Normalization

Grigorios Loukides
Email: grigorios.loukides@kcl.ac.uk

Review

What contains the OLD.column in a DELETE trigger?

Review

When we cannot modify the NEW value in an UPDATE trigger?

Session Objectives

In this session, you will learn:

- The purpose of normalization.
- The concept of functional dependency, which describes the relationship between attributes.
- How to undertake the process of normalization
- First Normal Form (1NF), Second Normal Form (2NF), and Third Normal Form (3NF).

Introduction

- Normalization is a technique for producing a set of **suitable** relations that support the data requirements of an enterprise:
  - the minimal number of attributes necessary to support the data requirements
  - attributes with a close logical relationship are in the same relation
  - minimal redundancy with each attribute represented only once with the important exception of attributes that form all or part of foreign keys

Purpose of Normalization

Suitable relations:

- Are easier for the user to access and maintain the data
- Take up minimal storage space on the computer

Data Redundancy Problems

Relations that contain redundant information may potentially suffer from update anomalies.

- Insertion
- Deletion
- Modification

Example

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

StaffBranch relation has redundant data:
the details of a branch are repeated for every member of staff

Example: Insert Anomalies

StaffBranch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

- To insert the details of new staff located at branch number *B*007, we must enter the correct details of branch number *B*007 so that the branch details are consistent with values for branch *B*007 in other tuples
- To insert details of a new branch that currently has no members of staff, it is necessary to enter nulls into the attributes for staff, which is not possible due to the integrity constraint

Example: Deletion Anomalies

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

- If we delete a tuple that represents the last member of staff located at a branch, the details about that branch are also lost from the database.

Example: Modification Anomalies

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

- If we want to change the value of one of the attributes of a particular branch we must update the tuples of all staff located at that branch. If this modification is not carried out on all the appropriate tuples, the database will become inconsistent

Example Solution: Normalization

12/56

Introduction
**Purpose of
Normalization**

**Functional
Dependencies**
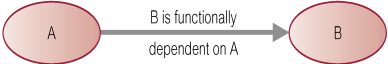Identifying
Functional
Dependencies

Normalization
UNF
1NF
2NF
3NF

**Tutorial
Exercises**

Conclusion
Suggested
Readings

Staff

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005

Branch

branchNo	bAddress
B005	22 Deer Rd, London
B007	16 Argyll St, Aberdeen
B003	163 Main St, Glasgow

Functional Dependency

- Important concept associated with normalization.
- Functional dependency describes relationship between attributes.
- For example, if A and B are attributes of relation R, B is functionally dependent on A (denoted $A \to B$), if each value of A in R is associated with exactly one value of B in R
- Diagrammatic representation:

- The *determinant* of a functional dependency refers to the attribute or group of attributes on the left-hand side of the arrow

Functional Dependency: Example
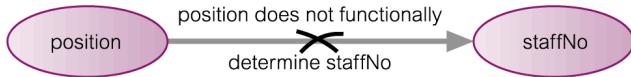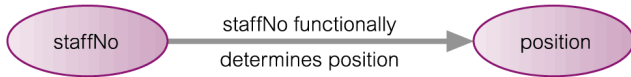
Example

Customer_No	Name	Address	Current_Balance
124	Sally Adams	5 Bertram Road	418.50
260	Mary Nelson	3 London Road	178.70
564	Sally Adams	185 Lark Lane	498.00

- *Name* is functionally dependent on *Customer_No*. We can also say that *Customer_No* functionally determines *Name*:

$$Customer\_No \rightarrow Name$$

- *Address* is NOT functionally dependent on *Name*

Full Functional Dependency

Full Functional Dependency

Full functional dependency indicates that if A and B are attributes of a relation, B is fully functionally dependent on A, if B is functionally dependent on A, but not on any proper subset of A

Partial Dependency

Partial Dependency

Partial dependency indicates that if A and B are attributes of a relation, B is partially dependent on A, if B is functionally dependent on A and on a proper subset of A

Transitive Dependency

Transitive Dependency

Transitive dependency describes a condition where A, B, and C are attributes of a relation such that if $A \rightarrow B$ and $B \rightarrow C$, then C is transitively dependent on A via B (provided that A is not functionally dependent on B or C).

