บทที่ 3

แนวคิดและทฤษฎีที่เกี่ยวข้อง

ข้อมูล และสารสนเทศ (สุโขทัยธรรมาธิราช, 2534)

ข้อมูล (Data) หมายถึง ข้อเท็จจริงต่าง ๆ ที่มีอยู่ในธรรมชาติ เป็นกลุ่มสัญลักษณ์แทนปริมาณหรือการกระทำต่าง ๆ ที่ยังไม่ผ่านการประมวลผล ข้อมูลอาจจะอยู่ในรูปของตัวเลข ตัวหนังสือ และท้ายที่สุดข้อมูลก็คือวัตถุดิบของสารสนเทศ

สารสนเทศ (Information) หมายถึง ข้อมูลที่ผ่านการประมวลผลด้วยวิธีการต่าง ๆ เพื่อให้ได้มาซึ่งบางสิ่งบางอย่างตามความต้องการของผู้ใช้ เป็นส่วนผลลัพธ์ของระบบการประมวลผลข้อมูล เป็นสิ่งซึ่งสื่อความหมายให้ผู้รับเข้าใจ และสามารถนำไปกระทำกิจกรรมใดกิจกรรมหนึ่งโดยเฉพาะได้ หรือเพื่อเป็นการย้ำความเข้าใจที่มีอยู่แล้วให้มีมากยิ่งขึ้นและเป็นผลลัพธ์ของระบบสารสนเทศ

ความสัมพันธ์ของข้อมูล สารสนเทศ และระบบสารสนเทศ

ระบบสารสนเทศ (Information System) ประกอบไปด้วยส่วนนำเข้า ส่วนกระบวนการหรือส่วนประมวลผล และส่วนผลลัพธ์ ซึ่งตัวข้อมูลจะเป็นวัตถุดิบของระบบในส่วนนำเข้าเพื่อประมวลผลข้อมูลด้วยวิธีต่าง ๆ กัน และได้สารสนเทศเป็นผลลัพธ์ของระบบสารสนเทศ ระบบสารสนเทศในปัจจุบันได้นำเอาเครื่องคอมพิวเตอร์เข้ามาใช้ เพื่อให้การประมวลผลทำได้รวดเร็วขึ้น และเพื่อให้สารสนเทศที่ได้มีคุณภาพน่าเชื่อมากขึ้น

คุณสมบัติของสารสนเทศที่ดี

สารสนเทศที่จัดว่าเป็นสารสนเทศที่ดี ควรจะมีคุณสมบัติที่สำคัญคือ *ถูกต้อง ทันต่อการใช้งาน ครบถ้วนสมบูรณ์ กระทัดรัด และตรงกับความต้องการ* ซึ่งคุณสมบัติเหล่านี้จะเป็นตัวบ่งบอกถึงคุณภาพและคุณค่าของสารสนเทศนั้น ในการนำสารสนเทศไปใช้งานครั้งหนึ่ง ๆ สารสนเทศควรจะมีคุณสมบัติดังกล่าวนี้เป็นปริมาณที่มากหรือน้อยเพียงใดนั้น ขึ้นอยู่กับความ

ต้องการของผู้ใช้ซึ่งจะเป็นผู้พิจารณาถึงความคุ้มค่าของสารสนเทศที่ต้องการกับต้นทุนหรือค่าใช้จ่ายที่จะให้ได้มาซึ่งสารสนเทศนั้น

นอกจากคุณสมบัติของสารสนเทศดังกล่าวแล้ว ยังมีคุณสมบัติที่แอบแฝงของสารสนเทศอีกบางลักษณะที่สัมพันธ์กับระบบสารสนเทศ และวิธีดำเนินงานของระบบสารสนเทศ นอกจากคุณสมบัติเหล่านี้จะมีความสำคัญแตกต่างกันไปตามลักษณะงานเฉพาะอย่าง คุณสมบัติเหล่านี้ได้แก่ *ความละเอียดแม่นยำ คุณสมบัติเชิงปริมาณ ความยอมรับได้ การใช้ได้ง่าย ความไม่ลำเอียง ชัดเจน*

