

บทที่ 2

ความรู้เบื้องต้นเกี่ยวกับระบบฐานข้อมูล

ปัจจุบันระบบข้อมูลและระบบสารสนเทศ เป็นปัจจัยสำคัญหนึ่งสำหรับการดำเนินงานขององค์กรอย่างมีประสิทธิภาพ ระบบฐานข้อมูลเป็นแนวความคิดของการดำเนินงานกับข้อมูล (data) และสารสนเทศ (information) ฐานข้อมูลขององค์กรต่างๆ ที่ใช้กันอยู่ในปัจจุบันมีหลายรูปแบบ ขึ้นอยู่กับระบบการจัดการกับฐานข้อมูลที่ใช้ งาน เช่น DB2 หรือ ORACLE เป็นระบบจัดการฐานข้อมูลสำหรับฐานข้อมูลแบบเชิงสัมพันธ์ IMS (Information Management System) เป็นระบบจัดการฐานข้อมูลสำหรับฐานข้อมูลแบบลำดับขั้น เป็นต้น

ในบทนี้ จะกล่าวถึง นิยามของคำศัพท์ต่างๆ ที่เกี่ยวข้องกับระบบฐานข้อมูล เพื่อความเข้าใจที่ตรงกันในการอ่านเนื้อหาบทต่อไป

2.1 ระบบฐานข้อมูล

ระบบฐานข้อมูล เป็นคำศัพท์ที่หมายถึงระบบงานซึ่งมีการนำเทคโนโลยีฐานข้อมูลมาใช้เพื่อเพิ่มประสิทธิภาพของการดำเนินงานกับข้อมูล โดยใช้เครื่องคอมพิวเตอร์เป็นเครื่องมือ ระบบฐานข้อมูลมีส่วนประกอบสำคัญสามส่วนคือ

1. ฐานข้อมูล

ฐานข้อมูล คือข้อมูลหลักของหน่วยงานหรือองค์กรใดๆ พร้อมทั้งความสัมพันธ์ระหว่างข้อมูลหลักเหล่านี้ ซึ่งจะถูกรวบรวมเข้าไว้ด้วยกันอย่างมีระบบ ไม่ซ้ำซ้อนกัน โดยมีเป้าหมายหลักคือ ให้ผู้ใช้ทุกคนสามารถใช้อยู่ร่วมกันได้ สามารถมองข้อมูลที่ใช้ร่วมกันนี้ได้ในลักษณะที่แตกต่างกันตามความต้องการของผู้ใช้แต่ละคน หรือที่เรียกว่า view ได้ และเพื่อให้ข้อมูลต่างๆ ในฐานข้อมูลถูกนำไปใช้ได้ง่าย สะดวก รวดเร็ว ทันสมัย และถูกต้องอยู่เสมอ

2. ระบบจัดการฐานข้อมูล

ระบบจัดการฐานข้อมูล คือ ซอฟต์แวร์ (software) ที่ใช้เป็นตัวกลางระหว่างผู้ใช้กับฐานข้อมูล เพื่ออำนวยความสะดวกต่อผู้ใช้งานข้อมูล เช่น การสร้าง การปรับปรุงแก้ไข หรือการเรียกใช้ข้อมูลในฐานข้อมูล โดยผู้ใช้ไม่ต้องรับรู้เกี่ยวกับรายละเอียดภายในของการจัดเก็บข้อมูลในฐานข้อมูล เนื่องจากระบบจัดการฐานข้อมูลจะไม่อนุญาตให้ผู้ใช้เรียกใช้ข้อมูลจากฐานข้อมูลได้โดยตรง

การทำงานกับข้อมูลในฐานข้อมูลจะผ่านระบบจัดการฐานข้อมูลเสมอ และโดยทั่วไประบบจัดการฐานข้อมูลประกอบด้วยโครงสร้างพื้นฐาน สองส่วนคือ

- ส่วนกำหนดโครงสร้างฐานข้อมูล รายละเอียดและเงื่อนไขบังคับของข้อมูลที่จะจัดเก็บ และการดำเนินงานพื้นฐานในการรับข้อมูลเข้า การลบข้อมูลออก หรือปรับปรุงเปลี่ยนแปลงแก้ไขข้อมูลในฐานข้อมูล
- ส่วนประมวลผลเพื่อให้ได้ผลลัพธ์ซึ่งเป็นสารสนเทศที่ต้องการ

