

บทที่ 3

ทฤษฎีระบบฐานข้อมูลแบบกระจาย

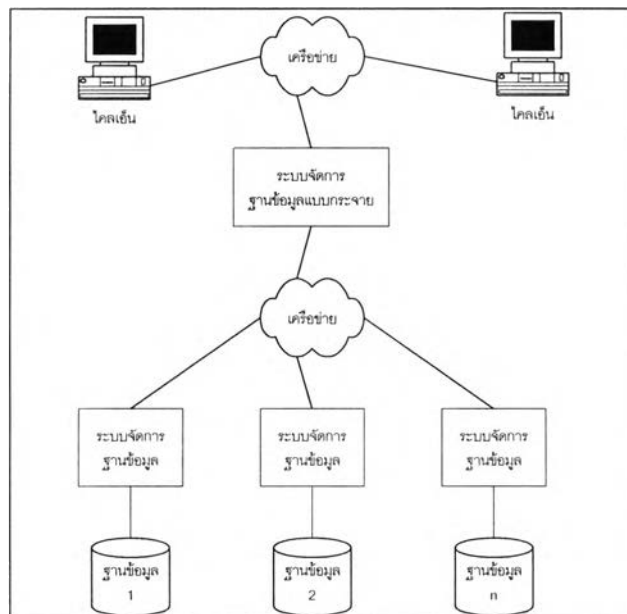
ระบบฐานข้อมูลแบบกระจาย (Distributed Database System:DDS)

ระบบฐานข้อมูลแบบกระจาย คือ ฐานข้อมูลที่ถูกเก็บไว้บนเครื่องคอมพิวเตอร์หลายๆ เครื่อง โดยที่เครื่องคอมพิวเตอร์แต่ละเครื่องติดตั้งอยู่ตามที่ตั้งต่างๆ และมีการติดต่อสื่อสารกันเป็นเครือข่าย (Network) ผู้ใช้สามารถใช้ฐานข้อมูลในเครื่องคอมพิวเตอร์แต่ละแห่งได้ โดยไม่จำเป็นต้องรู้ว่าข้อมูล ที่ใช้ขณะนั้น อยู่ที่ฐานข้อมูลในเครื่องคอมพิวเตอร์เครื่องใด การเข้าใช้ข้อมูลของผู้ใช้หรืองานประยุกต์ ต่างๆ จะต้องผ่านทางระบบจัดการฐานข้อมูลแบบกระจาย

โครงสร้างของระบบฐานข้อมูลแบบกระจาย (David and Jane, 1992)

ระบบฐานข้อมูลแบบกระจายจะประกอบด้วย

1. ระบบจัดการฐานข้อมูล
2. ระบบจัดการฐานข้อมูลแบบกระจาย
3. โปรแกรมจัดการเครือข่าย (Network Software)



รูปที่ 3-1 โครงสร้างของระบบฐานข้อมูลแบบกระจาย

ขั้นตอนการออกแบบฐานข้อมูลแบบกระจาย

1. การออกแบบระบบฐานข้อมูล

ขั้นตอนการออกแบบระบบฐานข้อมูลประกอบด้วย 2 ขั้นตอน คือ

- สร้างโมเดลข้อมูลเชิงตรรก (Logical Data Modeling :LDM)
- เปลี่ยนจากโมเดลข้อมูลเชิงตรรก ไปเป็นฐานข้อมูลแบบเชิงสัมพันธ์

1.1 สร้างโมเดลข้อมูลเชิงตรรก (Flemming and Von Halle, 1989)

LDM 1 กำหนดเอนทิตีหลัก

LDM 2 กำหนดความสัมพันธ์ระหว่างเอนทิตี

ความสัมพันธ์ คือ ลักษณะของความเกี่ยวข้องระหว่างเอนทิตี โดยที่ความสัมพันธ์ประกอบด้วยทิศทางของความสัมพันธ์ (Direction) และ สัดส่วนของความสัมพันธ์ (Cardinality Ratio) โดยสามารถแบ่งความสัมพันธ์ได้เป็น 3 ลักษณะดังนี้

