บทที่ 3

แนวความคิดและทฤษฎีที่เกี่ยวข้อง

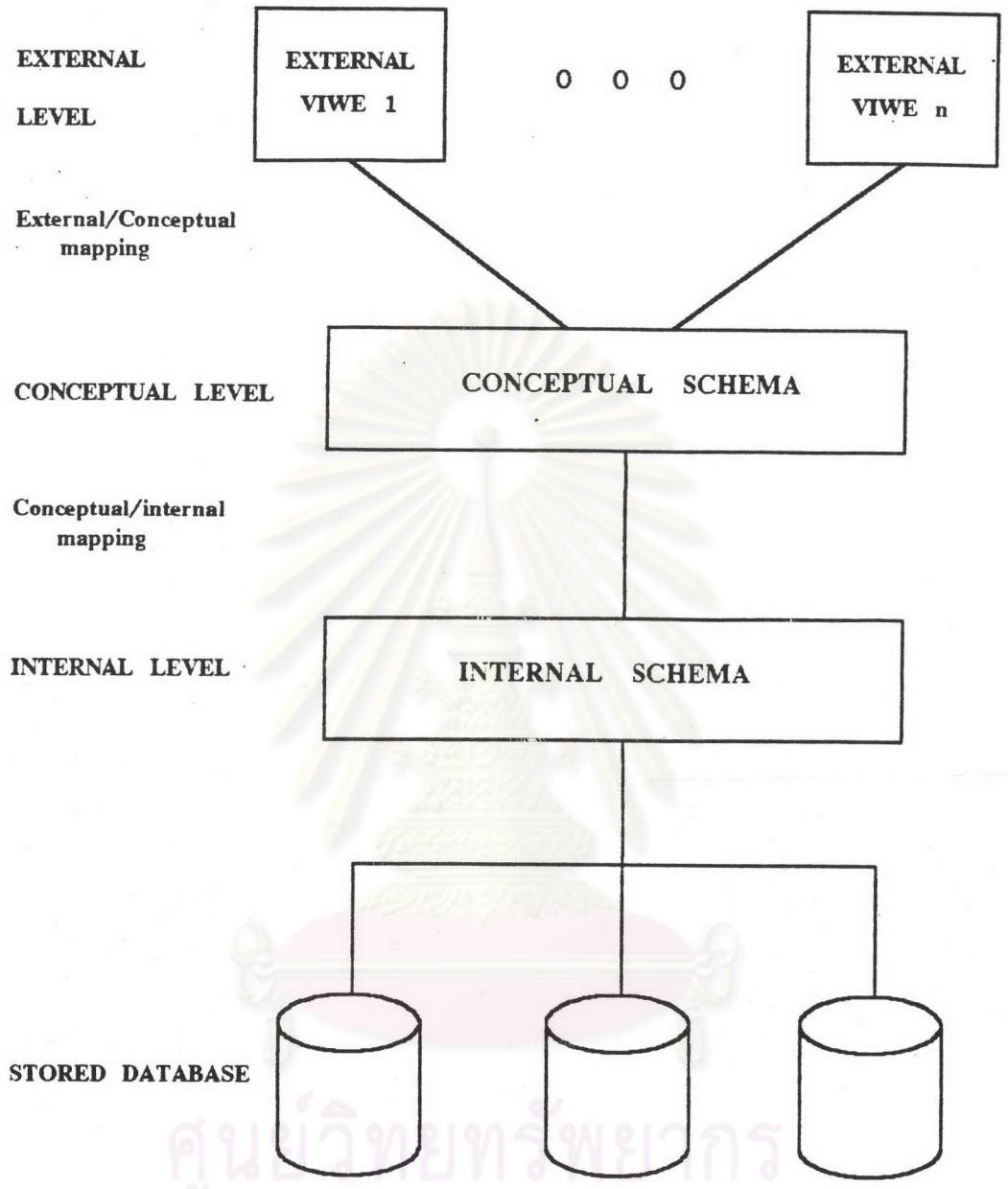
ระบบฐานข้อมูลแบบกระจาย เป็นระบบกระจาย การจัดเก็บข้อมูล และ กระจายการประมวลผลไปยังที่ต่าง ๆ ตามความจำเป็น และความต้องการของผู้ใช้ ดังนั้นระบบนี้จำเป็นต้องอาศัยระบบเครือข่ายคอมพิวเตอร์เป็นอย่างมาก หากไม่มีระบบเครือข่าย หรือยังไม่พัฒนาจนเป็นที่เชื่อถือได้ ระบบฐานข้อมูลแบบกระจาย ก็ยากที่จะดำเนินการได้ นอกจากนี้ระบบจัดการฐานข้อมูลแบบกระจาย ต้องมีความสามารถควบคุม จัดการฐานข้อมูลเฉพาะที่ได้อย่างทั่วถึง ดังนั้นจึงสามารถกล่าวได้ว่า ระบบฐานข้อมูลแบบกระจายยังประกอบด้วย ระบบฐานข้อมูลเฉพาะที่ (Local Database System) และระบบเครือข่ายทางคอมพิวเตอร์ (Computer Network) หรือระบบการสื่อสารข้อมูล (Data Communication)

ระบบฐานข้อมูล (Database System)

เป็นที่ทราบกันแล้วว่า ระบบฐานข้อมูลคือการนำข้อมูลทั้งหมดในระบบ ที่มีความเกี่ยวข้องกันมารวมกันไว้อย่างเป็นระบบ โดยผู้ใช้ฐานข้อมูลแต่ละคนจะมองข้อมูลในแง่มุมที่แตกต่างกัน ตามจุดประสงค์ของการประยุกต์ใช้งาน นอกจากนี้ระบบจัดการฐานข้อมูล ยังได้อำนวยความสะดวกให้แก่ผู้ใช้ โดยผู้ใช้ไม่จำเป็นต้องสนใจ วิธีการ ความยุ่งยากในการจัดเก็บข้อมูล เพราะระบบได้แบ่งระดับการมองเห็น (Date, 1986) และซ่อนรายละเอียดส่วนนี้ไว้ดัง รูป 3.1

แบบจำลองฐานข้อมูล (Database Model)

การจัดประเภทของแบบจำลอง ในระบบฐานข้อมูลจะยึดหลัก ความแตกต่างของระบบจัดการฐานข้อมูล การใช้งาน (Operation) และโครงสร้างของระบบข้อมูล ซึ่งในปัจจุบันสามารถแบ่งแบบจำลองเป็น 3 ชนิด (Henry and Siberschatz, 1986) ได้แก่ แบบจำลองเชิงสัมพันธ์ (Relational Model) แบบจำลองแบบแตกสาขา (Hierarchical Model) และแบบจำลองแบบเครือข่าย (Network Model) ในที่นี้เราจะสนใจเฉพาะแบบจำลองเชิงสัมพันธ์ เพราะเป็นแบบจำลองที่ได้รับความนิยมแพร่หลาย และมีความเหมาะสม ในการอธิบาย



รูป 3.1 สถาปัตยกรรมของระบบฐานข้อมูล

เกี่ยวกับมโนทัศน์ (Concept) ของระบบฐานข้อมูลกระจายดีกว่าแบบจำลองอื่น นอกจากนี้แบบจำลองเชิงสัมพันธ์ ยังมีข้อได้เปรียบกว่าอีกสองแบบจำลอง ดังจะสรุปได้ดังต่อไปนี้

1) เป็นแบบจำลองที่มีโครงสร้างข้อมูล คล้ายกับสภาพความเป็นจริงจึงทำให้เข้าใจง่าย ไม่สลับซับซ้อน

2) เป็นแบบจำลองที่มีเครื่องมือช่วยให้ผู้ใช้สามารถปฏิบัติการยาก ๆ กับข้อมูลด้วยคำสั่งง่าย ๆ และยังช่วยให้ผู้ออกแบบ ค้นหาจุดบกพร่อง และช่วยให้ง่ายต่อการแก้ไขข้อผิดพลาดจากการออกแบบ นอกจากนี้ยังมีระบบจัดการฐานข้อมูลหลายชนิด มากกว่าแบบจำลองอื่น ๆ เช่น ดีบี/2 (DB/2) บนเครื่องคอมพิวเตอร์ขนาดใหญ่ (Main-frame) เอสคิวแอล ดาต้าเบส (SQL Database) บนเครื่อง เอเอส/400, โอเอส/2 (AS/400,OS/2) เอสคิวแอล เบส (SQL Base) โอราเคิล (ORACLE) อินเกรส (INGRES) โปรเกส (PROGESS) อินฟอร์มิกซ์ (INFORMIX) ฟอกซ์โปร (FOXPRO) บนยูนิกซ์ (UNIX), บนดอส(DOS) กูปตา (GUPTA) เอสคิวแอล เซอร์เวอร์(SQL Server) บนวินโดว์(Microsoft Windows) ฯลฯ
3) ส่วนของการจัดเก็บข้อมูลจริงกับส่วนข้อมูลทางตรรกมีความอิสระจากกัน ดังนั้นแบบจำลองนี้จึงสอดคล้องกับหลักการของฐานข้อมูล ที่จะให้ผู้ใช้ไม่ต้องกังวลเกี่ยวกับการจัดเก็บข้อมูล

แบบจำลองเชิงสัมพันธ์ (Relational Model)

เป็นแบบจำลองที่มีการจัดเก็บข้อมูลในลักษณะ ตาราง ซึ่งอาจประกอบด้วยตารางหลายตารางที่มีความสัมพันธ์ต่อกัน ดังนั้นจึงนิยมเรียกตารางเหล่านี้ว่า ตารางความสัมพันธ์ (Relation) ตารางความสัมพันธ์เป็นตาราง 2 มิติ การจัดเก็บข้อมูลลงแต่ละช่องของตาราง จะบรรจุได้เพียงค่าเดียวโดยมี ช่วงค่า (Domain) กำหนดค่าทั้งหมดที่เป็นไปได้ของข้อมูล ข้อมูลในแนวนอนหมายถึงระเบียนข้อมูล (Record) เรียกว่า ทูเปิล (Tuple) แต่ละ ทูเปิล จะต้องมีค่าไม่ซ้ำกัน ความมาก่อนหลังไม่สำคัญ (Henry and Siberschatz, 1986) สำหรับข้อมูลในแนวสดมภ์ เป็นเขตข้อมูล (Field) เรียกว่าแอตตรีบิว (Attribute) แต่ละแอตตรีบิวต้องมีชื่อไม่ซ้ำกัน เช่นเดียวกับข้อมูลในแนวนอน แต่ละสดมภ์ของแอตตรีบิวต้องมีข้อมูลไม่ซ้ำกัน ส่วนอันไหนจะมาก่อนมาหลังไม่สำคัญ

<u>การออกแบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database Design)</u>

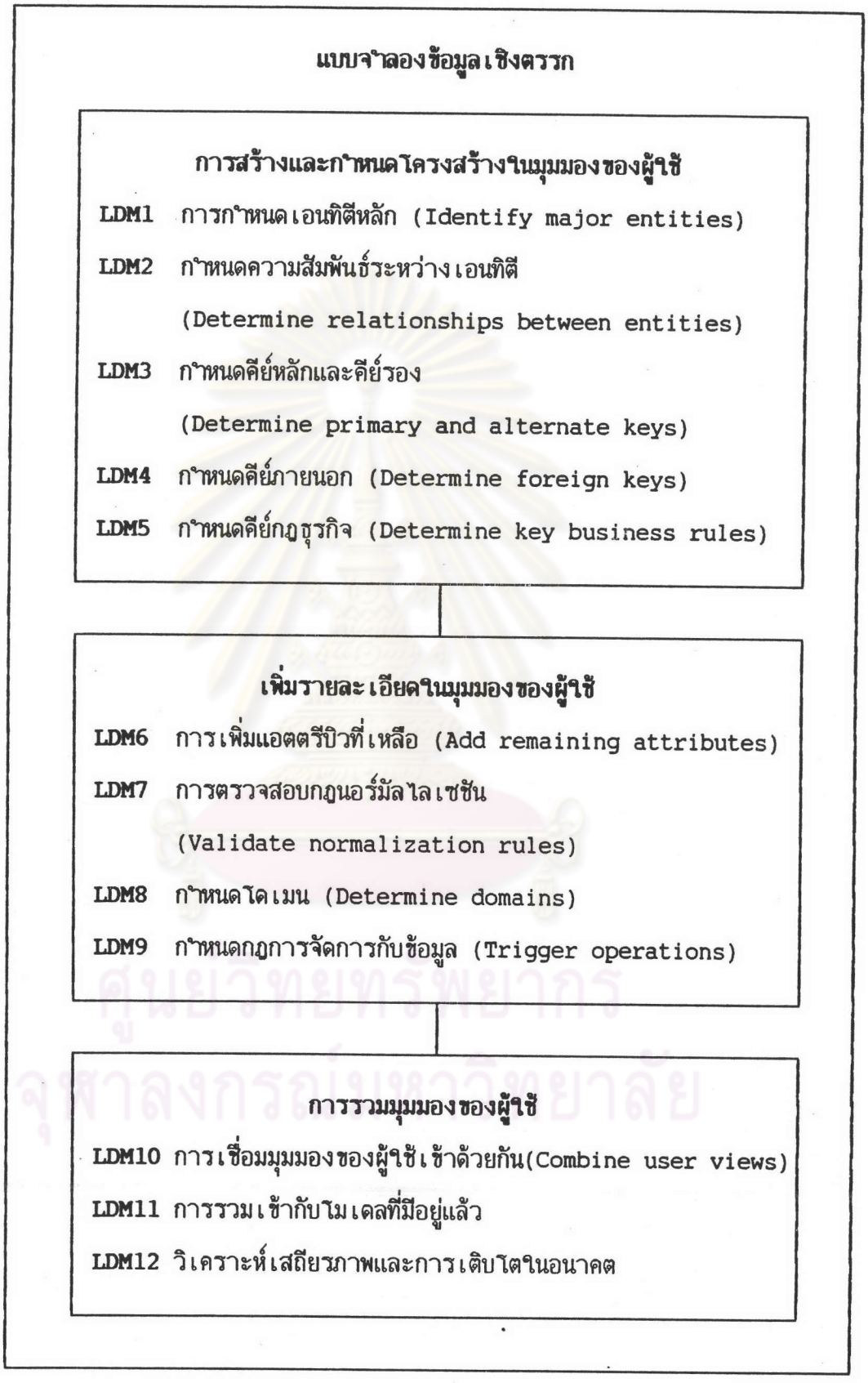
การออกแบบ ฐานข้อมูลเชิงสัมพันธ์ มีลักษณะคล้าย การออกแบบฐานข้อมูลชนิดอื่น ๆ (Fleming, Candace and Von Halle, 1989) โดยสามารถแบ่งขั้นตอนการออกแบบได้ดังต่อไปนี้

