# Chapter 5: Normalization

Objectives:

1. **Able to define the normalization and its importance.**

2. **Able to identify the insertion, deletion and modification anomalies (problems) in the unnormalized table.**

3. **Able to normalize the source document (form) to few normalized tables that store in database.**
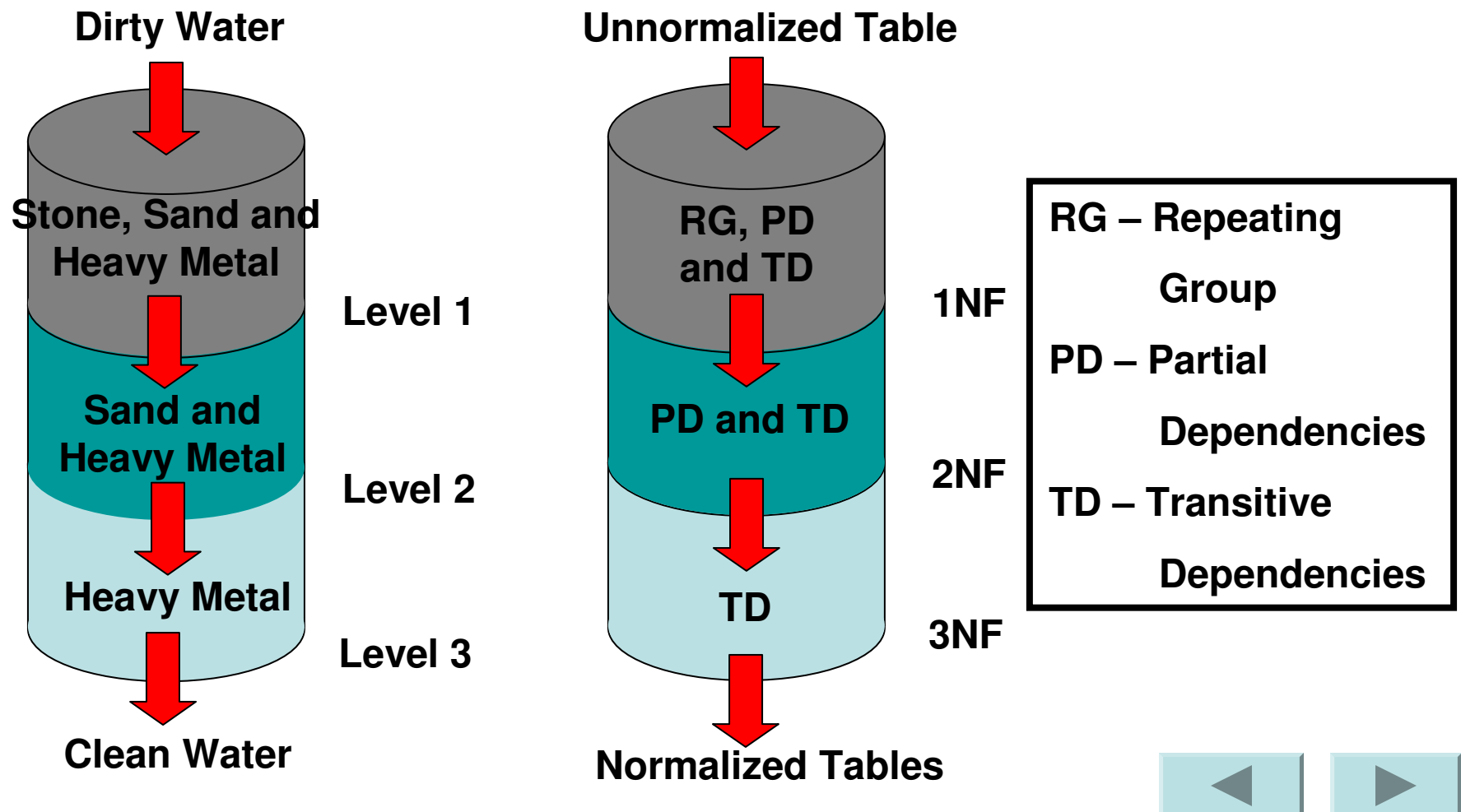
Outlines:

1. **Definition of normalization.**

2. **The five importance of normalization:**
   - **Reduce the data redundancy.**
   - **Reduce the storage space.**
   - **Eliminate the insertion anomaly.**
   - **Eliminate the deletion anomaly.**
   - **Eliminate the modification anomaly.**