Identifying Functional Dependencies

- Identifying all functional dependencies between a set of attributes is relatively simple if the meaning of each attribute and the relationships between the attributes are well understood

- However, if the users are unavailable for consultation and/or the documentation is incomplete then depending on the database application it may be necessary for the database designer to use their common sense and/or experience to provide the missing information

Example

StaffBranch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

Assume that position held and branch determine a member of staff's salary

Example

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

$staffNo \rightarrow sName, position, salary, branchNo, bAddress$

$branchNo \rightarrow bAddress$

$bAddress \rightarrow branchNo$

$branchNo, position \rightarrow salary$

$bAddress, position \rightarrow salary$

Exercise: Using Sample Data to Determine Functional Dependencies

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

Exercise: Using Sample Data to Determine Functional Dependencies

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

Exercise: Using Sample Data to Determine Functional Dependencies

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

Exercise: Using Sample Data to Determine Functional Dependencies

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

Exercise: Using Sample Data to Determine Functional Dependencies

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

Exercise: Using Sample Data to Determine Functional Dependencies

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

fd1

Exercise: Using Sample Data to Determine Functional Dependencies

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

Exercise: Using Sample Data to Determine Functional Dependencies

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

Exercise: Using Sample Data to Determine Functional Dependencies

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

fd1

fd2

Exercise: Using Sample Data to Determine Functional Dependencies

Introduction
Purpose of
Normalization

Functional
Dependencies
Identifying
Functional
Dependencies

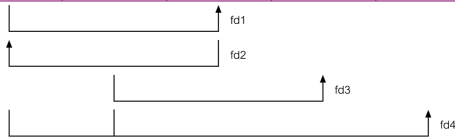Normalization
UNF
1NF
2NF
3NF

Tutorial
Exercises

Conclusion
Suggested
Readings

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

fd1

fd2

fd3

fd4

Exercise: Using Sample Data to Determine Functional Dependencies

Determine all functional dependencies:

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t

$$A \rightarrow C \qquad \text{(fd1)}$$
$$C \rightarrow A \qquad \text{(fd2)}$$
$$B \rightarrow D \qquad \text{(fd3)}$$
$$A, B \rightarrow E \qquad \text{(fd4)}$$

Normalization

- Formal technique for analysing a relation based on its primary key and the functional dependencies between the attributes of that relation.
- Often executed as a series of steps. Each step corresponds to a specific normal form, which has known properties

The Process of Normalization

The Process of Normalization

Data sources

Users

Users' requirements
specification

Forms/reports that are
used or generated by the
enterprise (as described in
this chapter and the next)

Sources describing the
enterprise such as data
dictionary and corporate
data model

Transfer attributes into table format (Described in Section 13.6)

Unnormalized Form (UNF)

Remove repeating groups (Described in Section 13.6)

First Normal Form (1NF)

Remove partial dependencies (Described in Section 13.7)

Second Normal Form (2NF)

Remove transitive dependencies (Described in Section 13.8)

Third Normal Form (3NF)
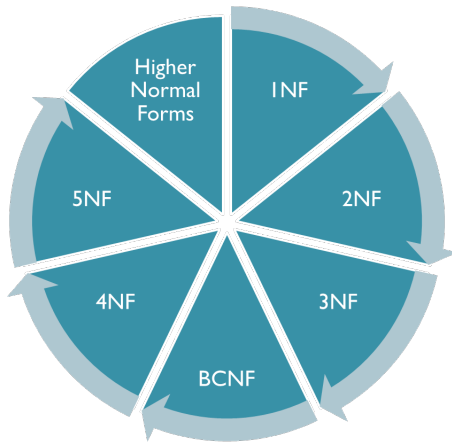
Unnormalized Form (UNF)