1) การวิเคราะห์ความต้องการ (Requirement Analysis) เป็นขั้นตอนรวบรวมพิจารณา และวิเคราะห์ข้อมูล โครงสร้างข้อมูล กระบวนการ (Process) ต่าง ๆ ในการดำเนินงาน การประยุกต์ใช้งานทั้งที่ใช้อยู่ในปัจจุบัน และคาดว่าจะเกิดขึ้นในอนาคต ตลอดจนการพิจารณาถึง ระเบียบ กฎเกณฑ์ ความต้องการของผู้ใช้ จุดประสงค์ของหน่วยงาน

2) การออกแบบฐานข้อมูลในระดับหลักการ (Conceptual Design) เป็นการสร้างรูปแบบจำลองข้อมูลเชิงตรรก (Logic Data Model) โดยในขั้นแรก จะทำการกำหนดเอนทิตี (Entity) ที่จำเป็นในระบบ พิจารณาความสัมพันธ์ระหว่างเอนทิตี วิเคราะห์คุณสมบัติของแต่ละเอนทิตีประกอบด้วยแอตตรีบิวอะไร และกำหนดคีย์ต่างๆ ที่เกี่ยวข้อง เช่น คีย์หลัก (Primary key) คีย์รอง (Secondary key) คีย์นอก (Foreign Key) หลังจากนั้นจะทำการ นอร์มัลไลซ์ (Normalize) เพื่อให้ได้แบบจำลองข้อมูลที่มีความเหมาะสม ถูกต้อง ไม่ขัดแย้งกัน ลดความซ้ำซ้อนและนำไปใช้งานได้ง่ายไม่ยุ่งยาก ด้วยเหตุนี้จึงควรทำนอร์มัลไลซ์ไม่ต่ำกว่าระดับที่ 3 (3NF) สำหรับขั้นตอนการออกแบบจำลองข้อมูลเชิงตรรกนี้พอสรุปได้ดัง รูป 3.2

3) การเลือกระบบจัดการฐานข้อมูล (Determination DBMS) การดำเนินงานการเลือกระบบจัดการฐานข้อมูลนี้ นอกจากจะพิจารณาเลือกประเภท ระบบจัดการฐานข้อมูล อันได้แก่ระบบจัดการฐานข้อมูล แบบแตกสาขา แบบความสัมพันธ์ และแบบเครือข่าย ยังต้องพิจารณาถึงเทคนิค และค่าใช้จ่ายต่าง ๆ ในการดำเนินงานดังนี้

   - ราคาซอฟต์แวร์ ซึ่งจะต้องพิจารณาถึงจำนวนผู้ใช้ (user) ความสะดวกในการเรียกใช้ข้อมูล ความสามารถในการจัดการรักษาความลับความปลอดภัยของข้อมูล ความเชื่อถือไว้วางใจได้ของระบบ ตลอดจนการประยุกต์ใช้งานต่าง ๆ และการใช้ร่วมกับโปรแกรมภาษาระดับสูงทั่วไป
   - ราคาของฮาร์ดแวร์ซึ่งต้องพิจารณาถึงอุปกรณ์ต่างๆ ทั้งอุปกรณ์หลักและอุปกรณ์ประกอบรอบด้านเช่นคอมพิวเตอร์ เทอร์มินัล หน่วยความจำหลัก หน่วยความจำรอง เครื่องพิมพ์ ฯลฯ นอกจากนี้ยังต้องพิจารณาถึงอายุการใช้งาน และการใช้งานร่วมกับระบบเก่ามากน้อยเพียงใด
   - ค่าใช้จ่ายในการ ปรับปรุง บำรุง รักษา ระบบ
   - ค่าใช้จ่ายในการสร้างระบบใหม่ และการเปลี่ยนจากระบบเก่าไปสู่ระบบใหม่
   - ค่าใช้จ่ายทางด้านบุคลากร รวมถึงการฝึกหัดอบรมต่าง ๆ



รูป 3.2 ขั้นตอนการออกแบบจำลองข้อมูลเชิงตรรก

4) การออกแบบการติดตั้ง (Implementation Design) เป็นการพิจารณา และวางแนวทางในการแปลงหรือวาง (Mapping) ระบบฐานข้อมูลในระดับหลักการให้เป็นระบบฐานข้อมูลจริง ซึ่งการดำเนินงานในขั้นตอนนี้ อันดับแรกจะทำการแปลงแบบจำลองฐานข้อมูลโดยอิสระจากชนิดของระบบจัดการฐานข้อมูล คือทำการแปลงให้อยู่ในแบบฐานข้อมูลที่เลือกไว้ในตอนต้น โดยไม่สนใจข้อจำกัด หรือเงื่อนไขพิเศษของชนิด หรือลักษณะของฐานข้อมูล จากนั้นจึงทำการปรับปรุงแก้ไขให้สอดคล้องและใช้ได้กับระบบจัดการฐานข้อมูล ตามที่เลือกไว้ดังกล่าว ตัวอย่างเช่น การแปลงแบบจำลองข้อมูล ซึ่งประกอบด้วย เอนทิตี ทูเปิล แอตตรีบิว ความสัมพันธ์ระหว่างเอนทิตี ฯลฯ เมื่อแปลงเข้าสู่ระบบฐานข้อมูลเชิงสัมพันธ์ก็จะได้ ตาราง (Table) แถวข้อมูล (Row) หลักข้อมูล (Column) คีย์ต่าง ๆ เป็นต้น

5) การออกแบบในระดับกายภาพ (Physical Design) เป็นการออกแบบ ที่พิจารณาถึงประสิทธิภาพ (Performance) ของระบบเช่นเวลาในการตอบสนอง (Response time) ปริมาณงานที่สามารถดำเนินการสำเร็จในหนึ่งหน่วยเวลา (Throughput) การใช้ที่ว่างในหน่วยความจำ (Space Utilization) การใช้หน่วยประมวลผลกลาง (CPU Utilization) และอื่นๆ นอกจากนี้ยังต้องพิจารณากำหนดทำเลที่ตั้งของระบบ เพื่ออำนวยความสะดวกต่อผู้ใช้ ทั้งในยามปกติ และยามฉุกเฉินจำเป็น

<u>การนอร์มัลไลเซชัน (Normalization)</u>

การนอร์มัลไลเซชัน เป็นขบวนการพัฒนา ตรวจสอบ และกลั่นกรองแบบจำลองข้อมูลในรูปแบบของตารางความสัมพันธ์ หรือบางครั้งมักเรียกโดยย่อว่า **ความสัมพันธ์** ให้เป็นไปตามนิยามของการนอร์มัลไล จุดประสงค์เพื่อ

- ลดปัญหาการซ้ำซ้อนของข้อมูล
- ลดที่ว่างที่ต้องใช้ในการจัดเก็บข้อมูล
- ลดความผิดพลาด ความไม่ตรงกันของข้อมูลในฐานข้อมูล
- ลดปัญหาการเกิดอะนอร์มัลไลของการลบ และการแก้ไขข้อมูล
- เพิ่มความคงทนให้แก่โครงสร้างของฐานข้อมูล

การนอร์มัลไลเซชันแบ่งได้เป็นหลายระดับ ซึ่งแต่ละระดับจะเป็นไปตามนิยาม (Date,1986) ดังนี้

นอร์มัลฟอร์มระดับที่ 1 (First Normal Form : 1NF)

ความสัมพันธ์ใดที่มีทุกทูเปิล ในความสัมพันธ์มีค่าแอตตรีบิว เป็นค่าเดียวโดด ๆ ไม่เป็นกลุ่มหรืออยู่ซ้ำซ้อนกัน แต่อาจไม่มีค่าเลย (Null) ก็ได้ ยกเว้นแอตตรีบิวที่เป็นคีย์หลัก จะเรียกความสัมพันธ์นี้ว่าความสัมพันธ์นอร์มัลฟอร์มระดับที่ 1 (1NF)

นอร์มัลฟอร์มระดับที่ 2 (Second Normal Form : 2NF)

ความสัมพันธ์ใดที่อยู่ในนอร์มัลฟอร์มระดับที่ 1 และมีทุกแอตตรีบิวที่ไม่ใช้คีย์หลักขึ้นตรงต่อคีย์หลักอย่างสมบูรณ์ จะเรียกความสัมพันธ์นี้ว่า ความสัมพันธ์นอร์มัลฟอร์มระดับที่ 2

นอร์มัลฟอร์มระดับที่ 3 (Third Normal Form : 3NF)

ความสัมพันธ์ที่อยู่ในนอร์มัลฟอร์มระดับที่ 3 จะต้องเป็นความสัมพันธ์นอร์มัลฟอร์มระดับที่ 2 และมีแอตตรีบิวทุกแอตตรีบิวที่เป็น ตัวเลือก (Determinant) จะต้องเป็นคีย์คู่แข่ง (Candidate Key) ตัวอย่างเช่น ความสัมพันธ์

นักศึกษา-อ.ที่ปรึกษา(<u>รหัสนักศึกษา</u>, ชื่อ-สกุลนักศึกษา, รหัส อ.ที่ปรึกษา, ชื่อ-สกุล อ.ที่ปรึกษา)

โดยกำหนดเงื่อนไขให้นักศึกษามีอาจารย์ที่ปรึกษาเพียงคนเดียวและไม่มี ชื่อ-สกุล ของนักศึกษาหรืออาจารย์ที่ปรึกษาซ้ำกัน ดังนั้นทุกแอตตรีบิว จึงขึ้นกับคีย์หลักคือรหัสนักศึกษา หรือจะสรุปได้ว่าความสัมพันธ์นี้อยู่ใน นอร์มัลฟอร์มระดับที่ 2 แต่จะสังเกตเห็นว่า รหัสอาจารย์ที่ปรึกษาสามารถระบุ หรือเลือก (Determine) ชื่อ-สกุลอาจารย์ที่ปรึกษาได้ ดังนั้นรหัสอาจารย์ที่ปรึกษา จึงเป็นตัวเลือก และรหัสอาจารย์ที่ปรึกษา ก็ไม่ใช่คีย์คู่แข่งเนื่องจากขาดคุณสมบัติ ฟังก์ชันการขึ้นต่อกัน เช่น คีย์หลัก ดังนั้นหากต้องการให้ความสัมพันธ์นี้ อยู่ในนอร์มัลฟอร์มระดับที่ 3 ต้องแตกเป็น 2 ความสัมพันธ์คือ

นักศึกษา(<u>รหัสนักศึกษา</u>, ชื่อ-สกุลนักศึกษา, รหัส อ.ที่ปรึกษา)

อ.ที่ปรึกษา(<u>รหัส อ.ที่ปรึกษา</u>, ชื่อ-สกุล อ.ที่ปรึกษา)

สำหรับนอร์มัลฟอร์มระดับที่ 3 นี้ บอยค์-คอดต์ เป็นผู้กำหนดนิยาม และได้บัญญัติชื่อขึ้นมาใหม่ว่า บอยค์คอดด์นอร์มัลฟอร์ม (Boyce/Codd Normal Form : BCNF) หรือการนอร์มัลฟอร์มในระดับ BCNF นั้นเอง

นอร์มัลฟอร์มระดับที่ 4 (Fourth Normal Form : 4NF)

ในบางครั้งความสัมพันธ์ แม้จะอยู่ในนอร์มัลฟอร์มระดับที่ 3 แต่แอตตรีบิว ภายในความสัมพันธ์ยังมีลักษณะการขึ้นต่อกันเชิงกลุ่ม (Multivalued Dependency) เช่น ความสัมพันธ์

นักศึกษา(<u>รหัสนักศึกษา</u>, <u>อ.ที่ปรึกษาวิทยานิพนธ์</u>, <u>รหัสวิชาลงทะเบียน</u>)

ซึ่งจะเห็นว่าจะเป็นปัญหาในการ เพิ่มเติม แก้ไข หรือลบข้อมูล ดังนั้นจึงแตกความสัมพันธ์เป็นสอง