3. **The stages of normalization:**
   - **Convert source document to unnormalized table.**
   - **First Normal Form (1 NF)**
   - **Second Normal Form (2 NF)**
   - **Third Normal Form (3 NF)**
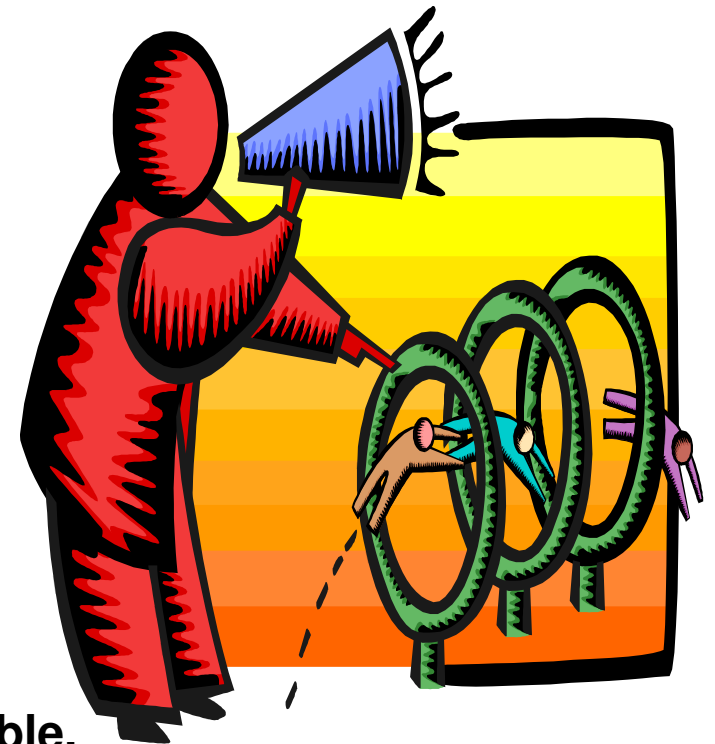
1. Definition of normalization

- **Normalization is a technique for producing a set of tables with desirable properties, given the data requirements of an enterprise.**

Dirty Water	Unnormalized Table	
Stone, Sand and Heavy Metal — Level 1	RG, PD and TD — 1NF	RG – Repeating Group
Sand and Heavy Metal — Level 2	PD and TD — 2NF	PD – Partial Dependencies
Heavy Metal — Level 3	TD — 3NF	TD – Transitive Dependencies
Clean Water	Normalized Tables	

- **First Normal Form (1 NF) – remove repeating group**
  # Select Primary Key (PK)
  # Break into non-repeating group and repeating group tables. (1:M relationship)

- **Second Normal Form (2 NF) – remove partial dependency (PD)**
  # Determinant of PD based on part of PK.

- **Third Normal Form (3 NF) – remove transitive dependency (TD)**
  # Determinant of TD not based on PK.

2. The five importance of normalization:

- **Reduce the data redundancy.**
- **Reduce the storage space.**
- **Eliminate the insertion anomaly.**
- **Eliminate the deletion anomaly.**
- **Eliminate the modification anomaly.**

Before Normalization:

DeliveryOrder

DO_No	DelDate	CustNo	CustName	CustPhone	ItemNo	Desc	UnitPrice	Qty
DO0001	23/8/05	C0002	Eric	03-61381111	I0007	Pencil	3.00	200
DO0001	23/8/05	C0002	Eric	03-61381111	I0010	Paper	8.00	100
DO0002	23/8/05	C0003	John	03-41490888	I0003	Gum	1.00	350
DO0002	23/8/05	C0003	John	03-41490888	I0007	Pencil	3.00	50
DO0002	23/8/05	C0003	John	03-41490888	I0010	Paper	8.00	500

After Normalization:

DeliveryItem

DO_No	ItemNo	Qty
DO0001	I0007	200
DO0001	I0010	100
DO0002	I0003	350
DO0002	I0007	50
DO0002	I0010	500