- ความสัมพันธ์แบบหนึ่งต่อหนึ่ง (One-to-One Relationships)
- ความสัมพันธ์แบบหนึ่งต่อกลุ่ม (One-to-Many Relationships)
- ความสัมพันธ์แบบกลุ่มต่อกลุ่ม (Many-to-Many Relationships)

LDM 3 กำหนดคีย์หลักและคีย์รอง

LDM 4 กำหนดคีย์นอก

LDM 5 กำหนดกฎการจัดการกับข้อมูล ประกอบด้วย

- โดเมน (Domain) คือกรอบของค่าต่าง ๆ ที่เป็นไปได้ของแอททริบิวต์ เช่น เฟส สามารถมีค่าได้เป็น 1 และ 3 เท่านั้น
- กฎการเพิ่ม และลบข้อมูล (Insert and Delete Rule)
- ทริกเกอร์ดำเนินการ (Trigger Operation) เป็นข้อกำหนดถึงผลกระทบที่ต้องดำเนินการกับเอนทิตีอื่น หรือเอนทิตีเดียวกัน เมื่อมีการเพิ่ม ลบ แก้ไข และเรียกใช้ข้อมูล

LDM 6 เพิ่มแอททริบิวต์ลงในเอนทิตี

LDM 7 ตรวจสอบความถูกต้องของมุมมองผู้ใช้โดยใช้ทฤษฎีนอร์มัลไลเซชัน

LDM 8 กำหนดโดเมนของแอททริบิวต์ ประกอบด้วยรายละเอียดดังนี้

- ชนิดของข้อมูล เช่น จำนวนเต็ม, ตัวอักษร
- รูปแบบข้อมูล เช่น DD/MM/YYYY (วันที่)

- ค่าที่ยอมรับได้ เช่น เฟลมีค่าเท่ากับ 1 หรือ 3
- ความหมาย
- ความยาว
- มีค่าซ้ำกันได้หรือไม่
- เป็นค่าว่างได้หรือไม่

LDM 9 กำหนดทริกเกอร์ดำเนินการ

LDM 10 รวมมุมมองของผู้ใช้เข้าด้วยกันเพื่อลดความซ้ำซ้อน

LDM 11 รวบโมเดลที่ได้เข้ากับโมเดลที่มีอยู่เดิม

LDM12 วิเคราะห์ความเสถียรภาพ และการเติบโตในอนาคต

1.2 เปลี่ยนจากโมเดลข้อมูลเชิงตรรก ไปเป็นฐานข้อมูลแบบเชิงสัมพันธ์ (Flemming and Von Halle, 1989)

เป็นหลักการใช้ในการเปลี่ยนโมเดลเชิงตรรก ไปเป็นโมเดลข้อมูลเชิงกายภาพ เพื่อให้ใช้งานจริงบนฐานข้อมูลเชิงสัมพันธ์ ซึ่งประกอบด้วยโครงสร้างข้อมูล (Data Structure) และความถูกต้องของข้อมูล (Data Integrity)

ขั้นตอนการเปลี่ยนโมเดลเชิงตรรก ไปเป็นฐานข้อมูลเชิงสัมพันธ์มีดังนี้

RDD 1 เปลี่ยนเอนติตีของโมเดลข้อมูลเชิงตรรกให้อยู่ในรูปตารางโดย 1 เอนติตี ต่อ 1 ตาราง

RDD 2 แปลง 1 แอททริบิวต์ของโมเดลข้อมูลเชิงตรรกเป็น 1 คอลัมน์

RDD 3 ปรับโครงสร้างของโมเดลที่ได้จาก 1.2.1 และ 1.2.2 ให้เหมาะกับการติดตั้งระบบตามผลิตภัณฑ์ที่เลือกใช้

RDD 4 ออกแบบกฎการจัดการข้อมูลกับเอนติตี

RDD 5 ออกแบบกฎการจัดการข้อมูลกับความสัมพันธ์

RDD 6 ออกแบบกฎการจัดการข้อมูลกับแอททริบิวต์ที่เหลือ

2. การออกแบบการจัดเก็บข้อมูลในฐานข้อมูลแบบกระจาย (David and Jane, 1992)