ความสัมพันธ์คือ

นักศึกษา-อ.ที่ปรึกษา(รหัสนักศึกษา,อ.ที่ปรึกษาวิทยานิพนธ์)

นักศึกษา-การลงทะเบียน(รหัสนักศึกษา,รหัสวิชาลงทะเบียน)

ซึ่งความสัมพันธ์ที่อยู่ในนอร์มัลฟอร์มระดับที่ 3 และไม่มีลักษณะการขึ้นต่อกันเชิงกลุ่มนี้เรียกว่า ความสัมพันธ์นอร์มัลฟอร์มระดับที่ 4

นอร์มัลฟอร์มระดับที่ 5 (Fifth Normal Form : 5NF)

ความสัมพันธ์ที่อยู่ใน นอร์มัลฟอร์มระดับที่ 5 คือความสัมพันธ์ ที่ไม่สามารถนำไปสร้างความสัมพันธ์ใหม่ โดยการเชื่อม 2 ความสัมพันธ์ด้วยคีย์ที่ต่างกัน ซึ่งปกติมักจะเป็นปัญหาสำหรับคีย์หลัก ที่เป็นคีย์ประกอบ ขั้นตอนนี้เป็นขั้นตอนพิจารณาได้ยาก และมักเกิดระเบียนข้อมูลใหม่ที่ไม่มีอยู่จริงขึ้นมา เมื่อนำเอนทิตีมารวมกัน

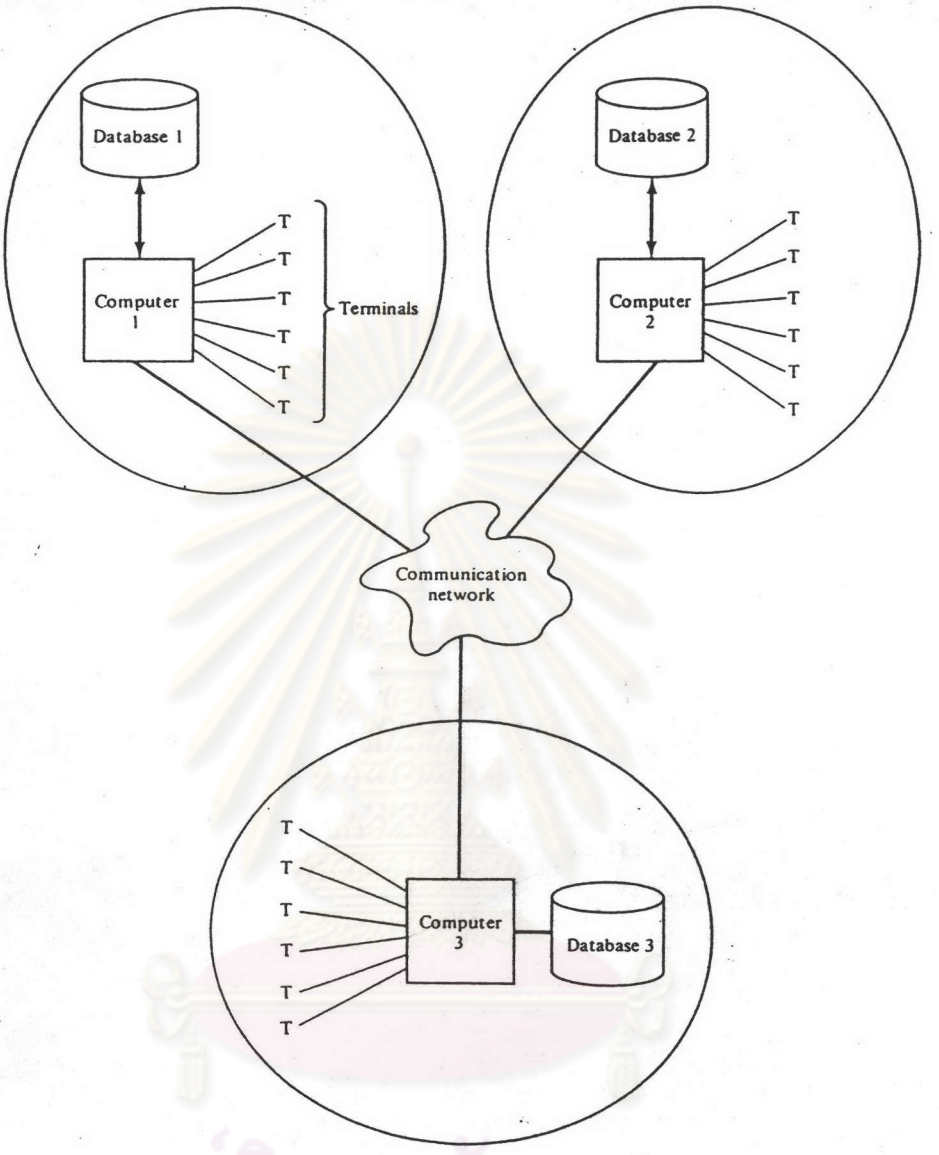
ระบบฐานข้อมูลแบบกระจาย (Distributed Database System)

ระบบฐานข้อมูลแบบกระจาย เป็นระบบกระจายข้อมูล ไปจัดเก็บตามสาขาต่าง ๆ ของแหล่งข้อมูล หรือตามความต้องการของผู้ใช้ และอาจเก็บข้อมูลซ้ำในหลาย ๆ สาขาเพื่อสะดวกในการเรียกใช้ โดยที่แหล่งข้อมูลเหล่านี้ยังมีระบบการจัดการฐานข้อมูล (DBMS) ของตัวเอง และสามารถเชื่อมโยงกับระบบฐานข้อมูลอื่นผ่านการจัดการของ ระบบการจัดการฐานข้อมูลแบบกระจาย (DDBMS) และระบบเครือข่ายของการสื่อสารข้อมูลดัง รูป 3.3

ลักษณะระบบฐานข้อมูลแบบกระจายสามารถแบ่งออกได้หลายชนิดขึ้นอยู่กับองค์ประกอบต่าง ๆ ดังนี้ (Ceri and Pelagatti, 1985)

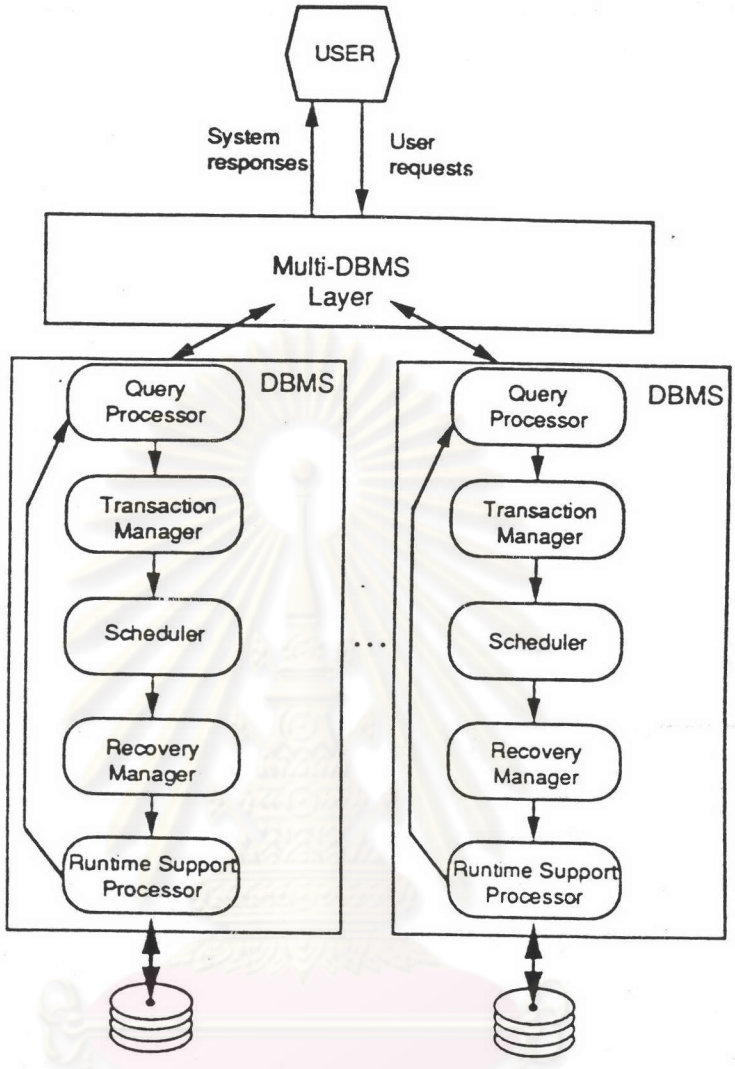
องค์ประกอบที่ 1 พิจารณาถึงความแตกต่างของระบบจัดการฐานข้อมูล ถ้าระบบจัดการฐานข้อมูลที่ติดตั้งตามที่ต่าง ๆ มีลักษณะเหมือนกันหมด ระบบฐานข้อมูลนี้จะเรียกว่า ระบบเดี่ยว (Homogeneous) การจัดการเกี่ยวกับระบบคงไม่ซับซ้อนมากนัก ถ้าระบบมีการใช้ ระบบจัดการฐานข้อมูลต่างชนิดกันจะเรียกว่า ระบบต่างกัน (Heterogeneous) ซึ่งถ้าระบบจัดการฐานข้อมูลที่ติดตั้งตามที่ต่าง ๆ มีความต่างกันมาก ๆ ระบบเกี่ยวกับการจัดการฐานข้อมูลแบบกระจาย ก็จะยิ่งทำงานซับซ้อนมากขึ้น

องค์ประกอบที่ 2 เป็นการพิจารณาถึงลักษณะการจัดเก็บข้อมูล ความอิสระในการเรียกใช้ข้อมูลในที่ต่าง ๆ การจัดเก็บข้อมูลสำหรับผู้ใช้เฉพาะสาขา (Local user) เป็นการจัดเก็บ



รูป 3.3 ระบบฐานข้อมูลแบบกระจาย

เพื่อให้ผู้ใช้ประจำแต่ละสาขา เรียกใช้เฉพาะในส่วนของตน (Local Application) โดยไม่มีการจัดเก็บข้อมูลซ้ำไว้ที่อื่นเลย สำหรับการจัดเก็บข้อมูลเพื่อเรียกใช้ร่วมกัน (Global Application) ผู้ใช้จะต้องเกี่ยวข้องกับฐานข้อมูลสาขาอื่น ในกรณีนี้ข้อมูลจะไม่จัดเก็บเฉพาะในสาขาของตน แต่จะจัดเก็บตามที่ต่าง ๆ ตามความเหมาะสม ส่วนความอิสระในการเรียกใช้ข้อมูล ขึ้นอยู่กับการจัดการของระบบฐานข้อมูลแบบกระจาย จะยินยอมให้แต่ละสาขามีอำนาจในการจัดการข้อมูลใน



รูป 3.4 ระบบฐานข้อมูลแบบกระจายที่มีระบบจัดการฐานข้อมูลหลายชนิด

ส่วนของตนมากน้อยเพียงใด ถ้าให้อำนาจเต็มที่ แต่ละสาขาจะมีผู้ดูแลฐานข้อมูลของตนเอง (Data base Administrator) สามารถตั้งกฎเกณฑ์ในการเรียกใช้ ดำเนินงานเอง โดยไม่ต้องผ่านระบบจัดการฐานข้อมูลแบบกระจาย ซึ่งระบบแบบนี้เรียกว่า ระบบแบบสหพันธ์ (Federate DDBMS หรือ Multi Database System) ระบบสหพันธ์จะดำเนินงานเหมือนระบบฐานข้อมูลรวม (Centralized Database System) ทุกประการ สำหรับการใช้ภายในสาขา และจะดำเนินงานแบบกระจายก็ต่อเมื่อใช้งานร่วมกัน

องค์ประกอบที่ 3 จะพิจารณาถึงความสามารถในการมองผ่าน (Transparency) ของระบบจัดการฐานข้อมูลแบบกระจาย ว่าสามารถช่วยอำนวยความสะดวกต่อผู้ใช้มากน้อยเพียงใด ระบบจัดการฐานข้อมูลแบบกระจาย บางระบบ ผู้ใช้จำเป็นต้องรับทราบข้อมูลที่ต้องการอยู่ที่สาขาใด และต้องระบุแหล่งข้อมูลลงในคำสั่งเรียกใช้ บางระบบผู้ใช้ไม่จำเป็นต้องรับทราบ ระบบจะจัดการดำเนินงานให้ ผู้ใช้จะเห็นเสมือนเป็นระบบฐานข้อมูลรวม นอกจากนี้ระบบจัดการฐานข้อมูลแบบกระจาย ยังมีความสามารถแตกต่างกันในด้านการจัดการ ดำเนินงานตามคำสั่งต่าง ๆ ระบบที่ดีกว่าจะสามารถเลือกเส้นทาง วิธีการที่ดีกว่า และมีการเคลื่อนย้ายข้อมูลน้อยกว่า หรือที่เรียกว่ามีออพทิไมเซชันทั้งในระดับสาขา และระดับทั่วไป (Local and Global Optimization) ดีกว่านั้นเอง

โครงสร้างทางสถาปัตย์ของระบบฐานข้อมูลแบบกระจาย (Distributed DBMS Architecture)

