บทที่ 2
แนวคิดและทฤษฎีที่เกี่ยวข้อง

ความหมายของคลังข้อมูล

ตามคำกล่าวของ W.H. INMON [1] คลังข้อมูล (DATA WAREHOUSE) หมายถึง ฐานข้อมูลที่รวบรวมข้อมูลที่ไม่สามารถสรุปได้อีกหรือข้อมูลที่สรุปพอประมาณ เพื่อสนับสนุนระบบงานเพื่อการตัดสินใจ ซึ่งประกอบด้วยลักษณะดังนี้

1. อิงตามชุดของเทเบิล (TABLE) ที่สัมพันธ์กันในคลังข้อมูล (SUBJECT ORIENTED)
2. มีการรวมข้อมูลให้เป็นหนึ่งเดียว
3. ไม่ค่อยเปลี่ยนแปลง
4. มีข้อมูลหลายงวดเวลา สามารถเรียกใช้เมื่อไรก็ได้ซึ่งข้อมูลจะถูกต้องตามเวลานั้นเสมอ

ตามคำกล่าวของ SAM ANAHORY, DENNIS MURRAY [2] คลังข้อมูล หมายถึง ชุดสะสมข้อมูลที่สำคัญที่ใช้ในการจัดการและชี้นำธุรกิจให้มีกำไรมากที่สุด

ในที่นี้จะยึดตามความหมายของ W.H. INMON และมีคำอธิบายลักษณะของคลังข้อมูลที่กล่าวข้างต้นดังนี้

1. ชุดของเทเบิลที่สัมพันธ์กันในคลังข้อมูล

ในที่นี้อิงตามระบบงานหลักที่ประกอบกันขึ้นเป็นภาพลักษณ์ขององค์กร ดังรูปที่ 3.2 คือ

ระบบการจัดหาพลังงาน
ระบบการจ่ายไฟ
ระบบการบริการผู้ใช้ไฟ
ระบบบัญชีและการเงิน
ระบบมิเตอร์
ระบบหม้อแปลง
ระบบงานพัสดุคงคลัง
ระบบงานก่อสร้าง
ระบบทรัพย์สินและเครื่องมือเครื่องใช้
ระบบงานบริหารบุคคล

2. มีการรวมข้อมูลให้เป็นหนึ่งเดียว เช่น ข้อมูลตัวเดียวกันและความหมายเดียวกันแต่มีค่าหลายค่าและหลายประเภทแล้วแต่การกำหนดในระบบต้นแบบต่าง ๆ ให้มีค่าเดียวกันและประเภทเดียวกันในคลังข้อมูล เช่น เพศหญิง มีค่า "F"/"1"/"Y" จะกำหนดให้เป็นค่า "F" ประเภทตัวอักษรสำหรับคลังข้อมูล

3. ไม่ค่อยเปลี่ยนแปลง

4. มีข้อมูลหลายงวดเวลา สามารถเรียกใช้เมื่อไรก็ได้ซึ่งข้อมูลจะถูกต้องตามงวดเวลานั้นเสมอ เช่น ข้อมูลรายได้ของเดือนมกราคม 2541 และรายได้ของเดือนกุมภาพันธ์ 2542 จะมีค่าต่างกัน แต่สามารถเรียกใช้ ณ วันเวลาใดก็ได้ ก็ยังคงเป็นข้อมูล ณ วันเวลาดังกล่าว เสมอ

คลังข้อมูลจะให้คำตอบเกี่ยวกับข้อมูลสรุป ข้อมูลสถิติ (HISTORICAL DATA) ข้อมูลเปรียบเทียบเพื่อการวิเคราะห์ ซึ่งเกิดจากการเชื่อมโยงข้อมูลจากระบบต้นแบบหลาย ๆ ระบบทั่วทั้งองค์กร คลังข้อมูลจัดได้ว่าเป็นฐานข้อมูลอย่างหนึ่งแต่เป็นฐานข้อมูลที่สนับสนุนการวิเคราะห์และตัดสินใจ ไม่ได้เป็นฐานข้อมูลที่สนับสนุนงานประจำวันที่เรียกว่าฐานข้อมูลทำการ (OPERATIONAL DATABASE)

