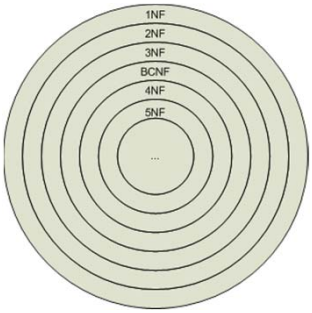
IT2201 / IT2601 / IT2564 / IT2621 / IT2521 / IT2323

Database Management Systems



Unit 6
Normalization

1

Unit Objectives

- At the end of this topic, you should be able to
  - Understand why and when to normalize tables.
  - Understand why redundant data can cause update anormalies.
  - Define functional dependency and apply it in normalization.
  - Perform normalization to place tables in third normal form (3NF).

Normalization ELO

2

Introduction to Normalization

- When to Normalize
  - If not sure ER model is correct
    - Used to validate the design of the tables resulting from the ER modeling.
  - If there has been no time to build an ER model and a set of tables is already available.
    - Used to validate the design of the existing tables.
  - If the application is simple
    - Used as a bottom-up approach to design the database for the application.

3

Introduction to Normalization

- What to Normalize
  - The major subject of normalization is tables, not entities.
- Objective of Normalization
  - To remove **redundant data** from tables in order to increase the integrity of the database design and to maximize flexibility of data storage.

4

Redundant Data

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

Figure 1- Table with redundant data

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

Figure 2 – Tables without redundant data

5

Update Anomalies

- Redundant data can cause problems called *update anomalies*, which are classified as
  - Insertion anomalies
  - Deletion anomalies
  - Modification anomalies
- Normalization and Update Anomalies
  - Normalization aims to remove redundant data, which helps to eliminate update anomalies.

6

Insertion anomalies

- **Adding a new staff member forces user to create branch data.**
  - è Eg Add a new staff member SG14
- **Insertion of a new branch is not allowed if there is no staff member yet.**
  - è Eg Add a new branch B008

Question: Can we set the staff information to NULL ?

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

7

Deletion anomaly

- Deleting rows may cause a loss of data that would be needed for other future rows.
  - Eg Deletion of staff member SA9 will result in branch details of B007 being lost.

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

8

Modification Anomaly

- Changing data in a row forces changes to other rows because of duplication

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

Need to update 3 tuples

9

Normalization

- Normalization vs. Normalized Data
  - Normalization is the activity, the process that leads to a normalized data structure to achieve normalized data.
  - Normalized data is data that contains no redundancies.
- Normal Forms
  - Normalization consists of **a series of rules** that must be **applied in steps** to reach a "higher" level of normalization. These levels are called normal forms.

10

Normalization

- There are many normal forms:
  - 1NF -> 2NF -> 3NF -> BCNF -> 4NF -> 5NF
- In general, the IT industry considers normalization to the 3NF an acceptable level to remove redundancy.
- Normal forms higher than the 3NF deal with more subtle anomalies.

11

Functional Dependency

- Functional dependency describes the relationship between columns in a table.
- Consider the following table:

Student_no	Student_name	Student_address
023344C	Kenny Ang	28 Woodland Rd
023817F	Kenny Ang	211 Jurong Rd
028893E	Jenny Huang	11 Ang Mo Kio Ave 5

- student_no → student_name, student_address
  - Given a student number, you can only find a value for student name and a value for student address.
  - Student_name and student_address are functionally dependent on student_number.
- student_name → student_no True/False

12

Types of Functional Dependency

- Full dependency (Full functional dependency)
  - Every non-key column is functionally dependent on **all parts** of the primary key.
  - Example, PartNo, SuppNo → Cost

Part_Supplier

PartNo	SuppNo	Cost
P1	S1	200
P1	S2	220
P2	S1	300
P2	S2	350

13

Types of Functional Dependency

- Partial dependency
  - There is non-key column that depends on **some part** of the primary key.
  - Example, SuppNo → Sname, SAddr

