

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the left and right sides of the slide, framing the central text area.

DBS211

Week 10 - Normalization

What is Normalization?

- ▶ Normalization is a series of steps used to evaluate and modify table structures to ensure that every non-key column in every table is directly dependent on the primary key.
- ▶ The results of normalization are reduced redundancies, fewer anomalies and improved efficiencies.

Two purposes of normalization

- ▶ Purposes of normalization
 - ▶ *Eliminate redundant data* (the same data stored in more than one table)
 - ▶ Eliminating redundant data is achieved by splitting tables with redundant data into two or more tables without the redundancy
 - ▶ *Ensure the data within a table are related*
- ▶ Normalization involves the process of applying rules called **normal forms** to table structures that produce a design that is free of data redundancy problems.

Problems with Un-normalized Data

- ▶ Update Problem
- ▶ Data Inconsistency Problem
- ▶ Data Redundancy Problem
- ▶ Insert Problem
- ▶ Deletion Problem

Modification Anomaly

Inventory

<u>product_id</u>	<u>whse_id</u>	product_desc	bin	qty	whse_address	city	prov	pcode
145	122	Saw	136	40	122 Peter St.	Newmarket	Ont	L4T5Y6
145	322	Saw	175	25	4433 Oak Ave	Oakville	Ont	L5T6R5
355	122	Screwdriver	111	55	122 Peter St.	Newmarket	Ont	L4T5Y6
130	322	Hammer	98	35	4433 Oak Ave	Oakville	Ont	L5T6R5

- The Update Problem
 - The need to perform the same update in several locations of the database because the same data is repeated
 - Oakville warehouse is moved to Burlington
 - We will have to make more than one change to the database

Data Inconsistency

Inventory

<u>product_id</u>	<u>whse_id</u>	product_desc	bin	qty	whse_address	city	prov	pcode
145	122	Saw	136	40	122 Peter St.	Newmarket	Ont	L4T5Y6
145	322	Saw	175	25	4433 Oak Ave	Oakville	Ont	L5T6R5
355	122	Screwdriver	111	55	122 Peter St.	Newmarket	Ont	L4T5Y6
130	322	Hammer	98	35	4433 Oak Ave	Oakville	Ont	L5T6R5

- The Data Inconsistency Problem
 - When the same data is repeated in several records, they can be inconsistent

Data Redundancy

Inventory

<u>product_id</u>	<u>whse_id</u>	product_desc	bin	qty	whse_address	city	prov	pcode
145	122	Saw	136	40	122 Peter St.	Newmarket	Ont	L4T5Y6
145	322	Saw	175	25	4433 Oak Ave	Oakville	Ont	L5T6R5
355	122	Screwdriver	111	55	122 Peter St.	Newmarket	Ont	L4T5Y6
130	322	Hammer	98	35	4433 Oak Ave	Oakville	Ont	L5T6R5

- The Data Redundancy Problem
 - the unnecessary repetition of data in the database of non-key fields.
 - While it is fine to repeat Primary Keys and Foreign Keys, we do not want to repeat data fields.

The Insert Problem

- ▶ Only one table for storing information, STUDENT
- ▶ Let us say we have just hired a new teacher: Mr. Vert. We have no way to put him into the database as he has no students yet.

STUDENT(<u>Student-Num</u> , Student-Name, Teacher, Student-Age)			
1243658712	Tom Blu	Ms.Green	14
2343216578	Jill Fall	Mr.Brown	14
3214325436	Jack Pail	Ms.Green	14

The Deletion Problem

- ▶ If there is no teacher table, and if a teacher's students all go to high school, then the teacher will disappear from our database.

STUDENT(<u>Student-Num</u> , Student-Name, Teacher, Student-Age)			
1243658712	Tom Blu	Ms.Green	14
2343216578	Jill Fall	Mr.Brown	14
3214325436	Jack Pail	Ms.Green	14

Normal Forms

- ▶ Several normal forms exist
 - ▶ 1NF
 - ▶ 2NF
 - ▶ 3NF
- ▶ Each normal form addresses the potential for a particular type of redundancy.
- ▶ A table is said to be in one of the normal forms if it satisfies the rules required by that form.

DBDL

DBDL or DataBase Definition Language is a standardized way of describing entities in a relational database in a written format. It is very similar to the entity format from ERDs, except written in line as a paragraph of text would be written.

