

บทที่ 3 การทำนอร์มัลไลเซชัน (Normalization)

การออกแบบฐานข้อมูลด้วย E-R Model นั้นมีจุดมุ่งหมายเพื่อนำเสนอข้อเท็จจริงต่างๆ ยังไม่ได้ตรวจสอบปัญหาทางด้านความซ้ำซ้อนของข้อมูล และความถูกต้องของข้อมูล ดังนั้น การแก้ไขปัญหเหล่านี้ จะทำได้ด้วยวิธีการ Normalization (นอร์มัลไลเซชัน)

3.1 Normalization

นอร์มัลไลเซชัน เป็นวิธีการตรวจสอบและแก้ไขปัญหาด้านความซ้ำซ้อนของข้อมูล เป็นการทำให้ข้อมูลในตารางอยู่ในหน่วยที่เล็กที่สุดที่ไม่สามารถแตกแยกเป็นหน่วยย่อยๆ ได้อีก โดยยังคงความสัมพันธ์ระหว่างข้อมูลไว้ดังเดิม

จุดประสงค์ของการนอร์มัลไลเซชัน

- 1.ลดเนื้อที่ในการจัดเก็บข้อมูล กระบวนการนอร์มัลไลเซชันเป็นการออกแบบเพื่อลดความซ้ำซ้อน (Redundancy) ในข้อมูล ดังนั้น การลดความซ้ำซ้อนในข้อมูลก็ย่อมทำให้ลดเนื้อที่ในการจัดเก็บข้อมูลด้วย
- 2.ลดปัญหาความไม่ถูกต้องของข้อมูล เมื่อข้อมูลไม่เกิดความซ้ำซ้อน ในการปรับปรุงข้อมูลก็สามารถปรับปรุงข้อมูลได้จากแหล่งข้อมูลเพียงแหล่งเดียว จึงช่วยลดความผิดพลาดที่อาจเกิดขึ้นจากการปรับปรุงข้อมูล (Update Anomalies) ซึ่งประกอบด้วย
 - ข้อผิดพลาดจากการเพิ่มข้อมูล (Insertion Anomalies)
 - ข้อผิดพลาดจากการลบข้อมูล (Deletion Anomalies)
 - ข้อผิดพลาดจากการเปลี่ยนแปลงข้อมูล (Modification Anomalies)

ขั้นตอนการทำนอร์มัลไลเซชัน

การทำ นอร์มัลไลเซชัน ต้องทำตามลำดับขั้นตอน ดังนี้

1. (1 NF) First Normal Form
2. (2 NF) Second Normal Form
3. (3 NF) Third Normal Form
- (BCNF) Boyce-Codd Normal Form
4. (4 NF) Fourth Normal Form
5. (5 NF) Fifth Normal Form

3.2 Functional Dependency (FDs)

[FD : determinant-attribute ----- dependency-attribute]

หมายถึง Det-att (Key) ที่ระบุค่าใดค่าหนึ่งแล้ว จะสามารถแสดงค่าของ Dep-att ซึ่งสัมพันธ์กันออกมา เช่น

ความสัมพันธ์ระหว่าง Person_ID และ Person_Name

Citizen_ID	
Person_ID	Person_Name
13045678927 →	สมชาย
13024893450	สมถวิล
12349874562	เจริญ
39409123894 ←×	สมชาย
32495098234	นิวัติ

จึงเขียนได้ว่า
FD : Person_id -> Person_name



Functional Dependency

ฟังก์ชันการขึ้นต่อกัน คือความสัมพันธ์ระหว่างแอตทริบิวต์หนึ่งหรือแอตทริบิวต์หนึ่งตั้งแต่หนึ่งตัวขึ้นไปมาประกอบกันแล้วสามารถทำการระบุค่าของแอตทริบิวต์อื่นๆในทูปเฟิลเดียวกันของรีเลชันนั้นได้