ถึงแม้ว่าระบบฐานข้อมูลแบบกระจาย ที่มีอยู่ในปัจจุบันจะมีลักษณะที่แตกต่างกัน แต่โดยทั่วไประบบฐานข้อมูลแบบกระจาย จะมีโครงสร้างทางสถาปัตย์ที่คล้ายกันดัง รูป 3.5 (Ceri and Pelagatti, 1985) และมีองค์ประกอบที่สำคัญดังนี้

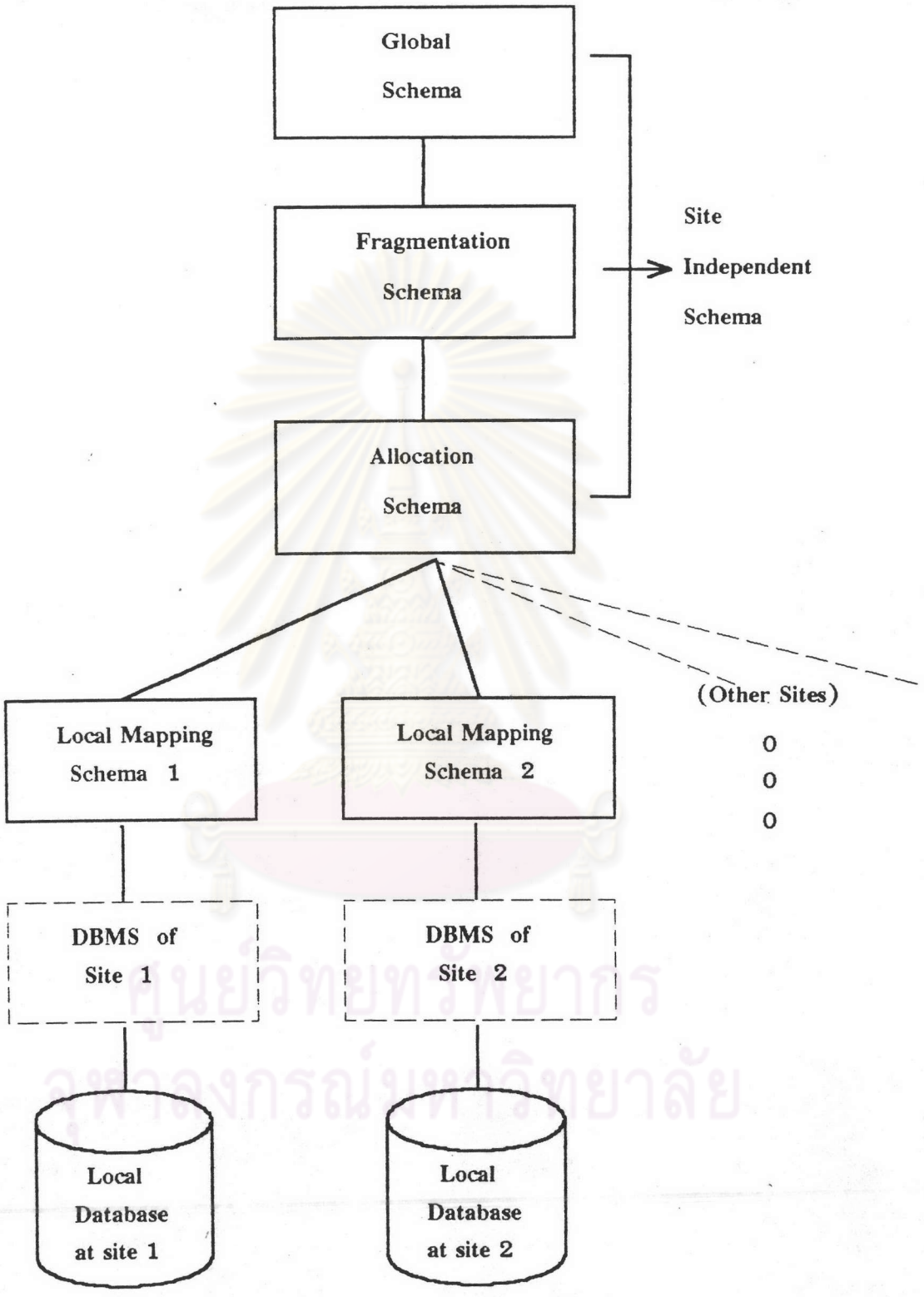
เค้าร่างข้อมูลรวม (Global Schema) : เป็นการนิยามเค้าร่างข้อมูลทั้งหมด ในระบบฐานข้อมูล เช่น การกำหนดลักษณะ องค์ประกอบความสัมพันธ์ของข้อมูล

เค้าร่างข้อมูลส่วนกระจาย (Fragmentation Schema) : เป็นการนิยามส่วนกระจายของเค้าร่างข้อมูล ที่สามารถแยกส่วน ออกจากเค้าร่างข้อมูลรวมของทั้งระบบดัง รูป 3.6 ซึ่งในการแยกส่วนเค้าร่างข้อมูลนี้ สามารถแยกตาม แนวตั้ง(Vertical Fragmentation) ตามแนวนอน (Horizontal Fragmentation) หรือทั้งแนวตั้ง และแนวนอน (Mixed Fragmentation)

เค้าร่างการจัดเก็บข้อมูล (Allocation Schema) : เป็นการกำหนดการจัดเก็บ ข้อมูลส่วนกระจาย (Fragmentation) ตามที่ต่าง ๆ ซึ่งข้อมูลส่วนกระจายหนึ่ง อาจมีการจัดเก็บซ้ำซ้อนในหลายสาขา

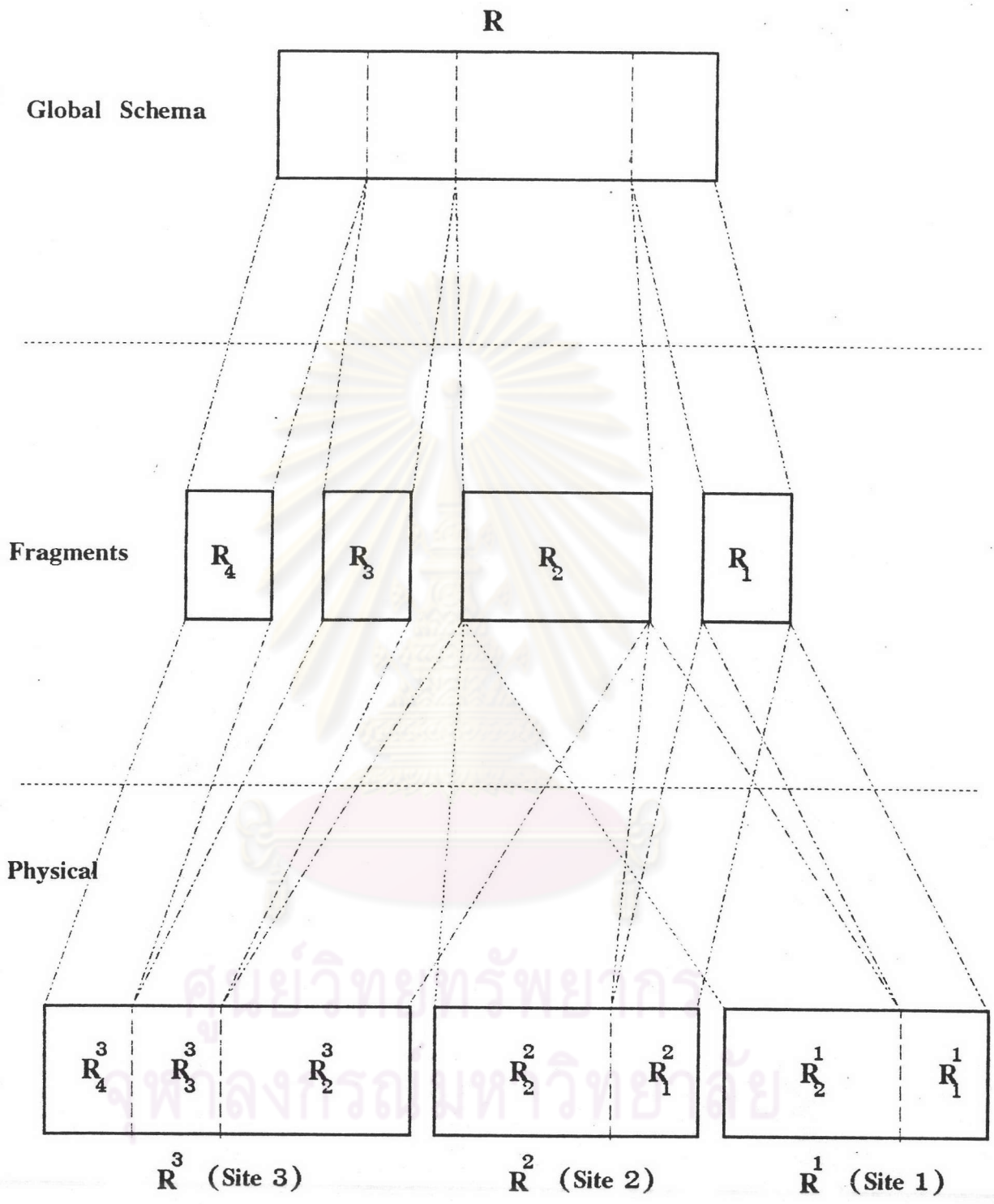
เค้าร่างข้อมูลเฉพาะสาขา (Local Schema): เป็นการนิยามเค้าร่างข้อมูลตามฐานข้อมูลประจำสาขา ซึ่งอาจประกอบด้วยหลาย ๆ ข้อมูลส่วนกระจาย

พจนานุกรมข้อมูลหรือสารบบข้อมูลของทั้งระบบ (Global Directory/Dictionary : GD/D): เป็นที่รวบรวมรายละเอียด ตารางข้อมูล ประเภทตัวแปร ตลอดจนสถานที่จัด



รูป 3.5 สถาปัตยกรรมของระบบฐานข้อมูลแบบกระจาย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูป 3.6 เค้าร่างข้อมูลและเค้าร่างข้อมูลส่วนกระจาย

เก็บข้อมูล เพื่อใช้ในการอ้างอิง การเข้าถึงทางกายภาพของข้อมูล

พจนานุกรมข้อมูลหรือสารบบข้อมูลประจำสาขา (Local Directory/Dictionary : LD/D) : เป็นบางส่วน หรือทั้งหมดของพจนานุกรมข้อมูลของระบบ ซึ่งจะมีการกระจายการจัดเก็บตามที่ต่าง ๆ และอาจมีการจัดเก็บซ้ำซ้อนขึ้นอยู่กับความเหมาะสมความต้องการในการเรียกใช้ข้อมูล

เค้าร่างเครือข่าย (Network Schema) : เป็นการ นิยามเค้าร่างเครือข่ายในการติดต่อสื่อสาร รวมถึงรูปแบบข้อตกลงในการรับส่งข่าวสารข้อมูล

ระบบจัดการฐานข้อมูลแบบกระจาย (Distributed Database Management System : DDBMS) : เป็นซอฟต์แวร์ที่ใช้ในการจัดการฐานข้อมูลแบบกระจาย

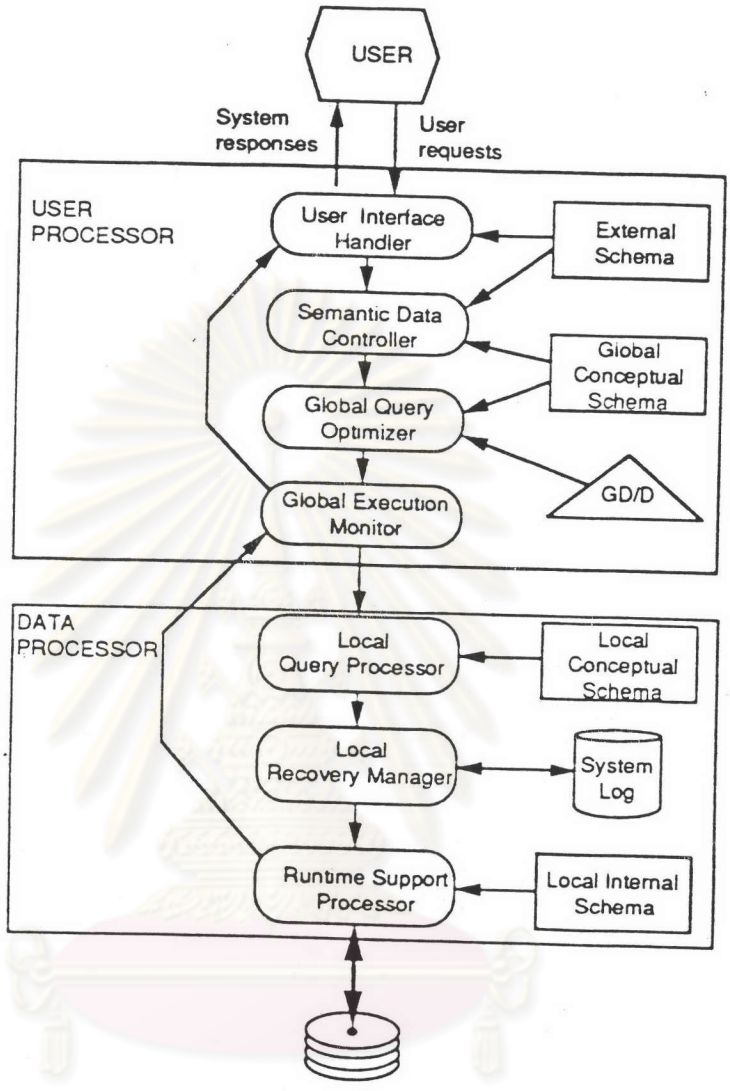
ระบบจัดการฐานข้อมูล (Database Management System : DBMS) : เป็นซอฟต์แวร์ที่ใช้ในการจัดการฐานข้อมูลแบบรวม (Centralized Database System) หรือเฉพาะสาขาในระบบฐานข้อมูลแบบกระจาย