ระบบฐานข้อมูล (Database System)

ระบบฐานข้อมูล คือ โครงสร้างสารสนเทศ ที่ประกอบด้วยข้อมูลที่มีความสัมพันธ์กัน เพื่อให้ผู้ใช้และโปรแกรมประยุกต์ต่าง ๆ สามารถดำเนินการกับข้อมูลเหล่านั้นได้ โดยอาศัยระบบจัดการฐานข้อมูล (Database Management System : DBMS) ในการควบคุมดูแลการสร้างและการเรียกใช้ฐานข้อมูล

ประโยชน์จากการประมวลผลด้วยฐานข้อมูล (ดวงแก้ว สวามิภักดิ์, 2534)

1. ลดความซ้ำซ้อนของข้อมูล (Redundancy can be reduced)
2. สามารถหลีกเลี่ยงความขัดแย้งของข้อมูลได้ในระดับหนึ่ง (Inconsistency can be reduced ... to some extent)
3. สามารถใช้ข้อมูลร่วมกันได้ (The data can be shared)
4. สามารถควบคุมความเป็นมาตรฐานได้ (Standards can be enforced)
5. สามารถจัดหาความปลอดภัยที่รัดกุมได้ (Security restrictions can be applied)
6. สามารถควบคุมความคงสภาพของข้อมูลได้ (Integrity can be maintained)
7. สามารถสร้างสมดุลในความขัดแย้งของความต้องการได้ (Conflicting requirements can be balanced)
8. เกิดความเป็นอิสระของข้อมูล (Data independence)

สถาปัตยกรรมของระบบจัดการฐานข้อมูล (Elmasri, Navathe, 1989)

1. ระดับภายใน (Internal Level) เป็นระดับต่ำสุดที่เกี่ยวกับเรื่องของการจัดเก็บข้อมูลภายในสื่อคอมพิวเตอร์ โดยพิจารณาถึงรูปแบบข้อมูลที่ถูกบันทึกและชนิดของดัชนีที่ใช้เชื่อมโยงภายใน รวมทั้งโครงสร้างและวิธีการเข้าถึงข้อมูล

2. ระดับหลักการ (Conceptual Level) เป็นระดับที่เกี่ยวกับเรื่องของการกำหนดความสัมพันธ์ของแบบจำลองข้อมูลเชิงตรรก เช่น กฎของการตั้งชื่อ ชนิด และลักษณะการใช้งานของข้อมูลที่เป็นส่วนประกอบในฐานข้อมูลนั้น

3. ระดับภายนอก (External Level) เป็นระดับที่เกี่ยวกับเรื่องของการแสดงข้อมูลในรูปแบบเฉพาะในมุมมองของผู้ใช้งานแต่ละคน หรือแต่ละกลุ่มเท่านั้น โดยที่ผู้ใช้งานแต่ละคนหรือแต่ละกลุ่ม สามารถเรียกใช้ข้อมูลหรือมองเห็นโครงสร้างข้อมูลได้เฉพาะในส่วนที่ถูกกำหนดไว้ให้ใช้เท่านั้น

การทำงานส่วนใหญ่จะให้ผู้ใช้งานอ้างถึงเฉพาะความต้องการในระดับภายนอกของตนเองเท่านั้น ระบบจัดการฐานข้อมูลจะทำการเปลี่ยนความต้องการที่ระบุนั้นให้อยู่ในรูปของระดับหลักการ และระดับภายในตามลำดับ

ความเป็นอิสระของข้อมูล (Data Independence) (ดวงแก้ว สวามิภักดิ์, 2534)