องค์ประกอบของคลังข้อมูล

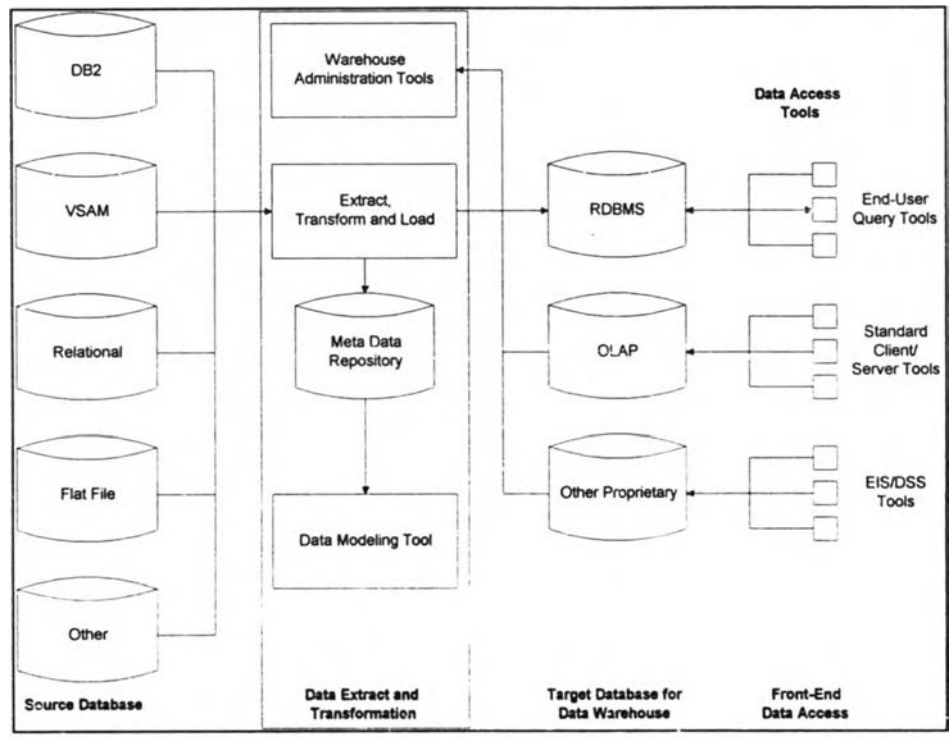
องค์ประกอบของคลังข้อมูล ประกอบด้วย 3 ส่วน (ดูรูปที่ 2.1) คือ

1. ฐานข้อมูลสำหรับคลังข้อมูล (TARGET DATABASE FOR DATA WAREHOUSE)

2. การนำข้อมูลจากระบบต้นแบบเข้าคลังข้อมูล (DATA EXTRACTION AND TRANSFORMATION)

3. การเรียกใช้ข้อมูลจากคลัง (FRONT-END DATA ACCESS)

ฐานข้อมูลสำหรับคลังข้อมูล เป็นไปได้ทั้งระบบจัดการฐานข้อมูลเชิงสัมพันธ์ (RDBMS : RELATIONAL DATABASE MANAGEMENT SYSTEM), โอแอลเอพี (OLAP : ON-LINE ANALYTICAL PROCESSING) หรืออื่น ๆ



รูปที่ 2.1 องค์ประกอบของคลังข้อมูล [3]

การนำข้อมูลจากระบบต้นแบบเข้าสู่คลังข้อมูล เป็นการคัดเลือกข้อมูลจากระบบต้นแบบเปลี่ยนรูปแบบแล้วนำข้อมูลนี้เข้าสู่คลัง เพื่อให้ได้ข้อมูลที่มีความหมายต่อผู้ใช้สุดท้าย (END USER) ซึ่งอาจใช้เครื่องมืออำนวยความสะดวกของระบบจัดการฐานข้อมูลเชิงสัมพันธ์มาตรฐาน หรือใช้เครื่องมือที่ใช้ดึงข้อมูลนำเข้าโดยเฉพาะ หรือจัดทำโปรแกรมขึ้นเองก็ได้ ข้อมูลจากระบบงานต้นแบบจะมีโครงสร้างแบบใดก็ได้ เช่น แฟลทไฟล์ (FLAT FILE), วีแซมไฟล์ (VSAM FILE) , เทเบิลที่สัมพันธ์กัน (RELATIONAL TABLE) หรือแบบใดก็ได้ ฐานข้อมูลของคลังข้อมูลจะเป็นฐานข้อมูลเชิงสัมพันธ์ หรือ แบบใดก็ได้ สิ่งสำคัญที่ผู้ดูแลคลังข้อมูลต้องตระหนักก็คือ ข้อมูลในคลังกับข้อมูลจากระบบต้นแบบจะต้องมีความหมายและโดเมน (DOMAIN) ของค่าเดียวกัน ในการจัดทำคลังข้อมูลที่สมบูรณ์แบบควรจะมีข้อมูลที่อธิบายความหมาย, นโยบายทางธุรกิจที่เกี่ยวกับข้อมูล, ที่มา, การเก็บข้อมูลในคลังข้อมูล และความสัมพันธ์ระหว่างข้อมูลด้วยกันของข้อมูลในคลังข้อมูล เพื่ออำนวยความสะดวกให้กับผู้ดูแลคลังข้อมูลและช่วยให้ผู้ใช้เข้าใจความหมายของข้อมูลในคลังข้อมูลด้วย ข้อมูลที่อธิบายข้อมูลในคลังข้อมูลนี้ เรียกว่า เมตะดาตา (METADATA) เมตะดาตาควรเก็บไว้ที่แหล่งเก็บส่วนกลางของคลังข้อมูลที่เรียกว่า เมตะดาตารีโพสิทอรี (METADATA REPOSITORY)