ในระบบฐานข้อมูลแบบกระจายจะมีการกระจายข้อมูลไปอยู่ตามฐานข้อมูลในเครื่องคอมพิวเตอร์ต่าง ๆ โดยมีการเชื่อมโยงข้อมูลกันผ่านทางเครือข่าย โดยมุมมองของผู้ใช้ที่มองระบบฐานข้อมูลแบบกระจายเสมือนหนึ่งว่าเป็นระบบฐานข้อมูลแบบรวมศูนย์ ซึ่งลักษณะการจัดเก็บข้อมูลในฐานข้อมูลแบบกระจายสามารถแบ่งได้ออกเป็น 2 เทคนิคด้วยกันคือ

2.1 การจัดเก็บข้อมูลแบบแตกกระจาย

การจัดเก็บข้อมูลแบบแตกกระจาย เป็นเทคนิคการจัดแบ่งกลุ่มของข้อมูลออกไปไว้ตามฐานข้อมูลในเครื่องคอมพิวเตอร์ต่าง ๆ ตามลักษณะของข้อมูล โดยที่จะไม่มีข้อมูลในแต่ละฐานข้อมูลจะไม่มีข้อมูลที่ซ้ำซ้อนกัน การจัดเก็บข้อมูลแบบแตกกระจาย สามารถแบ่งออกได้เป็น 2 วิธีด้วยกันคือ

2.1.1 การแตกกระจายตามแนวตั้ง (Vertical Fragmentation)

การแตกกระจายตามแนวตั้ง เป็นการแตกกระจายข้อมูลในลักษณะตามกลุ่มของข้อมูล

เช่น

รหัสการไฟฟ้า	สาย	หมายเลขผู้ใช้ไฟ	ชื่อ-นามสกุล	พีอีไอ	ประเภทมิเตอร์
0806101	007	084109	จอ.องอาจ สิทราทิพย์	0049031	1
0806101	158	919200	นายพรชัย ทองศิริรักษ์	029842	1
0806101	043	530800	นายศักดิ์ศรีพี กล้ากลิ่น	029913	1
1010101	098	222100	นายประกันภัย แป้นเพชรห้วย	002897	1
1010101	979	036000	นางปรีญา ฉิมโฉม	003019	1
1010101	048	587300	นางนงเยาว์ นฤมิตรเลขการ	0048946	1