Customer

CustNo	CustName	CustPhone
C0002	Eric	03-61381111
C0003	John	03-41490888

DeliveryOrder

DO_No	DelDate	CustNo
DO0001	23/8/05	C0003
DO0002	23/8/05	C0002

Item

ItemNo	Desc	UnitPrice
I0003	Gum	1.00
I0007	Pencil	3.00
I0010	Paper	8.00

3. The stages of normalization:

- **Stage 1: Convert source document to unnormalized table.**

Speed Delivery: Delivery Order Form

DO_No: DO0001
DelDate: 23/8/05
CustNo: C0002
CustName: Eric
CustPhone: 03-61381111

ItemNo	Desc	UnitPrice	Qty
I0007	Pencil	3.00	200
I0010	Paper	8.00	100

Speed Delivery: Delivery Order Form

DO_No: DO0002
DelDate: 23/8/05
CustNo: C0003
CustName: John
CustPhone: 03-41490888

ItemNo	Desc	UnitPrice	Qty
I0003	Gum	1.00	350
I0007	Pencil	3.00	50
I0010	Paper	8.00	500

Nonrepeating Group (1) — left portion
Repeating Group (M) — right portion

DeliveryOrder

DO_No	DelDate	CustNo	CustName	CustPhone	ItemNo	Desc	UnitPrice	Qty
DO0001	23/8/05	C0002	Eric	03-61381111	I0007	Pencil	3.00	200
					I0010	Paper	8.00	100
DO0002	23/8/05	C0003	John	03-4149888	I0003	Gum	1.00	350
					I0007	Pencil	3.00	50
					I0010	Paper	8.00	500

Nonrepeating Group (1)

Repeating Group (M)

- **Stage 2: Unnormalized table to First Normal Form (1NF) Table**

DeliveryOrder

DO_No	DelDate	CustNo	CustName	CustPhone	ItemNo	Desc	UnitPrice	Qty
DO0001	23/8/05	C0002	Eric	03-61381111	I0007	Pencil	3.00	200
					I0010	Paper	8.00	100
DO0002	23/8/05	C0003	John	03-4149888	I0003	Gum	1.00	350
					I0007	Pencil	3.00	50
					I0010	Paper	8.00	500

3

Nonrepeating Group (1)   **Repeating Group (M)**

DeliveryOrder

DO_No	DelDate	CustNo	CustName	CustPhone	ItemNo	Desc	UnitPrice	Qty
DO0001	23/8/05	C0002	Eric	03-61381111	I0007	Pencil	3.00	200
DO0001	23/8/05	C0002	Eric	03-61381111	I0010	Paper	8.00	100
DO0002	23/8/05	C0003	John	03-4149888	I0003	Gum	1.00	350
DO0002	23/8/05	C0003	John	03-4149888	I0007	Pencil	3.00	50
DO0002	23/8/05	C0003	John	03-4149888	I0010	Paper	8.00	500

1

Priorities to select the unique candidate key (composite key):

2 Codes → 1 Code + 1 Date → 2 Codes + 1 Date → 1 Code + 1 Date + 1 Time

Priorities to select the unique candidate key (composite key):
2 Codes → 1 Code + 1 Date → 2 Codes + 1 Date / 3 Codes → 1 Code + 1 Date + 1 Time

Nonrepeating Group (1)

Repeating Group (M)

Appointment

DocID	DocName	PatID	PatName	AppDate	AppTime	SurgeryRoomNo
D001	Eric	P0003	John	23/8/05	10:00-12:00	SR101
D001	Eric	P0003	John	23/8/05	14:00-16:00	SR303
D001	Eric	P0004	May	24/8/05	10:00-12:00	SR205
D001	Eric	P0003	John	25/8/05	10:00-12:00	SR101
D002	Jessie	P0004	May	26/8/05	08:00-10:00	SR402
D002	Jessie	P0002	Lily	26/8/05	10:00-12:00	SR402

2 Codes ✗

1 Code + 1 Date ✗

2 Codes + 1 Date ✗ 3 Codes (Nested 1:M relationship)

1 Code + 1 Date + 1 Time ✓

Example of Nested 1:M relationship:
1 sale have many furniture and 1 furniture have many component.
1 sale have many figure and 1 figure have many part.

INF (Remove Repeating Group)

DeliveryOrder (DO_No, DelDate, CustNo, CustName, CustPhone, ItemNo, Desc, UnitPrice, Qty)

DeliveryOrder (DO_No, DelDate, CustNo, CustName, CustPhone)

DeliveryItem (DO_No*, ItemNo, Desc, UnitPrice, Qty)

For Repeating Group, MUST be joined key.