3. ผู้ใช้

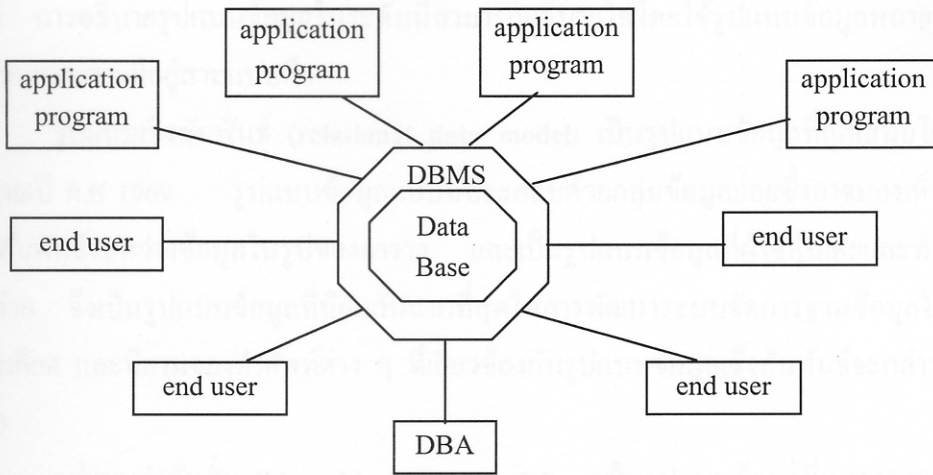
ผู้ใช้ในที่นี้จะหมายถึงผู้ที่เกี่ยวข้องกับฐานข้อมูลตามอำนาจหน้าที่รับผิดชอบ และความสามารถในการใช้งานฐานข้อมูล โดยทั่วไปแบ่งได้สามประเภทดังนี้

ผู้บริหารฐานข้อมูล (Database Administrator หรือ DBA) คือ บุคคลหรือกลุ่มบุคคลที่มีหน้าที่รับผิดชอบต่อการออกแบบและสร้างฐานข้อมูลในองค์กรให้มีจำนวนชิ้นข้อมูล (data item) และความหมาย (semantics) เป็นไปตามความต้องการของเจ้าของข้อมูลในแต่ละหน่วยงานย่อยขององค์กร ดำเนินการและจัดการให้ข้อมูลในฐานข้อมูลมีความถูกต้องอยู่เสมอ รวมถึงการเข้าถึงและการรักษาความปลอดภัยของข้อมูลด้วย ผู้บริหารฐานข้อมูลจึงมีความสำคัญต่อระบบฐานข้อมูลมาก ผู้บริหารฐานข้อมูลจึงควรเป็นผู้มีความรู้ และมีอำนาจหน้าที่รับผิดชอบสูงทั้งในด้านการบริหารงาน และด้านวิชาการ เพื่อให้การใช้งานฐานข้อมูลเป็นไปได้ด้วยดี [Teorey and Fry (1982)]

ผู้ใช้ทั่วไป (end users) คือ ผู้ใช้ที่อาจไม่มีความรู้ด้านคอมพิวเตอร์เลยก็ได้หรือ มีความรู้บ้าง ผู้ใช้ประเภทนี้ส่วนใหญ่จะติดต่อใช้งานฐานข้อมูลโดยการใช้เมนูคำสั่งของโปรแกรมประยุกต์ที่ถูกพัฒนาไว้ให้ใช้หรือใช้ภาษาสอบถาม (query language) ของระบบจัดการฐานข้อมูล

ผู้เขียนโปรแกรมประยุกต์ (application programmer) คือบุคคลหรือกลุ่มบุคคลที่มีความรู้ความสามารถในการติดต่อใช้งานกับระบบจัดการฐานข้อมูลเองได้โดยตรงด้วยการเขียนโปรแกรมประยุกต์

ผู้ใช้ทั้งสามประเภทดังกล่าวเมื่อต้องการใช้ข้อมูลในฐานข้อมูล ต้องใช้ผ่านระบบจัดการฐานข้อมูลทุกครั้งเสมอ ความสัมพันธ์ระหว่างฐานข้อมูล ระบบจัดการฐานข้อมูล และผู้ใช้แสดงได้ดังภาพประกอบที่ 2.1