ENTITY_NAME [KEY FIELD, field2, FK field3, (field4, field5, field6), ... fieldn]

The above is the basic syntax for a basic DBDL entity. A few points to note:

The entity name is written in ALL-CAPS outside the square brackets []

The Primary, or better referred to as Candidate, Key is underlined. If underlining is not possible, we write either PK or CK in front of it indicating Primary or Candidate Key.

Fields that are related to fields in other entities are indicated using FK, for Foreign Key.

The Round Brackets are used to distinguish groups of related data that may result in repeating groups (described in more detail below)

Ma & Pa's Bake Shop
Best Donuts & Cakes in Town!

Server: KEN DOB: 10/9/2018
06:44 PM 10/9/2018
18/1 1/10156

SALE

VISA 1048614
Card XXXXXXXXXXXX2227
Card Present: JOEY SMITH
Card Entry Method: S

Approval: 535665

1	Dozen Donuts	\$10.99
1	Box 0 Joe	\$6.99

Amount: \$17.98
Tax: \$1.35
+ Tip: _____
= Total: _____

Tipping Suggestion
Tip 15% Tip 18% Tip 20%
\$2.90 \$3.48 \$3.87

I agree to pay the above
total amount according to the
card issuer agreement

X _____
JOEY SMITH

CUSTOMER COPY

RECEIPT [receiptID, company_name,
purchase_date]

RECEIPT [receiptID, company_name,
purchase_date]

RECEIPT [receiptID, company_name, purchase_date, employee_id, server_name]

RECEIPT [receiptID, company_name,
purchase_date, employee_id,
server_name, product_name, quantity,
unit_price]

First Normal Form (1NF)

1

- Two dimensional table format

2

- no repeating groups - each row/column intersection only contains one value

3

- Primary key is identified

1NF (Example)

Inventory

product_id	product_desc	whse_id	bin	qty	whse_address	city	prov	pcode
145	Saw	122	136	40	122 Peter St.	Newmarket	Ont	L4T5Y6
		322	175	25	4433 Oak Ave	Oakville	Ont	L5T6R5
355	Screwdriver	122	111	55	122 Peter St.	Newmarket	Ont	L4T5Y6
130	Hammer	322	98	35	4433 Oak Ave	Oakville	Ont	L5T6R5

- Product no 145 is stored in two different warehouses.
- Having more than one value at the intersection of a row and a column is referred to as having a repeating group.
- A table that contains a repeating group is called an un-normalized table (UNF)
- Inventory [product_id, product_desc(whse_id,bin,qty,whse_address,city,prov,pcode)]

1NF (Example)

Inventory

product_id	product_desc	whse_id	bin	qty	whse_address	city	prov	pcode
145	Saw	122	136	40	122 Peter St.	Newmarket	Ont	L4T5Y6
145	Saw	322	175	25	4433 Oak Ave	Oakville	Ont	L5^6R5
355	Screwdriver	122	111	55	122 Peter St.	Newmarket	Ont	L4T5Y6
130	Hammer	322	98	35	4433 Oak Ave	Oakville	Ont	L5T6R5

- ▶ The repeating groups can be eliminated by filling in the values in vacant cells of the table.
- ▶ Each row/column intersection must contain only a single value to satisfy the first rule for 1NF.
- ▶ What should the primary key be?

1NF (Example)

Inventory

<u>product_id</u>	<u>whse_id</u>	product_desc	bin	qty	whse_address	city	prov	pcode
145	122	Saw	136	40	122 Peter St.	Newmarket	Ont	L4T5Y6
145	322	Saw	175	25	4433 Oak Ave	Oakville	Ont	L5T6R5
355	122	Screwdriver	111	55	122 Peter St.	Newmarket	Ont	L4T5Y6
130	322	Hammer	98	35	4433 Oak Ave	Oakville	Ont	L5T6R5

- ▶ The primary Key should be product_id concatenated to whse_id.
- ▶ Inventory [product_id, whse_id, product_desc, bin, qty, whse_address, city, prov, pcode]

Second Normal Form (2NF)

1

- 1NF

2

- no partial dependencies; all non-key columns are fully dependent on the entire primary key

2NF

Inventory

<u>product_id</u>	<u>whse_id</u>	product_desc	bin	qty	whse_address	city	prov	pcode
145	122	Saw	136	40	122 Peter St.	Newmarket	Ont	L4T5Y6
145	322	Saw	175	25	4433 Oak Ave	Oakville	Ont	L5T6R5
355	122	Screwdriver	111	55	122 Peter St.	Newmarket	Ont	L4T5Y6
130	322	Hammer	98	35	4433 Oak Ave	Oakville	Ont	L5T6R5

► Is this table in 2NF?

