_number	event_number	event_location
C1024	H25	Queens
C1024	H25	Queens
C1025	H4	Yonkers
C1025	H4	Yonkers
C1026	H25	Queens

iii. 3NF: To convert this data into 3NF, we must eliminate transitive dependencies. Looking at the three tables:

1. work: **hours** relates directly to **PK contract_number** and **PK employee_number**
2. employee**: employee_name** relates directly to **PK employee_number**
3. contract_info: **event_number** relates directly to **PK contract_number** but **event_location** is related to **event_number and contract_number**

Since contract_info has a transitive dependency, we must split the contact_info table into **PK contract_number, event_number** table and a **PK event_number, event_location table.**

3NF

work

employee_number	**contract_number**	hours
1135	C1024	16
1057	C1024	24
1068	C1025	28
1135	C1025	15
1135	C1026	10

employee

employee_number	employee_name
1135	Smith J
1057	Hocine D
1068	White T
1135	Smith J
1135	Smith J

contract_info

contract_number	event_number
C1024	H25
C1024	H25
C1025	H4
C1025	H4
C1026	H25

event_info

event_number	event_location
H25	Queens
H25	Queens
H4	Yonkers

H4	Yonkers
H25	Queens