Introduction
Purpose of
Normalization

Functional
Dependencies

Identifying
Functional
Dependencies

Normalization
UNF
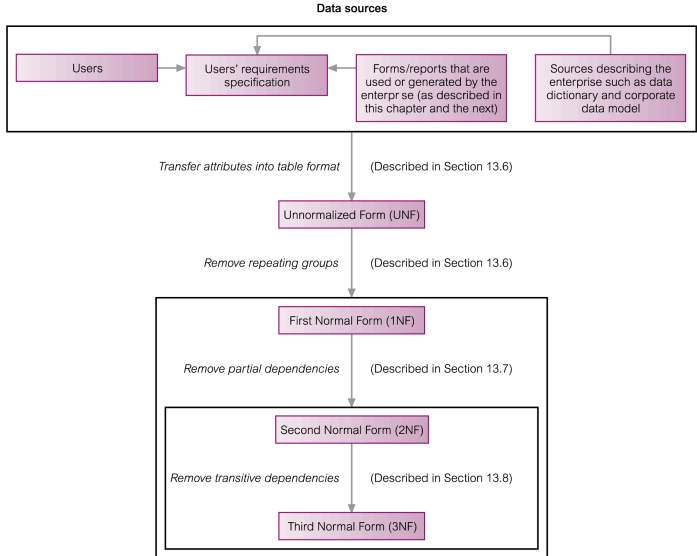1NF
2NF
3NF

Tutorial
Exercises

Conclusion
Suggested
Readings

- A table that contains one or more repeating groups.
- This happens usually when date from a form is transformed into a table

ClientRental

clientNo	cName	propertyNo	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	John Kay	PG4	6 Lawrence St, Glasgow	1-Jul-03	31-Aug-04	350	CO40	Tina Murphy
		PG16	5 Novar Dr, Glasgow	1-Sep-04	1-Sep-05	450	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Glasgow	1-Sep-02	10-June-03	350	CO40	Tina Murphy
		PG36	2 Manor Rd, Glasgow	10-Oct-03	1-Dec-04	375	CO93	Tony Shaw
		PG16	5 Novar Dr, Glasgow	1-Nov-05	10-Aug-06	450	CO93	Tony Shaw

First Normal Form (1NF)

First Normal Form (1NF)

A relation in which the intersection of each row and column
contains one and only one value

UNF to 1NF

Introduction
Purpose of
Normalization

Functional
Dependencies

Identifying
Functional
Dependencies

Normalization
UNF
1NF
2NF
3NF

Tutorial
Exercises

Conclusion
Suggested
Readings

1 Nominate an attribute or group of attributes to act as the key for the unnormalized table.

2 Identify the repeating group(s) in the unnormalized table which repeats for the key attribute(s).

3 Remove the repeating group by

- Entering appropriate data into the empty columns of rows containing the repeating data (flattening the table).
- Or by
- Placing the repeating data along with a copy of the original key attribute(s) into a separate relation.

UNF to 1NF: Flattening Example

ClientRental

clientNo	cName	propertyNo	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	John Kay	PG4	6 Lawrence St, Glasgow	1-Jul-03	31-Aug-04	350	CO40	Tina Murphy
		PG16	5 Novar Dr, Glasgow	1-Sep-04	1-Sep-05	450	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Glasgow	1-Sep-02	10-June-03	350	CO40	Tina Murphy
		PG36	2 Manor Rd, Glasgow	10-Oct-03	1-Dec-04	375	CO93	Tony Shaw
		PG16	5 Novar Dr, Glasgow	1-Nov-05	10-Aug-06	450	CO93	Tony Shaw

↓

ClientRental

clientNo	propertyNo	cName	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-03	31-Aug-04	350	CO40	Tina Murphy
CR76	PG16	John Kay	5 Novar Dr, Glasgow	1-Sep-04	1-Sep-05	450	CO93	Tony Shaw
CR56	PG4	Aline Stewart	6 Lawrence St, Glasgow	1-Sep-02	10-Jun-03	350	CO40	Tina Murphy
CR56	PG36	Aline Stewart	2 Manor Rd, Glasgow	10-Oct-03	1-Dec-04	375	CO93	Tony Shaw
CR56	PG16	Aline Stewart	5 Novar Dr, Glasgow	1-Nov-05	10-Aug-06	450	CO93	Tony Shaw

UNF to 1NF: Example

UNF to 1NF: Exercise

```
HILLTOP ANIMAL HOSPITAL        DATE: JAN 13/2002
INVOICE # 987

MR. RICHARD COOK
123 THIS STREET
MY CITY, ONTARIO
Z5Z 6G6

PET          PROCEDURE              AMOUNT

ROVER        RABIES VACCINATION     30.00
MORRIS       RABIES VACCINATION     24.00

                   TOTAL            54.00
                   TAX (8%)          4.32

             AMOUNT OWING           58.32
```

invoice (invoice_no, invoice_date, cust_name, cust_addr,
(pet_name, procedure, amount))

Second Normal Form (2NF)

Second Normal Form (2NF)