ความเป็นอิสระของข้อมูล คือความสามารถในการเปลี่ยนแปลงแก้ไขข้อมูลในระดับที่ต่ำกว่า โดยไม่มีผลกระทบกับระดับที่เหนือกว่า มี 2 ประเภท คือ

1. ความเป็นอิสระแบบตรรก (Logical Data Independence) คือลักษณะของการเปลี่ยนแปลงข้อมูลในระดับหลักการโดยที่ไม่มีผลกระทบต่อระดับผู้ใช้ภายนอก เช่น อาจเพิ่มเอนติตี (entity) ชนิดใหม่ลงไปในฐานข้อมูล หรืออาจเพิ่มแอตทริบิวต์ (attribute) เข้าไปใหม่เป็นต้น

2. ความเป็นอิสระแบบกายภาพ (Physical Data Independence) คือลักษณะของการเปลี่ยนแปลงข้อมูลในระดับภายใน จะไม่มีผลกระทบต่อโครงสร้างของระดับหลักการและระดับภายนอก ตัวอย่างของการเปลี่ยนแปลงแก้ไขในระดับภายใน เช่น การเปลี่ยนแปลงวิธีการจัดสร้างข้อมูลจากลำดับเชิงดัชนี (Index Sequential) เป็นแบบสุ่ม (Direct Access) ซึ่งในการจัดเก็บจริง ๆ นั้นอาจจะหมายถึงการเปลี่ยนแปลงจากการใช้ลิสต์ มาใช้ Inverted File แทน เป็นต้น ทั้งนี้เพื่อมุ่งหวังให้การเรียกใช้ข้อมูลมีประสิทธิภาพมากขึ้น แต่ในระดับของผู้ใช้ หรือภาพรวมในระดับหลักการจะไม่มีผลกระทบใด ๆ ทั้งสิ้น

การออกแบบฐานข้อมูล (Database Design) (จรณิต แก้วกังวาน, 2536)

การออกแบบฐานข้อมูลเพื่อนำมาใช้งานภายในองค์กร จำแนกได้เป็น 2 วิธีใหญ่ ๆ คือ วิธีการอุปมัย (Bottom-up) และวิธีการนิรมัย (Top-down)

1. วิธีการอุปมัย (Bottom-up หรือ Inductive Approach) เป็นการออกแบบสร้างฐานข้อมูลโดยอาศัยวิธีการรวบรวมข้อมูลหรือโปรแกรมที่มีอยู่แล้วจากหลายๆ หน่วยงานในองค์กร แนวคิดพื้นฐานของการออกแบบประเภทนี้ถือว่าลักษณะงานของแต่ละหน่วยงานมีความซับซ้อนสมบูรณ์แตกต่างกัน ดังนั้นรูปแบบของฐานข้อมูลที่ดีจึงต้องเป็นรูปแบบที่รวบรวมเอาข้อดีของข้อมูลหรือโปรแกรมต่าง ๆ ที่มีอยู่แล้วในหน่วยงานต่าง ๆ มารวบรวมเป็นรูปแบบขนาดใหญ่ทั้งหมด ข้อเสียของวิธีการนี้คือ การรวมวิธีการย่อย ๆ เข้าด้วยกันทำได้ไม่ง่ายนัก และเสียเวลามากในการที่จะออกแบบและสร้างระบบฐานข้อมูลที่สมบูรณ์

2. วิธีการนิรมัย (Top-down หรือ Deductive Approach) เป็นวิธีการที่เหมาะสมกับการออกแบบระบบฐานข้อมูล ที่มีการออกแบบเป็นครั้งแรก โดยเลือกเอาผู้ที่เข้าใจระบบทั้งหมด ศึกษาว่าองค์กรมีข้อมูลอะไรบ้าง แล้วจึงนำมาออกแบบเป็นโครงสร้างทั้งหมดของฐานข้อมูล ปัญหาของวิธีการนี้คือ จะต้องได้ผู้ที่ศึกษาและเข้าใจระบบทั้งหมดจริง แต่เป็นวิธีการที่เหมาะสมสำหรับการจัดวางระบบฐานข้อมูลที่มีความยุ่งยาก มีความหลากหลายของหน่วยงานต่าง ๆ ภายในองค์กร