ตารางที่ 3-1 ตารางก่อนแตกข้อมูลแบบการแตกกระจายตามแนวตั้ง

จากตารางที่ 3-1 สามารถแบ่งตารางออกเป็น 2 กลุ่ม ตามการจัดกลุ่มข้อมูล คือ

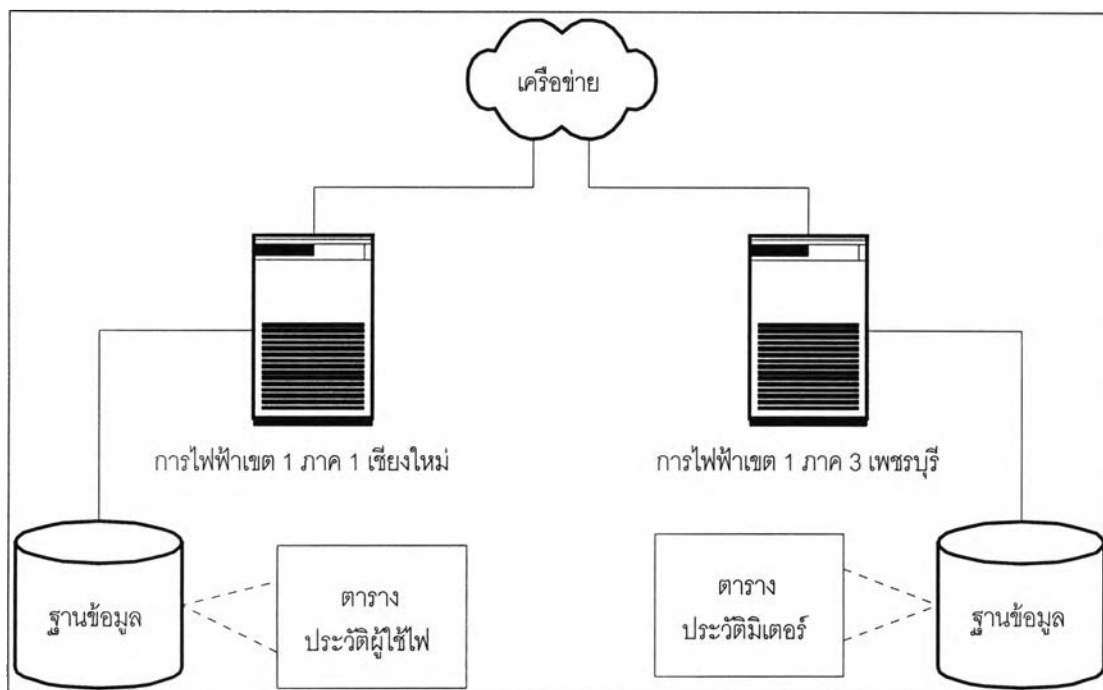
- กลุ่มประวัติผู้ใช้ไฟ (ตารางที่ 3-2)
- กลุ่มประวัติมิเตอร์ (ตารางที่ 3-3)

รหัสการไฟฟ้า	สาย	หมายเลขผู้ใช้ไฟ	ชื่อ-นามสกุล	พีอีเอ
0806101	007	084109	จอ.องอาจ สิทธิพิทย	0049031
0806101	158	919200	นายพรชัย ทองศิริรักษ์	029842
0806101	043	530800	นายศักดิ์รพี กล้ากลิ่น	029913
1010101	098	222100	นายประกันภัย เป้นเพชรห้วย	002897
1010101	979	036000	นางปรียา ฉิมโคม	003019
1010101	048	587300	นางนงเยาว์ นฤมิตรเลขการ	0048946

ตารางที่ 3-2 ตารางประวัติผู้ใช้ไฟ

พีอีเอ	ประเภทมิเตอร์
0049031	1
029842	1
029913	1
002897	1
003019	1
0048946	1

ตารางที่ 3-3 ตารางประวัติมิเตอร์



รูปที่ 3-3 การแตกกระจายตามแนวตั้ง

2.1.2 การแตกกระจายตามแนวนอน (Horizontal Fragmentation)

การแตกกระจายตามแนวนอน เป็นการแตกกระจายข้อมูลในลักษณะตามแนวนอน โดยแต่ละฐานข้อมูลในเครื่องคอมพิวเตอร์ต่าง ๆ จะมีตารางข้อมูลเหมือนกัน แต่ข้อมูลที่จัดเก็บไม่เหมือนกัน โดยจะจัดเก็บข้อมูลเฉพาะของแต่ละฐานข้อมูลเท่านั้น อาจแบ่งตามรหัสสถานที่ตั้ง

เช่น

รหัสการไฟฟ้า	สาย	หมายเลขผู้ใช้ไฟ	ชื่อ-นามสกุล	พีอีไอ	ประเภทมิเตอร์
0806101	007	084109	จอ.องอาจ สิทราทิพย์	0049031	1
0806101	158	919200	นายพรชัย ทองศิริรักษ์	029842	1
0806101	043	530800	นายศักดิ์รพี กล้ากลิ่น	029913	1
1010101	098	222100	นายประกันภัย แป้นเพชรห้วย	002897	1
1010101	979	036000	นางปรียา ฉิมโสม	003019	1
1010101	048	587300	นางนงเยาว์ นฤมิตรเลขการ	0048946	1