@DBDL – underline PK (Primary key) and put * for FK (foreign key)

- **Stage 3: First Normal Form (1NF) Tables to Second Normal Form (2NF) Tables**

DO_No	ItemNo	Desc	UnitPrice	Qty
DO0001	I0007	Pencil	3.00	200
DO0001	I0010	Paper	8.00	100
DO0002	I0003	Gum	1.00	350
DO0002	I0007	Pencil	3.00	50
DO0002	I0010	Paper	8.00	500

2NF (Remove Partial Dependency)

DeliveryOrder (DO_No, DelDate, CustNo, CustName, CustPhone)

DeliveryItem (DO_No*, ItemNo*, Qty)

Item (ItemNo, Desc, UnitPrice)

ItemNo → Desc, UnitPrice   --- Depend Single Key

DO_No, ItemNo → Qty   --- Depend Composite Key

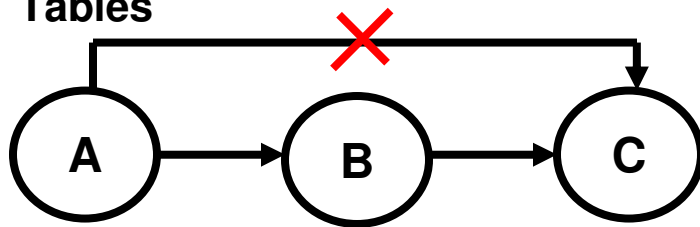@DBDL – underline PK (Primary key) and put * for FK (foreign key)

2NF (Remove Partial Dependency)

DeliveryOrder (DO_No, DelDate, CustNo, CustName, CustPhone)

DeliveryItem (DO_No, ItemNo, Qty)

Item (ItemNo, Desc, UnitPrice)

- **Stage 4: Second Normal Form (2NF) Tables to Third Normal Form (3NF) Tables**

- A is primary key (underline key).
- B normally is code or name.
- A and B are two different objects (persons, events, things, etc).
- B and C must logical related but cannot relate to A.
- Extra item(s) (not belong to A, B or C) put between A to B.

C

A   **B**

DeliveryOrder (DO_No, DelDate, CustNo, CustName, CustPhone)

3NF (Remove Transitive Dependency)

DeliveryOrder (DO_No, DelDate, CustNo*)   A→ B

A   B

C

B

Customer (CustNo, CustName, CustPhone)   B → C

Do_No → DelDate, CustNo
CustNo → CustName, CustPhone

Any remaining column put back A to B

B to C is transitive dependency

DeliveryItem (DO_No*, ItemNo*, Qty)

Item (ItemNo, Desc, UnitPrice)

@DBDL – underline PK (Primary key) and put * for FK (foreign key)

Model Question and Answer:

A1) Normalize the following DeliveryOrder table into 1NF, 2NF and 3NF.

DeliveryOrder

DO_No	DelDate	CustNo	CustName	CustPhone	ItemNo	Desc	UnitPrice	Qty
DO0001	23/8/05	C0002	Eric	03-61381111	I0007	Pencil	3.00	200
DO0001	23/8/05	C0002	Eric	03-61381111	I0010	Paper	8.00	100
DO0002	23/8/05	C0003	John	03-41490888	I0003	Gum	1.00	350
DO0002	23/8/05	C0003	John	03-41490888	I0007	Pencil	3.00	50
DO0002	23/8/05	C0003	John	03-41490888	I0010	Paper	8.00	500

INF (Remove Repeating Group)

DeliveryOrder (DO_No, DelDate, CustNo, CustName, CustPhone, ItemNo, Desc, UnitPrice, Qty)

DeliveryOrder (DO_No, DelDate, CustNo, CustName, CustPhone)

DeliveryItem (DO_No*, ItemNo, Desc, UnitPrice, Qty)

@DBDL – underline PK (Primary key) and put * for FK (foreign key)

2NF (Remove Partial Dependencies)

DeliveryOrder (<u>DO_No</u>, DelDate, CustNo, CustName, CustPhone)

DeliveryItem (<u>DO_No</u>*, <u>ItemNo</u>*, Qty)

Item (<u>ItemNo</u>, Desc, UnitPrice)

3NF (Remove Transitive Dependencies)

DeliveryOrder (<u>DO_No</u>, DelDate, CustNo*)

Customer (<u>CustNo</u>, CustName, CustPhone)

DeliveryItem (<u>DO_No</u>*, <u>ItemNo</u>*, Qty)

Item (<u>ItemNo</u>, Desc, UnitPrice)

@DBDL – underline PK (Primary key) and put * for FK (foreign key)

Model Question and Answer:

B1) Normalize the following StudentGrade table into 1NF, 2NF and 3NF.