ข้อมูล (Data) : เป็นข้อเท็จจริง ซึ่งอาจเป็นได้ทั้งรูปธรรม หรือนามธรรมที่สามารถนำมาดำเนินการแล้วได้คำตอบตามความต้องการของผู้ใช้

จากรูปที่ 3.5 จะเห็นว่า 3 ระดับบนสุดจะเป็นอิสระจากสาขาย่อยของระบบ และจะอยู่ภายใต้การกำกับดูแลของระบบจัดการฐานข้อมูลแบบกระจาย สำหรับระดับล่างถัดลงไปจะเกี่ยวข้องเฉพาะสาขา และมีหน้าที่ในการดำเนินงาน เพื่อให้ได้คำตอบตามความต้องการของผู้ใช้ ขบวนการนี้เรียกว่าการแปลงลงสู่สาขา (Local Mapping) ดำเนินการโดยระบบจัดการฐานข้อมูล ซึ่งแต่ละระบบจัดการฐานข้อมูล อาจมีลักษณะที่แตกต่างกันตามแต่ละสาขา สำหรับการดำเนินงานในระบบฐานข้อมูลแบบกระจาย (Tramer and Patrick, 1991) พอสรุปได้ดัง รูป 3.7 และการดำเนินงานนี้ จะเป็นไปอย่างมีประสิทธิภาพก็ต่อเมื่อ ระบบจัดการฐานข้อมูลแบบกระจายจำเป็นต้องมีความสามารถในการมองผ่าน (Transparency) ซึ่งมีรายละเอียดดังต่อไปนี้

ความอิสระจากข้อมูล (Data Independent): เป็นการมองผ่านในระดับพื้นฐานของระบบจัดการฐานข้อมูล โดยผู้ใช้ไม่จำเป็นต้องรับทราบวิธี และลักษณะการจัดเก็บข้อมูล อีกทั้งการเปลี่ยนแปลงแก้ไขทางตรรก ในการดำเนินงานของผู้ใช้ ก็ไม่มีผลประการใดต่อข้อมูล ทั้งนี้เพราะระบบจัดการฐานข้อมูลจะเป็นผู้ดำเนินการให้

การมองผ่านข้อมูลส่วนการกระจาย (Fragmentation Transparency) : ผู้ใช้จะมีความรู้สึกเสมือนหนึ่งว่าข้อมูลทั้งหมดที่กำลังใช้อยู่ไม่มีการแบ่งแยกและรวมอยู่ที่เดียวกัน



รูป 3.7 กระบวนการดำเนินงานของระบบฐานข้อมูลแบบกระจาย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การมองผ่านการจัดเก็บซ้ำซ้อน (Replication Transparency): ในบางครั้งข้อมูลชุดเดียวกันอาจจัดเก็บไว้ในหลาย ๆ ที่ เพื่อความสะดวกรวดเร็วในการเรียกใช้ แต่สำหรับผู้ใช้ จะไม่รับรู้ในการจัดเก็บข้อมูล ซ้ำซ้อนเหล่านี้

การมองผ่านสถานที่ตั้ง (Location Transparency) : เป็นการมองผ่านสถานที่จัดเก็บข้อมูลที่กระจายตามที่ต่าง ๆ ผู้ใช้จะมีความรู้สึกเหมือนว่า เขากำลังใช้ข้อมูลที่เก็บในที่เขากำลังทำงานอยู่

การมองผ่านระบบจัดการฐานข้อมูล (Local Mapping Transparency) : เป็นการมองผ่านความแตกต่างของระบบจัดการฐานข้อมูล ที่ติดตั้งตามที่ต่างๆ ของระบบฐานข้อมูลแบบกระจาย ซึ่งระบบจัดการฐานข้อมูลแบบกระจาย มีหน้าที่ในการดำเนินงานให้ระบบจัดการฐานข้อมูลเหล่านี้สามารถเชื่อมต่อและรับส่งข้อมูลข่าวสารระหว่างกันได้

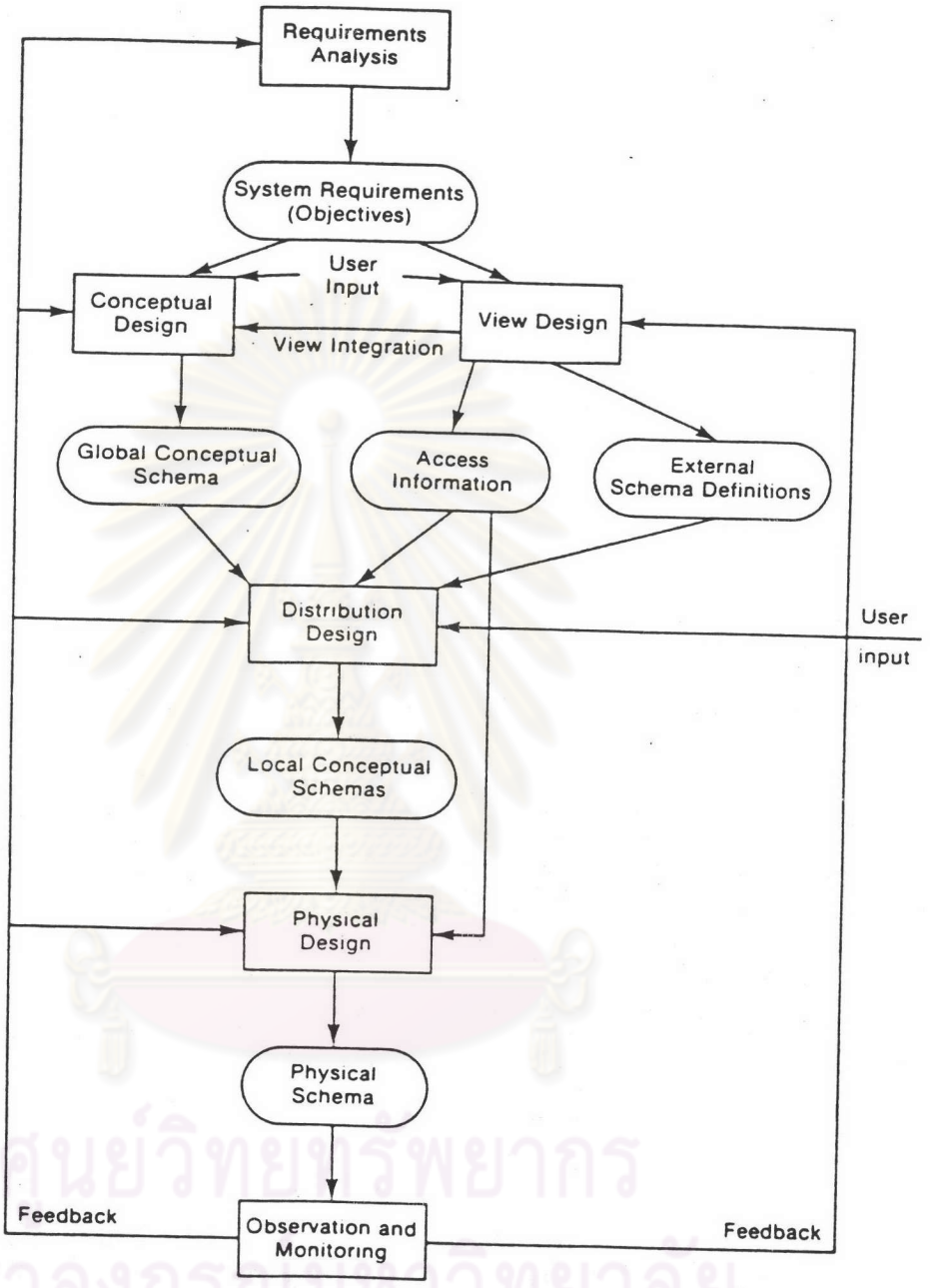
การออกแบบฐานข้อมูลกระจาย (Distribution Database Design)

การออกแบบฐานข้อมูลแบบกระจาย มีลักษณะที่กว้างขวาง และเกี่ยวข้องกับสิ่งแวดล้อมรอบด้านมากกว่าฐานข้อมูลแบบรวม ในการออกแบบฐานข้อมูลแบบรวมจะมีการดำเนินการออกแบบในระดับหลักการและในระดับกายภาพดังได้กล่าวมาแล้ว สำหรับการออกแบบฐานข้อมูลแบบกระจายก็มีลักษณะ เทคนิค และวิธีการคล้ายกัน จะมีส่วนที่เพิ่มเติมคือการออกแบบแยกส่วนกระจายของข้อมูลและการพิจารณากำหนดสถานที่ในการจัดเก็บข้อมูล การออกแบบฐานข้อมูลแบบกระจายแบ่งออกได้เป็น 2 ลักษณะ (Tramer and Patrick, 1991) คือ

การออกแบบฐานข้อมูลแบบบนลงล่าง (Top-Down Design)

การออกแบบฐานข้อมูลแบบ บนลงล่าง เป็นการออกแบบจาก ส่วนรวมของทั้งระบบ สู่รายละเอียดปลีกย่อยตามแต่ละสาขา การออกแบบลักษณะนี้เหมาะสำหรับฐานข้อมูลใหม่ หรือระบบฐานข้อมูลที่มีลักษณะเหมือนกันทุกที่ (Homogeneous) รายละเอียดการดำเนินการแสดงดังรูป 3.8 และมีขั้นตอนพอสรุปได้ดังนี้

1) การวิเคราะห์ความต้องการ (Requirement Analysis): เป็นขั้นตอนแรกในการดำเนินงานซึ่งเป็นขั้นรวบรวมข้อมูล พิจารณากระบวนการทั้งหมดที่ผู้ใช้ดำเนินการอยู่ การวิเคราะห์ความต้องการ เป็นบรรทัดฐานในการวิเคราะห์และออกแบบฐานข้อมูลมาตรฐาน ที่มีประสิทธิภาพ (Performance) ความพร้อม ความเชื่อถือ (Reliability and Availability) ความยืดหยุ่น (Flexibility) และยังประหยัดค่าใช้จ่ายในการดำเนินการจริง
2) การออกแบบในระดับมุมมองของผู้ใช้ (View Design): เป็นขั้นตอนนำเอาผลในข้อหนึ่งมาวิเคราะห์ และนิยามความต้องการของผู้ใช้ และส่วนที่ผู้ใช้เกี่ยวข้อง เช่น ขั้นตอนการปฏิบัติ ขอบเขตหน้าที่ความรับผิดชอบ และอำนาจในการเข้าถึง หรือการ



รูป 3.8 กระบวนการออกแบบฐานข้อมูลจากบนลงล่าง

แก้ไขข้อมูล

3) การออกแบบในระดับหลักการ(Conceptual Design): เป็นการพิจารณาโครงสร้าง ลักษณะข้อมูลมี เอนทิตีอะไร ประกอบด้วย แอตตรีบิวอะไร ตลอดจนหน้าที่ ความสัมพันธ์ระหว่างเอนทิตีเหล่านี้ ขั้นตอนนี้จะดำเนินควบคู่พร้อมกับการออกแบบในระดับ

มุมมองของผู้ใช้ เพื่อออกแบบ พัฒนาเค้าร่างข้อมูลทั้งหมด (Global Conceptual Schema) ให้สอดคล้องตามความต้องการของผู้ใช้ และเพื่อให้ครอบคลุมทุกกระบวนการที่เกิดขึ้นทั้งในปัจจุบัน และอนาคต

4) การออกแบบการกระจาย(Distributed Design): เป็นขั้นตอนที่สำคัญในการออกแบบ ฐานข้อมูลแบบกระจาย ขั้นตอนนี้จะนำเค้าร่างข้อมูลทั้งระบบ มาพิจารณาว่าสามารถกระจายเป็น ข้อมูลส่วนกระจาย ได้อย่างไร ซึ่งอาจกระจายตาม แนวตั้ง แนวนอน หรือทั้งสองแนว ตามความเหมาะสม แต่ทั้งหมดนี้ยังคงรักษาไว้ซึ่งความสัมพันธ์ ตามลักษณะความสัมพันธ์ของทั้งระบบ (Global Relation) และเมื่อรวมข้อมูลส่วนกระจายเหล่านี้จะได้เค้าร่างข้อมูลของทั้งระบบเหมือนเดิม หลังจากนั้นเป็นการพิจารณาและกำหนดสถานที่จัดเก็บข้อมูล ซึ่งในแต่ละที่อาจเก็บหลายๆ ข้อมูลส่วนกระจาย และประกอบกันขึ้นเป็นเค้าร่างข้อมูลเฉพาะที่ (Local Conceptual Schema)