ภาพประกอบที่ 2.1 ระบบฐานข้อมูล

2.2 รูปแบบข้อมูล

รูปแบบข้อมูล (data model) เป็นภาพโครงสร้างโดยรวมของข้อมูลพร้อมทั้งความสัมพันธ์ของข้อมูลที่ผู้ออกแบบได้รวบรวมมาเพื่อสร้างเป็นฐานข้อมูลที่ใช้ทุกคนของหน่วยงานย่อยต่างๆ ในองค์กรสามารถใช้ข้อมูลร่วมกันได้ ระบบจัดการฐานข้อมูลทุกระบบต้องมีการจัดการและเก็บข้อมูลภาพโครงสร้างโดยรวมนี้ของข้อมูลในรูปแบบใดรูปแบบหนึ่งเสมอ

Korth และ Silberschatz [Korth and Silberschatz (1991)] ได้จำแนกรูปแบบข้อมูลที่ใช้ในการออกแบบฐานข้อมูลออกเป็นสามระดับดังนี้

1. Object-based logical models

รูปแบบข้อมูลระดับนี้ใช้อธิบายข้อมูล และความสัมพันธ์ของข้อมูลในระดับเชิงวัตถุ ซึ่งเป็นระดับที่ห่างจากตัวเครื่องคอมพิวเตอร์และใกล้ชิดกับผู้ใช้มากที่สุด มักเป็นรูปแบบของข้อมูลที่สามารถทำความเข้าใจความหมายและความสัมพันธ์ของข้อมูลได้โดยง่าย ลักษณะการอธิบายข้อมูลและความสัมพันธ์ของข้อมูลจึงใช้แผนภาพในการอธิบาย ได้มีผู้เสนอรูปแบบการอธิบายโครงสร้างข้อมูลระดับนี้หลายรูปแบบ เช่น แผนภาพโออาร์ (Object Relationship (O-R) diagram) แผนภาพอีอาร์ (Entity Relationship (E-R) diagram) เป็นต้น

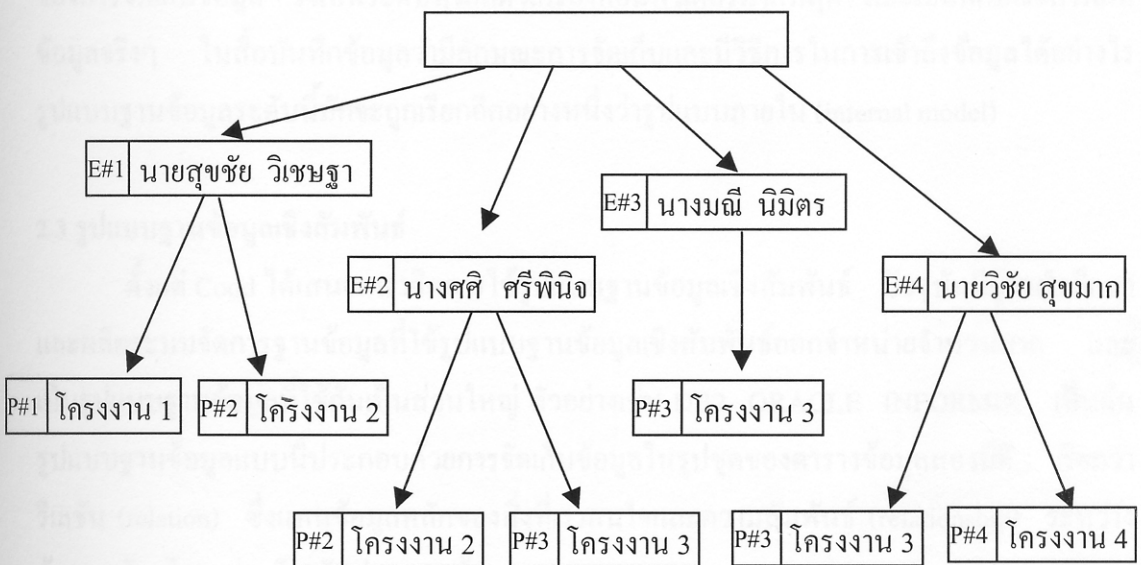
2. Record-based logical models

รูปแบบข้อมูลระดับนี้ใช้อธิบายข้อมูล และความสัมพันธ์ของข้อมูลในระดับระเบียบข้อมูล ซึ่งเป็นระดับที่ใกล้ตัวเครื่องคอมพิวเตอร์มากกว่ารูปแบบข้อมูลแบบ object-based logical models และเป็นรูปแบบข้อมูลที่สามารถนำไปใช้ในการเขียนโปรแกรมเพื่อปฏิบัติการกับ