► Fix this

2NF

Product

product_id	product_desc
145	Saw
355	Screwdriver1
130	Hammer

Warehouse

<u>whse_id</u>	whse_address	city	prov	pcode
122	122 Peter St.	Newmarket	Ont	L4T5Y6
322	4433 Oak Ave	Oakville	Ont	L5T6R5

Inventory

<u>product_id</u>	<u>whse_id</u>	bin	qty
145	122	145	40
145	322	175	25
355	122	111	55
130	322	98	35

Let's write this out in a text based form called Relational Schema:

Capitalize the entity name

Put attributes in parenthesis

Bold and underline the primary key

For later - Italicize the foreign key or use FK

2NF

Product

product_id	product_desc
145	Saw
355	Screwdriver1
130	Hammer

Warehouse

whse_id	whse_address	city	prov	pcode
122	122 Peter St.	Newmarket	Ont	L4T5Y6
322	4433 Oak Ave	Oakville	Ont	L5T6R5

Inventory

product_id	whse_id	bin	qty
145	122	145	40
145	322	175	25
355	122	111	55
130	322	98	35

PRODUCT (product_id, prod_desc)

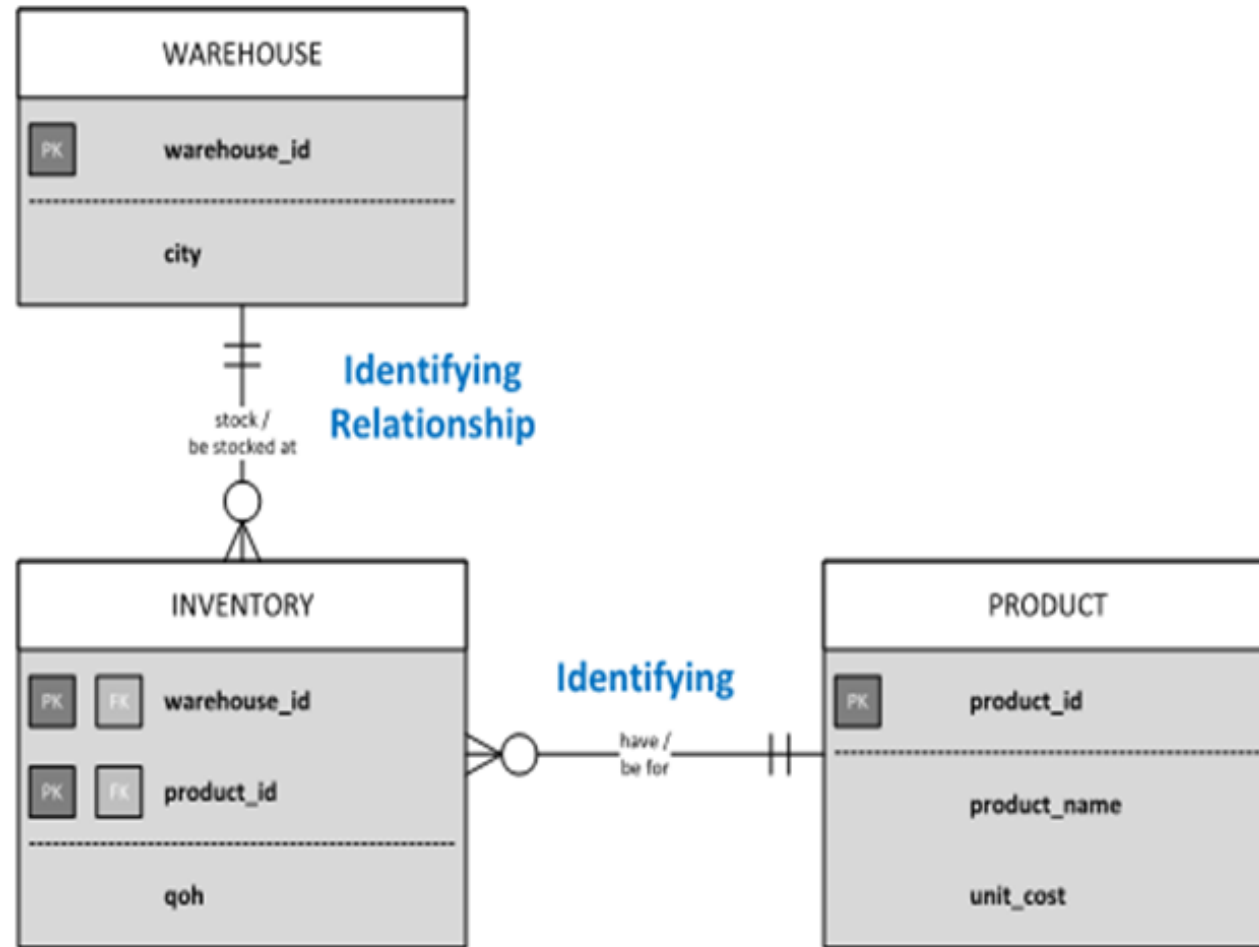
WAREHOUSE (whse_id, whse_address, city, prov, pcode)