Part_Supplier

PartNo	SuppNo	SName	SAddr	Cost
P1	S1	Bright	Jurong	200
P1	S2	SuperGood	Senkang	220
P2	S1	Bright	Jurong	300
P2	S2	SuperGood	Senkang	350

14

Types of Functional Dependency

- Transitive dependency
  - There is non-key column that depends on another **non-key column**.
  - Example

    Part

partNo	partDesc	storeman	SContact
P1	Printer	Ken	98553344
P2	Speaker	Ken	98553344
P3	Ribbon	Alvin	97334400
P4	Cartridge	Alvin	97334400

15

Repeating Group

- A repeating group is an attribute, or group of attributes, that occurs with multiple values for a single occurrence of another attribute(s).

Part_Inventory

part No	pDesc	suppNo	sName	sAddr	cost
P1	**Printer**	S1 S2	Bright SuperGood	Jurong Senkang	200 220
P2	**Speaker**	S1 S2	Bright SuperGood	Jurong Senkang	300 350

Multiple occurrences of vendors for each single occurrence of part.

Repeating group =(suppNo, sName, sAddr, cost)

16

Exercise 1

- Is custName functionally dependent on State?
- Is custName functionally dependent on custNo?

Cust No	custName	company	address	city	state	Zip Code
101	Ludwig Pauli	All Sports Supplies	213 Erstwild Court	Sunnyvale	CA	94086
102	Carole Sadler	Sports Spot	785 Geary St	San Francisco	CA	94117
103	Philip Currie	Phil's Sports	654 Poplar	Palo Alto	NJ	94303
104	Anthony Higgins	Play Ball!	East Shopping Cntr.	Redwood City	CA	94026
105	Raymond Vector	Los Altos Sports	1899 La Loma Drive	Los Altos	NJ	94022

17

Normalization Rules

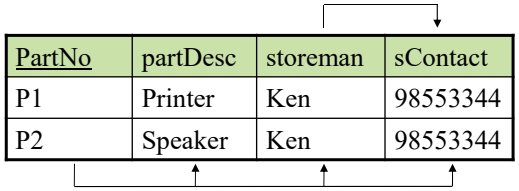
Normal Form Rule	Description
First Normal Form (1NF)	The table must express a set of unordered, two-dimensional tables. The table cannot contain repeating groups.

PartNo	suppNo	sName	sAddr	Cost
P1	V1	Bright	Jurong	200
P1	V2	SuperGood	Senkang	220

18

Normalization Rules

Normal Form Rule	Description
Second Normal Form (2NF)	The table must be in 1NF. Every non-key column must be functionally dependent on all parts of the primary key.

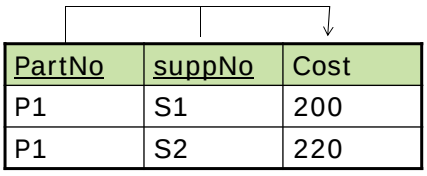


PartNo	partDesc	storeman	sContact
P1	Printer	Ken	98553344
P2	Speaker	Ken	98553344

19

Normalization Rules

Normal Form Rule	Description
Third Normal Form (3NF)	The table must be in 2NF. No non-key column may be functionally dependent on another non-key column.