โดยแอตทริบิวต์ที่เป็นตัวระบุค่าในแอตทริบิวต์อื่นๆ เรียกว่า determinant

แอตทริบิวต์ที่ถูกระบุค่าจาก determinant เรียกว่า dependent

3.3 ประเภท Functional Dependency (FDs)

1. FD ที่มีความสัมพันธ์ **อย่างละ 1 ค่า**

FD : Person_id \rightarrow Person_name

2. FD ที่มีความสัมพันธ์ **มากกว่า 1 ค่า**

FD : Person_id \rightarrow Fname, Lname, Address, Birth_date

3. FD ที่มี **ความสัมพันธ์ 2 ทาง**

FD : Project_no \rightarrow Student /or/ FD : Student \rightarrow Project_no

4. FD ที่ต้องใช้ **Determinant มากกว่า 1 ค่า**

FD : Student_id, subject_id \rightarrow Grade



3.4 Fully Functional Dependency

ได้แก่ FD ที่มี Determinant ขนาดเล็กที่สุด ที่สามารถระบุถึง Dependency ได้ เช่น FD 2 แบบนี้

D1 : Person_id Address , Road , City , Post , Tel

D2 : Person_id , Prduct_Id Quantity

คือ รีเลชันจะต้องไม่มีความสัมพันธ์ระหว่างแอตทริบิวต์บางส่วน (Partially Dependent)

3.5 Multi-value Dependency

หมายถึง Determinant 1 ค่า สามารถระบุค่าของ Dependency ที่สัมพันธ์กับ Det. ได้มากกว่า 1 ค่า เช่น พนักงาน สังกัดหลายแผนก และหลายโครงการ

MVD : Employee# \twoheadrightarrow Department#, Project#

Employee	Department	Project
11001	01	P1
11002	02	P3
11001	03	P2
12003	03	P7
12114	01	p1

3.6 Relation Key

หมายถึง Attribute จำนวน 1 หรือมากกว่า ที่สามารถบ่งชี้ และแยกแยะข้อมูลในแต่ละ Tuple ได้ แสดงโดยการใช้ขีดเส้นใต้ชื่อ Attribute ดังในตาราง

Employee				
<u>EmpID</u>	Name	Sex	Military_Status	Salary
00001	สมชาย นิลกลัด	M	1	8,500
00003	เจริญ ก้าวหน้า	M	0	12,000
00005	นิวัติ เหล่าสรวรณ	M	2	25,000

3.7 First Normal Form (1 NF)

นิยามคือ ทุกๆ Attribute ใน Relation จะต้องมีความสมบัติ Atomicity คือจะต้องไม่มีข้อมูลแบบ Repeating Group เช่น ใบสั่งซื้อสินค้า ที่ลูกค้าหนึ่งคน แต่ส่งของหลายรายการ ดังนี้

ใบสั่งซื้อสินค้า				
บริษัท สร้างเงินเทคโนโลยีจำกัด				
รหัสลูกค้า	C001	ชื่อ นารี เกิดสว่าง	ที่อยู่ จ.อยุธยา	เขต 001
รายการสินค้า		จำนวน		
P001		24		
P003		30		
P004		50		

ทำข้อมูลให้อยู่ในรูปของขั้นตอน 1NF

ทำการแปลงข้อมูล Repeating Group

ให้เป็น Atomicity โดยการให้ข้อมูลการซื้อแต่ละรายการไปขึ้นกับลูกค้าจะทำได้ดังนี้

Order					
Cust_id	Cust_name	City	Zone	Product_id	Order_Qty
C001	นารี เกิดสว่าง	อยุธยา	001	P001	24
C001	นารี เกิดสว่าง	อยุธยา	001	P003	30
C001	นารี เกิดสว่าง	อยุธยา	001	P004	50
C002	สลักจิต สว่างภพ	ศรีสะเกษ	002	P001	29