A relation that is in 1NF and every non-primary-key attribute is fully functionally dependent on the primary key

1NF to 2NF

1NF to 2NF

1 Identify the primary key for the 1NF relation.

2 Identify the functional dependencies in the relation.

3 If partial dependencies exist on the primary key remove them by placing then in a new relation along with a copy of their determinant

1NF to 2NF: Example

ClientRental

clientNo	propertyNo	cName	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	PG4	John Kay	6 Lawrence St., Glasgow	1-Jul-03	31-Aug-04	350	CO40	Tina Murphy
CR76	PG16	John Kay	5 Novar Dr, Glasgow	1-Sep-04	1-Sep-05	450	CO93	Tony Shaw
CR56	PG4	Aline Stewart	6 Lawrence St, Glasgow	1-Sep-02	10-Jun-03	350	CO40	Tina Murphy
CR56	PG36	Aline Stewart	2 Manor Rd, Glasgow	10-Oct-03	1-Dec-04	375	CO93	Tony Shaw
CR56	PG16	Aline Stewart	5 Novar Dr, Glasgow	1-Nov-05	10-Aug-06	450	CO93	Tony Shaw

↓

Client

clientNo	cName
CR76	John Kay
CR56	Aline Stewart

Rental

clientNo	propertyNo	rentStart	rentFinish
CR76	PG4	1-Jul-03	31-Aug-04
CR76	PG16	1-Sep-04	1-Sep-05
CR56	PG4	1-Sep-02	10-Jun-03
CR56	PG36	10-Oct-03	1-Dec-04
CR56	PG16	1-Nov-05	10-Aug-06

PropertyOwner

propertyNo	pAddress	rent	ownerNo	oName
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw

1NF to 2NF: Tips

- Look at entities with composite primary keys (formed by more than one attribute)
- Find any attributes which are functionally dependant on only part of the key. Remove these into a separate entity
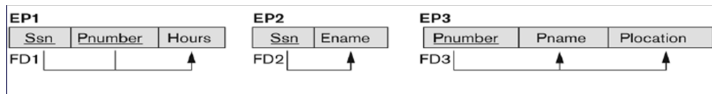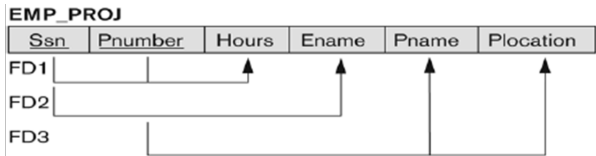
1NF to 2NF: Exercise

HILLTOP ANIMAL HOSPITAL   DATE: JAN 13/2002
INVOICE # 987

MR. RICHARD COOK
123 THIS STREET
MY CITY, ONTARIO
Z5Z 6G6

PET   PROCEDURE   AMOUNT

ROVER   RABIES VACCINATION   30.00
MORRIS   RABIES VACCINATION   24.00

                      TOTAL   54.00
                      TAX (8%)   4.32

             AMOUNT OWING   58.32

invoice(invoice_no, invoice_date, cust_name, cust_addr)
invoice_pet(invoice_no, pet_id, pet_name, procedure, amount)

Third Normal Form (3NF)

Third Normal Form (3NF)

A relation that is in 1NF and 2NF and in which no non-primary-key attribute is transitively dependent on the primary key

2NF to 3NF

1 Identify the primary key in the 2NF relation.

2 Identify functional dependencies in the relation.

3 If transitive dependencies exist on the primary key remove them by placing them in a new relation along with a copy of their determinant.

2NF to 3NF: Example

Client

clientNo	cName
CR76	John Kay
CR56	Aline Stewart

Rental

clientNo	propertyNo	rentStart	rentFinish
CR76	PG4	1-Jul-03	31-Aug-04
CR76	PG16	1-Sep-04	1-Sep-05
CR56	PG4	1-Sep-02	10-Jun-03
CR56	PG36	10-Oct-03	1-Dec-04
CR56	PG16	1-Nov-05	10-Aug-06

PropertyOwner

propertyNo	pAddress	rent	ownerNo	oName
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw

↓

PropertyForRent

propertyNo	pAddress	rent	ownerNo
PG4	6 Lawrence St, Glasgow	350	CO40
PG16	5 Novar Dr, Glasgow	450	CO93
PG36	2 Manor Rd, Glasgow	375	CO93