ตารางที่ 3-4 ตารางก่อนแตกข้อมูลแบบการแตกกระจายตามแนวนอน

จากตารางที่ 3-4 สามารถแบ่งตารางออกเป็น 2 กลุ่ม ตามสถานที่ข้อมูลอยู่ คือ

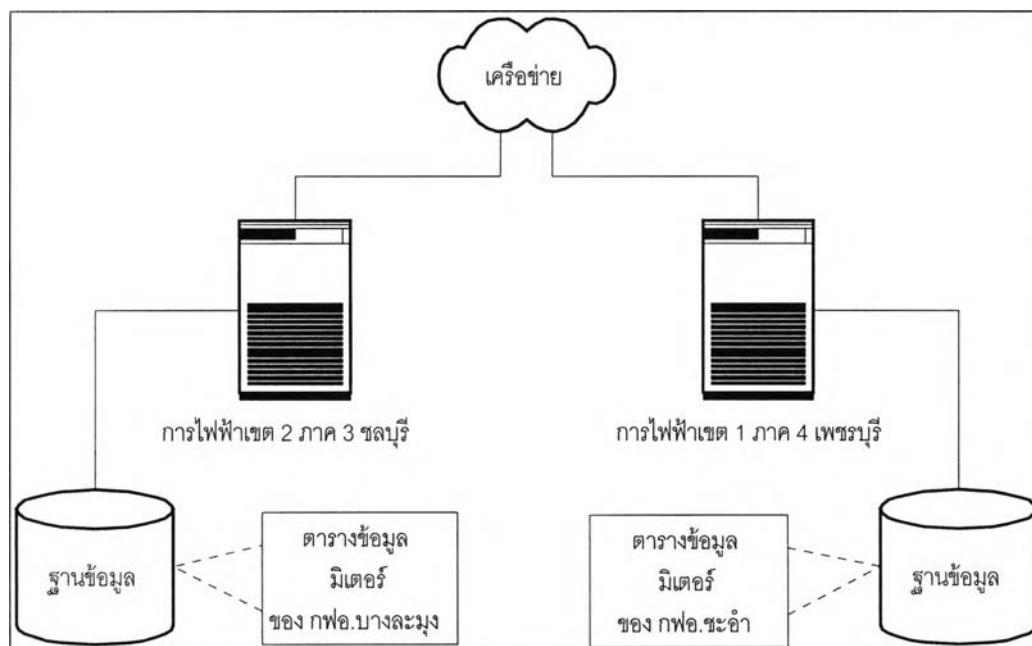
- กลุ่มข้อมูลของการไฟฟ้าอำเภอบางละมุง (ตารางที่ 3-5)
- กลุ่มข้อมูลของการไฟฟ้าอำเภอชะอำ (ตารางที่ 3-6)

รหัสการไฟฟ้า	สาย	หมายเลขผู้ใช้ไฟ	ชื่อ-นามสกุล	พีอีเอ	ประเภทมิเตอร์
0806101	007	084109	จอ.องอาจ สิทธาทิพย์	0049031	1
0806101	158	919200	นายพรชัย ทองศิริรักษ์	029842	1
0806101	043	530800	นายศักดิ์รพี กล้ากลิ่น	029913	1

ตารางที่ 3-5 ตารางของการไฟฟ้าอำเภอบางละมุง

รหัสการไฟฟ้า	สาย	หมายเลขผู้ใช้ไฟ	ชื่อ-นามสกุล	พีอีเอ	ประเภทมิเตอร์
1010101	098	222100	นายประกันภัย เป้นเพชรห้วย	002897	1
1010101	979	036000	นางปรียา ฉิมโฉม	003019	1
1010101	048	587300	นางนงเยาว์ นฤมิตรเลขการ	0048946	1