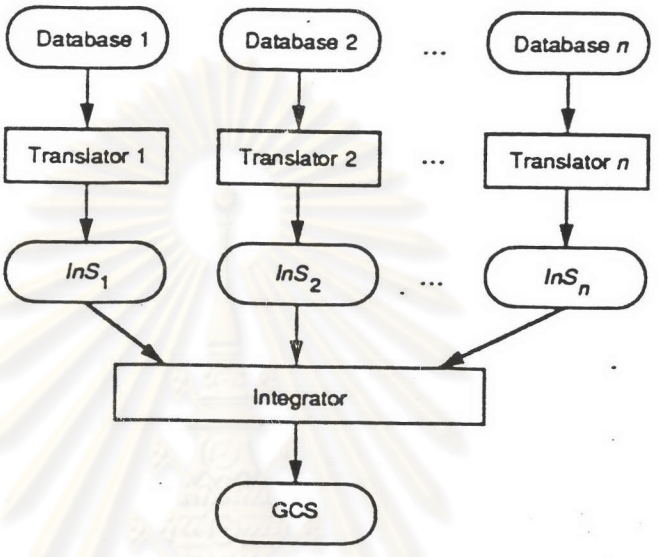
5) การออกแบบทางกายภาพ (Physical Design) : เป็นขั้นตอนการพิจารณานำเครื่องมือต่าง ๆ มาใช้ในการดำเนินงาน และการจัดเก็บข้อมูลในที่ต่าง ๆ (Map the Local Conceptual to Physical)ให้สอดคล้องตามเค้าร่างข้อมูลเฉพาะสาขาที่ได้ออกแบบมาแล้ว

6) การตรวจสอบและทบทวน (Observation and Monitering): ทบทวนขั้นตอนที่ผ่านมาหาจุดบกพร่อง และความไม่เหมาะสม นำไปปรับปรุงแก้ไขระบบให้สอดคล้องตามความต้องการ และจุดประสงค์ที่วางในตอนแรก

การออกแบบฐานข้อมูล แบบล่างขึ้นบน (Bottom-Up Design)

ในระบบงานที่มีฐานข้อมูลตามที่ต่าง ๆ อยู่แล้ว การออกแบบฐานข้อมูลกระจายแบบล่างขึ้นบน จะมีความสะดวก และเหมาะสมมากกว่า และถ้าระบบจัดการฐานข้อมูลที่ใช้ตามฐานข้อมูลมีลักษณะแตกต่างกัน การใช้วิธีการอื่นออกแบบจะเป็นสิ่งที่ยากลำบากยิ่ง การออกแบบจากล่างขึ้นบนเป็นการพิจารณา เค้าร่างข้อมูลในแต่ละสาขา มารวมกันเป็นเค้าร่างข้อมูลของทั้งระบบ ซึ่งในขั้นตอนนี้ อันดับแรกต้องพิจารณารายละเอียดโครงสร้าง ของข้อมูลส่วนกระจาย มีลักษณะอะไรร่วมกัน มีความสัมพันธ์กันอย่างไร และมีส่วนใดซ้ำซ้อน แล้วนำมาแปลง (Translation) ให้เป็นเค้าร่างที่มีลักษณะกลาง ๆ (Intermediate Schema : IS) ไม่ขึ้นอยู่กับระบบจัดการฐานข้อมูล หรือ

ลักษณะของแต่ละฐานข้อมูล ขั้นตอนนี้ไม่จำเป็นต้องดำเนินการ ถ้าฐานข้อมูลมีลักษณะเหมือน ๆ กันทุกที่ จากนั้นนำเค้าร่างที่ได้จากการแปลงมารวมกันเป็นเค้าร่างข้อมูลของทั้งระบบ โดยตัดส่วนที่ซ้ำซ้อนออก และสร้างโครงสร้างความสัมพันธ์ใหม่ให้ทั้งระบบ



รูป 3.9 กระบวนการออกแบบฐานข้อมูลจากล่างขึ้นบน

การออกแบบการกระจายการจัดเก็บข้อมูล (The Design of Distributed Data Allocation) (CODASYL, 1982)

การออกแบบฐานข้อมูล ไม่ว่าจะเป็นแบบบนลงล่าง หรือล่างขึ้นบน เมื่อสร้างแบบจำลองข้อมูลและแยกส่วนกระจายข้อมูลเป็นที่เรียบร้อย ขั้นตอนต่อไปคือการกำหนดสถานที่จัดเก็บฐานข้อมูลตามที่ต่าง ๆ ซึ่งลักษณะการจัดเก็บมีด้วยกัน 4 รูปแบบดังนี้

1) การจัดเก็บฐานข้อมูลแบบรวมศูนย์ (Centralized Database) เป็นการจัดเก็บฐานข้อมูลทั้งหมดไว้ ณ ที่เดียวกัน ซึ่งปกติจะจัดเก็บไว้ที่คอมพิวเตอร์ศูนย์กลาง โดยมีผู้ใช้กระจายตามที่ต่าง ๆ ตามความเหมาะสม และความจำเป็น ผู้ใช้สามารถเรียกใช้ข้อมูลผ่านทางระบบเครือข่าย และการเข้าถึงข้อมูลอาจแบ่งตามระดับผู้ใช้หรือตามลำดับการเข้าถึงก่อนหลัง ทั้งนี้ขึ้นอยู่กับนโยบายของแต่ละหน่วยงาน การจัดเก็บในลักษณะนี้ มีข้อดีคือสามารถควบคุมการใช้งานและการเข้าถึงข้อมูลได้ดีกว่า

I16726765

วิธีการอื่น อีกทั้งยังไม่จำเป็นต้องใช้เจ้าหน้าที่ดูแลรักษาฐานข้อมูลมากนัก แต่ลักษณะนี้ก็มีข้อเสียคือ หามีผู้ใช้เป็นจำนวนมาก ความเชื่อถือ ความไว้วางใจ รวมทั้งประสิทธิภาพในการดำเนินงานจะด้อยกว่าวิธีการอื่น และที่สำคัญคือหากที่ศูนย์กลางมีปัญหาจะทำให้ทั้งระบบหยุดชะงักงัน

2) การแยกการจัดเก็บฐานข้อมูลบางส่วน (Partitioned Database) เป็นการกระจายส่วน ฐานข้อมูลไปจัดเก็บตามที่ต่าง ๆ ตามความจำเป็นและความเหมาะสมในการใช้งาน โดยที่แต่ละส่วนจะไม่มีข้อมูลซ้ำซ้อน หรือเหมือนกัน การจัดเก็บข้อมูลในลักษณะนี้ เป็นการแก้ปัญหาที่เกิดขึ้น ในการจัดเก็บแบบรวมศูนย์ รวมถึงเพื่อลดค่าใช้จ่ายในการสื่อสารข้อมูล

3) การจัดเก็บฐานข้อมูลแบบเก็บซ้ำ (Replicated Database) เนื่องจากการจัดเก็บฐานข้อมูลบางส่วน ไม่สามารถรับรองได้ว่า ข้อมูลที่จัดเก็บตามที่ต่าง ๆ จะถูกเรียกใช้เฉพาะในส่วนของตน ดังนั้นหากมีการเรียกใช้จากที่อื่น ซึ่งอาจอยู่ห่างไกลมาก ปัญหาที่ตามมาคือความเชื่อถือ ความไว้วางใจ ประสิทธิภาพในการดำเนินงานจะลดน้อยถอยลง โดยเฉพาะปัญหา เวลาหน่วง (Delay Time) ในการสื่อสารข้อมูลย่อมเกิดขึ้นแน่นอน เพื่อแก้ปัญหาเหล่านี้ โดยเฉพาะปัญหาการหยุดชะงักงันของทั้งระบบ จึงให้มีการจัดเก็บฐานข้อมูลทั้งระบบไปยังที่ต่าง ๆ วิธีการนี้จะทำให้ระบบดำเนินงานได้อย่างมี ประสิทธิภาพ แต่สิ่งที่ต้องระวังเป็นอย่างยิ่ง คือ ปัญหาความสอดคล้อง และความคงสภาพของข้อมูล นอกจากนั้นการจัดเก็บข้อมูลในลักษณะนี้จะต้องใช้เนื้อที่หน่วยความจำและเจ้าหน้าที่ดูแลรักษาฐานข้อมูลมากกว่าวิธีการอื่น

4) การจัดเก็บฐานข้อมูลแบบผสม (Hybrid Database) เป็นลักษณะการจัดเก็บฐานข้อมูลแบบผสม ระหว่างการจัดเก็บแบบบางส่วนและการจัดเก็บแบบเก็บซ้ำ กล่าวคือ ข้อมูลส่วนใดใช้เฉพาะที่ ก็เก็บไว้เฉพาะ ในที่ของตน และหากส่วนใดใช้ร่วมกันก็จะเก็บซ้ำซ้อนไปยังที่ต่าง ๆ ตามความจำเป็น การจัดเก็บข้อมูลในลักษณะนี้ถือได้ว่าเป็นวิธีที่ก่อให้เกิดประสิทธิภาพที่สุดในการดำเนินงาน แต่การจัดเก็บตามลักษณะนี้ก็มีข้อเสียคือ ควบคุมข้อมูลยาก และมีความซับซ้อนมากกว่าวิธีการอื่น สำหรับการดำเนินงานครั้งนี้ การกำหนดการจัดเก็บฐานข้อมูล จะใช้หลักวิธีการนี้ เนื่องจากในระบบทะเบียนของสหวิทยาลัยรัตนโกสินทร์ จะมีข้อมูลบางส่วนใช้ร่วมกัน เช่นรายวิชาและหลักสูตร ฯลฯ ในขณะที่มีข้อมูลบางส่วนใช้เฉพาะในที่ของตน เช่น บุคลากร อาคารสถานที่ เป็นต้น

ผลดีผลเสียของระบบฐานข้อมูลแบบกระจายเมื่อเทียบกับฐานข้อมูลแบบรวม

(ดวงแก้ว สวามิภักดิ์, 2534; Ceri and Pelagatti, 1985)

ผลดี

1) ความสะดวกในการจัดการและประหยัดค่าใช้จ่าย (Organization and Economic) : ลดค่าใช้จ่ายในการดำเนินงาน เช่น การรับส่งข้อมูลข่าวสาร ค่าใช้จ่ายในการจัดหาเครื่องมือขนาดใหญ่ เพื่อใช้ในการดำเนินงานทั้งระบบ นอกจากนี้ยังสามารถรองรับกับการเติบโตของข้อมูล และความเจริญก้าวหน้าทางเทคโนโลยีในอนาคต

2) มีประสิทธิภาพมากกว่า (Improved Performance) : เนื่องจากมีทรัพยากรและข้อมูลใช้มากกว่า และถ้าการออกแบบฐานข้อมูล มีความสอดคล้องกับลักษณะการใช้งานจริง การเรียกใช้ข้อมูลก็ยิ่ง สะดวก รวดเร็ว เพราะข้อมูลที่ต้องการจะอยู่ในที่ของตน หรืออยู่ในที่ข้างเคียง

3) มีความพร้อม และความเชื่อถือมากกว่า (Improved Reliability/Availability): เพราะในระบบฐานข้อมูลแบบรวม หากคอมพิวเตอร์ส่วนกลางชำรุดเสียหาย จะมีผลทำให้งานทั้งระบบหยุดชะงัก เมื่อเทียบกับระบบแบบกระจาย ความเสียหายจะมีผลกระทบเฉพาะในสาขาที่คอมพิวเตอร์เกิดปัญหา

4) มีความยืดหยุ่นมากกว่า (Expandability) : สะดวกในการขยายขนาดของระบบ เช่น เมื่อต้องการเพิ่มฐานข้อมูล หรือเมื่อฐานข้อมูลมีขนาดใหญ่ขึ้น การขยายระบบคอมพิวเตอร์ เพื่อรองรับปริมาณงานที่มากขึ้น จะกระทำได้สะดวก โดยไม่กระเทือนงานในส่วนรวม

5) มีความอิสระมากกว่า (Local Autonomy) : สามารถควบคุมการใช้ข้อมูลภายในเฉพาะที่ดีกว่า เนื่องจากแต่ละสาขาจะจัดเก็บข้อมูลส่วนใหญ่ไว้ ณ ที่ของตน ดังนั้นการควบคุมการใช้ในส่วนของสาขาตน (Local Control) จึงดีกว่าการขึ้นกับระบบรวม

6) สามารถใช้ทรัพยากร และข้อมูลรวมกันดีกว่า (Shareability) : ถึงแม้ในระบบฐานข้อมูลแบบกระจาย อุปกรณ์และข้อมูลจะกระจายตามที่ต่าง ๆ แต่ก็มีการเชื่อมต่อและสามารถแลกเปลี่ยน ข้อมูลข่าวสารได้ตลอดเวลา ภายใต้การดูแลของระบบจัดการฐานข้อมูลแบบกระจาย