ในขั้นตอนการคัดเลือกข้อมูลนำเข้า จะต้องทราบความหมายของรายการข้อมูลแต่ละตัว (DATA ELEMENT) โดยมีขั้นตอนดังนี้ (วิธีการนี้สามารถนำไปใช้ในขั้นตอนใด ๆ ของคลังข้อมูลได้ด้วย)

1. ค้นหาความหมายของข้อมูลแต่ละตัว แหล่งกำเนิด ค่าที่เป็นไปได้ และความหมายของแต่ละค่า (ข้อมูลตัวเดียวกันอาจมีซ้ำกันได้ในหลาย ๆ ระบบต้นแบบ และค่าของข้อมูลนี้อาจอยู่ในโดเมนเดียวกันหรือต่างโดเมนก็ได้ เช่น คำว่า "เดือน" และ "งวด" "เดือน" ของระบบบริการผู้ใช้ไฟ หมายถึง วันที่ 14 เดือนปัจจุบันถึงวันที่ 13 ของเดือนถัดไป โดยเริ่มนับจากวันที่ 14 บวกไป 7 วัน เป็นงวดที่ 1 เดือนหนึ่งมี 4 งวดหรือมากกว่า "เดือน" ของระบบบัญชี หมายถึง วันที่ 1 ถึงวันสิ้นเดือน โดยเริ่มนับวันที่ 1-7 เป็นงวดที่ 1 เดือนหนึ่งมี 4 งวด เป็นต้น)

2. แยกประเภทข้อมูล เช่น แยกตามวิธีแก้ปัญหาที่ข้อมูลไม่สมบูรณ์

3. ตัดสินใจแก้ปัญหา เช่น อาจเปลี่ยนนโยบายธุรกิจหรือ ขั้นตอนทางธุรกิจบางส่วน เปลี่ยนการทำงานในระบบต้นแบบ เปลี่ยนวิธีการนำข้อมูลเข้า หรือวิธีอื่นที่สามารถแก้ปัญหาได้

4. สร้างโปรแกรมแปรรูปและขัดเกลาข้อมูล

5. แปรรูปข้อมูลให้เป็นไปตามกฎหรือเงื่อนไขที่ตั้งไว้ ถ้าไม่เป็นไปตามกฎหรือเงื่อนไข ก็ให้นำข้อมูลนั้นไปแยกเก็บไว้อีกที่ ทำการแก้ไขให้ถูกต้องแล้วจึงนำข้อมูลเข้าคลัง

6. ตรวจสอบข้อมูลกับเมตะดาตาเสมอ เพื่อป้องกันกรณีที่ทำให้ข้อมูลบางรายการในคลังเปลี่ยนแปลงความหมายและโดเมน (DOMAIN) ของค่าไปจากเดิมโดยสิ้นเชิง มีผลให้ข้อมูลในคลังสับสน เช่น กรณีที่ผู้พัฒนาระบบต้นแบบอาจเปลี่ยนแปลงโดเมนของค่าของข้อมูลบางรายการโดยรู้เท่าไม่ถึงการณ์ หรือไม่สนใจว่าจะเกิดผลกระทบต่อคลังข้อมูล

การเรียกใช้ข้อมูลจากคลัง อาจเลือกใช้การสอบถามเชิงสัมพันธ์ (RELATIONAL QUERY), ภาพข้อมูลหลายมิติ (MULTI-DIMENSION VIEWS OF RELATIONAL DATA), การสร้างโปรแกรมเอง หรือใช้โปรแกรมสำหรับดีเอสเอส (DECISION SUPPORT SYSTEM) หรือ อีไอเอส (EXECUTIVE INFORMATION SYSTEM)

ความหมายของเมตะดาตา

ตามคำกล่าวของ W.H. INMON [1] เมตะดาตา (META DATA) หมายถึง ข้อมูลเกี่ยวกับข้อมูล เมตะดาตาจัดเก็บข้อมูลเกี่ยวกับ

1. โครงสร้างข้อมูล สำหรับผู้พัฒนาโปรแกรม
2. โครงสร้างข้อมูล สำหรับผู้วิเคราะห์ระบบเพื่อการตัดสินใจ
3. ข้อมูลที่นำเข้าคลัง
4. การแปรรูปข้อมูลจากระบบต้นแบบเข้าสู่คลัง
5. แบบจำลองข้อมูล
6. ความสัมพันธ์ระหว่างแบบจำลองข้อมูลกับคลังข้อมูล
7. ประวัติการดึงข้อมูล

ตามคำกล่าวของ SAM ANAHORY, DENNIS MURRAY [2] เมตะดาตา หมายถึง ข้อมูลเกี่ยวกับข้อมูล หรือคล้าย ๆ กับบัตรดัชนีที่อธิบายว่าสารสนเทศมีโครงสร้างและถูกจัดเก็บอย่างไรในคลังข้อมูล