เครื่อง การอธิบายรูปแบบข้อมูลในระดับนี้สามารถอธิบายได้โดยใช้รูปแบบข้อมูลหลายอย่าง ที่รู้จักกันแพร่หลายมีอยู่สามแบบคือ

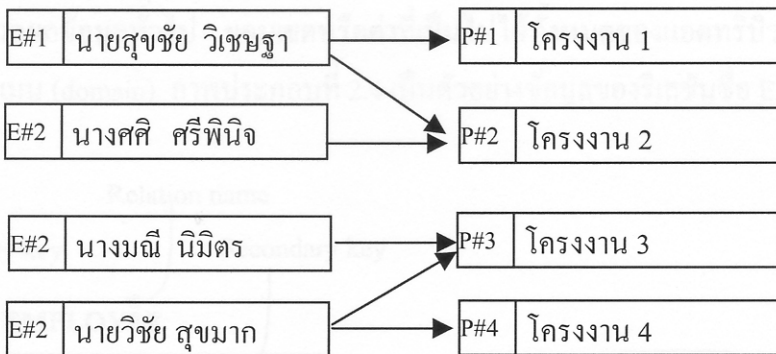
รูปแบบเชิงสัมพันธ์ (relational data model) เป็นรูปแบบข้อมูลที่ถูกเสนอโดย Codd ประมาณปี ค.ศ 1969 รูปแบบข้อมูลแบบนี้ประกอบด้วยกลุ่มข้อมูลย่อยซึ่งอาจมองเห็นลักษณะความสัมพันธ์ระหว่างข้อมูลในรูปของตาราง และเป็นรูปแบบข้อมูลที่ใช้คุ้นเคยและทำความเข้าใจได้ง่าย จึงเป็นรูปแบบข้อมูลที่นิยมกันมากที่สุดในการพัฒนาาระบบจัดการฐานข้อมูลในปัจจุบัน รายละเอียด และนิยามของคำศัพท์ต่าง ๆ ที่เกี่ยวข้องกักรูปแบบข้อมูลเชิงสัมพันธ์จะกล่าวถึงในหัวข้อ 2.3

รูปแบบลำดับชั้น (hierachical data model) เป็นรูปแบบข้อมูลที่แสดงความสัมพันธ์ของข้อมูลในลักษณะเป็นลำดับชั้นเหมือนโครงสร้างแบบต้นไม้ (tree structure) โดยเริ่มจากจุดแรกๆ ที่เรียกว่า โหนดราก (root node) ซึ่งจะมีโหนดรากเพียงโหนดเดียวเป็นหลักแล้วจึงแตกกิ่งก้านสาขาออกไปเหมือนต้นไม้ ดังภาพประกอบที่ 2.2 แต่ละกิ่งสามารถแตกกิ่งย่อยออกไปได้อีกเรื่อยๆ โดยมีความสัมพันธ์กันเป็นระดับ เรียกข้อมูลบนกิ่งย่อยก่อนที่จะแตกกิ่งในระดับย่อยถัดไปว่า โหนดแม่ (parent node) และเรียกข้อมูลบนปลายกิ่งย่อยที่แตกออกมาว่า โหนดลูก (child node) โดยโหนดลูกจะเกิดจากโหนดแม่ได้เพียงโหนดแม่เดียวเท่านั้น แต่โหนดแม่สามารถแตกกิ่งออกมาเป็นโหนดลูกได้มากกว่าหนึ่งโหนด



ภาพประกอบที่ 2.2 ตัวอย่างรูปแบบข้อมูลแบบลำดับชั้น

รูปแบบเครือข่าย (network data model) เป็นรูปแบบข้อมูลที่มีความสัมพันธ์ของข้อมูลคล้ายรูปแบบลำดับขั้น แต่มีการปรับปรุงความสัมพันธ์ให้ยืดหยุ่นมากขึ้น โดยข้อมูลในแต่ละระดับสามารถเกิดจากโหนดแม่ได้มากกว่าหนึ่ง โหนดแม่จึงมีลักษณะ โครงสร้างเป็นเครือข่ายดังภาพประกอบที่ 2.3 รูปแบบเครือข่ายเหมาะกับระบบข้อมูลที่มีความสัมพันธ์ระหว่างข้อมูลโดยไม่จำกัดลักษณะความสัมพันธ์