ผลเสีย

1) มีความยุ่งยากซับซ้อนมากกว่า (Complexity) : เนื่องจากระบบฐานข้อมูลแบบกระจายเป็นการดำเนินงานร่วมกันของฐานข้อมูลหลายแห่งหลายขบวนการ จากผู้ใช้หลายคน หลายที่ในเวลาเดียวกัน ดังนั้นความยุ่งยากซับซ้อน จึงเป็นปัญหาสำคัญในระบบนี้ ปัญหาความยุ่งยากซับซ้อนสามารถแยกเป็นประเภทต่าง ๆ ดังนี้
   - ความยุ่งยากซับซ้อนในการประมวลผล และการเรียกใช้ข้อมูล
   - ความยุ่งยากซับซ้อนในการควบคุม การจวบกัน (Concurrency Control) ของขบวนการต่าง ๆ
   - ความยุ่งยากซับซ้อนในการแก้ไขข้อมูลที่เก็บซ้ำซ้อน
   - ความยุ่งยากซับซ้อนในการฟื้นสภาพ (Recovery)
   - ความยุ่งยากซับซ้อนในการดูแลจัดการ พจนานุกรมฐานข้อมูล
   - ความยุ่งยากซับซ้อนในการรักษาความลับความปลอดภัย (Security) ของข้อมูล
   - ความยุ่งยากซับซ้อนในการแปลง ฐานข้อมูลแบบรวม ซึ่งปกติจะมีอยู่แล้ว ให้เป็นฐานข้อมูลแบบกระจาย
   - ความยุ่งยากซับซ้อนในการ ออกแบบฐานข้อมูลกระจาย ให้มีประสิทธิภาพ และมีคุณสมบัติตามความต้องการ
2) ค่าใช้จ่ายและการลงทุน (Cost) : ในระบบฐานข้อมูลแบบกระจายจะเกี่ยวข้องกับอุปกรณ์ทั้งทางด้านฮาร์ดแวร์ และซอฟต์แวร์หลายอย่างหลายชนิด ดังนั้นในระยะแรกจะเสียค่าใช้จ่ายมาก โดยเฉพาะค่าใช้จ่ายทางด้านซอฟต์แวร์ซึ่งขณะนี้ยังมีราคาค่อนข้างสูง และมีการพัฒนาตลอดเวลา ทำให้ต้องเพิ่มค่าใช้จ่ายในการ ปรับปรุง พัฒนาให้ทันต่อการเปลี่ยนแปลงและความต้องการที่เพิ่มขึ้น

ระบบเครือข่ายสื่อสารข้อมูล (Data Communication Network)

ดังที่ได้กล่าวไว้แล้วว่า ระบบฐานข้อมูลแบบกระจายเป็นระบบที่กระจายการจัดเก็บข้อมูลและกระจายการประมวลผลไปยังที่ต่างๆ ตามความต้องการของผู้ใช้ ดังนั้นระบบนี้จำเป็นต้องพึ่งพาอาศัยระบบเครือข่ายสื่อสารข้อมูลเป็นอย่างมาก อาจกล่าวได้ว่าระบบฐานข้อมูลแบบกระจายไม่อาจดำเนินการได้ หากไม่มีระบบเครือข่าย หรือระบบเครือข่ายยังไม่พัฒนาจนเป็นที่เชื่อถือ หรือไว้วางใจได้ ระบบเครือข่ายสื่อสารข้อมูล หรือบางครั้งมักเรียกว่า เครือข่าย (Network) เป็นการนำ

คอมพิวเตอร์ และอุปกรณ์ต่อพ่วงหลาย ๆ เครื่อง มาต่อรวมกันเป็น เครือข่ายคอมพิวเตอร์ เพื่อให้สามารถแลกเปลี่ยนข้อมูลข่าวสาร และใช้งานร่วมกันหลาย ๆ คนในเวลาเดียวกัน สำหรับระบบเครือข่ายในปัจจุบันสามารถแบ่งได้เป็น 3 ประเภทใหญ่ ๆ (Halsall, 1988) คือ

1) **เครือข่ายระยะไกล (Wide Area Network : WAN)**: เป็นเครือข่ายคอมพิวเตอร์ที่ครอบคลุมพื้นที่บริเวณกว้างไกลในระดับ ประเทศ ทวีป หรือระดับโลก การสื่อสารข้อมูลจะอาศัยเครือข่ายต่าง ๆ ทางชุมสายสาธารณะ คลื่นไมโครเวฟ หรือการสื่อสารผ่านดาวเทียมเป็นต้น

2) **เครือข่ายระดับเมือง (Metropolis Area Network : MAN)** : เป็นเครือข่ายคอมพิวเตอร์ที่ครอบคลุมพื้นที่ระดับเมือง นคร การสื่อสารข้อมูลจะอาศัยเครือข่ายทางสาธารณะ เช่น สายเช่า คลื่นไมโครเวฟ คลื่นวิทยุความถี่สูง ฯลฯ

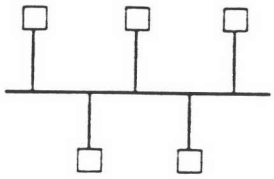
3) **เครือข่ายท้องถิ่น (Local Area Network : LAN)**: เป็นเครือข่ายคอมพิวเตอร์เฉพาะท้องถิ่นซึ่งมีขนาดเล็กกว่า 2 ประเภทแรก ครอบคลุมพื้นที่ตั้งแต่ไม่กี่ตารางฟุต จนถึงหลายตารางกิโลเมตร การสื่อสารข้อมูลอาจเป็นการเชื่อมต่อถึงกันโดยตรง ในลักษณะส่วนตัว หรืออาจผ่านเครือข่ายสาธารณะ เช่นเดียวกับเครือข่ายระดับเมือง

เนื่องจากปัจจุบัน เทคโนโลยีทางการสื่อสารโทรคมนาคม และการสื่อสารข้อมูล ผ่านเครือข่ายทางคอมพิวเตอร์มีความเจริญก้าวหน้ารวดเร็วมาก จนยากต่อการกำหนดหรือนิยามแค่ไหนคือเครือข่ายท้องถิ่น แค่ไหนคือเครือข่ายระดับเมือง อย่างเช่นการสื่อสารข้อมูลผ่านใยแก้วนำแสง (Fiber Optic) สามารถรับส่งข้อมูลในอัตรานับพันล้านบิต เป็นระยะทางนับร้อยกิโลเมตร โดยไม่จำเป็นต้องใช้ตัวเพิ่มสัญญาณ (Repiter) และยังมีแนวโน้มว่าจะพัฒนารุดหน้าอย่างไม่หยุดยั้ง

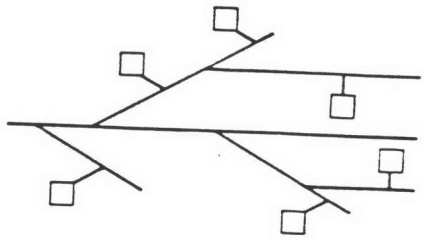
โทโพโลยี่ (Topology)

โทโพโลยี่ เป็นการจัดการเกี่ยวกับการเชื่อมต่ออุปกรณ์ทั้งหมด ของระบบเครือข่าย เช่น อุปกรณ์ทางการสื่อสาร เครื่องคอมพิวเตอร์ ตลอดจนอุปกรณ์ประกอบต่าง ๆ (Peripheral Devices) ซึ่งมีโครงสร้างพื้นฐานเชื่อมต่อ 4 รูปแบบ (Helmer, 1989) ด้วยกันคือ

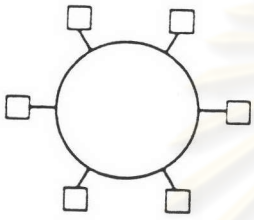
1) **โทโพโลยี่แบบบัส (Bus Topology)** เป็นโทโพโลยี่ที่เชื่อมต่อทุกสถานีเข้าหาบัส ซึ่งอาศัยสายเคเบิลหลัก (Back bone) โดยทุกสถานีจะใช้ร่วมกันรับ/ส่งข้อมูลข่าวสาร ผ่านสายเคเบิลนี้ ดังนั้นขณะใดขณะหนึ่งจะสามารถส่งข้อมูลข่าวสารได้เพียงสถานีเดียว การเชื่อมต่อลักษณะนี้มีด้วยกัน 2 รูปแบบดัง รูป 3.10 ก และ 3.10 ข ข้อดีของบัสโทโพโลยี่ คือ ขยายระบบง่าย ใช้สายเคเบิลน้อย แต่โทโพโลยี่นี้ก็มีข้อเสียคือ เกิดการชนกันของข้อมูล การจัดลำดับความสำคัญ



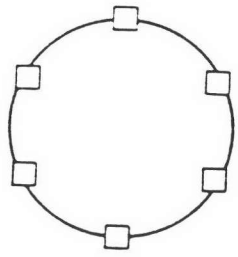
(ก) Bus



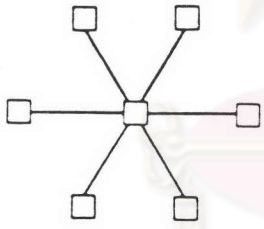
(ข) Branching tree



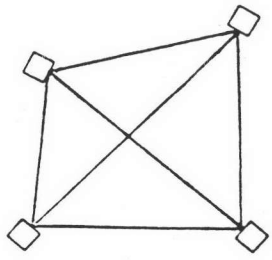
(ค) Ring



(ง) Ring



(จ) Star



(ฉ) Point-to-point

รูป 3.10 โทโพโลยี่พื้นฐานสำหรับเครือข่าย

(Priority) ของการส่งข้อมูลยุ่งยาก และการต่อแบบแยกสาขาหากมีการชัดข้องจะทำให้ทั้งระบบล้มเหลว

2) **โทโพโลยี่แบบวงแหวน (Ring Topology)**. เป็นโทโพโลยี่ที่เชื่อมต่อกันเป็นวงปิด ดังรูป 3.10 ค เมื่อสถานีใดต้องการส่งข้อมูลจะต้องรอจนได้รับสิทธิ์ในการส่งข้อมูลก่อน จากนั้นจึงกำหนดแอดเดรส ทั้งต้นทาง และปลายทางแล้วส่งข้อมูลเป็นแพคเก็ต (Packet) วนไปในวงแหวน

และเมื่อข้อมูลผ่านสถานีใดก็จะทำการตรวจสอบแอดเดรส หากตรงกับของตนจะคัดลอกข้อมูลไป แต่ถ้าไม่ตรงก็ส่งผ่านไปยังสถานีถัดไป ข้อดีของโทโพโลยี่นี้คือ มีประสิทธิภาพกว่า ข้อมูลผิดพลาดน้อย และไม่มีโอกาสชนกัน สำหรับข้อเสียคือ หากจุดใดจุดหนึ่งของระบบชำรุด การตรวจสอบจะยุ่งยาก เพราะไม่ทราบเสียหายตรงจุดใดและยังทำให้เครือข่ายทั้งระบบหยุดชะงัก อีกทั้งการขยายเพิ่มเติมจะยุ่งยากซับซ้อนกว่า

3) โทโพโลยี่แบบดาว (Star Topology) เป็นโทโพโลยี่ที่เชื่อมต่อทุกสถานีเข้ากับศูนย์สลับช่องสัญญาณ (Switching Center) การติดต่อระหว่างสถานี จะอาศัยวงจรสลับเหมือนชุมสายโทรศัพท์ ซึ่งปกติตัวสลับสัญญาณนี้จะมีความยุ่งยากซับซ้อน และไม่คล่องตัวเท่าที่ควร ถ้าระบบมีสถานีจำนวนมาก ๆ การติดต่อลักษณะนี้ต้องใช้สายเป็นจำนวนมาก และถ้าศูนย์สลับสัญญาณชัดข้องระบบทั้งระบบจะใช้การไม่ได้ แต่โทโพโลยี่ นี้ก็มีข้อดีคือ ติดตั้งง่าย สถานีใดชัดข้องจะไม่มีผลต่อเครือข่าย และสามารถส่งข้อมูลข่าวสารให้ ศูนย์กลางสลับได้ทุกเวลา

4) โทโพโลยี่แบบจุดต่อจุด (Point-to-Point Topology) เป็นโทโพโลยี่ที่เชื่อมต่อทุกสถานีเข้าหากันดังรูป 3.10 ฉ จำนวนสายที่ใช้เชื่อมโยงสามารถคำนวณได้จากสมการ $n(n-1)/2$ โดย n คือจำนวนสถานี การเชื่อมต่อลักษณะนี้บางครั้งเรียกว่าเมชโทโพโลยี่ (Mesh Topology) ข้อดีของโทโพโลยี่นี้อยู่ตรงที่ เชื่อมต่อง่าย ติดต่อถึงกันโดยตรง ดังนั้นจะรวดเร็วกว่าแบบอื่น แต่ระบบนี้ต้องเสียค่าใช้จ่ายสูงกว่าทุกระบบ ยิ่งระบบเครือข่ายขนาดใหญ่ จะต้องเพิ่มสายเชื่อมโยงจำนวนมาก และระบบซอฟต์แวร์ควบคุมเครือข่ายก็ยุ่งยากซับซ้อนมากกว่า

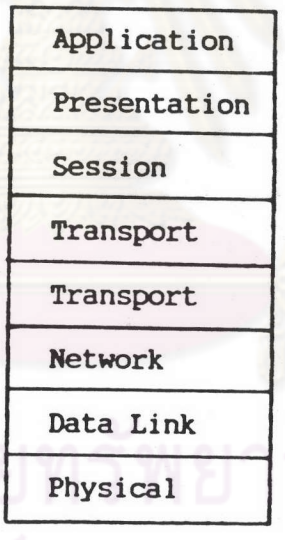
สถาปัตยกรรมเครือข่าย (Network Architecture)