เมตะดาตาใช้ในวัตถุประสงค์ดังนี้

1. การดึงข้อมูลและการนำข้อมูลเข้าคลัง
2. การจัดการข้อมูล
3. การเรียกใช้ข้อมูล

ตามคำกล่าวของ RAMON BARQUIN, HERB EDELSTEIN AND OTHERS [3] เมตะดาตา รีพอสิทอรี หมายถึง ฐานข้อมูลที่บอกให้ทราบเกี่ยวกับข้อมูลหลากชนิดที่เก็บในคลังข้อมูล

เมตะดาตา หมายถึง ข้อมูลเกี่ยวกับข้อมูล

เมตะดาตามีประโยชน์ต่อผู้บริหารคลังข้อมูลและผู้ใช้คลังข้อมูล ดังนี้

1. จับคู่ข้อมูลในระบบต้นแบบกับข้อมูลในคลังข้อมูล
2. สร้างชุดคำสั่งในการดึงข้อมูล, แปรรูปข้อมูล, และโหลดข้อมูลเข้าคลัง
3. เป็นเทเบิลเวลาสำหรับงานนำข้อมูลเข้าคลัง
4. ช่วยเหลือผู้ใช้คลังข้อมูลในการค้นพบข้อมูลในคลัง
5. ช่วยเหลือผู้ใช้คลังข้อมูลในการสร้างคำถามเรียกใช้ข้อมูลได้ตามความต้องการ

ในที่นี้จะยึดตามหลักการเมตะดาตาของ W.H. INMON และ RAMON BARQUIN, HERB EDELSTEIN AND OTHERS คือ

1. การจัดการข้อมูล
2. จับคู่ข้อมูลในระบบต้นแบบกับข้อมูลในคลังข้อมูล

3. สร้างชุดคำสั่งในการดึงข้อมูล, แปรรูปข้อมูล, และนำข้อมูลเข้าคลัง

4. เป็นเทเบิลเวลาสำหรับงานนำข้อมูลเข้าคลัง

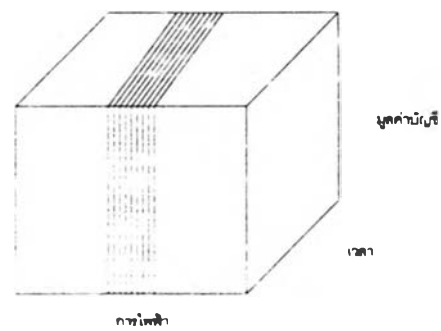
5. ช่วยเหลือผู้ใช้คลังข้อมูลในการค้นพบข้อมูลในคลัง

6. ช่วยเหลือผู้ใช้คลังข้อมูลในการสร้างคำถามเรียกใช้ข้อมูลได้ตามความต้องการ

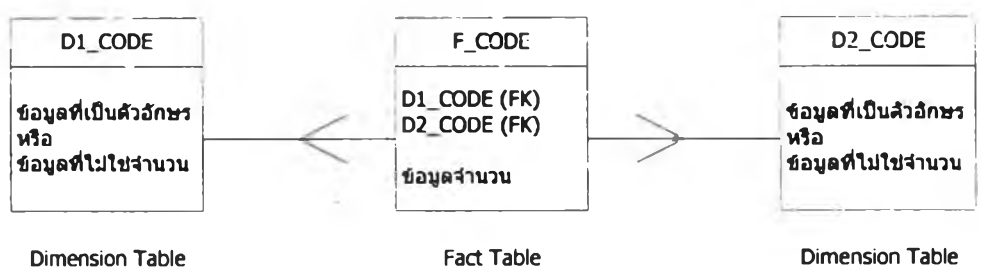
การสร้างแบบจำลองข้อมูลของคลังข้อมูล

การสร้างแบบจำลองข้อมูล (DATA MODELING) ของคลังข้อมูล ในที่นี้จะทำในรูปของข้อมูลหลายมิติ (MULTIDIMENSIONAL DATA MODELING) (ดูรูปที่ 2.2 ประกอบ) และเป็นระดับชั้น (HIERARCHY) เพื่อให้สามารถสอบถามข้อมูลสรุปแล้วค่อยเจาะลึกลงรายละเอียดหรือดริลดาวน์ (DRILL DOWN)

การสร้างแบบข้อมูลหลายมิติ จะใช้ระบบการจัดการฐานข้อมูลที่สนับสนุนข้อมูลหลายมิติ หรือจะใช้หลักการของฐานข้อมูลเชิงสัมพันธ์โดยกำหนดให้มีการเชื่อมโยงเทเบิลแบบดาว (STAR SCHEMA) ดังรูปที่ 2.3 ในที่นี้จะใช้หลักการของฐานข้อมูลเชิงสัมพันธ์