ภาพประกอบที่ 2.3 ตัวอย่างรูปแบบข้อมูลแบบเครือข่าย

3. Physical data models

รูปแบบข้อมูลระดับนี้ใช้อธิบายข้อมูลและความสัมพันธ์ของข้อมูลในระดับทางกายภาพของการจัดเก็บข้อมูล ซึ่งเป็นระดับที่ใกล้ตัวเครื่องคอมพิวเตอร์มากที่สุด และเป็นลักษณะการเก็บข้อมูลจริงๆ ในสื่อบันทึกข้อมูลว่ามีลักษณะการจัดเก็บและมีวิธีการในการเข้าถึงข้อมูลได้อย่างไร รูปแบบฐานข้อมูลระดับนี้มักจะถูกเรียกอีกอย่างหนึ่งว่ารูปแบบภายใน (internal model)

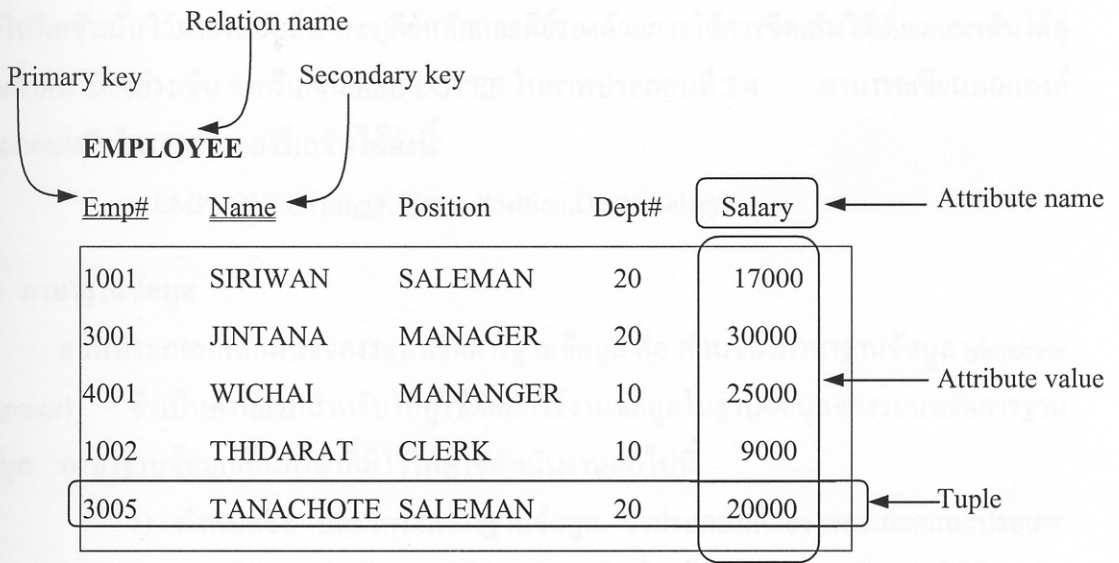
2.3 รูปแบบฐานข้อมูลเชิงสัมพันธ์

ตั้งแต่ Codd ได้เสนอแนวคิดการใช้รูปแบบฐานข้อมูลเชิงสัมพันธ์ ปัจจุบันมีผู้สนใจศึกษาและผลิกระบบจัดการฐานข้อมูลที่ใช้รูปแบบฐานข้อมูลเชิงสัมพันธ์ออกจำหน่ายจำนวนมาก และเป็นรูปแบบฐานข้อมูลที่ใช้กันเป็นส่วนใหญ่ ตัวอย่างเช่น DB2 ORACLE INFORMIX เป็นต้น รูปแบบฐานข้อมูลแบบนี้ประกอบด้วยการจัดเก็บข้อมูลในรูปชุดของตารางข้อมูลสองมิติ เรียกว่า รีเลชัน (relation) ซึ่งแทนข้อมูลหลักของสิ่งที่เราสนใจและความสัมพันธ์ (relationship) ระหว่างข้อมูลหลัก โดยแต่ละรีเลชันประกอบด้วย

ชื่อรีเลชัน (relation name) หรือชื่อตาราง แทนชื่อสิ่งที่เราสนใจจะบันทึกข้อมูล

ชื่อแอตทริบิว (attribute name) แทนคุณลักษณะ (attribute) หรือรายละเอียดของสิ่งที่เราสนใจจะบันทึกข้อมูล ซึ่งจะอยู่ที่ส่วนหัวของตาราง และมีความหมายทำนองเดียวกับฟิลด์ (field) ของแฟ้มข้อมูล (file)