StudentGrade

StuNo	StuName	StuPhone	SubjectNo	SubDesc	LecID	LecLocation	Grade
S0001	Eric	03-61381111	S007	Database	1234	Block B	A
S0001	Eric	03-61381111	S010	Accounting	3344	Block A	C
S0002	John	03-41490888	S003	Management	5666	Block C	A
S0002	John	03-41490888	S007	Database	1234	Block B	C
S0002	John	03-41490888	S010	Accounting	3344	Block A	B

INF (Remove Repeating Group)

StudentGrade(StuNo, StuName, StuPhone, SubjectNo, SubDesc, LecID, LecLocation, Grade)

Student(StuNo, StuName, StuPhone, SubjectNo*)
StudentGrade(SubjectNo, SubDesc, LecID, LecLocation, Grade)

2NF (Remove Partial Dependencies)

StudentGrade(StuNo*, SubjectNo, StuName, StuPhone, Grade)
Subject(SubjectNo, SubDesc, LecID*)
Lecturer(LecID, LecLocation)

3NF (Remove Transitive Dependencies)

StudentGrade(StuNo*, SubjectNo*, Grade)
Student(StuNo, StuName, StuPhone)
Subject(SubjectNo, SubDesc, LecID*)
Lecturer(LecID, LecLocation)

@DBDL – underline PK (Primary key) and put * for FK (foreign key)

Model Question and Answer:

A2) Explain the THREE (3) data anomaly and provide ONE (1) example for each of them.

DeliveryOrder

DO_No	DelDate	CustNo	CustName	CustPhone	ItemNo	Desc	UnitPrice	Qty
DO0001	23/8/05	C0002	Eric	03-61381111	I0007	Pencil	3.00	200
DO0001	23/8/05	C0002	Eric	03-61381111	I0010	Paper	8.00	100
DO0002	23/8/05	C0003	John	03-41490888	I0003	Gum	1.00	350
DO0002	23/8/05	C0003	John	03-41490888	I0007	Pencil	3.00	50
DO0002	23/8/05	C0003	John	03-41490888	I0010	Paper	8.00	500

Insertion Anomaly: Inability to add new records without adding redundant data.

It is not possible to add the new item unless that new item is ordered by customer.

Modification anomaly : Inability to accurately and efficiently maintain data. A change to an attribute's values in one record requires changes in multiple records.

When we update the item description (I0007) from Pencil to 2B Pencil, we need also update the similar item description in another row, if not it will cause data inconsistency.

Deletion anomaly: Inability to delete unwanted data without deleting data that you need to retain.

When we delete the delivery order (DO0002) record we also will delete the item details: Gum (I0003).

◀ ▶

Model Question and Answer:

B2) Explain the THREE (3) data anomaly and provide ONE (1) example for each of them.

StudentGrade

StuNo	StuName	StuPhone	SubjectNo	SubDesc	LecID	LecLocation	Grade
S0001	Eric	03-61381111	S007	Database	1234	Block B	A
S0001	Eric	03-61381111	S010	Accounting	3344	Block A	C
S0002	John	03-41490888	S003	Management	5666	Block C	A
S0002	John	03-41490888	S007	Database	1234	Block B	C
S0002	John	03-41490888	S010	Accounting	3344	Block A	B

Insertion Anomaly: Inability to add new records without adding redundant data.

It is not possible to insert a new subject unless that subject already belongs to a Student.

Modification anomaly : Inability to accurately and efficiently maintain data. A change to an attribute's values in one record requires changes in multiple records.

When we change the SubDesc (S007) from Database to Database 2, we also need to update the similar SubDesc in another row, if not, it will cause data inconsistency.

Deletion anomaly: Inability to delete unwanted data without deleting data that you need to retain.

When we delete the StuNo (S0002), we also delete the SubDesc: Management (S003).

◀ ▶