รูปที่ 2.2 มิติข้อมูลของคลังข้อมูล



รูปที่ 2.3 ตัวอย่างการเชื่อมโยงข้อมูลแบบดาวเพื่อสร้างข้อมูลหลายมิติ

การเชื่อมโยงแบบดาว

การเชื่อมโยงแบบดาว [2] ประกอบด้วย เทเบิลแฟคท์ (FACT TABLE) อยู่ตรงกลาง ล้อมรอบด้วยเทเบิลอ้างอิงที่เรียกว่าเทเบิลไดเมนชัน (DIMENSION TABLE) และนำเทเบิลทั้งสองประเภทมาเชื่อมโยงกันในลักษณะของดาว โดยมีตัวอย่างความสัมพันธ์ดังรูปที่ 2.3 และอธิบายรายละเอียดในบทที่ 4

ข้อมูลของคลังข้อมูลแบ่งได้เป็น 2 ประเภทใหญ่ ๆ คือ ข้อมูลที่เป็นจริง (ข้อมูลแฟคท์) และข้อมูลอ้างอิง (ข้อมูลไดเมนชัน) ข้อมูลแฟคท์เป็นข้อมูลหลักของคลังข้อมูล ซึ่งเปรียบได้กับข้อมูลทรานแซคชัน (TRANSACTION) ที่เกิดขึ้น ณ จุดเวลาหนึ่งและไม่ค่อยเปลี่ยนแปลงตลอดชีวิตของคลังข้อมูล ในขณะที่ข้อมูลไดเมนชันเป็นข้อมูลที่ใช้สำหรับการวิเคราะห์ทรานแซคชัน และสามารถเปลี่ยนแปลงได้เสมอซึ่งอาจเปลี่ยนไปตามธุรกิจ ข้อมูลแฟคท์จะถูกจัดเก็บไว้ในเทเบิลที่เรียกว่าเทเบิลแฟคท์ และข้อมูลไดเมนชันจะถูกจัดเก็บไว้ในเทเบิลไดเมนชัน ซึ่งเทเบิลทั้งสองชนิดมีลักษณะดังนี้

<u>ลักษณะของเทเบิลแฟคท์</u>

1. จำนวนแถวมากกว่า 1 ล้านแถว หรือ BILLION แถว
2. มีฟอเรนคีย์ (FOREIGN KEY) หลายตัว
3. เก็บข้อมูลตัวเลข
4. ข้อมูลไม่เปลี่ยนแปลง

<u>ลักษณะของเทเบิลไดเมนชัน</u>

1. จำนวนแถว 2-3 ล้านแถว
2. มี ไพรมารีคีย์ (PRIMARY KEY) ตัวเดียว
3. เก็บข้อมูลที่เป็นข้อความหรือข้อมูลอ้างอิง
4. มีการแก้ไขข้อมูลบ่อย

ในที่นี้ ผู้เขียนได้ดัดแปลงมาจากหนังสือ DATA WAREHOUSE IN THE REAL WORLD [2] ในเรื่องของขั้นตอนในการกำหนดแฟคท์และไดเมนชัน วิธีการออกแบบเทเบิลแฟคท์ และวิธีการออกแบบเทเบิลไดเมนชัน ได้ดังนี้

ขั้นตอนในการกำหนดแฟคท์และไดเมนชัน

1. เลือกข้อมูลทรานแซคชันของการดำเนินธุรกิจ เพื่อประกอบการคัดเลือกข้อมูลแฟคท์

2. ระบุข้อมูลไดเมนชันที่ใช้อ้างอิงข้อมูลแฟคท์ที่เลือกไว้แต่ละตัว

3. ตรวจสอบว่าข้อมูลแฟคท์ที่เลือกไว้ ไม่ใช่ข้อมูลไดเมนชันที่แฝงไว้ในข้อมูลแฟคท์ ถ้าไม่ผ่านให้ทำข้อ 2 ใหม่

4. ตรวจสอบว่าข้อมูลไดเมนชันที่เลือกไว้ ไม่ใช่ข้อมูลแฟคท์ ถ้าไม่ผ่านให้ทำข้อ 2 ใหม่

วิธีการออกแบบเทเบิลแฟคท์

1. ระบุงวดเวลาที่ใช้ในการวิเคราะห์

2. ระบุตัวอย่างข้อมูลรายละเอียดที่ใช้ในการวิเคราะห์ เพื่อใช้ประกอบการพิจารณากำหนดระดับของความละเอียดของข้อมูล เพื่อลดจำนวนข้อมูลรายละเอียดในคลังข้อมูล เช่น ผู้บริหารจะวิเคราะห์ข้อมูลทางการเงินในระดับปี ไตรมาส เดือน และงวด 7 วัน แต่ไม่จำเป็นต้องย่อยถึงระดับวัน เป็นต้น ในที่นี้ผู้เขียนได้ออกแบบเทเบิลเวลาให้มีรายละเอียดถึงระดับวัน โดยมีการนำข้อมูลเข้าคลังทุก 7 วัน ซึ่งก็สามารถนำไปสรุปเป็นเดือน ไตรมาส และปีได้