INVENTORY (whse_id (FK), product_id (FK), bin, qty)

Functional Dependency

- ▶ A functional dependency occurs when one or more attributes in a table uniquely determines another attribute
- ▶ $\text{product_id} \rightarrow \text{prod_desc}$
- ▶ $\text{whse_id}, \text{product_id} \rightarrow \text{bin}, \text{qty}$
- ▶ $\text{whse_id} \rightarrow \text{whse_address}, \text{city}, \text{prov}, \text{pcode}$

ERD



Partial Dependency

- ▶ Partial dependency is where a non-key column is dependent on part of the primary key but is not dependent on the entire primary key.
[product_id, whse_id, product_desc, bin, qty, whse_address, city, prov, pcode]
- ▶ The prod_desc column is dependent on the product_id key but is not determined by the whse_id key.
- ▶ The whse_address column is dependent on the whse_id key but is not related to the product_id key
- ▶ We only need to look for this and solve it when we have a concatenated key

Third Normal Form (3NF)

1

- 2NF

2

- A non-key column cannot determine the value of another non-key column. Every non-key column must depend directly on the primary key

3NF

Warehouse

<u>whse_id</u>	whse_address	city	prov	pcode
122	122 Peter St.	Newmarket	Ont	L4T5Y6
322	4433 Oak Ave	Oakville	Ont	L5T6R5

To satisfy the second rule of 3NF, the warehouse table can be split into two tables.

3NF

Warehouse

<u>whse_id</u>	whse_address	pcode
122	122 Peter St.	L4T5Y6
322	4433 Oak Ave	L5T6R5

PostalCode

pcode	city	prov
L4T5Y6	Newmarket	Ont
L5T6R5	Oakville	Ont

Postal Code determines city and province.

Why is this table not in 3NF?

order_id	cust_id	product_id	quantity	unit_price	total_amt
101	141	222	11	25.86	284.46
102	156	333	54	35.77	1931.58
103	322	654	33	22.10	729.30
104	445	321	87	91.55	7968.33

ORDER in 3NF

order_id	cust_id	product_id	quantity	unit_price
101	141	222	11	25.86
102	156	333	54	35.77
103	322	654	33	22.10
104	445	321	87	91.55

Property	UNF	1NF	2NF	3NF
Each Entity has a Candidate Primary Key	✓	✓	✓	✓
Repeating Groups Indicated by Parenthesis	✓	✓	✓	✓
Repeating Groups Eliminated		✓	✓	✓
Atomic Fields Only		✓	✓	✓
No Partial Dependencies			✓	✓
No Transitive Dependencies				✓

Summary

- ▶ Rules for reducing data redundancy and related problems called normal forms
- ▶ UNF?
- ▶ 1NF - no repeating groups
- ▶ 2NF - no partial dependencies
- ▶ 3NF - no transitive dependencies
- ▶ Functional dependency?
- ▶ Partial dependency?