3.8 Second Normal Form (2 NF)

1.จะต้องมีคุณสมบัติของ 1NF

2.ทุก Attribute ต้องมีคุณสมบัติ Fully Function Dependency

จากตารางที่ผ่านมาจะได้ FD ว่า

D1 : Cust_no , Product-id Order_Qy

D2 :Cust_no Cust_ame , City , Zoe

นั่นคือ มี FFD 2 ตัว ในตารางนี้

ทำให้อยู่ในรูปของ 2 NF

- โดยการแยกตารางออกมาตาม D1 และ D2 ดังนี้

Customer

Cust_no	Cust_name	City	Zone
C001	นารี เกิดสว่าง	อยุธยา	001
C002	สลักจิต สว่างภพ	ศรีสะเกษ	002
C003	สุทิดา แจ็กสกุล	เชียงใหม่	004
C009	ฟ้า เพิ่มพร	ศรีสะเกษ	002

Cust_Order

Cust_no	Product_id	Order_Qty
C001	P001	24
C001	P003	30
C001	P004	50
C002	P001	29
C002	P004	40

3.9 Third Normal Form (3 NF)

1. ต้องมีคุณสมบัติของ 2 NF
2. ต้องไม่มี FD เกิดขึ้น ระหว่าง Nonkey Attri ด้วยกัน

จากขั้นตอนที่ 2 จะเห็นว่ามี City และ Zone สามารถระบุค่าระหว่างกันได้ เช่น ถ้าทราบเลข Zone จะทราบชื่อเมือง ลักษณะนี้เรียกว่า Transitive Dependency ซึ่งสัมพันธ์กันเองระหว่าง Attribute

ทำให้อยู่ในรูปของ 3 NF

โดยการแยก Transitive
Dependency ออกมาเป็น
ตาราง City ดังนี้

City

City	Zone
อยุธยา	001
ศรีสะเกษ	002
สงขลา	003
เชียงใหม่	004

Cust_Order

Cust_no	Product_id	Order_Qty
C001	P001	24
C001	P003	30
C001	P004	50
C002	P001	29
C002	P004	40

Customer

Cust_no	Cust_name	Zone
C001	นารี เกิดสว่าง	001
C002	สลักจิต สว่างภาพ	002
C003	สุทิดา แจกสกุล	004
C009	ฟ้า เพิ่มพร	002

ยังมี Att Zone ไว้
เพื่อคงความสัมพันธ์
ไว้ดังเดิม

3.10 Boyce-Codd Normal Form (BCNF)

- 1.ต้องมีคุณสมบัติของ 3NF
- 2.Attribute ที่เป็น Determinant จะต้องเป็น Primary Key

City		
Zone	City	
001	อยุธยา	
002	ศรีสะเกษ	
003	สงขลา	
004	เชียงใหม่	

Cust		
Cust_no	Cust_name	Zone
C001	นารี เกิดสว่าง	001
C002	สลักจิต สว่างภพ	002
C003	สุทิศา แจกสกุล	004
C009	ฟ้า เพิ่มพร	002

Cust_Order		
Cust_no	Product_id	Order_Qty
C001	P001	24
C001	P003	30
C001	P004	50
C002	P001	29
C002	P004	40

3.11 Fourth Normal Form (4 NF)

- 1.ต้องมีคุณสมบัติของ BCNF
- 2.ต้องไม่มีความสัมพันธ์แบบ Multi-value Dependency

Employee	Department	Project
11001	01	P1
11002	02	P3
11001	02	P2
12003	03	P7
12114	01	p1

ลักษณะข้อมูลแบบ Multi-value Dependency

Multivalued Dependency เป็นความสัมพันธ์ระหว่างค่าของแอททริบิวต์แบบหลายค่าที่มักเกิดขึ้นกับรีเลชันที่มี Attribute อย่างน้อย 3 Attribute และมี Attribute หนึ่งสามารถระบุค่า Attribute อื่นๆ ในรีเลชันได้มากกว่า 1 ค่า