3. กำหนดคอลัมน์ โดยที่คอลัมน์ต้องแสดงถึงเหตุการณ์จริง ไม่สามารถคำนวณ/รับมาจากแหล่งอื่น และต้องเป็นที่ลนใจของธุรกิจด้วย ไม่ใช่ใช้ในการควบคุมอย่างเดียว

4. ลดขนาดคอลัมน์ ซึ่งจะต้องคำนึงว่าทุกตัวอักษรของคอลัมน์ต้องใช้ประโยชน์ ไม่กำหนดอย่างฟุ่มเฟือย เนื่องจากมีข้อมูลจำนวนมหาศาลในคลังข้อมูล

5. ออกแบบการใส่เวลาในเทเบิลแฟคท์ โดยในที่นี้ใส่วันที่จริง

วิธีการออกแบบเทเบิลไดเมนชัน

1. สร้างไดเมนชัน โดยการทำดีนอร์มอลไลซ์ (DENORMALIZATION) ข้อมูลอ้างอิงกลุ่มเดียวกันแต่ต่างระดับชั้นกันให้อยู่ในเทเบิลเดียวกัน เพื่อลดการเชื่อมโยงเทเบิล เช่น ไดเมนชันเวลาสามารถแบ่งย่อยเป็น ปี ไตรมาส เดือน สัปดาห์ และวัน ก็ทำดีนอร์มอลไลซ์ให้อยู่ในเทเบิลเดียวกัน ซึ่งการทำดีนอร์มอลไลซ์นี้จะอธิบายให้ชัดเจนต่อไป

2. ถ้าลักษณะคำถามเปลี่ยนไปในอนาคต ให้สร้างระดับชั้นใหม่(ที่ถูกเข้าถึงบ่อย ๆ)และเพิ่มคอลัมน์เข้าไป

3. ไดเมนชันที่เปลี่ยนแปลงบ่อย ๆ

4. ใส่ช่วงวันที่ในไดเมนชัน เพื่อให้รองรับคำถามที่เปรียบเทียบข้อมูลในอดีต

5. หลีกเลี่ยงการเก็บข้อมูลไดเมนชันไว้ในเทเบิลแฟคท์ เพื่อป้องกันไม่ให้ข้อมูลแฟคท์เปลี่ยนแปลงในกรณีที่มีการเปลี่ยนแปลงข้อมูลไดเมนชัน

6. คำนึงว่าข้อมูลไดเมนชัน ควรมีการเปลี่ยนแปลงตามกฎระเบียบพื้นฐานและโครงสร้างของคลังข้อมูล เพื่อลดค่าใช้จ่ายในการเปลี่ยนแปลงข้อมูล และให้แยกไดเมนชันออกจากแฟคท์ และกำหนดฟอเรนคีย์ให้กับเทเบิลแฟคท์

การทำดีนอร์มอลไลซ์

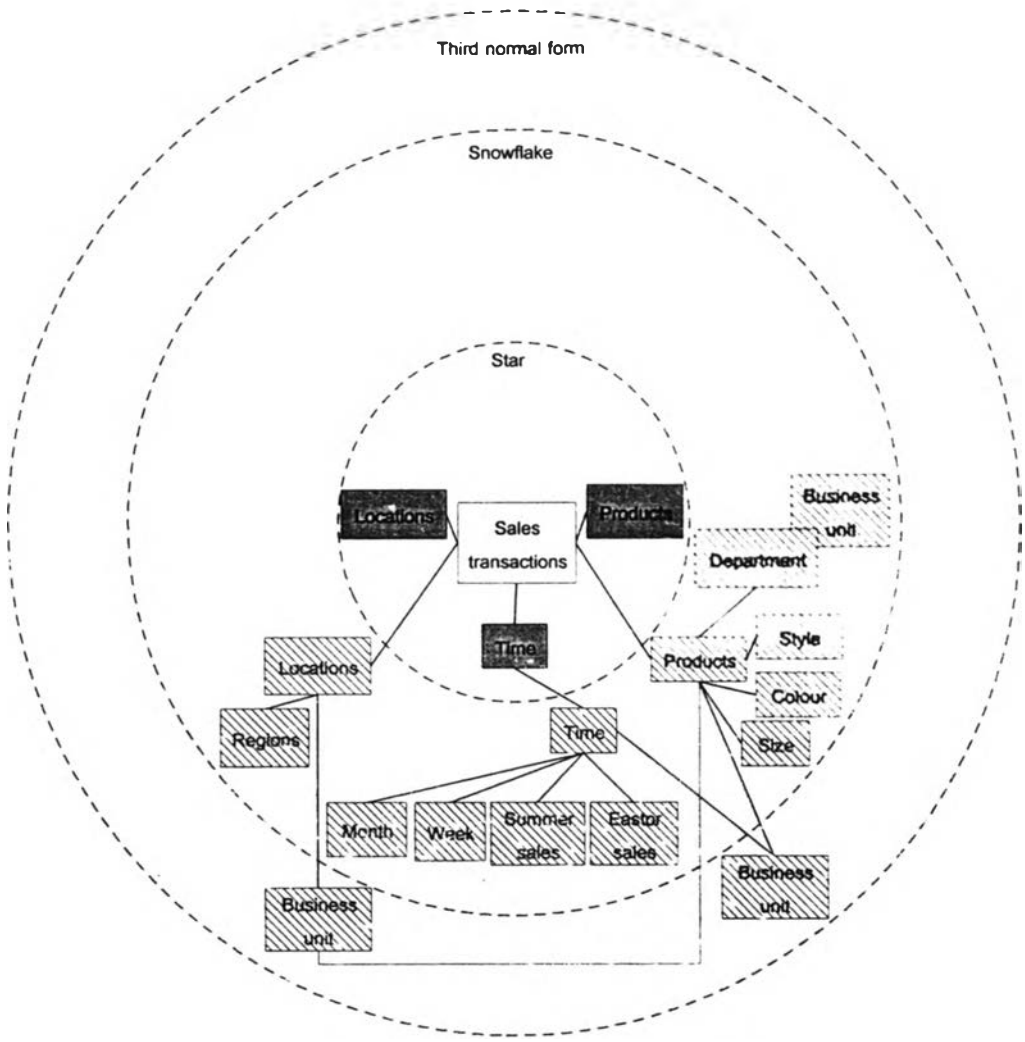
การทำดีนอร์มอลไลซ์ [4] เป็นกระบวนการที่ทำลายกฎของการทำนอร์มอลไลซ์ (NORMALIZATION) โดยการทำให้ข้อมูลที่ได้รับการออกแบบโดยวิธีลดความซ้ำซ้อนให้อยู่ในรูปที่ไม่อยู่ในกฎของการทำนอร์มอลไลซ์ โดยมีวัตถุประสงค์เพื่อเพิ่มสมรรถนะ (PERFORMANCE) เช่น ลดการเชื่อมโยงระหว่างเทเบิลทำให้เวลาในการประมวลผลเร็วขึ้น