ฐานข้อมูลเชิงสัมพันธ์ (Relational Database)

ฐานข้อมูลเชิงสัมพันธ์ จะแทนข้อมูลในฐานข้อมูลในลักษณะของตาราง และแถวต่าง ๆ ในตารางจะแสดงค่าของข้อมูลที่มีความสัมพันธ์กัน ชื่อของตารางและชื่อของสดมภ์ จะใช้ในการแปลความหมายของค่าในแต่ละแถวของตาราง (Flemming, Halle, 1989)

ในฐานข้อมูลเชิงความสัมพันธ์ ได้กำหนดคำเพื่ออธิบายตาราง ดังนี้

รีเลชั่น (Relation)	หมายถึง	ตาราง (Table)
ทูเปิล (Tuple)	หมายถึง	แถว (Row) หรือ ระเบียน (Record)
แอตทริบิวต์	หมายถึง	สดมภ์ (Column) หรือ เขตข้อมูล (Field)
โดเมน (Domain)	หมายถึง	ค่าที่เป็นไปได้ของข้อมูลในแต่ละสดมภ์

การออกแบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database Design)

เป็นขั้นตอนการแปลงจากแบบจำลองข้อมูลเชิงตรรก เข้าสู่ระบบฐานข้อมูลเชิงสัมพันธ์อย่างมีเสถียรภาพภายใต้ระบบจัดการฐานข้อมูลที่ติดตั้งบนเครื่องคอมพิวเตอร์นั้น ๆ

การออกแบบฐานข้อมูลเชิงสัมพันธ์ ประกอบด้วย 2 ขั้นตอนดังนี้

1. การออกแบบจำลองข้อมูลเชิงตรรก (Logical Data Modeling)

แบบจำลองข้อมูลเชิงตรรก (Logical Data Model) เป็นแบบจำลองที่สร้างขึ้นเพื่อใช้อธิบายลักษณะโครงสร้างข้อมูล และความสัมพันธ์ระหว่างข้อมูล รวมถึงการปฏิบัติการในการเรียกใช้ข้อมูลและแก้ไขข้อมูล ตามความต้องการของผู้ใช้งานในลักษณะที่ถูกต้อง ก่อนที่จะทำการแปลงให้อยู่ในรูปของฐานข้อมูลเชิงความสัมพันธ์ต่อไป การออกแบบในขั้นตอนนี้จะต้องมีการกำหนดสิ่งที่เป็นพื้นฐานได้แก่

1.1 เอนติตี (Entity) คือสิ่งต่าง ๆ ในระบบ อาจเป็นสิ่งที่มีอยู่จริงในทางกายภาพ หรือสิ่งที่มีอยู่ทางความคิด เช่น อาจารย์ นิสิต รายวิชา เป็นต้น

1.2 แอตทริบิวต์ คือข้อมูลที่แสดงลักษณะและคุณสมบัติของเอนติตี

1.3 ความสัมพันธ์ (Relationship) คือความสัมพันธ์ระหว่างเอนติตี

2. การออกแบบจำลองข้อมูลเชิงสัมพันธ์ (Relational Data Modeling)

แบบจำลองข้อมูลเชิงสัมพันธ์ (Relational Data Model) เป็นแบบจำลองข้อมูลที่เปลี่ยนรูปแบบมาจากแบบจำลองข้อมูลเชิงตรรก เพื่อเป็นแนวคิดที่แสดงให้เห็นถึงข้อมูลที่ถูกมองเห็นโดยผู้ใช้งาน โดยจะแสดงรายละเอียดของข้อมูลที่ปรากฏแก่ผู้ใช้ การดำเนินการกับข้อมูล และกฎเกณฑ์ต่าง ๆ แบบจำลองข้อมูลเชิงสัมพันธ์มีส่วนประกอบดังนี้