ตัวอย่างข้อมูลของรีเลชัน ณ ขณะเวลาหนึ่ง (occurrence) ถ้าระบุข้อมูลตามแนวตั้งแต่ละสดมภ์ (column) จะเรียกว่าค่าของแอตทริบิว (attribute value) ถ้าระบุข้อมูลตามแนวนอนแต่ละแถวจะเรียกว่า ทูเปิล (tuple) ซึ่งมีความหมายทำนองเดียวกับระเบียบ (record) ของแฟ้มข้อมูล ในการประมวลผลข้อมูลทั่วไป ขอบเขตหรือค่าที่เป็นไปได้ทั้งหมดของแอตทริบิวหนึ่งในรีเลชันจะเรียกว่า โดเมน (domain) ภาพประกอบที่ 2.4 เป็นตัวอย่างข้อมูลของรีเลชันชื่อ EMPLOYEE



ภาพประกอบที่ 2.4 ตัวอย่างของข้อมูลในรีเลชันชื่อ EMPLOYEE ณ ขณะเวลาหนึ่ง

คุณสมบัติสำคัญของข้อมูลภายในรีเลชันหนึ่งๆ มีดังนี้

- 1) แต่ละค่าของแอตทริบิวจะต้องเป็นค่าโดด (atomic value) ซึ่งแทนค่าข้อมูลหนึ่งค่าเสมอ เช่น ค่าแทนเลขจำนวน แทนชื่อคน แทนสัตว์ หรือแทนสิ่งของ เป็นต้น
- 2) ข้อมูลในแต่ละทูเปิลของรีเลชันเดียวกันจะมีค่าต่างกันเสมอ
- 3) มีค่าของแอตทริบิวหนึ่ง หรือหลายแอตทริบิวที่สามารถใช้เป็นตัวแทนในการเข้าถึงข้อมูลแต่ละทูเปิลในรีเลชัน เรียกค่าแอตทริบิวนี้ว่า candidate key หรือเรียกสั้นๆ ว่า คีย์ (key) และเรียกคีย์ที่ประกอบด้วยค่าของหลายแอตทริบิวว่า compound key ส่วนค่าที่ไม่สามารถเป็นตัวแทนในการเข้าถึงข้อมูลแต่ละทูเปิลของรีเลชันได้จะเรียกว่า non-key ในรีเลชันใด ๆ อาจมีคีย์ได้มากกว่าหนึ่งคีย์

ผู้ออกแบบฐานข้อมูลจะต้องเป็นผู้ระบุว่า จะใช้คีย์ใดเป็นตัวแทนในการเข้าถึงข้อมูลแต่ละทิวเปิลของรีเลชัน คีย์ตัวที่ถูกเลือกจะเรียกว่าคีย์หลัก (primary key) ส่วนคีย์ตัวที่ไม่ได้ถูกเลือกจะเรียกว่าคีย์รอง (secondary key) จากภาพประกอบที่ 2.4 คีย์หลักของรีเลชัน EMPLOYEE ก็คือ ค่าของ Emp# (ระบุโดยการขีดเส้นใต้เดี่ยว) และคีย์รองของรีเลชัน EMPLOYEE คือค่าของ Name (ระบุโดยการขีดเส้นใต้คู่) และสามารถเข้าถึงข้อมูลใดๆ ในรีเลชันได้ โดยวิธีระบุค่าคีย์ของข้อมูลที่ต้องการ

ส่วนการแทนรีเลชันใดๆเฉพาะองค์ประกอบหรือ โครงสร้างหลักโดยไม่ต้องกล่าวถึงข้อมูลจริงที่เกิดขึ้นเลยนั้นสามารถเขียนได้โดยการเขียนชื่อรีเลชัน ตามด้วยการเขียนชื่อแอตทริบิวต์ทั้งหมดที่มีในรีเลชันนั้นไว้ภายในวงเล็บ ระบุคีย์หลักและคีย์รองด้วยการใช้การขีดเส้นใต้เดี่ยวและเส้นใต้คู่ตามลำดับ ตัวอย่างเช่น จากรีเลชัน EMPLOYEE ในภาพประกอบที่ 2.4 สามารถเขียนบอกองค์ประกอบหรือ โครงสร้างของรีเลชัน ได้ดังนี้