การทำดีนอร์มอลไลซ์ สามารถทำได้หลายวิธี ดังนี้

1. ทำให้ข้อมูลซ้ำซ้อน โดยการเพิ่มคอลัมน์ให้ซ้ำกันหรือกำหนดให้มีการสรุปหรือคำนวณ
2. เปลี่ยนรายการของคอลัมน์ โดยการรวมคอลัมน์ หรือลดจำนวนคอลัมน์
3. รวมหลาย ๆ เทเบิลให้เป็นเทเบิลเดียว
4. ถอดแบบเทเบิลทั้งหมดหรือบางส่วนของเทเบิล แล้วแบ่งเทเบิลออกเป็นหลาย ๆ เทเบิล

ในที่นี้ ใช้กระบวนการดีนอร์มอลไลซ์กับเทเบิลไดเมนชัน ซึ่งออกแบบให้อยู่ในรูปของเทเบิลที่ทำการดีนอร์มอลไลซ์ โดยยอมให้มีกลุ่มที่ซ้ำกัน (REPEATING GROUP) ได้ เช่น เทเบิลเวลามีกลุ่มที่ซ้ำกัน ได้แก่ ปี เดือน ไตรมาส

เพื่อให้เห็นภาพชัดเจนมากขึ้นเกี่ยวกับการสร้างการเชื่อมโยงแบบดาว รูปแบบนอร์มอล (NORMAL FORM) และการทำดีนอร์มอลไลซ์ จะแสดงดังรูปที่ 2.4 จะเห็นว่าการเชื่อมโยงแบบดาวนั้นเทเบิลไดเมนชันจะไม่เชื่อมโยงกับเทเบิลใด ๆ นอกจากเทเบิลแฟคท์ การเชื่อมโยงแบบสโนว์เฟลค (SNOWFLAKE) เทเบิลไดเมนชันจะมีเทเบิลย่อย และรูปแบบนอร์มอลระดับที่ 3 นั้น เทเบิลไดเมนชันจะมีการเชื่อมโยงกันเองระหว่างเทเบิลไดเมนชันด้วยกัน



รูปที่ 2.4 โครงสร้างแบบสตาร์เฟลค (STRUCTURE OF A STARFLAKE SCHEMA) [2]

คุณภาพของข้อมูล (Data Quality)

ในที่นี้ยึดตามสำนวน "ขยะเข้ามา ก็ได้ขยะออกไป" (GARBAGE IN GARBAGE OUT - GIGO) [5] หากสิ่งที่ป้อนเข้ามาเป็นอินพุต (INPUT) ไม่ถูกต้อง มีความผิดพลาดคลาดเคลื่อนหรือไม่สมบูรณ์ ผลลัพธ์ที่ได้ออกไปก็จะผิดพลาดคลาดเคลื่อนหรือไม่สมบูรณ์ตามไปด้วย