ตารางที่ 3-6 ตารางของการไฟฟ้าอำเภอชะอำ

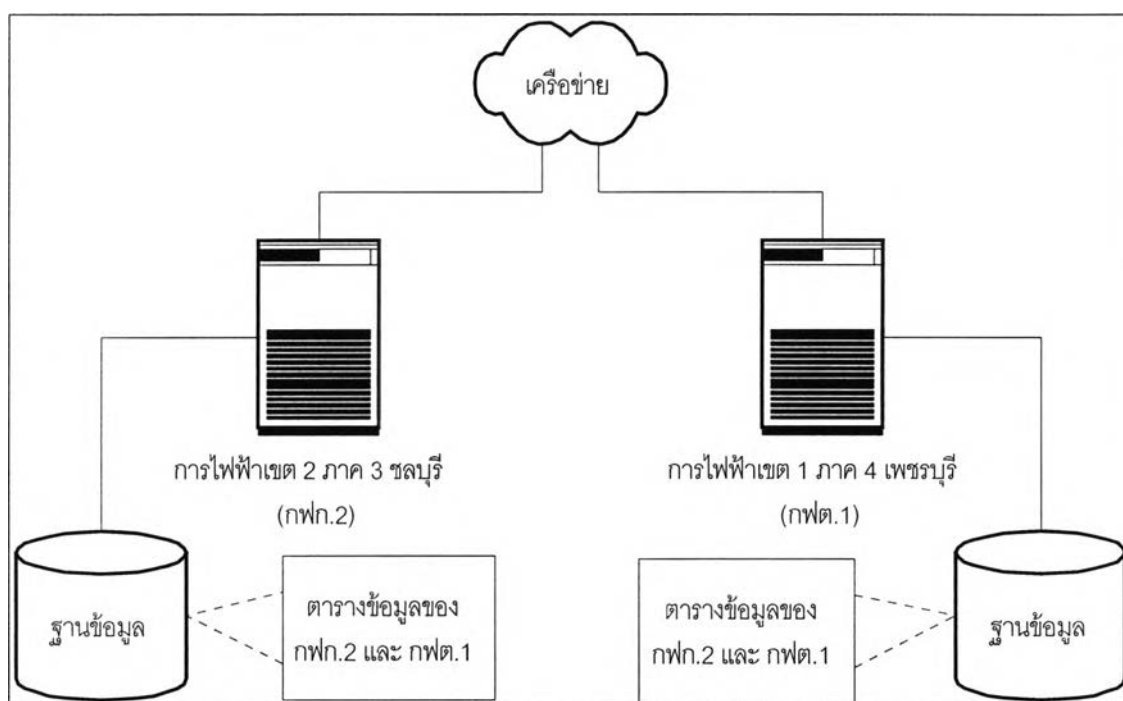


รูปที่ 3-4 การแตกกระจายตามแนวตั้ง

2.2 การเก็บซ้ำ

การเก็บซ้ำ หมายถึง การยอมให้มีการจัดเก็บข้อมูลที่เหมือนกัน กระจายไว้ในเครื่องคอมพิวเตอร์ต่างๆ โดยจะทำการสำรองข้อมูลจากฐานข้อมูลหลัก (Master Site) ไปยังฐานข้อมูลรอง (Snapshot Site) โดยที่ฐานข้อมูลสำรองสามารถมีได้มากกว่า 1 ฐานข้อมูลขึ้นไป เพื่อความรวดเร็วในการทำงาน หรือเมื่อคอมพิวเตอร์ที่เป็นที่เก็บฐานข้อมูลหลักมีปัญหา ฐานข้อมูลที่เป็นการเก็บซ้ำก็ยังสามารถทำงานได้

เช่น ฐานข้อมูลมิเตอร์ของ กฟน.1 เชียงใหม่ ทำการเก็บซ้ำไว้ที่ฐานข้อมูลของ กฟท.2 ชลบุรี และ กฟต.1 เพชรบุรี



รูปที่ 3-5 การเก็บซ้ำ

ประเภทของการจัดการระบบฐานข้อมูลแบบกระจาย (Devid and Jane, 1992)

1. การจัดการระบบฐานข้อมูลแบบกระจายที่มีฐานข้อมูลชนิดเดียวกัน (Homogeneous Distributed Database Management System)

คือระบบจัดการฐานข้อมูลแบบกระจายที่มีระบบจัดการฐานข้อมูลที่เป็นชนิดเดียวกัน

2. การจัดการระบบฐานข้อมูลแบบกระจายที่มีฐานข้อมูลต่างชนิดกัน (Heterogeneous Distributed Database Management System)

คือระบบจัดการฐานข้อมูลแบบกระจายที่มีระบบจัดการฐานข้อมูลต่างชนิดกัน