การกำหนดโครงสร้างของสถาปัตยกรรมเครือข่ายเป็นการกำหนด ขั้นตอน หน้าที่ ระดับชั้นความรับผิดชอบ รูปแบบข้อมูล ข้อตกลงในการรับส่งข้อมูล (Protocol) ตลอดจนมาตรฐานต่างๆ ในเรื่องของฮาร์ดแวร์ และซอฟต์แวร์ ดังนั้นสถาปัตยกรรมเครือข่ายจึงมีหลายรูปแบบขึ้นอยู่กับหน่วยงาน หรือสถาบันที่พัฒนาระบบเครือข่ายจะกำหนดมาตรฐาน หรือนิยาม ในที่นี้จะขอกล่าวถึงเฉพาะสถาปัตยกรรมที่สำคัญ ๆ

1) สถาปัตยกรรมการเชื่อมโยงระบบเปิด (Open Systems Interconection:OSI) เป็นรูปแบบที่สถาบันวิศวกรไฟฟ้าและอิเล็กทรอนิกส์ของสหรัฐ (IEEE) และสมาคมผู้ผลิตคอมพิวเตอร์แห่งยุโรป ได้ร่วมกันกำหนดมาตรฐาน การสื่อสารข้อมูลทางคอมพิวเตอร์ระบบเปิด ซึ่งเรียกว่า โอเอสไอ (OSI) ให้เป็นรูปแบบจำลองอ้างอิงหนึ่ง (Reference Model) ขององค์กรมาตรฐาน

นานาชาติ (International Standard Organization:ISO) ซึ่งแบ่งระบบการสื่อสารเป็น 7 ระดับชั้น (7-Layers) ดังรูป 3.11 (Helmer, 1989)

- **ระดับชั้นแอพพลิเคชัน (Application Layer)** เป็นระดับชั้นบนสุด มีหน้าที่ในการคอยให้การบริการแก่โปรแกรมประยุกต์ของผู้ใช้งาน เช่น การโอนข้อมูล การเรียกใช้ข้อมูลจากที่ห่างไกล การใช้เทอร์มินัลเสมือน เป็นต้น
- **ระดับชั้นพรีเซนเตชัน (Presentation Layer)** เป็นชั้นแปลงข้อมูลให้อยู่ในรูปแบบที่คอมพิวเตอร์แต่ละยี่ห้อสามารถเข้าใจได้ มีการเข้ารหัสข้อมูล (Data Encryption) และการลดจำนวนข้อมูล (Data Compression)
- **ระดับเซสชัน (Session Layer)** เป็นระดับชั้นเริ่มต้นของการติดต่อ และดูแลประสานงานการสื่อสารระหว่างเครือข่ายทั้งระบบของผู้ส่ง และระบบของผู้รับ



รูป 3.11 ระดับชั้นของมาตรฐาน โอเอสไอ

- **ระดับชั้นทรานสปอร์ต (Transport Layer)** ระดับชั้นนี้จะจัดการในด้านการส่งข้อมูล ระหว่างเครื่องคอมพิวเตอร์ รวมทั้งการจัดลำดับส่วนต่าง ๆ ของข้อความ อัตราเร็วในการส่ง และการตัดส่วนของข้อมูลที่ซ้ำกันทิ้งไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

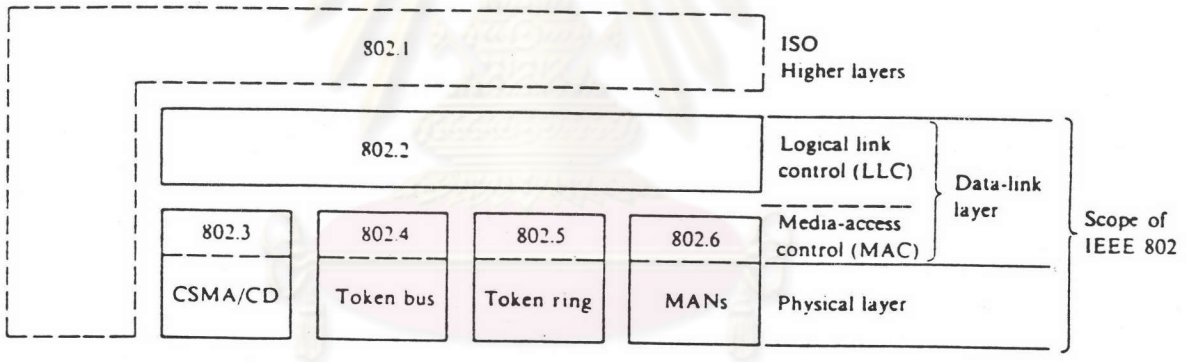
- **ระดับชั้นเน็ตเวิร์ค (Network Layer)** เป็นระดับชั้นที่มีหน้าที่ในการติดตั้งแอดเดรส และการกำหนดเส้นทางของกลุ่มข้อมูลต่าง ๆ ให้ไปถึงปลายทาง
- **ระดับชั้นดาต้าลิงค์ (Data Link Layer)** ทำหน้าที่ในการจัดการส่งข้อมูล และตรวจสอบความผิดพลาดของข้อมูลระหว่างข่ายงานที่ติดต่อกันอยู่
- **ระดับชั้นฟิสิคอล (Physical Layer)** เป็นระดับชั้นจัดการคุณสมบัติทางฮาร์ดแวร์ของสายส่ง และอุปกรณ์ในการส่ง ซึ่งการส่งข้อมูลระหว่างเครื่องจะมองในลักษณะรูปแบบที่เป็นบิต

2) สถาบัตยกรรม ไออีอีอี 802 (IEEE 802) เป็นสถาบัตยกรรมที่ สถาบันวิศวกรไฟฟ้าและอิเล็กทรอนิกส์ของสหรัฐ กำหนดขึ้นเพื่อการสื่อสารข้อมูลในเครือข่ายคอมพิวเตอร์เฉพาะบริเวณในระดับชั้นฟิสิคอลและระดับชั้นดาต้าลิงค์ ของมาตรฐาน ไอเอสไอ ดังรูป 3.12 (Helmer, 1989) ระดับชั้นบนสุดของ ไออีอีอี คือ โลจิคอลลิงค์คอนโทรล (Logical Link Control : LLC) มีหน้าที่ในการจัดเตรียมข้อมูล กำหนด หรือตรวจสอบแอดเดรส ในการรับส่งข้อมูล ระดับชั้นถัดลงมาคือ มีเดียมแอ็คเซสคอนโทรล (MEDIUM ACCESS CONTROL : MAC) คอยควบคุมการเข้าถึงสื่อกลางที่ใช้ในการรับส่งข้อมูลส่วนในระดับล่างสุดคือ ระดับชั้นฟิสิคอล มีหน้าที่เช่นเดียวกับมาตรฐาน ไอเอสไอ สำหรับชุดมาตรฐานของ 802 มีดังนี้

- ไออีอีอี 802.1 (IEEE 802.1) เป็นการกำหนดเกี่ยวกับการควบคุมและการจัดการระบบเครือข่าย
- ไออีอีอี 802.2 (IEEE 802.2) เป็นการบรรยายถึง ความหมาย หน้าที่ และโปรโตคอลของโลจิคอลลิงค์คอนโทรล (LLC) ของระบบเครือข่ายท้องถิ่น
- ไออีอีอี 802.3 (IEEE 802.3) มาตรฐานนี้จะบรรยายถึง วิธีการรับส่งข้อมูลระหว่างสถานี (Work Station) โดยใช้ บัส ร่วมกันในลักษณะ "ฟังขณะคุย" (CSMA/CD)
- ไออีอีอี 802.4 (IEEE 802.4) มาตรฐานนี้เกี่ยวข้องกับองค์ประกอบของ โทเกนพาสซิ่งบัส (TOKEN PASSING BUS) ซึ่งอธิบายถึง วิธีการใช้สายส่งข้อมูลสัญญาณทางกายภาพ และเทคโนโลยีของสื่อ ในการสื่อสารข้อมูล
- ไออีอีอี 802.5 (IEEE 802.5) มาตรฐานนี้เป็นการกำหนดรูปแบบ และโปรโตคอลที่ใช้ในการสื่อสารแบบ โทเกนริง (TOKEN RING)
- ไออีอีอี 802.6 (IEEE 802.6) เป็นการกำหนดมาตรฐานให้ระบบเครือข่าย

ระดับเมือง (MAN) ให้มีประสิทธิภาพมากขึ้น โดยใช้มาตรฐาน เอ็มเอซี (MAC) ตัวใหม่เพื่อใช้กับบัสในการรับส่งข้อมูล

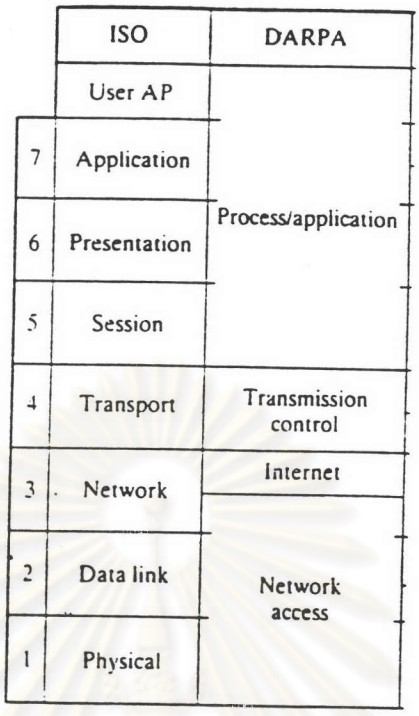
- ไออีอีอี 802.7 (IEEE 802.7) เป็นการกำหนดมาตรฐาน บรอดแบนด์ (Broadband) ซึ่งเป็นการส่งสัญญาณแบบ อนาลอก (Analog) และสามารถส่งหลายช่องสัญญาณในระบบเครือข่ายท้องถิ่น
- ไออีอีอี 802.8 (IEEE 802.8) เป็นการกำหนดมาตรฐานของการรับส่งสัญญาณบนใยแก้วนำแสง (Fiber Optic) เพื่อรองรับระบบการสื่อสารข้อมูลแบบรวมหลายประเภท
- ไออีอีอี 802.9 (IEEE 802.9) มาตรฐานนี้ยังอยู่ในขั้นเริ่มต้นซึ่งเป็นการกำหนดระบบในการสื่อสารทางด้าน เสียง ข้อมูล หรือภาพ บนสื่อเดียวกัน (ISDN)



รูป 3.12 สถาปัตยกรรม ไออีอีอี 802

3) สถาปัตยกรรมแบบดาร์ปา (Defense Advance Research Project Agency :DARPA) เป็นสถาปัตยกรรมเครือข่าย ที่พัฒนาโดยกระทรวงกลาโหมสหรัฐ จุดประสงค์เพื่อให้สามารถติดต่อสื่อสารกับทุกระบบเครือข่าย ซึ่งสถาปัตยกรรมนี้มีโครงสร้างดังรูป 3.13 (Helmer, 1989)

- ระดับชั้นอินเตอร์เน็ต (Internet Layer) ซึ่งเทียบเท่าระดับชั้นเน็ตเวิร์คในมาตรฐานโอเอสไอ มีหน้าที่ในการสร้างงานเกี่ยวกับการสื่อสารในระดับ อินเตอร์



รูป 3.13 เปรียบเทียบสถาปัตยกรรมแบบ ดารปา และ ไอเอสไอ

เน็ตเวิร์คให้ตัวกลาง และการบริการต่าง ๆ ที่เกี่ยวข้องในระดับชั้นนี้ด้วยอินเตอร์เน็ตโปรโตคอล หรือไอพี (Internet Protocol : IP) ในรูปแบบดาต้าแกรม (Datagram) ซึ่งเป็นโปรโตคอลหนึ่งของระบบแพ็คเกตสวิตชิ่ง เพื่อใช้ตรวจสอบความผิดพลาด จัดแพ็คเกต พร้อมทั้งกำหนดเส้นทาง และเวลาในการรับส่งแพ็คเกตระหว่างเน็ตเวิร์ค

- ระดับชั้นทรานสมิสชัน คอนโทรล (Transmission Control Layer) ซึ่งเทียบได้กับระดับชั้นทรานสปอร์ตของมาตรฐาน ไอเอสไอ ดำเนินงานด้วยหลักการ ทีซีพี (Transmission Control Protocal : TCP) มีหน้าที่เชื่อมต่อประสานงานอย่างใกล้ชิดกับ ไอพี คอยติดต่อตรวจสอบระหว่างผู้ส่ง และผู้รับ จัดแยกส่วนข้อมูล (Segment) กำหนดลำดับส่วนข้อมูล แจ้งข้อผิดพลาดให้ผู้ใช้ทราบ เป็นต้น

สำหรับระดับชั้นล่างถัดลงไปคือ ฟิสิคัล และเดต้าลิงค์ จะเป็นอะไรก็ได้ ดังนั้นเครือข่ายตามสถาปัตยกรรมนี้ จึงอิสระจากฮาร์ดแวร์ นอกจากนี้ยังสามารถส่งผ่าน (Transfer) ข้อมูลจากที่หนึ่ง ไปยังอีกที่หนึ่งได้อย่างรวดเร็ว ด้วยความเชื่อถือสูง ดังนั้นระบบเครือข่ายแบบ ดารปา จึงเป็นที่นิยม และแพร่หลายในปัจจุบัน