คลังข้อมูลจะสมบูรณ์ได้ ก็ต่อเมื่อมีองค์ประกอบครบทั้ง 3 ประการดังที่กล่าวมาก่อนหน้านี้ และข้อมูลที่นำเข้าคลังจะต้องมีคุณภาพ ถูกต้อง และสมบูรณ์ ผลลัพธ์ที่ใช้ในการตัดสินใจก็จึงจะสมบูรณ์ตามไปด้วย ในที่นี้จะใช้คุณภาพข้อมูลดังนี้

ลักษณะของคุณภาพข้อมูล [3]

1. ความถูกต้อง (ACCURACY) เมื่อเทียบกับระบบต้นแบบ
2. ความครบถ้วน (COMPLETENESS) เมื่อเทียบกับระบบต้นแบบ
3. ค่าของข้อมูลอยู่ในโดเมนที่ยอมรับได้ (VALIDITY)
4. ความคงที่ (CONSISTENCY) ถ้าข้อมูลเข้าเหมือนกันผลลัพธ์ก็ต้องเหมือนกันด้วย
5. อ้างอิงได้ (RELATABILITY)
6. ข้อมูลถูกต้อง ณ จุดเวลาเสมอ แม้นว่าเวลาจะผ่านไปนานแล้วก็ตาม (TIMELINESS)
7. ไม่ซ้ำกัน (UNIQUENESS)

แนวคิดเพิ่มเติมที่ใช้ในการออกแบบคลังข้อมูล

1. อิสระต่อกัน และ สามารถดัดแปลงแก้ไขได้
2. สามารถเพิ่มและลดรายการข้อมูลในคลังได้ทุกเมื่อ ในที่นี้จะกำหนดให้รายการข้อมูลให้เป็นรหัสและเพิ่มหรือลบรายการข้อมูลในแนวดิ่งแทนการเพิ่มและลดคอลัมน์ การสร้าง แก้ไข และลบเทเบิลจะทำผ่านเมตะดาตา ดังนั้นคลังข้อมูลจะสมบูรณ์หรือไม่จะขึ้นอยู่กับข้อมูลเมตะดาตา
3. สิ่งที่เห็นคือสิ่งที่ได้รับ ผู้ใช้สามารถทราบได้ว่ามีข้อมูลอะไรบ้างในคลังข้อมูล โดยในที่นี้ออกแบบให้มีการเก็บรายการข้อมูลทุกรายการไว้ในเมตะดาตารีโพสิทอรี ซึ่งการสร้าง แก้ไข และลบเทเบิลจะทำผ่านเมตะดาตาเสมอ

ขั้นตอนการดำเนินการ

1. กำหนดผู้ใช้ และสำรวจความต้องการของผู้ใช้
2. กำหนดระบบงานต้นแบบและรูปแบบของข้อมูล
3. สร้างภาพรวมของการใช้ข้อมูลร่วมกันทั้งองค์กร เพื่อเป็นแนวทางในการกำหนดเรื่องที่สนใจ
4. คัดเลือกและสร้างแบบ (MODEL) ของข้อมูลนำเข้าของคลังข้อมูล โดยคัดเลือกจากรูปแบบโครงสร้างข้อมูล (DATA STRUCTURE) ของระบบต้นแบบ พร้อมทั้งกำหนดวิธีการสรุป คำนวณข้อมูลบางรายการที่จำเป็นต้องใช้ยอดสรุปและผลลัพธ์ในการคำนวณเป็นข้อมูลนำเข้าของคลังข้อมูล และระบุเหตุผลในการคัดเลือกข้อมูลแต่ละรายการ เนื่องจากเป็นไปได้ที่จะมีข้อมูลเดียวกันซ้ำกันระหว่างระบบ
5. กำหนดข้อมูลในคลังข้อมูลและออกแบบการเชื่อมโยงของข้อมูลเหล่านี้

5.1 สร้างเมตะดาตา

5.2 ออกแบบฐานข้อมูลของคลังข้อมูลและการเชื่อมโยงแบบดาว

5.3 พัฒนาต้นแบบของคลังข้อมูล

6. กำหนดวิธีการและขั้นตอนแปรรูปข้อมูลที่คัดเลือกไว้ เพื่อเป็นข้อมูลนำเข้าของคลังข้อมูล พัฒนาโปรแกรม

7. กำหนดวิธีการและขั้นตอนการนำข้อมูลเข้าคลัง พร้อมทั้งวิธีการตรวจสอบความถูกต้องของข้อมูล พัฒนาโปรแกรม

8. กำหนดวิธีการเรียกใช้ข้อมูลในคลัง และพัฒนาโปรแกรมสร้างต้นแบบของการเรียกใช้ข้อมูล เพื่อแสดงว่าคลังข้อมูลนี้สามารถใช้งานได้จริง

9. กำหนดวิธีการเปลี่ยนแปลงข้อมูลในคลังเมื่อข้อมูลในระบบต้นแบบเปลี่ยนแปลง

10. กำหนดวิธีการรักษาความปลอดภัยของคลังข้อมูล

11. กำหนดวิธีการลบข้อมูลออกจากคลังและเวลาที่ควรลบ