Owner

ownerNo	oName
CO40	Tina Murphy
CO93	Tony Shaw

1NF to 2NF: Exercise

HILLTOP ANIMAL HOSPITAL   DATE: JAN 13/2002
INVOICE # 987

MR. RICHARD COOK
123 THIS STREET
MY CITY, ONTARIO
Z5Z 6G6

PET	PROCEDURE	AMOUNT
ROVER	RABIES VACCINATION	30.00
MORRIS	RABIES VACCINATION	24.00
	TOTAL	54.00
	TAX (8%)	4.32
	AMOUNT OWING	58.32

invoice(invoice_no, invoice_date, cust_name, cust_addr)
invoice_pet(invoice_no, pet_id, procedure, amount)
pet(pet_id, pet_name)

Exercise 1

Determine Functional Dependencies:

staffNo	dentistName	patNo	patName	appointment date time	surgeryNo
S1011	Tony Smith	P100	Gillian White	12–Sep–04 10.00	S15
S1011	Tony Smith	P105	Jill Bell	12–Sep–04 12.00	S15
S1024	Helen Pearson	P108	Ian MacKay	12–Sep–04 10.00	S10
S1024	Helen Pearson	P108	Ian MacKay	14–Sep–04 14.00	S10
S1032	Robin Plevin	P105	Jill Bell	14–Sep–04 16.30	S15
S1032	Robin Plevin	P110	John Walker	15–Sep–04 18.00	S13

Exercise 2

Normalise the following relation that is in 1NF:
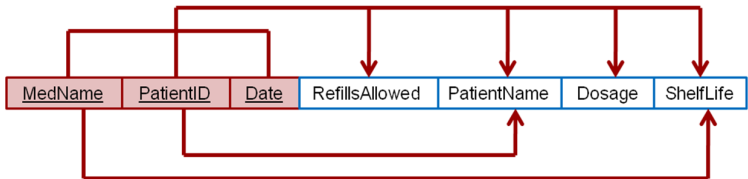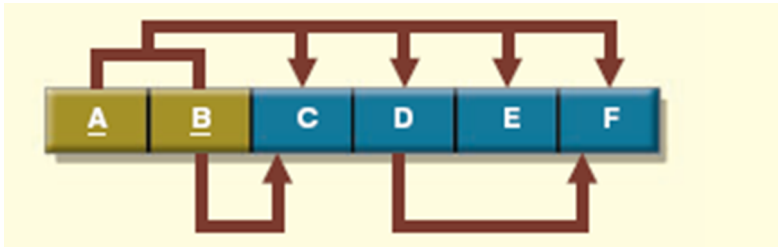
| MedName | PatientID | Date | RefillsAllowed | PatientName | Dosage | ShelfLife |

Exercise 3

Normalise the following relation into 3NF

Exercise 4

Normalise the following relation into 3NF:

ATTRIBUTE NAME	SAMPLE VALUE	SAMPLE VALUE	SAMPLE VALUE	SAMPLE VALUE	SAMPLE VALUE
INV_NUM	211347	211347	211347	211348	211349
PROD_NUM	AA-E3422QW	QD-300932X	RU-995748G	AA-E3422QW	GH-778345P
SALE_DATE	15-Jan-2008	15-Jan-2008	15-Jan-2008	15-Jan-2008	16-Jan-2008
PROD_LABEL	Rotary sander	0.25-in. drill bit	Band saw	Rotary sander	Power drill
VEND_CODE	211	211	309	211	157
VEND_NAME	NeverFail, Inc.	NeverFail, Inc.	BeGood, Inc.	NeverFail, Inc.	ToughGo, Inc.
QUANT_SOLD	1	8	1	2	1
PROD_PRICE	$49.95	$3.45	$39.99	$49.95	$87.75

(**INV_NUM**,PROD_NUM,SALE_DATE,PROD_LABEL,VEND_COD
E,VEND_NAME,QUANT_SOLD,PROD_PRICE)

Conclusion

In this session we have covered:

- Functional Dependencies
- Normalization
    - 1NF
    - 2NF
    - 3NF

Suggested Readings

- Chapter 15 of Fundamentals of Database Systems. Elmasri & Navathe.
- Chapter 13 of Database systems: a practical approach to design, implementation, and management. Connolly, Thomas M; Begg, Carolyn