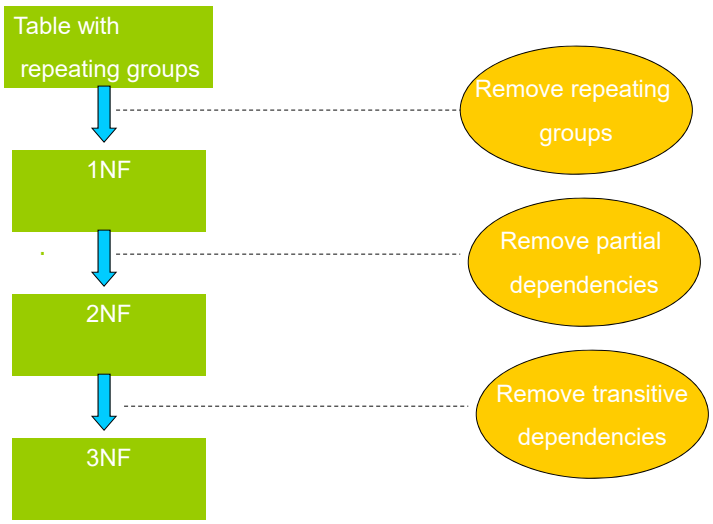


PartNo	suppNo	Cost
P1	S1	200
P1	S2	220

"Each non-primary key value MUST be dependent on the key, the whole key, and nothing but the key."

20

Normalization Process



21

Normalization Process

- Question:
  - How to remove repeating groups, partial dependencies and transitive dependencies?
- Answer:
  - Remove them by creating a new table

22

Exercise 2

- Identify functional dependencies in the Orders table.
- What normal form is this table in?
- Normalize this table to third normal form (3NF). Identify the primary and foreign keys in your 3NF relations.

Order Num	orderDate	Part Num	Desc	Num Ordered	Quoted Price
21608	20/10/04	AT94	Iron	11	21.95
21610	20/10/04	DR93	Gas Range	1	495.00
21610	20/10/04	DW11	Washer	1	399.99
21613	21/10/04	KL62	Dryer	4	329.95
21614	21/10/04	KT03	Dishwasher	2	595
21617	23/10/04	BV06	Home gym	2	794.95
21617	23/10/04	CD52	Oven	4	150

23

Exercise 2 Answers

Orders PD 1 PD 2

Order Num	orderDate	Part Num	Desc	Num Ordered	Quoted Price
21608	20/10/04	AT94	Iron	11	21.95
21610	20/10/04	DR93	Gas Range	1	495.00
21610	20/10/04	DW11	Washer	1	399.99
21613	21/10/04	KL62	Dryer	4	329.95
21614	21/10/04	KT03	Dishwasher	2	595
21617	23/10/04	BV06	Home gym	2	794.95
21617	23/10/04	CD52	Oven	4	150

- Identify functional dependencies in the Orders table.

24

Exercise 2 Answers

Orders PD 1 PD 2

Order Num	orderDate	Part Num	Desc	Num Ordered	Quoted Price
21608	20/10/04	AT94	Iron	11	21.95
21610	20/10/04	DR93	Gas Range	1	495.00
21610	20/10/04	DW11	Washer	1	399.99
21613	21/10/04	KL62	Dryer	4	329.95
21614	21/10/04	KT03	Dishwasher	2	595
21617	23/10/04	BV06	Home gym	2	794.95
21617	23/10/04	CD52	Oven	4	150

- What normal form is this table in?

25

Normalize this table to third normal form (3NF). Identify the primary and foreign keys in your 3NF relations.

Orders PD 1 PD 2

Order Num	Order Date	Part Num	Desc	Num Ordered	Quoted Price
21608	20/10/04	AT94	Iron	11	21.95
21610	20/10/04	DW11	Washer	1	399.99
21613	21/10/04	KL62	Dryer	4	329.95
21617	23/10/04	BV06	Home gym	2	794.95
21617	23/10/04	CD52	Oven	4	150

- **1NF →2NF (Remove partial dependency)**
- **2NF → 3NF (Remove transitive dependency)**

26

Summary

- Normalization is a technique for producing a set of tables with desirable properties that supports the requirements of a user or company.
- Tables that have redundant data may have problems called update anomalies, which are classified as insertion, deletion, or modification anomalies.
- Normalization aims to remove redundant data in order to eliminate update anomalies.
- Use normalization techniques to design simple database or to validate the design of tables.

27

Reference Materials

1. Database Systems, Connolly, Ch 14

2. Modern Database Management, Hoffer, Ch 5

28