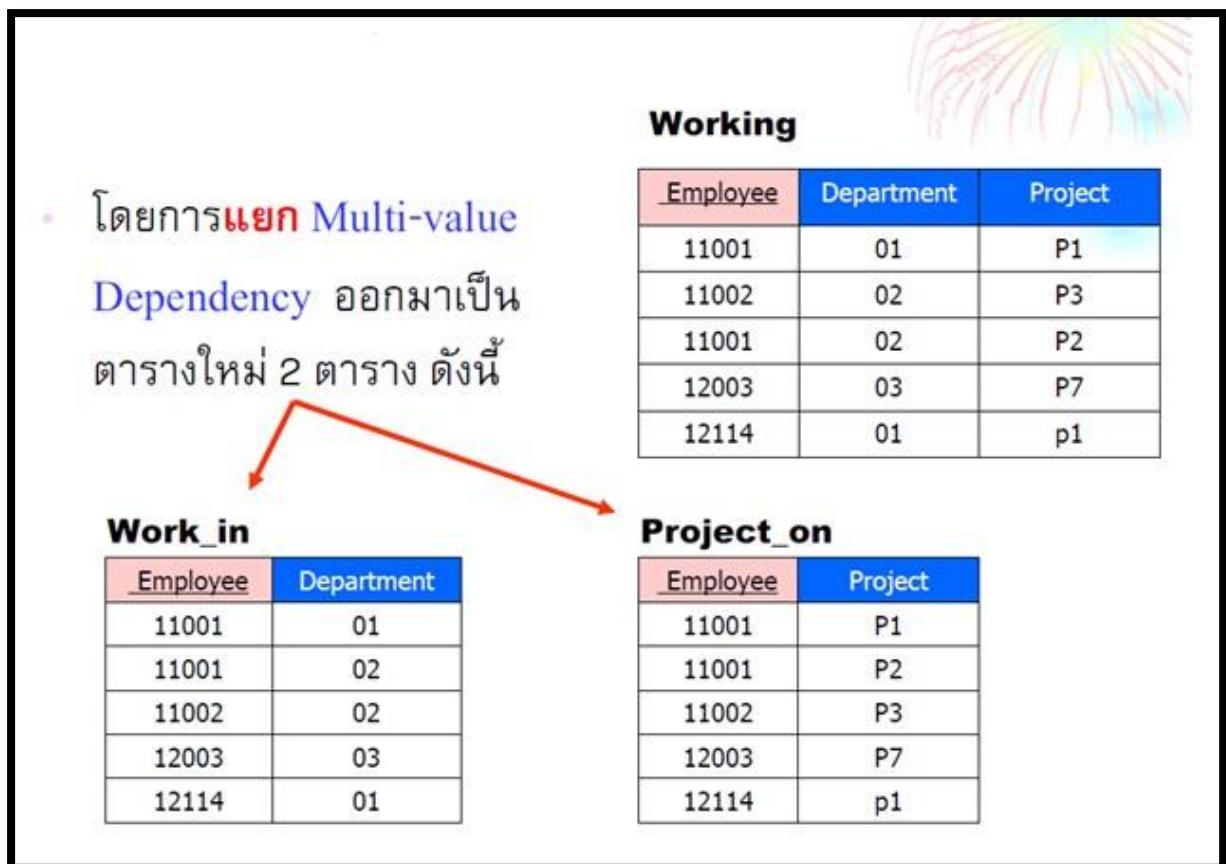
Employee	Department	Project
11001	01	P1
11002	02	P3
11001	02	P2
12003	03	P7
12114	01	p1

Employee สามารถระบุค่า

Department ได้หลายค่า

Department สามารถระบุค่า Project ได้หลายค่า

ทำให้อยู่ในรูปของ 4NF



3.12 Fifth Normal Form (5 NF)

1. ต้องมีคุณสมบัติของ 4NF
2. ต้องมีคุณสมบัติ Join Dependency