2.1 โครงสร้างข้อมูล (Data Structure) เป็นข้อมูลที่มองเห็นโดยผู้ใช้งานอยู่ในลักษณะของตารางความสัมพันธ์ (Relation)

2.2 การดำเนินการกับข้อมูล (Data Manipulation) เป็นการดำเนินการกับข้อมูลในตารางความสัมพันธ์

2.3 ความเป็นบูรณภาพของข้อมูล (Data Integrity) เป็นกฎข้อบังคับสำหรับข้อมูลในตารางความสัมพันธ์ (Business Rule)

ข้อดีของแบบจำลองข้อมูลเชิงสัมพันธ์

1. เป็นแบบจำลองข้อมูลที่เสนอต่อผู้ใช้งานในรูปแบบที่เข้าใจง่าย การทำงานเพื่อตอบสนองความต้องการของผู้ใช้จะเกี่ยวข้องเฉพาะข้อมูล โดยไม่ต้องคำนึงถึงความซับซ้อนของอุปกรณ์ทางฮาร์ดแวร์

2. แบบจำลองข้อมูลจะแยกโครงสร้างของหน่วยเก็บข้อมูล (Storage Structure) และวิธีการเข้าถึงข้อมูล (Access Strategy) ออกจากส่วนที่ติดต่อกับผู้ใช้ (User Interface) ซึ่งถือเป็นคุณสมบัติที่เหนือกว่าแบบจำลองข้อมูลชนิดอื่น

3. แบบจำลองข้อมูลเชิงสัมพันธ์นี้มีพื้นฐานบนทฤษฎีทางคณิตศาตร์ที่ได้รับการพัฒนามาเป็นอย่างดี และวิธีการออกแบบฐานข้อมูลโดยมีการทำนอร์มัลไลซ์ (Normalization) ทำให้ได้แบบจำลองที่มีพื้นฐานที่ดี

กฎของการนอร์มัลไลซ์ (Normalization)

การนอร์มัลไลซ์ เป็นการวิเคราะห์และแยกส่วนโครงสร้างข้อมูลออกเป็นชุดของความสัมพันธ์ที่มีคุณสมบัติตามที่ต้องการ ทำให้การออกแบบโมเดลข้อมูลมีความถูกต้อง คงที่ มีเสถียรภาพ และไม่ซับซ้อน (Date,1990)

1. นอร์มัลฟอร์มระดับ 1 (First Normal Form, 1NF)

ทุกสมาชิกในเอนติตีจะมีค่าแอตทริบิวต์หนึ่ง ๆ ได้เพียงค่าเดียว หรือกล่าวคือ จะมีกลุ่มของค่าในแอตทริบิวต์นั้นในสมาชิกหนึ่งไม่ได้

2. นอร์มัลฟอร์มระดับ 2 (Second Normal Form, 2NF)

ความสัมพันธ์จะมีคุณสมบัติเป็นนอร์มัลฟอร์มระดับ 2 เมื่อมีคุณสมบัติของนอร์มัลฟอร์มระดับ 1 และแอตทริบิวต์ทุกตัวที่ไม่ใช่คีย์จะต้องขึ้นตรงกับคีย์หลัก

3. นอร์มัลฟอร์มระดับ 3 (Third Normal Form, 3NF)

ความสัมพันธ์จะมีคุณสมบัติเป็นนอร์มัลฟอร์มระดับ 3 เมื่อมีคุณสมบัติของนอร์มัลฟอร์มระดับ 2 และแอตทริบิวต์ทุกตัวที่ไม่ใช่คีย์จะต้องไม่ขึ้นกับแอตทริบิวต์อื่น ๆ ที่ไม่ใช่คีย์หลักหรือกลุ่มของคีย์หลัก