EMPLOYEE (Emp#, Name, Position, Dept#, Salary)

2.4 ภาษาฐานข้อมูล

ส่วนประกอบหลักหนึ่งของระบบจัดการฐานข้อมูล คือ ส่วนของภาษาฐานข้อมูล (database language) ซึ่งเป็นด่านแรกสำหรับให้ผู้ใช้ติดต่อใช้งานข้อมูลในฐานข้อมูลของระบบจัดการฐานข้อมูล ภาษาฐานข้อมูลเป็นภาษาที่มีไว้ให้ผู้ใช้ดำเนินงานต่อไปนี้

- 1) กำหนดชื่อ และโครงสร้างฐานข้อมูล ว่าประกอบด้วยรายละเอียดและประเภทของข้อมูลอะไรบ้าง มีข้อกำหนดหรือเงื่อนไขอะไรในการที่จะทำให้ข้อมูลในฐานข้อมูลมีความถูกต้อง เป็นหนึ่งเดียวกันเสมอ ภาษาที่ใช้ในการทำงานลักษณะนี้ เรียกว่า Data Definition Language (DDL) หรือ Data Description Language
- 2) การเพิ่มข้อมูลเข้าไปในฐานข้อมูล (insert) ลบข้อมูลออกจากฐานข้อมูล (delete) ปรับปรุงเปลี่ยนแปลงแก้ไขข้อมูลเก่าในฐานข้อมูล (update) และการประมวลผลให้ได้ผลลัพธ์ตามที่ต้องการจากข้อมูลในฐานข้อมูลที่มีอยู่ ภาษาที่ใช้ในการทำงานส่วนนี้เรียกว่า Data Manipulation Language (DML)

จากความต้องการหลักอันหนึ่งในการใช้ระบบจัดการฐานข้อมูล คือ การค้นคืน (retrieval) ข้อมูลในฐานข้อมูลตามวัตถุประสงค์ของผู้ใช้ให้ได้อย่างรวดเร็ว ทันทีทันใด ระบบจัดการฐานข้อมูลในปัจจุบันจึงอำนวยความสะดวกในการทำงานลักษณะนี้ โดยมีภาษาพิเศษสำหรับใช้ในการค้นคืนข้อมูลต่างๆ ตามที่ต้องการจากข้อมูลในฐานข้อมูล ภาษาพิเศษที่อำนวยความสะดวกให้

เรียกว่า ภาษาสอบถาม ซึ่งในปัจจุบันมีภาษาสอบถามมาตรฐานที่นิยมใช้กันในระบบจัดการฐานข้อมูลเชิงสัมพันธ์คือ SQL และ QBE (Query-By-Example)

2.5 ปทานุกรมข้อมูล

ส่วนประกอบสำคัญที่สุดส่วนหนึ่งสำหรับการดำเนินงานของระบบจัดการฐานข้อมูล คือ ปทานุกรมข้อมูล ซึ่งเป็นส่วนประกอบสำหรับจัดเก็บรายละเอียดรูปแบบของข้อมูลต่างๆ ในฐานข้อมูลที่ถูกกำหนดโดยผู้ใช้ รายละเอียดที่เก็บอาจเป็นรายละเอียดของรูปแบบข้อมูลของฐานข้อมูลทั้งหมด และเงื่อนไขสำหรับการดำเนินงานกับข้อมูลในฐานข้อมูล เช่น ใครมีสิทธิ์ใช้ข้อมูลส่วนใดของฐานข้อมูล เป็นต้น รายละเอียดเหล่านี้คือ ข้อมูลเกี่ยวกับข้อมูล (data about data) ที่ถูกสร้างในฐานข้อมูล ดังนั้นโดยทั่วไปปทานุกรมข้อมูลสำหรับระบบจัดการฐานข้อมูลหนึ่ง ๆ มักจะใช้รูปแบบเดียวกับข้อมูลธรรมดาของฐานข้อมูล เช่น ระบบจัดการฐานข้อมูลเชิงสัมพันธ์ ก็จะมีปทานุกรมข้อมูลประกอบด้วยข้อมูลซึ่งอยู่ในรูปแบบเชิงสัมพันธ์เช่นกัน และผู้ใช้สามารถเรียกดูข้อมูลในส่วนของปทานุกรมข้อมูลนี้ได้ด้วยภาษาฐานข้อมูลเช่นเดียวกันกับที่ใช้กับข้อมูลธรรมดาในฐานข้อมูล